A Note on an LQG Regulator with Markovian Switching and Pathwise Average Cost

by M.K. Ghosh, A. Arapolasthis and S.I. Marcus
A Note on an LQG Regulator with Markovian Switching and Pathwise Average Cost

University of Maryland, Systems Research Center, College Park, MD, 20742

Approved for public release; distribution unlimited

see report

unclassified
unclassified
unclassified

7
A NOTE ON AN LQG REGULATOR WITH MARKOVIAN SWITCHING AND PATHWISE AVERAGE COST*

MRINAL K. GHOSH†, ARISTOTLE ARAPOSTATIS‡ AND STEVEN I. MARCUS§

ABSTRACT. We study a linear system with a Markovian switching parameter perturbed by white noise. The cost function is quadratic. Under certain conditions, we find a linear feedback control which is almost surely optimal for the pathwise average cost over the infinite planning horizon.

1. Introduction. We study a parameterized linear system perturbed by white noise. The parameters are randomly switching from one state to the other and are modeled as a finite state Markov chain; the values of the parameter and the state of the linear system are assumed to be known to the controller. The objective is to minimize a quadratic cost over the infinite planning horizon. Such dynamics arise quite often in numerous applications involving systems with multiple modes or failure modes, such as fault tolerant control systems, multiple target tracking, flexible manufacturing systems, etc. [3], [4], [9].

For a finite planning horizon, the problem is well understood, but difficulties arise when the planning horizon is infinite (very large in practice) and one looks for a steady state solution. Due to constant perturbation by the white noise, the total cost is usually infinite, rendering the total cost criterion inappropriate for measuring performance. In this situation, one studies the (long-run) average cost criterion. Here we study the optimal control problem of such a system with pathwise average cost. Pathwise results are very important in practical applications, since we often deal with a single realization. This problem for a very general hybrid system has been studied in [5], where we have established the existence of an almost surely optimal stationary Markov control and have characterized it as a minimizing selector of the Hamiltonian associated with the corresponding dynamic programming equations. When specializing to the LQG case, the existence results in [5] carry through with minor modifications, but the characterization results do not, since the boundedness of the drift is crucially used to derive the dynamic programming equations. In this note we sketch the derivation of the existence result for the LQG problem from [5]. We then characterize

*This work was supported in part by the Texas Advanced Research Program (Advanced Technology Program) under Grant No. 003658-186, in part by the Air Force Office of Scientific Research under Grants F49620-92-J-0045 and F49620-92-J-0083, and in part by the National Science Foundation under Grant CDR-8803012.
†Department of Mathematics, Indian Institute of Science, Bangalore 560012.
‡Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, Texas 78712.
§Electrical Engineering Department and Systems Research Center, University of Maryland, College Park, Maryland 20742.
the optimal control by solving the dynamic programming equations via Riccati equations. Similar dynamics, but with an overtaking optimality criterion, have been recently studied in [6].

Our paper is structured as follows. Section 2 deals with the problem description. The main results are contained in Section 3. Section 4 concludes the paper with some remarks.

2. Problem Description. For the sake of notational convenience, we treat the scalar case. The higher dimensional case can be treated in an analogous manner. Let \( S(t) \) be a (continuous time) Markov chain taking values in \( S = \{1, 2, \ldots, N\} \) with generator \( \Lambda = [\lambda_{ij}] \) such that \( \lambda_{ij} > 0, i \neq j \) (this condition can be relaxed. See Remark 3.1(v)). Let \( X(t) \) be a one-dimensional process given by

\[
\begin{align*}
    dX(t) &= [A(S(t))X(t) + B(S(t))u(t)]dt + \sigma(S(t))dW(t) \\
    X(0) &= X_0
\end{align*}
\]  

for \( t \geq 0 \), where \( W(\cdot) \) is a one-dimensional standard Brownian motion independent of \( S(\cdot) \) and \( X_0 \). \( u(t) \) is a real valued nonanticipative process satisfying

\[
\int_0^T u^2(t) dt < \infty \quad \text{a.s. (almost surely)}
\]  

for each \( T > 0 \), \( A(i), B(i), \sigma(i) \) are scalars such that \( B(i) \neq 0 \) and \( \sigma(i) > 0 \) for each \( i \). We will often write \( A_i, B_i, \) etc. instead of \( A(i), B(i) \). The nonanticipative process \( u(t) \) satisfying (2.2) is called an admissible control. It is called a stationary Markov control if \( u(t) = v(X(t), S(t)) \) for a measurable map \( v : \mathbb{R} \times S \to \mathbb{R} \). With an abuse of notation, the map \( v \) itself is called a stationary Markov control. A stationary Markov control \( v(x, i) \) is called a linear feedback control if \( v(x, i) = k_i x \), where \( k_i \) are scalars, \( i = 1, \ldots, N \).

Under a stationary Markov control \( v \), the hybrid process \( (X(t), S(t)) \) is a strong (time-homogeneous) Markov process [4]. Let \( v \) be a stationary Markov control. Under \( v \), the point \( (x, i) \in \mathbb{R} \times S \) is said to be recurrent under \( v \) if

\[
P_{x,i}^v(X(t_n) = x, S(t_n) = i, \text{ for some sequence } t_n \uparrow \infty) = 1,
\]

where \( P_{x,i}^v \) is the measure under \( v \) and with initial condition \( X(0) = x, S(0) = i \), in (2.1).

A point \((x, i)\) is transient under \( v \), if

\[
P_{x,i}^v(|X(t)| \to \infty, \text{ as } t \to \infty) = 1.
\]

For a space of dimension \( d > 1 \), the point \((x, i) \in \mathbb{R}^d \times S\), is said to be recurrent under \( v \), if given any \( \epsilon > 0 \),

\[
P_{x,i}^v(\|X(t_n) - x\| < \epsilon, S(t_n) = i, \text{ for some sequence } t_n \uparrow \infty) = 1.
\]

If all points \((x, i)\) are recurrent, then the hybrid process \((X(t), S(t))\) is called recurrent. It is shown in [5] that under our assumption, for any stationary Markov control, \((X(t), S(t))\)
is either recurrent or transient. A recurrent \((X(t), S(t))\) will admit a unique (up to a constant multiple) \(\sigma\)-finite invariant measure on \(\mathbb{R} \times \mathcal{S}\) (resp. \(\mathbb{R}^d \times \mathcal{S}\) for higher dimension). The hybrid process \((X(t), S(t))\) is called positive recurrent if it is recurrent and admits a finite invariant measure. For other details concerning recurrence and ergodicity of this class of hybrid systems we refer the reader to [5]. A stationary Markov control \(v\) is called \textit{stable} if the corresponding process \((X(t), S(t))\) is positive recurrent. The cost function \(c(x, i, u) : \mathbb{R} \times \mathcal{S} \times \mathbb{R} \to \mathbb{R}_+\) is given by
\[
(2.3) \quad c(x, i, u) = C(i)x^2 + D(i)u^2,
\]
where \(C_i > 0, D_i > 0\) for each \(i\). We say that an admissible policy \(u(\cdot)\) is a.s. optimal if there exists a constant \(\rho^*\) such that
\[
(2.4) \quad \limsup_{T \to \infty} \frac{1}{T} \int_0^T [C(S(t))X^2(t) + D(S(t))u^2(t)] dt = \rho^* \quad P^u \text{ a.s.},
\]
where \(X(t)\) is the solution of (2.1) under \(u(\cdot)\), and for any other admissible policy \(\bar{u}(\cdot)\)
\[
(2.5) \quad \limsup_{T \to \infty} \frac{1}{T} \int_0^T [C(S(t))\bar{X}^2(t) + D(S(t))\bar{u}^2(t)] dt \geq \rho^* \quad P^{\bar{u}} \text{ a.s.},
\]
where \(\bar{X}(t)\) is the solution of (2.1) under \(\bar{u}(\cdot)\), and initial condition \(\bar{X}(0) = X_0\). Note that in (2.4) and (2.5) the two measures with respect to which the 'a.s.' qualifies may be defined on different measurable spaces. Our goal is to show the existence of an a.s. optimal control and then find an a.s. optimal linear feedback control.

3. Main Results. We first note that the set of stable Markov policies is nonempty. The proof of this claim is rather standard and relies on a Lyapunov technique (see [10] for more general results).

For a stable stationary Markov control \(v\), there exists a unique invariant probability measure, denoted by \(\eta_v\), for the corresponding process \((X(t), S(t))\), and
\[
\rho_v := \lim_{T \to \infty} \frac{1}{T} \int_0^T \left[ C(S(t))X^2(t) + D(S(t))v^2(X(t), S(t)) \right] dt
\]
\[
= \sum_{i=1}^N \int_{\mathbb{R}} (C_i x^2 + D_i v^2(x, i)) \eta_v(dx, i) \quad \text{a.s.}
\]
If \(v\), as above, is a stable linear feedback, then it can be shown as in [11] that \(\rho_v < \infty\). Let
\[
\rho^* = \inf_v \rho_v
\]
where the infimum is over all stable linear feedback controls. Clearly, we look for a stable linear feedback control \(v^*\) such that \(\rho^* = \rho_{v^*}\) and \(v^*\) is a.s. optimal among all admissible controls. Since \(C_i > 0\), the cost function \(c\) in (2.3) penalizes the unstable behavior. In other words, the cost penalizes the drift of the process away from some compact set, requiring the optimal control to exert some kind of a "centripetal force" pushing the process back towards this compact set. Thus, the optimal control gains the desired stability property. In the framework of [5], it is easily seen that the penalizing condition (A5)
\[
\lim_{|x| \to \infty} \inf_u c(x, i, u) > \rho^*
\]
is satisfied. Thus by the results of [5], we have the following existence result.
Theorem 3.1. There exists a stable stationary Markov control which is a.s. optimal.

We now proceed to find a stable linear feedback control which is a.s. optimal. To this end, we study the dynamic programming equation given by

\[
(3.4) \quad \frac{1}{2} \sigma_i^2 V''(x, i) + \min_u \left[ (A_i x + B_i u) V'(x, i) + C(i) x^2 + D(i) u^2 \right] + \sum_{j=1}^{N} \lambda_{ij} V(x, j) = \rho,
\]

where \( \rho \) is a scalar and \( V : \mathbb{R} \times \mathcal{S} \to \mathbb{R} \). Since the drift is unbounded, the dynamic programming treatment of [5] is not applicable here. We look for a trial solution of (3.4), of the form

\[
(3.5) \quad V(x, i) = Q_i x^2 + R_i,
\]

where \( Q_i \) and \( R_i \) are to be determined. Following the usual procedure, we find that the minimizing selector in (3.4) is given by

\[
(3.6) \quad u(x, i) = -\frac{B_i Q_i x}{D_i}
\]

where the \( Q_i \)'s are the unique positive solution of the algebraic Riccati system

\[
(3.7) \quad 2A_i Q_i - \frac{B_i^2 Q_i^2}{D_i} + C_i + \sum_{j=1}^{N} \lambda_{ij} Q_j = 0
\]

and the \( R_i \)'s are given by

\[
(3.8) \quad \sigma_i^2 Q_i + \sum_{j=1}^{N} \lambda_{ij} R_j = \rho.
\]

Note that (3.8) is an underdetermined system of equations in \( R_i, \rho, i = 1, 2, \ldots, N \). Also, if the \( R_i \)'s satisfy (3.8), then so do \( R_i + k \) for any constant \( k \).

Lemma 3.1. Fix an \( i_0 \in \mathcal{S} \). Then there exists a unique solution \((R_i, \rho)\) to (3.8) satisfying \( R_{i_0} = 0 \).

Proof. We have for any \( T > 0 \),

\[
(3.9) \quad E R(S(T)) - E R(S(0)) = E \left[ \int_0^T \sum_{j=1}^{N} \lambda_{S(t), j} R_j \, dt \right]
\]

where the second equality follows from (3.8). Dividing (3.9) by \( T \), letting \( T \to \infty \), and using the fact that the chain \( S(t) \) is irreducible and ergodic, we have

\[
(3.10) \quad \rho = \lim_{T \to \infty} \frac{1}{T} E \left[ \int_0^T \sigma^2(S(t)) Q(S(t)) \, dt \right]
\]

\[
= \lim_{T \to \infty} \frac{1}{T} \int_0^T \sigma^2(S(t)) Q(S(t)) \, dt \quad \text{a.s.}
\]

\[
= \sum_{i=1}^{N} \sigma_i^2 Q_i \pi(i)
\]
where \( \pi = [\pi(1), \ldots, \pi(N)]' \) is the (unique) invariant probability measure of \( S(t) \). Let \( R_i \) be defined by

\[
R_i = E\left[ \int_0^{\tau_{io}} (\sigma^2(S(t))Q(S(t)) - \rho) dt \mid S(0) = i \right]
\]

(3.11)

where \( \tau_{io} = \inf \{ t > 0 \mid S(t) = i_0 \} \). Then, using Dynkin’s formula, it is seen that \( R_i \) satisfies (3.8) with \( R_{i_0} = 0 \). Let \( (R'_i, \rho') \) be another solution to (3.8) such that \( R'_{i_0} = 0 \). As before, we have

\[
\rho' = \sum_{i=1}^{N} \sigma_i^2 Q_i \pi(i) = \rho.
\]

Let \( M(t) = R(S(t)) - R'(S(t)) \). Then using (3.8), \( M(t) \) is easily seen to be a martingale which converges a.s. Since \( S(t) \) is positive recurrent, it visits every state infinitely often. Hence \( R_i - R'_i \) must be a constant. Thus \( R_i - R'_i = R_{i_0} - R'_{i_0} = 0 \). \( \square \)

In view of these results, using Ito’s formula and the pathwise analysis of [5], the following result is now apparent.

**Theorem 3.2.** For the LQG problem (2.1)–(2.3), the linear feedback control given by (3.6) is a.s. optimal, where the \( Q_i \)'s are determined by (3.8). The pathwise optimal cost is given by (3.9).

Some comments are in order now.

**Remark 3.1.**

(i) The condition \( B_i \neq 0 \) for each \( i \) can be relaxed. If for some \( i \), \( B_i = 0 \), then the above result will still hold if \( (A_i, B_i) \) are stochastically stabilizable in a certain sense [7], [9].

(ii) For the multidimensional case, if \( B_i, C_i, \sigma_i \) are positive definite, then all the above results will hold. But the positive definiteness of \( B_i \) is a very strong condition, since in many cases the \( B_i \)'s may not even be square matrices. In such a case if we assume that \( (A_i, B_i) \) are stochastically stabilizable, then the above results will again hold. Sufficient conditions for stochastic stabilizability are given in [2], [7], [9]. If \( C_i \) is not positive definite, then the cost does not necessarily penalize the unstable behavior, as discussed in the foregoing. Thus the condition (A5) of [5] is not satisfied. In this case under a further detectability condition, the optimality can be obtained in a restricted class of stationary Markov controls [2].

(iii) Let \( \rho \) be as in (3.9). Then for any admissible policy \( u(t) \) it can be shown by the pathwise analysis in [5] that

\[
\liminf_{T \to \infty} \frac{1}{T} \int_0^T \left[ C(S(t)) X^2(t) + D(S(t)) u^2(t) \right] dt \geq \rho \quad \text{a.s.}
\]

This establishes the optimality of the linear feedback control \( v \) (3.6) in a much stronger sense, viz. the most “pessimistic” pathwise average cost under \( v \) is no worse than the most “optimistic” pathwise average cost under any admissible control.
(iv) For \( T > 0 \), let \( V(x, i, T) \) denote the optimal expected cost for the finite horizon \([0, T]\). Then it can be shown as in [2] that

\[
\lim_{T \to \infty} \frac{1}{T} V(x, i, T) = \rho
\]

where \( \rho \) is as in (3.9). Thus the finite horizon value function approaches the optimal pathwise average cost as the length of the horizon increases to infinity. Thus for large \( T \), the linear feedback control (3.6) would be a reasonably good nearly optimal control for the finite horizon case. This would be particularly useful in practical applications since it is computationally more economical to solve the algebraic Riccati system than the Riccati system of differential equations.

(v) The condition \( \lambda_{ij} > 0 \) can be relaxed to the condition that the chain \( S(t) \) is irreducible (and hence ergodic). The existence part in [5] can be suitably modified to make the necessary claim here. In the dynamic programming part, the existence of a unique solution in Lemma 3.1 is clearly true under the irreducibility condition.

4. Conclusions. In this note, we have studied the pathwise optimality of an LQG regulator with Markovian switching parameters. We have assumed that the Markovian parameters are known to the controllers. This is an ideal situation. In practice the controllers may not have a complete knowledge of these parameters. In this case, one usually studies the corresponding minimum variance filter. Unfortunately, this filter is almost always infinite dimensional [9]. A computationally efficient suboptimal filter has been developed in [1], [8]. We hope that our results will be useful in the dual control problem arising in this situation.

REFERENCES

5. ______, Ergodic control of switching diffusions, preprint.