Piecewise-Deterministic Differential Games with Hybrid Controls

Eitan Altman
INRIA B.P. 93
2004 Route des Lucioles
06902 Sophia Antipolis Cedex
FRANCE

Tamer Başar[1]
Coordinated Science Lab.
University of Illinois
1308 West Main Street
Urbana, IL 61801 / USA

Zigang Pan[2]
EE Dept
Polytechnic Univ.
Six MetroTech Center
Brooklyn,NY 11201 / USA

Abstract

We consider a class of linear-quadratic differential games with piecewise deterministic dynamics, where the changes from one structure (for the dynamics) to another are governed by a finite-state Markov process. Player 1 controls the continuous dynamics, whereas Player 2 controls the rate of transition for the finite-state Markov process; both have access to the states of both processes. Player 1 wishes to minimize a given quadratic performance index, while Player 2 wishes to maximize the same quantity. For this class of differential games, we develop a theory of existence and uniqueness of a saddle point, and develop computational algorithms for solving it. Analytical (closed-form) solutions are obtained for the case when the continuous state is of dimension one. The paper also considers the modified problem where Player 2 is also a minimizer, that is when the underlying problem is a *team*, and illustrates the theoretical results in this context on a problem that arises in high speed telecommunication networks. This application involves combined admission and rate-based flow control, where the former corresponds to control of the finite-state Markov process, and the latter to control of the continuous linear system.

1 Introduction

We consider in this paper the optimal control of linear jump parameter systems under a quadratic cost criterion, when the transition probabilities associated with the underlying Markov chain are also controlled. We study two scenarios: The transition probabilities are controlled cooperatively (which leads to a team problem), and they are controlled antagonistically (which leads to a zero-sum differential game). In both cases, the control governing the switches takes values in a finite set, and is allowed to depend on the present and past values of the hybrid state – the same information

[1]September 30, 1996. Research supported by Grants NSF ECS 92-20632 and NSF ECS 93-12807 from the National Science Foundation.

[2]This work was done while this author was with the ECE Department of the University of California at Santa Barbara, Santa Barbara, CA, 93106.

that the "continuous" controller is endowed with. We treat both the finite- and infinite- horizon cases.

One of the difficulties associated with these classes of problems is that the associated dynamic programming equations cannot be solved explicitly, and may not even admit continuously-differentiable value functions. It turns out that they can be solved in the scalar case (for which we provide a complete set of results), and for some specific classes controlled rate matrices in the higher dimensional case. For the cases when they do not admit closed-form solutions, we provide schemes whereby they can be approximated, both from below and above, and suboptimal policies can be constructed. We illustrate the results obtained for the scalar case on a telecommunications problem that features admission and rate-based controls, where the former controls the probabilities of accepting different types of traffic to the network, and the latter controls the flow of the trafic through the network.

Research on control systems with Markovian switching structure was initiated more than thirty years ago by Krassovskii and Lidskii [10, 11] and Florentin [12], with follow-up work in the late sixties and early seventies addressing the stochastic maximum principle [13, 14], dynamic programming [14], and linear-quadratic control [15, 16] in this context. Late eighties and early nineties have witnessed renewed interest in the topic, with concentrated research on theoretical issues like controllability and stabilizability [17, 6] in linear-quadratic systems. Perhaps the first theoretical development in a differential games context was reported in [18], where a general model was adopted that allows in a multiple player environment the Markov jump process (also controlled by the players) to determine the mode of play, in addition to affecting the system and cost structure. Results in [18] also apply, as a special case, to zero-sum differential games with Markov jump parameters and state feedback information for both players. More recently, this framework has also been used in worst-case (H^∞) control design problems where the underlying system changes structure at random points of time according to a finite-state Markov process with a constant rate matrix, and with the system dynamics affected by an additional (continuous-valued) disturbance process; some selected publications on this topic are [19], [3], [7], and [20]

The paper is organized as follows. Section 2 introduces the general model studied, and the solution concepts used. Section 3 provides the associated dynamic programming equations, and discusses their solutions in some special cases. Sections 4 and 5 study respectively computational algorithms, and bounding and approximation techniques. The illustrative example involving a high speed communication network is presented in Section 6, and the paper ends with the concluding remarks of Section 7.

2 General Model

The continuous dynamics of the game under consideration evolve according to

$$\frac{dx}{dt} = A(\theta)x + B(\theta)u^1, \quad x(0) = x_0 \tag{1}$$

where $x \in I\!R^n$, x_0 is a fixed (known) initial state, u^1 is a control, applied by Player 1, taking values in $\mathbf{U}_1 = I\!R^r$, and $\theta(t)$ is a controlled, continuous time Markov process, taking values in a finite state space $\mathcal{S}$, of cardinality s. Transitions from state i to j occur at a rate controlled by Player 2, who chooses at time t an action $u^2(t)$ from among a finite set $\mathbf{U}_2(i)$ of actions available at state i. Let $\mathbf{U}_2 := \cup_{i\in\mathcal{S}}\mathbf{U}_2(i)$. The controlled rate matrix (of transitions within $\mathcal{S}$) is

$$\Lambda = \{\lambda_{i,a,j}\}, \qquad i,j \in S,\ a \in \mathbf{U}_2(i),$$

where henceforth we drop the "commas" in the subscripts of λ. The λ_{iaj}'s are real numbers such that for any $i \neq j$, and $a \in \mathbf{U}_2(i)$, $\lambda_{iaj} \geq 0$, and for all $a \in \mathbf{U}_2(i)$ and $i \in \mathcal{S}$, $\lambda_{iai} = -\sum_{j\neq i}\lambda_{iaj}$. Fix some initial state i_0 of the controlled Markov chain $\mathcal{S}$, and the final time t_f (which may be infinite). Consider the class of policies $\mu^k \in \mathcal{U}_k$ for Player k ($k = 1, 2$), whose elements (taking values in $\mathbf{U}_k$) are of the form

$$u^k(t) = \mu^k(t, x_{[0,t]}; \theta_{[0,t]}), \qquad t \in [0, t_f)\,. \tag{2}$$

Here, μ^k is taken to be piecewise continuous in its first argument, and piecewise Lipschitz continuous in its second argument.

Define $\mathcal{X} = I\!R^n \times \mathcal{S}$ to be the global state space of the system and $\mathcal{U} := \mathcal{U}_1 \times \mathcal{U}_2$ to be the class of multi-strategies $\mu := (\mu^1, \mu^2)$. Define the running (immediate) cost $L : \mathcal{X} \times \mathbf{U}_1 \to I\!R$ as:

$$L(x, i, u^1) = |x|^2_{Q(i)} + |u^1|^2_{R(i)}\,, \tag{3}$$

where $Q(\cdot) \geq 0$ and $R(\cdot) > 0$, and $|x|_Q$ denotes the Euclidean (semi-) norm.

To any fixed initial state (x_0, i_0) and a multi-strategy $\mu \in \mathcal{U}$, there corresponds a unique probability measure $P^{\mu}_{x_0,i_0}$ on the canonical probability space Ω of the states and actions of the players, equipped with the standard Borel σ-algebra. Denote by $\mathbf{E}^{\mu}_{x_0,i_0}$ the expectation operator corresponding to $P^{\mu}_{x_0,i_0}$. By a possible abuse of notation, denote by $(x(t), \theta(t)), u(t), t \in [0, t_f)$, the stochastic processes corresponding to the states and actions, respectively.

In terms of this notation and convention, and for each fixed initial state (x_0, i_0), multi-strategy $\mu \in \mathcal{U}$, and a horizon of duration t_f, we introduce the discounted (expected) cost function:

$$J_\beta(x_0, i_0, \mu; t_f) := \mathbf{E}^{\mu}_{x_0,i_0}\left\{|x(t_f)|^2_{Q_f(\theta(t_f))}e^{-\beta t_f} + \int_0^{t_f} e^{-\beta t}L(x(t), \theta(t), u^1(t))dt\right\}, \tag{4}$$

where $\beta \geq 0$ is a discount factor (allowed also to be *zero*), and $Q_f(\cdot) \geq 0$ is a terminal state weighting matrix. We further denote the *cost-to-go* from any time-state pair $(t; x, i)$ by

$$J_\beta(t; x, i, \mu; t_f) := \mathbf{E}^\mu_{x,i} \left\{ |x(t_f)|^2_{Q_f(\theta(t_f))} e^{-\beta(t_f - t)} + \int_t^{t_f} e^{-\beta(\sigma - t)} L(x(\sigma), \theta(\sigma), u^1(\sigma)) d\sigma \right\} . \tag{5}$$

For $t_f = \infty$, we shall simply drop t_f as an argument of J_β in both cases, and take $Q_f \equiv 0$. For this infinite-horizon problem, we have to ensure that the cost is finite for at least one stationary policy of Player 1, and for all possible realizations of the rate matrix. A sufficient condition for this is the following, which we invoke throughout our analysis, when dealing with infinite-horizon team and game problems.

Condition 2.1 *The pair $(A(\theta(t)), B(\theta(t)))$ is stochastically stabilizable for all x-independent stationary policies of Player 2, and $(A(i), Q(i))$ is observable for each $i \in \mathcal{S}$.*[3]

The problem addressed in this paper is the derivation of a solution to

$$\hat{J}_\beta(t; x, i; t_f) := \mathrm{val}_{\mu \in \mathcal{U}} J_\beta(t; x, i, \mu; t_f) \tag{6}$$

where $\mathrm{val}_{\mu \in \mathcal{U}}$ stands for the

$$\inf_{\mu^1 \in \mathcal{U}_1} \sup_{\mu^2 \in \mathcal{U}_2} \qquad \text{and} \qquad \inf_{\mu^1 \in \mathcal{U}_1} \inf_{\mu^2 \in \mathcal{U}_2}$$

operations for the game and team cases, respectively, with again t_f dropped as an argument in the infinite-horizon problem. For the team problem, $\hat{J}_\beta$ stands clearly for the global minimum of J_β, and in this case we also seek a multi-strategy $\mu^* \in \mathcal{U}$, such that

$$J_\beta(t; x, i, \mu^*; t_f) = \hat{J}_\beta(t; x, i; t_f) .$$

For the game problem, $\hat{J}_\beta$ stands for the upper value, and in this case we seek a $\mu^{1^*} \in \mathcal{U}_1$ such that

$$\sup_{\mu^2 \in \mathcal{U}_2} J_\beta(t; x, i, \mu^{1^*}, \mu^2; t_f) = \hat{J}_\beta(t; x, i; t_f) . \tag{7}$$

If the inf and sup operations can be interchanged, that is

$$\hat{J}_\beta(t; x, i; t_f) = \sup_{\mu^2 \in \mathcal{U}_2} \inf_{\mu^1 \in \mathcal{U}_1} J_\beta(t; x, i, \mu^1, \mu^2; t_f) ,$$

then $\hat{J}_\beta$ is the *value* of the game, and if furthermore (in addition to (7)),

$$\inf_{\mu^1 \in \mathcal{U}_1} J_\beta(t; x, i, \mu^1, \mu^{2^*}; t_f) = \hat{J}_\beta(t; x, i; t_f) , \tag{8}$$

for some $\mu^{2^*} \in \mathcal{U}_2$, then the pair (μ^{1^*}, μ^{2^*}) constitutes a *saddle-point* solution for the game.

[3]Stochastic stabilizability means that there exists a quadratic Lyapunov function for the linear jump parameter system (1) for each fixed achievable rate matrix. Together with observability, this implies that there exists a stabilizing controller that renders the quadratic terminal cost exponentially go to zero; see [6] and [7].

3 The Solution Process

Introduce a backward controlled Markov operator $\mathcal{A}^u$ associated with the system (1)-(4) as follows: For each $\psi(t,\cdot,i)$ such that $\psi(\cdot,\cdot,i) \in \mathcal{C}^1$ for all $i \in \mathcal{S}$ (where $\mathcal{C}^1$ is the vector space of all functions with continuous partial derivatives), and for each $u = (u^1, a) \equiv (u^1, u^2) \in \mathbf{U}(i)$,

$$\mathcal{A}^u \psi(t,x,i) := \frac{\partial \psi(t,x,i)}{\partial t} + [D_x \psi(t,x,i)]\ f(x,u^1,i) + \sum_{j \in \mathcal{S}} \lambda_{iaj} \psi(t,x,j)$$

where D_x stands for the gradient operator, and $f(x,u^1,i) := A(i)x + B(i)u^1$. Further introduce, for each function $\psi(x,i)$ for which $\psi(\cdot,i) \in \mathcal{C}^1$ for all $i \in \mathcal{S}$, and for each $u \in \mathbf{U}_1 \times \mathbf{U}_2(i)$, $i \in \mathcal{S}$,

$$G^u \psi(x,i) := [D_x \psi(x,i)]\ f(x,u^1,i) + \sum_{j \in \mathcal{S}} \lambda_{iaj} \psi(x,j).$$

Next we introduce the dynamic programming equations associated with the team and game problems above, where the continuous-differentiability of the value functions will be replaced simply by continuity, and their solutions interpreted in the generalized (viscosity) sense if differentiability fails.[4] The "val" operator below will have the same interpretation as before, but now applied to static game or team problems, over the product action space $\mathbf{U}_1 \times \mathbf{U}_2(i)$.

For finite t_f, consider, subject to the given boundary condition, the Hamilton-Jacobi-Bellman (HJB, in short) (or Isaacs) equation:

$$\beta \psi(t,x,i) = \mathrm{val}_{u^1,u^2} \left[\mathcal{A}^{(u^1,u^2)} \psi(t,x,i) + L(x,i,u^1) \right], \qquad \psi(t_f,x,i) = |x|^2_{Q_f(i)}. \tag{9}$$

For infinite t_f, consider its infinite-horizon counterpart:

$$\beta \psi(x,i) = \mathrm{val}_{u^1,u^2} \left[G^{(u^1,u^2)} \psi(x,i) + L(x,i,u^1) \right]. \tag{10}$$

Associate with (9) and (10) the corresponding sets, $\mathcal{C}$ and $\mathcal{C}_f$, respectively, of functions ψ continuous in (t,x) and x. The following theorems now relate optimal values and policies to the dynamic programming equation for the team and the game cases, respectively.

Theorem 3.1 Team problem
(i) For the case of finite t_f, the HJB equation (9) has a unique viscosity solution ψ in $\mathcal{C}_f$, and the value of (5) equals ψ. Moreover, any Markov policy that chooses at time t, for all $t \in [0,t_f]$, actions that achieve the argmin in (9), given that the state at that time is (x,i), is optimal.
(ii) Let Condition 2.1 be satisfied. Then, there exists a unique nonnegative viscosity solution ψ in $\mathcal{C}$ to (10), with the further properties that:

[4] See [5] or [8] for the notion of *viscosity solution* as it appears in this context.

(a) $\psi(x,i) \le J_\beta(x,i,\mu)$ *for every* $\mu \in \mathcal{U}$ *that satisfies*

$$\underline{\lim}_{t_1 \to \infty} e^{-\beta t_1} \mathbf{E}^{\mu}_{x,i} \psi(x(t_1), \theta(t_1)) \le 0. \tag{11}$$

(b) Any stationary policy g *that chooses at state* (x,i) *actions that achieve the argmin in (10) satisfies* $\psi(x,i) \ge J_\beta(x,i,g)$*, provided that*

$$\overline{\lim}_{t_1 \to \infty} e^{-\beta t_1} \mathbf{E}^{g}_{x,i} \psi(x(t_1), \theta(t_1)) \ge 0. \tag{12}$$

Proof. (i) We first perturb the linearly coupled HJB equations (9) by a "small" second-order term, arriving at:

$$\beta \psi(t,x,i,\epsilon) = \min_{u^1,u^2} \left[\mathcal{A}^{(u^1,u^2)} \psi(t,x,i,\epsilon) + L(x,i,u^1) + \epsilon \frac{\partial^2}{\partial x^2} \psi(t,x,i,\epsilon) \right], \quad \psi(t_f,x,i,\epsilon) = |x|^2_{Q_f(i)}; \tag{13}$$

where $\epsilon > 0$ is the small parameter. It follows from an extension of an existence and uniqueness result for second-order PDE's due to Friedman [9] that:

(1) The linearly coupled second-order PDE's (13) admit a unique set of $\mathcal{C}^2$ solutions $\psi(t,x,i,\epsilon)$, $i = 1,\ldots,s$, and

(2) The functions $\psi(t,x,i,\epsilon)$, $i = 1,\ldots,s$, are bounded and equicontinuous on compact sets, for $\epsilon \in [0,\epsilon_0]$, where $\epsilon_0 > 0$ is a known constant.

Then, we can conclude that there exists a set of viscosity solutions to the HJB equation (9) as follows: Because of the equicontinuity of $\psi(t,x,i,\epsilon)$, $i = 1,\ldots,s$, by the Ascoli-Arzela Theorem, there exists a sequence of $\{\epsilon_k\}_{k=1}^{\infty}$ such that

$$\lim_{k \to \infty} \epsilon_k = 0, \qquad \lim_{k \to \infty} \psi(t,x,i,\epsilon_k) = \psi^*(t,x,i) \qquad \forall (t,x,i) \in [0,\infty) \times \mathbb{R}^{\mathrm{n}} \times \mathcal{S}$$

Furthermore, the functions converge uniformly on compact sets, and the limit functions $\psi^*(t,x,i)$ are continuous.

For each fixed $i \in \mathcal{S}$, the value function $\psi(t,x,i,\epsilon_k)$ satisfies

$$\begin{aligned} \beta \psi(t,x,i,\epsilon_k) - \epsilon \frac{\partial^2}{\partial x^2} \psi(t,x,i,\epsilon_k) - \min_{u^1} &\left[\frac{\partial \psi(t,x,i,\epsilon_k)}{\partial t} + [D_x \psi(t,x,i,\epsilon_k)]\, f(x,u^1,i) + L(x,i,u^1) \right] \\ = \quad \min_{u^2} &\left[\sum_{j \in \mathcal{S}} \lambda_{iaj} \psi(t,x,j,\epsilon_k) \right], \qquad \psi(t_f,x,i,\epsilon_k) = |x|^2_{Q_f(i)}. \end{aligned} \tag{14}$$

Note that the right-hand-side of the above equation converges to a continuous function. Then, by Corollary 3.2 of [8] (p. 21), we have that $\psi(t,x,i,\epsilon)$ converges to the unique viscosity solution of (9), which is $\psi^*(t,x,i)$.

This completes the proof of the first statement of (i). The second statement, that is optimality of the Markov policy that minimizes the RHS of (9), follows from the results of [5], Chapters III.8-III.10.

(ii) Proof of this part is similar to that of (i) above, with Condition 2.1 guaranteeing boundedness of the infinite-horizon cost. The two properties given follow from the results given in Chapters III.8-III.10 of [5]. ∎

We now establish the counterparts of Theorem 3.1 for the game setting separately for finite and infinite horizons.

Theorem 3.2 Zero-sum game: Finite horizon

Let t_f be finite, and assume that (9) admits a unique viscosity solution ψ in $\mathcal{C}_f$, with "val" interpreted simultaneously as "inf sup" and "sup inf". Then:

(i) The dynamic game has a value $\hat{J}_\beta \equiv \psi$.

(ii) Let $\overline{\mu}^2$ be a policy that chooses at time t, for all $t \in [0, t_f)$, an action that achieves the maximum in the static game:

$$\sup_{u^2 \in \mathbf{U}_2(i)} \inf_{u^1 \in \mathbf{U}_1} \left[\mathcal{A}^{(u^1,u^2)} \psi(t,x,i) + L(x,i,u^1) \right] = \beta \psi(t,x,i) \tag{15}$$

given that the state at that time is (x,i).

If $\overline{\mu}^2$ is in $\mathcal{U}_2$ (i.e., if the continuity and Lipschitz conditions are satisfied), then it is an optimal policy for Player 2.

(iii) Any Markov policy $\overline{\mu}^1 \in \mathcal{U}_1$ for Player 1 that chooses at time t, for all $t \in [0, t_f)$, actions that achieve the minimum in the static game:

$$\inf_{u^1 \in \mathbf{U}_1} \sup_{u^2 \in \mathbf{U}_2(i)} \left[\mathcal{A}^{(u^1,u^2)} \psi(t,x,i) + L(x,i,u^1) \right] = \beta \psi(t,x,i) \tag{16}$$

given that the state at that time is (x,i), is optimal for Player 1.

Proof. (ii) Consider the Markov policy $\overline{\mu}^1$ given in part (iii). Let $\mu^2 \in \mathcal{U}_2$ be an arbitrary policy for Player 2. Let $u(\cdot) = (u^1(\cdot), u^2(\cdot))$ be the actions used by the players when $\mu = (\overline{\mu}^1, \mu^2)$ is adopted. Since the value of the game in (9) is $\beta\psi(t,x,i)$, and μ^2 is not necessarily maximizing, we have

$$\mathcal{A}^{(u(t))}\psi(t,x,i) + L(t,x,i,u^1(t)) \le \beta\psi(t,x,i).$$

It then follows from Dynkin's formula of integration that

$$\psi(t,x,i) \;=\; E^{\mu}_{x,i}\left[\int_t^{t_f} -e^{-\beta(s-t)}\mathcal{A}^{(u(s))}\psi(s,x(s),\theta(s))ds + e^{-\beta(t_f-t)}\psi(t_f,x(t_f),\theta(t_f))\right]$$

$$
\begin{aligned}
&\geq E^{\mu}_{x,i}\left[\int_t^{t_f} e^{-\beta(s-t)}L(x(s),\theta(s),u^1(s))ds + e^{-\beta(t_f-t)}\psi(t_f,x(t_f),\theta(t_f))\right] \\
&\equiv J_\beta(t;x,i,\mu;t_f).
\end{aligned}
$$

Hence, μ^2 cannot be a maximizing policy unless (15) is satisfied.

(iii) Let $\overline{\mu}^2$ be an optimal policy for Player 2, as described in (ii). Let $\mu^1 \in \mathcal{U}_1$ be an arbitrary policy for Player 1. Let $u(s) = (u^1(s), u^2(s))$ be the actions used by the players dictated by $\mu = (\mu^1, \overline{\mu}^2)$. Then it follows from a reasoning similar to that in part (ii) that

$$
\mathcal{A}^{(u(t))}\psi(t,x,i) + L(x,i,u^1(s)) \geq \beta\psi(t,x,i)\,.
$$

Again using Dynkin's formula, we obtain the following equality and inequality:

$$
\begin{aligned}
\psi(t,x,i) &= E^{\mu}_{x,i}\left[\int_t^{t_f} -e^{-\beta(s-t)}\mathcal{A}^{(u(s))}\psi(s,x(s),\theta(s))ds + e^{-\beta(t_f-t)}\psi(t_f,x(t_f),\theta(t_f))\right] \\
&\leq E^{\mu}_{x,i}\left[\int_t^{t_f} e^{-\beta(s-t)}L(x(s),\theta(s),u^1(s))ds + e^{-\beta(t_f-t)}\psi(t_f,x(t_f),\theta(t_f))\right] \\
&\equiv J_\beta(t;x,i,\mu;t_f).
\end{aligned}
$$

This again shows that μ^1 cannot be a minimizing policy for Player 1 unless (16) is satisfied.

(i) is obtained by combining (ii) and (iii). ∎

Theorem 3.3 Zero-sum game: Infinite horizon
Let $t_f = \infty$, and Condition 2.1 be satisfied. Assume that (10) admits a unique nonnegative viscosity solution ψ in $\mathcal{C}$, with "val" interpreted simultaneously as "inf sup" and "sup inf". Let $\overline{\mu}^1 \in \mathcal{U}_1$ be a stationary policy for Player 1 that chooses at state (x,i) actions that achieve the minimum in

$$
\inf_{u^1\in\mathbf{U}_1}\sup_{u^2\in\mathbf{U}_2(i)}\left[G^{(u^1,u^2)}\psi(x,i) + L(x,i,u^1)\right]. \tag{17}
$$

Let $\overline{\mu}^2 \in \mathcal{U}_2$ be a stationary policy for Player 2 that chooses at state (x,i) actions that achieve the maximum in

$$
\sup_{u^2\in\mathbf{U}_2(i)}\inf_{u^1\in\mathbf{U}_1}\left[G^{(u^1,u^2)}\psi(x,i) + L(x,i,u^1)\right]. \tag{18}
$$

Then:
(i) $\psi(x,i) \leq J_\beta(x,i,\mu)$ for every $\mu \in \mathcal{U}$ that satisfies $\mu^2 = \overline{\mu}^2$ and

$$
\underline{\lim}_{t_1\to\infty}\, e^{-\beta t_1}\mathbf{E}^{\mu}_{x,i}\psi(x(t_1),\theta(t_1)) \leq 0. \tag{19}
$$

(ii) $\psi(x,i) \geq J_\beta(x,i,\mu)$, provided that $\mu^1 = \overline{\mu}^1$ and

$$
\overline{\lim}_{t_1\to\infty}\, e^{-\beta t_1}\mathbf{E}^{\mu}_{x,i}\psi(x(t_1),\theta(t_1)) \geq 0. \tag{20}
$$

Proof. Choose an arbitrary $\mu^1 \in \mathcal{U}_1$, and pick $\mu^2 = \overline{\mu}^2$. With $u(t)$ denoting the actions of both players taken at time t, we have by (10):

$$G^{u(t)}\psi(x(t),\theta(t)) - \beta\psi(x(t),\theta(t)) + L(x(t),\theta(t),u^1(t)) \geq \beta\psi(x(t),\theta(t)). \tag{21}$$

By applying Dynkin's formula as in the proof of the previous theorem, we get (as in [5] p. 146),

$$\psi(x,i) \geq E^{\mu}_{x,i}\int_0^{t_1} e^{-\beta t}L(x(t),\theta(t),u^1(t))dt + e^{-\beta t_1}E^{\mu}_{x,i}\psi(x(t_1),\theta(t_1)).$$

(i) is obtained by taking the limit as $t_1 \to \infty$ along a sequence for which the last term tends to a non-positive value. (ii) is established by symmetrical arguments. ∎

Equations (9) and (10) do not admit closed-form solutions in all cases, but they do in some special cases. To investigate these cases, let us first consider the finite-horizon case, and stipulate a structure for $\psi(t,x,i)$ that is quadratic in x:

$$\psi(t,x,i) := x^T P(i,t)x, \quad t \in [0,t_f], \quad i \in \mathcal{S}, \tag{22}$$

where $P(i,t)$ is an $n \times n$ symmetric matrix for each $i \in \mathcal{S}$, $t \in [0,t_f]$. Substituting this structural form into (9), we obtain:

$$\begin{aligned}\beta x^T P(i,t)x \quad &= \quad x^T P_t(i,t)x + \min_{u^1}\left[2\left(A(i)x + B(i)u^1\right)^T P(i,t)x + |u^1|^2_{R(i)}\right] \\ &\quad + \mathrm{opt}_a \sum_{j\in\mathcal{S}} \lambda_{iaj}\, x^T P(j,t)x \end{aligned} \tag{23}$$

where "opt" stands for minimization in the team problem, and maximization in the game problem. Note that for the game problem there is a complete separation of the min and max operations, and hence a value exists in (9), provided of course that this quadratic structure is valid. We now introduce a useful definition.

Definition 3.1 *We shall say that for the team problem (respectively, game problem) a square symmetric matrix P_1 dominates another square symmetric matrix P_2 of the same dimension if the difference matrix $P_1 - P_2$ is nonpositive definite (respectively, nonnegative definite).*

Now continuing with the derivation, the unique minimizing control u^1 in (23) is

$$\mu^1_{opt}(x,i,t) = -R^{-1}(i)B^T(i)P(i,t)\,x, \tag{24}$$

whose substitution into (23) leads to:

$$\begin{aligned} 0 \quad = \quad & x^T\Big(P_t(i,t) - \beta P(i,t) + Q(i) + A^T(i)P(i,t) + P(i,t)A(i) \\ & - P(i,t)B(i)R^{-1}(i)B^T(i)P(i,t)\Big)\,x + \mathrm{opt}_{a\in\mathbf{U}_2(i)}\sum_{j\in\mathcal{S}}\lambda_{iaj}x^T P(j,t)x\,, \end{aligned}$$

Hence, the quadratic structure is the right one provided that the optimization over $u^2 = a$ is independent of x, i.e.

$$\text{opt}_{a\in\mathbf{U}_2(i)} \sum_{j\in\mathcal{S}} \lambda_{iaj}\, x^T P(j,t) x = x^T \left(\text{opt}_{a\in\mathbf{U}_2(i)} \sum_{j\in\mathcal{S}} \lambda_{iaj}\, P(j,t) \right) x \tag{25}$$

where the second optimization is in the sense of *domination of matrices* (cf. Def. 3.1), and $P(j,t)$'s are obtained as nonnegative-definite matrices satisfying the coupled matrix Riccati differential equations:

$$\begin{aligned} P_t(i,t) &= A^T(i)P(i,t) + P(i,t)A(i) - \beta P(i,t) + Q(i) \\ &\quad - P(i,t)B(i)R^{-1}(i)B^T(i)P(i,t) + \text{opt}_{a\in\mathbf{U}_2(i)} \sum_{j\in\mathcal{S}} \lambda_{iaj}\, P(j,t)\,; \qquad (26) \\ P(i,t_f) &= Q_f(i)\,, \quad i = 1,\ldots,s \end{aligned}$$

Of course, for the scalar case (i.e. $n = 1$), condition (25) is automatically satisfied. This now brings us to the following theorem.

Theorem 3.4 *(i) Let t_f be finite, and assume that there exist nonnegative-definite matrix-valued functions $P(i,t)$, $i \in \mathcal{S}$, $t \in [0,t_f)$, satisfying (25) and (26). Then,*

$$\psi(t,x,i) = x^T P(i,t) x\,. \tag{27}$$

Furthermore, a Markov policy μ^{2^*} that uses at time t an action (depending on i and t, but not on x) that achieves the optimum in (25) (minimum in the team case, and maximum in the game case) is overall optimal. A corresponding optimal Markov policy for Player 1 is

$$\mu^{1^*}(x,i,t) = -R^{-1}(i)B^T(i)P(i,t)\, x\,. \tag{28}$$

(ii) Let $t_f = \infty$, and assume that there exist nonnegative-definite matrices $P(i), i \in \mathcal{S}$, satisfying

$$\begin{aligned} \beta P(i) &= A^T(i)P(i) + P(i)A(i) + Q(i) - P(i)B(i)R^{-1}(i)B^T(i)P(i) \\ &\quad + \text{opt}_a \sum_{j\in\mathcal{S}} \lambda_{iaj} P(j), \quad i \in \mathcal{S}\,, \qquad (29) \end{aligned}$$

with the properties that there exist $f_i \in \mathbf{U}_2(i), i \in \mathcal{S}$, such that for all $a \in \mathbf{U}_2(i)$ and all $i \in \mathcal{S}$,

$$\sum_{j\in\mathcal{S}} (\lambda_{if_ij} - \lambda_{iaj})\, P(j) \;\geq\; 0 \qquad \text{if “opt” = “max”} \tag{30}$$

$$\sum_{j\in\mathcal{S}} (\lambda_{if_ij} - \lambda_{iaj})\, P(j) \;\leq\; 0 \qquad \text{if “opt” = “min”} \tag{31}$$

where ≥ 0 *(respectively,* ≤ 0*) stands for nonnegative-definite (respectively, nonpositive-definite),* **and** *for all* $i \in \mathcal{S}, x \in I\!R^n$,

$$\lim_{t_1 \to \infty} e^{-\beta t_1} \mathbf{E}^{\mu^*}_{x,i} x^T(t_1) P(i) x(t_1) = 0\,, \tag{32}$$

where $\mu^* = (\mu^{1*}, \mu^{2*})$, *with*

$$\mu^{1*}(x,i) = -R^{-1}(i) B^T(i) P(i)\, x\,, \quad \mu^{2*}(i) = f_i\,. \tag{33}$$

Under these conditions, (33) constitutes a pair of optimal (team-optimal or saddle-point) policies.

Proof. (i) follows directly from Theorem 3.1 (i) for the team problem, and from Theorem 3.2 for the game problem. (ii) follows from Theorem 3.1 (ii) for the team case, and Theorem 3.3 for the game case, by following the derivation given prior to the statement of the theorem. ∎

The statement of Theorem 3.4 can be strengthened for the scalar case, which we now discuss for the infinite-horizon problem. Let F denote the set of stationary policies for Player 2 that depend only on the current state of the Markov chain. With a fixed as $f \in F$, let us denote the solution of (29) by $P^f(i)$, $i \in \mathcal{S}$. For each such fixed $f \in F$, (29) is in fact precisely the set of coupled algebraic Riccati equations that arises in the quadratic optimal control of standard linear jump parameter systems with a constant rate matrix ([6, 7]). Using that standard theory, we can readily conclude that (29) for $n = 1$, i.e. the linearly coupled set of quadratic equations:

$$\beta P(i) = 2A(i)P(i) + Q(i) - [P(i)]^2 N(i) + \sum_{j \in \mathcal{S}} \lambda_{if(i)j}\, P(j)\,,\ i \in \mathcal{S}\,, \tag{34}$$

where

$$N(i) := [B(i)]^2 / R(i)\,,\ i \in \mathcal{S}\,, \tag{35}$$

admits a unique positive solution set $P^f(i)$, $i \in \mathcal{S}$, provided that $N(i) \neq 0$ (i.e. $B(i) \neq 0$) and $Q(i) > 0$ for all $i \in \mathcal{S}$ — conditions we henceforth assume to hold.

We now describe a policy iteration algorithm that will lead to a solution of (29) for $n = 1$. For a reason which will become clear shortly, we first rewrite (29) for $n = 1$ in the following equivalent form:

$$(\beta + \bar{\lambda}_i) P(i) = 2A(i)P(i) - [P(i)]^2 N(i) + Q(i) + \mathrm{opt}_{a \in \mathbf{U}_2(i)} \sum_{j \in \mathcal{S}} \lambda_{iaj}\, P(j) + \bar{\lambda}_i P(i)\,, \tag{36}$$

where

$$\bar{\lambda}_i := \max_{a \in \mathbf{U}_2(i)} \sum_{j \neq i} \lambda_{iaj} \equiv \max_{a \in \mathbf{U}_2(i)} |\lambda_{iai}|\,, \quad i \in \mathcal{S}\,. \tag{37}$$

Note that this re-organization of (29) makes the last two terms of (36) nonnegative for all $P(j)$, $j \in \mathcal{S}$.

Now, let $f \in F$ and $P^f > 0$, identified earlier, and to be re-labeled as (f^0, P^{f^0}), be the starting choices for the policy iteration algorithm (to be described next). Given such a pair (f^k, P^{f^k}) at the k-th step, a new (improved) pair at the $k+1$-st step can be generated as follows:

$$f^{k+1}(i) = \arg\operatorname{opt}_{a \in \mathbf{U}_2(i)} \sum_{j \in \mathcal{S}} \lambda_{iaj}\, P^{f^{(k)}}(j)\,,\ i \in \mathcal{S} \tag{38}$$

$$P^{f^{(k+1)}}(j),\ i \in \mathcal{S},\ \text{ solves (36) with } a \text{ fixed at } f^{(k+1)}\,. \tag{39}$$

Note that $f^{(k+1)}$ exists (but may not be unique) since $\mathbf{U}_2(i)$, $i \in \mathcal{S}$, is a finite set, and $P^{f^{(k+1)}}$ exists, and is unique, since $N(i) \neq 0$ and $Q(i) > 0$, as pointed out earlier. $P^{f^{(k+1)}}$ will have the further property that

$$\text{if} \quad \text{“opt”} = \text{“min”}: \qquad P^{f^{(k+1)}} \leq P^{f^{(k)}} \quad \forall\, i \in \mathcal{S} \tag{40}$$

$$\text{if} \quad \text{“opt”} = \text{“max”}: \qquad P^{f^{(k+1)}} \geq P^{f^{(k)}} \quad \forall\, i \in \mathcal{S} \tag{41}$$

To see this for the team problem (“opt” = “min”), let us introduce an *inner-loop* iteration, using (36). With a fixed at $f \in F$, denote the last three terms of (36) by $\tilde{Q}(f, P; i)$, and note that

$$P'(i) \geq P(i) \quad \forall\, i \in \mathcal{S} \quad \Rightarrow \quad \tilde{Q}(f, P'; i) \geq \tilde{Q}(f, P; i) \qquad \forall\, i \in \mathcal{S}\,. \tag{42}$$

Consider the iteration (*inner-loop*):

$$(\beta + \bar{\lambda}_i - 2A(i))\, P^k_{(\ell+1)}(i) = -[P^k_{(\ell+1)}(i)]^2 N(i) + \tilde{Q}(f, P^k_{(\ell)}; i)\,,\ \ell = 0, 1, \ldots\,; \quad P^k_{(0)}(i) = P^{f^{(k)}}(i)\,, \tag{43}$$

where $P^k_{(\ell+1)}(i) > 0$ exists, and is unique, for each $i \in \mathcal{S}$ (note that $\tilde{Q} > 0$). Now, since $P^{f(k)}$ solves (36) with $a = f^{(k)}$, and since, from the definition of $f^{(k+1)}$,

$$\tilde{Q}(f^{(k+1)}, P^k_{(0)}; i) \leq \tilde{Q}(f^{(k)}, P^k_{(0)}; i)\,, \tag{44}$$

we have

$$P^k_{(1)}(i) \leq P^k_{(0)}(i)\,. \tag{45}$$

Hence, the inner-loop iteration generates a monotonically nondecreasing positive sequence, which has to converge to the solution of (36) with a fixed at $a = f^{(k+1)}$, which is $P^{f^{(k+1)}}$. This then verifies (40), which says that the outer-loop iteration generates a monotonically nonincreasing positive sequence, which therefore should converge:

$$\lim_{k \to \infty} P^{f^{(k)}}(i) = P^{f^*}(i)\,, \quad i \in \mathcal{S}\,. \tag{46}$$

Since $\mathbf{U}_2(i)$'s are finite, this convergence is in fact achieved in only a finite number of steps. This also says that there should exist a minimizing stationary policy for Player 2:

$$\mu^{2^*} = f^* . \tag{47}$$

Identical arguments as above hold for the game case ("opt" = "max"), where now we have a monotonically nondecreasing sequence of $P^{f^{(k)}}$'s, which is bounded above because (36) admits a positive solution for every $a = f,\ f \in F$.

We now provide a precise statement of this result in the following theorem.

Theorem 3.5 *Consider the framework of Theorem 3.4(ii), but for the scalar case, with* $B(i) \neq 0,\ Q(i) > 0\ \forall\, i \in \mathcal{S}$. *Then:*
(i) The coupled set of equations (29) admits a unique positive solution, P^*, *which can be generated through the outer-loop algorithm (38).*
(ii) There exists a unique stationary optimal policy for Player 1, given by

$$\mu^{1^*}(x,i) = -\left[B(i)P^*(i)/R(i)\right]\, x\,, \quad i \in \mathcal{S}$$

(iii) There is an optimal (minimizing in the team case, and maximizing in the game case) stationary policy for Player 2, generated by the outer-loop algorithm (38).

Proof. The result follows from Theorem 3.4(ii) and the derivation given prior to the statement of the Theorem. Perhaps the only additional point that should be touched upon is that condition (32) is necessarily satisfied here since for $\mu^2 = f^*$ the problem involves a standard quadratic optimization for jump linear systems with a constant rate matrix, and the positivity of $Q(i)$'s and the nonsingularity of $B(i)$'s guarantee the cited property [6]. ∎

4 Additional Algorithms and Perspectives

4.1 Mathematical programming

We show here that the solution of (29) (whenever it exists and is smooth – which is certainly guaranteed for the scalar case) can be obtained as the value of a (finite) quadratic program. As a byproduct, we obtain an alternative proof of existence and uniqueness for the scalar case. First we introduce some useful definitions:

Definition 4.1 *A function* $\phi(\cdot,i) : I\!R^n \to \mathrm{I\!R}$, $\phi(\cdot,i) \in \mathcal{C}$, *is a* **superharmonic function** *if it satisfies the following inequality in the viscosity sense, for all* $x \in I\!R^n$, $i \in \mathcal{S}$, *and* $u^1 \in \mathbf{U}_1$ *and*

$a \in \mathbf{U}_2(i)$:

$$\beta\phi(x,i) \leq [D_x\phi(x,i)]\ f(x,i,u^1) + \sum_{j\in\mathcal{S}} \lambda_{iaj}\ \phi(x,j) + L(x,i,u^1)\,. \tag{48}$$

Definition 4.2 *A function $\phi(\cdot,i) : I\!R^n \to \mathbb{R}$, $\phi(\cdot,i) \in \mathcal{C}$, is a* **subharmonic function** *if it satisfies the following inequality in the viscosity sense, for all $x \in I\!R^n$, $i \in \mathcal{S}$, $a \in \mathbf{U}_2(i)$, and for some stationary control policy $\mu^1 \in \mathcal{U}_1$ (for Player 1):*

$$\beta\phi(x,i) \geq [D_x\phi(x,i)]\ f(x,i,u^1) + \sum_{j\in\mathcal{S}} \lambda_{iaj}\ \phi(x,j) + L(x,i,\mu^1(x,i))\,. \tag{49}$$

Theorem 4.1 *Let Condition 2.1 be satisfied.*
(i) Consider the infinite-horizon team problem, and assume that it admits an optimal stationary team policy μ^*. Then, its value function $\hat{J}_\beta(x,i)$ is the largest superharmonic function (componentwise) among functions $\phi \in \mathcal{C}$ that satisfy the additional condition:

$$\lim_{t_1\to\infty} e^{-\beta t_1}\, \mathbf{E}^{\mu^*}_{x,i}\phi(x(t_1),\theta(t_1)) = 0\,. \tag{50}$$

(ii) If the value function $\hat{J}_\beta(x,i)$ of the zero-sum game exists and is $\mathcal{C}$ in x, then it is the smallest nonnegative subharmonic function among those that correspond to finite-cost[5] policies of Player 1.

Proof. (i) Consider an arbitrary superharmonic function ϕ satisfying the given condition (50), and integrate both sides of (48) from $t = 0$ to $t = t_1 > 0$. Take the expected value of both sides under μ^*, and conditioned on $(x(0) = x, \theta(0) = i)$, and then apply Dynkin's formula of integration (see [5], p. 146, eq.(9.7)) to get

$$\phi(x,i) \leq \mathbf{E}^{\mu^*}_{x,i} \int_0^{t_1} e^{-\beta t}\, L(x(t),\theta(t),u^1(t))\, dt + e^{-\beta t_1}\, \mathbf{E}^{\mu^*}_{x,i}\phi(x(t_1),\theta(t_1)).$$

Since μ^* is optimal, in view of the condition (50) imposed on ϕ the second term above goes to zero as $t_1 \to \infty$, and hence we arrive at the inequality:

$$\phi(x,i) \leq \mathbf{E}^{\mu^*}_{x,i} \int_0^{\infty} e^{-\beta t}\, L(x(t),\theta(t),u^1(t))dt = \hat{J}_\beta(x,i).$$

Since $\hat{J}_\beta(x,i)$ also satisfies (48), as well as (50) under the given stabilizability and observability conditions,, it follows that it is the largest superharmonic function among those satisfying (50).

[5]A policy $\mu^1 \in \mathcal{U}_1$ is a *finite-cost* policy for Player 1, if under it the infinite-horizon cost is finite for each $\mu^2 \in \mathcal{U}_2$.

(ii) Let $\bar{\mu}^1$ be a fixed stationary finite-cost control policy for Player 1, and consider a corresponding arbitrary subharmonic nonnegative function ϕ. Let $\bar{\mu} := (\bar{\mu}^1, \mu^2)$, where $\mu^2 \in \mathcal{U}_2$ is an arbitrary stationary policy for Player 2. Then, following the first step in the proof of part (i) above, we have the inequality:

$$\begin{aligned} \phi(x,i) &\geq E^{\overline{\mu}}_{x,i} \int_0^{t_1} e^{-\beta t} L(x(t), \theta(t), u^1(t))dt + e^{-\beta t_1} \mathbf{E}^{\overline{\mu}}_{x,i} \phi(x(t_1), \theta(t_1)) \\ &\geq E^{\overline{\mu}}_{x,i} \int_0^{t_1} e^{-\beta t} L(x(t), \theta(t), u^1(t))dt \end{aligned}$$

where the second line has followed because ϕ is nonnegative. Taking the limit as t_1 goes to infinity, we get

$$\phi(x,i) \geq J_\beta(x,i;\bar{\mu}) \geq \hat{J}_\beta(x,i)\,,$$

where the second inequality has followed by picking μ^1 and μ^2 as saddle-point policies. Since $\hat{J}_\beta(x,i)$ is a subharmonic function, the result follows. ∎

Theorem 4.1 enables us to formalize a mathematical program to compute the solution $P(\cdot)$ of (29) in the general dimensional case, provided that it exists:[6]

Theorem 4.2 *Assume that Condition 2.1 holds. Consider equation (29), and assume that it admits a nonnegative-definite solution $P(i)$, $i \in \mathcal{S}$, for both the team and game cases. Fix an initial state (x_0, k).*

(i) For the team problem, $\hat{J}_\beta(x_0,k)$ is given by the value of the following quadratic program:

QP1(k): *Find $P(i) \geq 0$, $i \in \mathcal{S}$, so as to maximize $x_0^T P(k) x_0$ subject to*

$$\begin{aligned} 0 \;\leq\; & -\beta P(i) + A^T(i)P(i) + P(i)A(i) - P(i)B(i)R^{-1}(i)B^T(i)P(i) + Q(i) \\ & + \sum_{j\in\mathcal{S}} \lambda_{iaj} P(j), \qquad \forall i \in \mathcal{S},\; a \in \mathbf{U}_2(i) \end{aligned} \tag{51}$$

(ii) For the game problem, $\hat{J}_\beta(x_0,k)$ is obtained as the value of the following quadratic program:

QP2(k): *Find $P(i) \geq 0$, $i \in \mathcal{S}$, so as to minimize $x_0^T P(k) x_0$ subject to*

$$\begin{aligned} 0 \;\geq\; & -\beta P(i) + A^T(i)P(i) + P(i)A(i) - P(i)B(i)R^{-1}(i)B^T(i)P(i) + Q(i) \\ & + \sum_{j\in\mathcal{S}} \lambda_{iaj} P(j),\; \forall i \in \mathcal{S}, a \in \mathbf{U}_2(i) \end{aligned} \tag{52}$$

Proof. This is a direct consequence of Theorem 4.1, obtained by specializing it to functions ϕ of the form $\phi(x,i) = x^T P(i) x$. More precisely, consider the team problem. Theorem 3.4(ii) implies that

[6] We should note that a unique nonnegative-definite solution exists for each fixed rate matrix, and under the given stochastic stabilizability and observability conditions, as it has already been demonstrated in the literature; see [6, 7].

$\hat{J}_\beta$ has the form $x^T P(i)x$.[7] Theorem 4.1 (i) implies that it is the largest superharmonic function among those which have this form. By the definition of superharmonic functions, this means that $\hat{J}_\beta$ is the largest function among those that satisfy the constraints

$$\beta x^T P(i)x \le 2(A(i)x + B(i)u^1)^T P(i)x + \sum_{j\in\mathcal{S}} \lambda_{iu^2j} x^T P(j)x + x^T Q(i)x + {u^1}^T R(i)u^1$$

for all x, i, u^1, u^2. The above constraint is equivalent to

$$\beta x^T P(i)x \le \inf_{u^1} \left[2(A(i)x + B(i)u^1)P(i) + \sum_{j\in\mathcal{S}} \lambda_{iu^2j} x^T P(j)x + x^T Q(i)x + {u^1}^T R(i)u^1 \right],$$

for all x, i and u^2, which is further equivalent to (by carrying out the minimization in u^1):

$$\beta x^T P(i)x \le x^T \left(A^T(i)P(i) + P(i)A(i) - P(i)B(i)R^{-1}(i)B^T(i)P(i) + Q(i) \right) x + \sum_{j\in\mathcal{S}} \lambda_{iu^2j} x^T P(j)x\,,$$

for all x, i and u^2. This is, finally, equivalent to the constraint (51).

(ii) Consider the class of functions F_1 of the form $\phi(x,i) = x^T P(i)x$. Since the value function is the smallest subharmonic function, it is, in particular, the smallest among F_1. It is moreover, the smallest in the class of policies $F_2 \subset F_1$, defined as follows: $\phi \in F_2$ if (49) holds for all $x \in \mathbb{R}^n$, $i \in \mathcal{S}$, $a \in \mathbf{U}_2(i)$, and for $u^1(x,i)$ given in (33). This is equivalent to the following: The value function is the smallest function of the form $x^T P(i)x$ satisfying, for all x, i and u^2,

$$\beta x^T P(i)x \ge x^T \left(A^T(i)P(i) + P(i)A(i) - P(i)B(i)R^{-1}(i)B^T(i)P(i) + Q(i) \right) x + \sum_{j\in\mathcal{S}} \lambda_{iu^2j} x^T P(j)x\,.$$

This is, finally, equivalent to the constraint (52). ∎

It is appropriate to list here some useful properties of the mathematical program $\mathbf{QP1}(k)$. First note that the feasible region determined by the constraints (51) is nonempty; this follows from the existence hypothesis of the theorem, and also directly from the fact that $P(i) = 0$, $i \in \mathcal{S}$, is feasible. Moreover it is a closed region. Second, if $P^*(i)$ is the optimal solution of $\mathbf{QP1}(i)$, for $i \in \mathcal{S}$, then $P^*(i)$, $i \in \mathcal{S}$, are feasible for $\mathbf{QP1}(k)$ for any $k \in \mathcal{S}$ (which follows from Theorem 4.1). Consequently, if optimal solutions $P^*(i)$ have already been computed for $i \in \mathcal{S}' \subset \mathcal{S}$, then one can substitute these for $P(i)$, $i \in \mathcal{S}'$, in (51), when computing $P^*(j)$ for $j \notin \mathcal{S}'$.

Similar properties can be observed in connection with the mathematical program (52). The feasible region satisfying the constraints in (52) is nonempty by the existence hypothesis of the theorem. Let $P^*(i)$ be optimal solutions of $\mathbf{QP2}(i)$, $i \in \mathcal{S}$. Then $P^*(i)$, $i \in \mathcal{S}$ are feasible for

[7] This is of course true provided that conditions (30)-(31) hold.

QP2(k) for any $k \in \mathcal{S}$ (this follows from Theorem 4.1). Consequently, if optimal solutions $P^*(i)$'s have been already computed for $i \in \mathcal{S}' \subset \mathcal{S}$, then one can readily substitute these for $P(i)$, $i \in \mathcal{S}'$, in (52), when computing $P^*(j)$ for $j \notin \mathcal{S}'$.

4.2 Value iteration

We now restrict our discussion to the scalar case ($n = 1$), for which Theorem 3.5 already presented the complete solution, which also provided an iterative procedure for numerical computation. Here, we present another iterative algorithm to compute the solution, which is based on value iteration:

(A1) Set $P_0(j) := 0$, $j \in \mathcal{S}$. Introduce constants $\overline{\lambda}_i$, $i \in \mathcal{S}$, as in (37):

(A2) Compute $P_\ell(i)$, $i \in \mathcal{S}$, $\ell \geq 1$, iteratively by solving the following set of quadratic equations:

$$\begin{aligned} 0 &= -(\overline{\lambda}_i + \beta)P_{\ell+1}(i) + Q(i) + 2A(i)P_{\ell+1}(i) - [P_{\ell+1}(i)]^2 N(i) \\ &\quad + \mathrm{opt}_{a \in \mathbf{U}_2(i)} \sum_{j \in \mathcal{S}} \lambda_{iaj} P_\ell(j) + \overline{\lambda}_i P_\ell(i)\,, \qquad i \in \mathcal{S}\,. \end{aligned} \tag{53}$$

Theorem 4.3 *Consider either the team or the game problem, for the scalar case, and under the conditions $B(i) \neq 0$, $Q(i) > 0$, $i \in \mathcal{S}$. Positive solutions $P(j), j \in \mathcal{S}$, of (29) can be obtained as the limit of the set of nondecreasing sequences $\{P_\ell(j)\}_{\ell \geq 0}$.*

Proof. The proof uses ideas similar to those employed in the result that led to Theorem 3.5, but some of the details are different. To see that the $P_{(\ell)}(j)$'s are nondecreasing, write (53) as

$$0 = -(\overline{\lambda}_i + \beta)P_{\ell+1}(i) + Q_\ell(i) + 2A(i)P_{\ell+1}(i) - P_{\ell+1}(i)^2 N(i) \tag{54}$$

where

$$Q_\ell(i) := Q(i) + \mathrm{opt}_{a \in \mathbf{U}_2(i)} \sum_{j \in \mathcal{S}} \lambda_{iaj} P_\ell(j) + \overline{\lambda}_i P_\ell(i). \tag{55}$$

Set $Q_0(i) = Q(i)$. It follows from (55) that for any i, $\ell \geq 0$ (compare this with 42):

$$\text{if } \forall j \in \mathcal{S} \;\; P_{\ell+1}(j) \geq P_\ell(j), \;\; \textit{then} \;\; Q_{\ell+1}(i) \geq Q_\ell(i)\,. \tag{56}$$

This is because under **(A1)** coefficients of the $P_\ell(j)$'s in

$$m(a; P_\ell) := \sum_{j \in \mathcal{S}} \lambda_{iaj} P_\ell(j) + \overline{\lambda}_i P_\ell(i) \tag{57}$$

are each nonnegative, for every $a \in \mathbf{U}_2(i)$. Hence, under the hypothesis of (56),

$$m(a; P_{\ell+1}) \geq m(a; P_\ell) \quad \text{for each } a \in \mathbf{U}_2(i)\,, \tag{58}$$

from which (56) follows. Now, (54) is a standard Riccati equation that corresponds to a system that always remains in state i, and where the weighting on the quadratic cost for the state is $Q_\ell(i)$. Therefore, its solution $P_{\ell+1}(i)$ will be increasing in $Q_\ell(i)$. It then follows that

$$\textit{if} \quad Q_{\ell+1}(i) \geq Q_\ell(i), \quad \textit{then} \quad P_{\ell+2}(i) \geq P_{\ell+1}(i). \tag{59}$$

In view of the fact $P_1(i) \geq P_0(i) = 0$, (56)-(59) establish by induction the desired result that the sequences $\{P_\ell(i)\}$ and $\{Q_\ell(i)\}$ are nondecreasing for each $i \in \mathcal{S}$, and therefore have respective limits, with the former satisfying (29). ∎

Next, we present a monotone nonincreasing iterative scheme to compute P for the game problem, based on a feasible solution of the mathematical program. This is an alternative way to establish that the solution P of (29) exists and is finite.

Lemma 4.1 *Fix an arbitrary $k \in \mathcal{S}$. Let P_0 be an arbitrary vector that is feasible in the quadratic program (52) for that k. Define P_ℓ, $\ell = 1, 2, \ldots$ to be generated by the iterations (53). Then $P(j)$, $j \in \mathcal{S}$, are obtained as the limit of the nonincreasing sequences $\{P_\ell(j)\}$.*

Proof. We write (53) as

$$0 = -(\overline{\lambda}_i + \beta)P_{\ell+1}(i) + Q_\ell(i) + 2A(i)P_{\ell+1}(i) - [P_{\ell+1}(i)]^2 N(i) \tag{60}$$

where

$$Q_\ell(i) := Q(i) + \max_{a \in \mathbf{U}_2(i)} \sum_{j \in \mathcal{S}} \lambda_{iaj} P_\ell(j) + \overline{\lambda}_i P_\ell(i). \tag{61}$$

Introduce

$$\Delta(i) := -\beta P_0(i) + Q(i) + 2A(i)P_0(i) - [P_0(i)]^2 N(i) + \max_{v \in \mathbf{U}_2(i)} \sum_{j \in \mathcal{S}} \lambda_{ivj} P_0(j)\,,$$

and let $\tilde{Q}(i) = Q_0(i) + \Delta(i)$. It follows from (52) that $\Delta(i) \leq 0$, hence $\tilde{Q}(i) \leq Q_0(i)$, and that $P_0(i)$ is the solution of the standard Riccati equation

$$0 = -(\overline{\lambda}_i + \beta)P_0(i) + \tilde{Q}(i) + 2A(i)P_0(i) - [P_0(i)]^2 N(i)$$

corresponding to a system that always remains in state i, and where the weighting on the quadratic cost for the state is $\tilde{Q}(i)$.

Since $\tilde{Q}(i) \leq Q_0(i)$, the above discussion leads to the conclusion that $P_1(i) \leq P_0(i)$. This further implies that $Q_1(i) \leq Q_0(i)$, which by the same argument leads to $P_2(i) \leq P_1(i)$. We may thus continue iteratively and conclude that the sequence $P_n(i)$ is monotonically nonincreasing. Since

$P_\ell(i)$'s are positive, they converge to a finite limit P. It is now easy to check that the limit indeed satisfies (29). ∎

In the above, we made use of the monotonicity to establish the convergence of the iterative scheme. It turns out that the iterative scheme would converge (though not necessarily monotonically) when starting from any initial condition P_0. The reason is that any initial condition P_0 can be lower bounded by $\underline{P}_0 = 0$ and upper bounded by some $\overline{P}_0$ that is a feasible vector in the quadratic program (52). It then follows that for any ℓ,

$$\underline{P}_\ell \leq P_\ell \leq \overline{P}_\ell.$$

The convergence of P_ℓ to the solution P of (29) (for $n=1$) now follows from the fact that both $\underline{P}_\ell$ and $\overline{P}_\ell$ converge to P, as we established above.

5 Non-Quadratic Value Functions and Quadratic Bounds

When the continuous part of state is not of dimension one (i.e. $n > 1$), conditions (30) and (31) will, in general, not hold. We discuss in this section some computable bounds on the value function and some heuristics for obtaining sub-optimal policies that are easy to compute and to implement – all for the infinite-horizon case. We assume throughout that the stochastic stabilizability and observability conditions of Theorem 4.1 hold.

We first show that the value is bounded by quadratic functions – both upper and lower bounds for the team problem as well as for the game. Then, we devise a scheme for obtaining an approximation to the non-quadratic value, and a corresponding sequence of suboptimal policies.

5.1 Quadratic bounds

For the team problem, a simple **upper bound** for the value is obtained by restricting Player 2 to policies that are *not functions of the x component of the state*, but are possibly dependent on the initial state x_0, and are furthermore stationary with respect to the state of the Markov process. Denote the class of such policies (for Player 2) by M_2^0, and note that this set is finite. Hence, for any initial state pair (x_0, i_0),

$$\hat{J}_\beta(x_0, i_0) \leq \min_{\mu^2 \in M_2^0} \inf_{\mu^1 \in \mathcal{U}_1} J_\beta(x_0, i_0, \mu) =: \bar{J}_\beta(x_0, i_0).$$

When Player 2 uses such a policy $\mu^2 \in M_2^0$, Player 1 is faced with a linear-quadratic control problem with a constant (control-independent) rate matrix, for which we have (as it has already been discussed in previous sections):

$$\min_{\mu^1 \in \mathcal{U}_1} J_\beta(x_0, i_0, \mu) = x_0^T P(i, \mu^2) x_0 \,, \tag{62}$$

where $P(i,\mu^2)$, $i \in \mathcal{S}$, is the set of unique nonnegative-definite solutions of the following linearly coupled Riccati equations:

$$\begin{aligned}\beta P(i) &= A^T(i)P(i) + P(i)A(i) - P(i)B(i)R^{-1}(i)B^T(i)P(i) \\ &\quad + Q(i) + \sum_{j\in\mathcal{S}} \lambda_{i\mu^2(i)j} P(j),\ i \in \mathcal{S}. \end{aligned} \tag{63}$$

Let $\hat{\mu}^2 = \hat{\mu}^2(i,x_0)$ be a policy for Player 2 that minimizes (62) subject to (63), and define

$$\hat{\mu}^1(x,i;x_0) = -R^{-1}(i)B^T(i)P(i,\hat{\mu}^2)\,x\,,$$

which is the control that achieves the unique minimum in (62) with $\mu^2 = \hat{\mu}^2$. Let $\hat{\mu} := (\hat{\mu}^1, \hat{\mu}^2)$ denote the resulting composite policy. Clearly, this policy achieves the bound $\bar{J}_\beta(x_0,i_0)$.

For the game case, by following a similar procedure, this time a **lower bound** for the value can be obtained, by again restricting the Player 2 to policies in M_2^0. In this case, $\hat{\mu}^2$ will be obtained by maximizing (62) subject to (63).

Next, we obtain a **lower bound** on the value of the team. From Theorem 3.1, it follows that the value $\hat{J}_\beta(x,i)$ is given by $\psi(x,i)$, where the latter satisfies:

$$\beta\psi(x,i) = \min_{u^1}\left[\left(A(i)x + B(i)u^1\right)^T D_x\psi(x,i) + |u^1|^2_{R(i)}\right] + |x|^2_{Q(i)} + \min_a \sum_{j\in\mathcal{S}} \lambda_{iaj}\psi(x,j)\,. \tag{64}$$

Since $\lambda_{iai} < 0$ and $\lambda_{iaj} \geq 0$, for $j \neq 0$, the following inequality readily follows from (64):

$$(\beta + \bar{\lambda}_i)\psi(x,i) \geq \min_{u^1}\left[\left(A(i)x + B(i)u^1\right)^T D_x\psi(x,i) + |u^1|^2_{R(i)}\right] + |x|^2_{Q(i)} + \sum_{j\in\mathcal{S}, j\neq i} (\min_a \lambda_{iaj})\psi(x,j)\,, \tag{65}$$

where

$$\bar{\lambda}_i := \max_{a\in\mathbf{U}_2} |\lambda_{iai}|\,,\quad i \in \mathcal{S}\,,$$

as already defined earlier by (37). Consider now the following equality version of (65):

$$\begin{aligned}\beta\tilde{\psi}(x,i) &= \min_{u^1}\left[\left(A(i)x + B(i)u^1\right)^T D_x\tilde{\psi}(x,i) + |u^1|^2_{R(i)}\right] + |x|^2_{Q(i)} \\ &\quad + \sum_{j\in\mathcal{S}, j\neq i} (\min_a \lambda_{iaj})\tilde{\psi}(x,j) - \bar{\lambda}_i\tilde{\psi}(x,i). \end{aligned} \tag{66}$$

This admits a nonnegative quadratic solution, which lower-bounds $\hat{J}_\beta(x,i)$ – as we next show. First note that (66) is similar to the HJB equation of a jump linear-quadratic control problem with a constant rate matrix, with the only difference being that the rates do not necessarily add up to *zero*, i.e.

$$\sum_{j\in\mathcal{S},j\neq i}\min_a \lambda_{iaj} - \bar{\lambda}_i \neq 0$$

But this can be fixed by adding to both sides of (66) the quantity

$$\left(\bar{\lambda}_i - \sum_{j\in\mathcal{S},j\neq i}\min_a \lambda_{iaj}\right)\tilde{\psi}(x,i)$$

This then makes (66) precisely the HJB equation associated with a legitimate jump linear-quadratic control problem, with a new discount factor

$$\beta + \bar{\lambda}_i - \sum_{j\in\mathcal{S},j\neq i}\min_a \lambda_{iaj}$$

which is indeed nonnegative. Hence, using the standard theory (under the given stochastic stabilizability and observability conditions), one can conclude that (66) admits a quadratic solution of the form $\psi'(x,i) = x^T P'(i)x$, where P' is the unique nonnegative-definite solution set to the following linearly coupled Riccati equations:

$$\begin{aligned}\beta P'(i) &= A^T(i)P'(i) + P'(i)A(i) - P'(i)B(i)R^{-1}(i)B^T(i)P'(i)\\ &\quad + Q(i) + \sum_{j\in\mathcal{S},j\neq i}\left(\min_a \lambda_{iaj}\right)P'(j) - \bar{\lambda}_i\,P'(i),\ i\in\mathcal{S}.\end{aligned} \tag{67}$$

We now show that this quadratic function $\psi'(x,i)$ lower-bounds $\hat{J}_\beta(x,i)$ for all x and i. Toward this end we first observe that, for each fixed $i\in\mathcal{S}$, (66) can also be viewed as the HJB equation for a deterministic linear-quadratic control problem whose system dynamics are given by (1) permanently locked to state i, whose immediate cost is $L_{\tilde{\psi}}(x,i,u)$, where

$$L_{\tilde{\psi}}(x,i,u) := |u^1|^2_{R(i)} + |x|^2_{Q(i)} + \sum_{j\in\mathcal{S},j\neq i}(\min_a \lambda_{iaj})\,\tilde{\psi}(x,j)\,, \tag{68}$$

and for which the discount factor is $\beta+\bar{\lambda}_i$. Let $\psi^0 := \hat{J}_\beta$, and define $\psi^\ell\in\mathcal{C}$, $\ell = 1,2,\ldots$ recursively to be the unique nonnegative viscosity solutions of the following sequence of HJB equations:

$$\begin{aligned}(\beta+\bar{\lambda}_i)\psi^\ell(x,i) &= \min_{u^1}\left[\left(A(i)x + B(i)u^1\right)^T D_x\psi^\ell(x,i) + |u^1|^2_{R(i)}\right] + |x|^2_{Q(i)}\\ &\quad + \sum_{j\in\mathcal{S},j\neq i}(\min_a \lambda_{iaj})\psi^{\ell-1}(x,j)\,,\quad i\in\mathcal{S}\,.\end{aligned} \tag{69}$$

We note that an inductive argument would show that such solutions necessarily exist, since the instantaneous cost is continuous in each case.[8] We further observe that (64), of which $\hat{J}_\beta$ is the solution, can be viewed as the HJB equation of the same system as above, i.e. a system that is always in state i, with the same discount factor, but with an immediate cost of

$$L(x,i,u) = |u^1|^2_{R(i)} + |x|^2_{Q(i)} + \min_a \sum_{j\in\mathcal{S}} \lambda_{iaj}\hat{J}_\beta(x,j) + \bar{\lambda}_i \hat{J}_\beta(x,j)\,.$$

Since $0 \leq L_{\psi^0}(x,i,u) \leq L(x,i,u)$, it follows that the corresponding values are ordered accordingly, i.e. $\psi^1 \leq \psi^0 = \hat{J}_\beta$. This implies that $L_{\psi^1}(x,i,u) \leq L_{\psi^0}(x,i,u)$, and thus $\psi^2 \leq \psi^1$. Proceeding in this fashion, it follows by induction that $\{\psi^\ell\}$ is indeed a nonincreasing sequence, which therefore has a limit for each $x \in I\!R^n$ and $i \in \mathcal{S}$. Finally, $\psi' = \lim_{\ell\to\infty} \psi^\ell$ is easily seen to be the (classical) solution of (66).

Remark 5.1 *(i) An alternative way to show that $\psi' \leq \hat{J}_\beta$ is to note that ψ' is a superharmonic function, and then apply Theorem 4.1(i).*
(ii) The bounds that we obtained above are based on transforming an LQ (linear-quadratic) problem with controlled jumps into one with control-free jumps. In this transformation we did not make use of the LQ properties, so that the bounds are valid for other dynamics and cost structures as well (provided that both f and L are continuous, L is also nonnegative, and the infinite-horizon discounted cost is finite for some stationary controller).

For the game case we can obtain an **upper bound** ψ' on the value by employing the above approach, where ψ' is now the solution of

$$\begin{aligned}(\beta + \min_a |\lambda_{iai}|)\psi(x,i) \quad &= \quad \min_{u^1}\left[\left(A(i)x + B(i)u^1\right)^T D_x\psi(x,i) + |u^1|^2_{R(i)}\right] + |x|^2_{Q(i)} \\ &\quad + \sum_{j\in\mathcal{S}, j\neq i} (\max_a \lambda_{iaj})\psi(x,j). \qquad (70)\end{aligned}$$

5.2 A dynamic approximation, and suboptimal team policies

We now describe a procedure for obtaining an approximation to the value function of the n-dimensional team problem, which involves solutions to a sequence of team problems of the type that were using in arriving at the upper bound $\bar{J}_\beta$ in the previous subsection. Consider a partition of the time interval into finite subintervals, determined by the increasing sequence of natural numbers $0 < t_1 < t_2 < t_3 < \ldots$. First solve the original team problem by restricting Player 2 to policies

[8] Note that the term that starts the recursion, $\hat{J}_\beta$, is continuous in x, as per Theorem 3.1.

that are open-loop in x and stationary closed-loop in θ. Such policies were already introduced in the previous subsection, in connection with the derivation of an upper bound for the team problem. The class of such policies were denoted by M_2^0, and the team optimal controllers for the two players (which necessarily exist) were denoted by $\hat{\mu}$. Here we add a subscript to this notation, i.e. $\hat{\mu}_0$, where the subscript '0' signifies that these policies were derived by fixing the initial value of x at time $t = 0$ at x_0. Let us recall that the derivation of these policies followed a fairly simple construction, as it involved the solution of a finite optimization problem, subject to constraints imposed by s linearly coupled Riccati equations. For each $j = 1, 2, \ldots$, let us now introduce the notation $\hat{\mu}_j$ to denote a similar team-optimal policy pair, obtained over the interval $[t_j, \infty)$, with the initial state now being $x = x(t_j) =: x_j$. We denote the class to which the corresponding policy for Player 2 (i.e. $\hat{\mu}_j^2$) belongs by M_2^j. Note that elements of M_2^j are restricted to the interval $[t_j, \infty)$.

We now construct a sequence of policies (with improving performance) by way of concatenation as follows:

Consider the policy μ_j^* that uses $\hat{\mu}_0$ on $[0, t_1)$, $\hat{\mu}_1$ on $[t_1, t_2)$, $\ldots$, $\hat{\mu}_{j-1}$ on $[t_{j-1}, t_j)$, and $\hat{\mu}_j$ on $[t_j, \infty)$. Then, we have the following:

Lemma 5.1 *The sequence of corresponding costs,*

$$\{\hat{J}_\beta(x, i; \mu_j^*)\}_{j \geq 0} \,,$$

is nonincreasing.

Proof. For any $j = 1, 2, \ldots$,

$$\begin{aligned}
\hat{J}_\beta(x, i; \mu_j^*) &= E_{x,i}^{\mu_j^*} \int_0^{t_{j-1}} e^{-\beta t} L(x(t), \theta(t), u^1(t)dt + E_{x,i}^{\mu_j^*} \int_{t_{j-1}}^{t_j} e^{-\beta t} L(x(t), \theta(t), u^1(t)dt \\
&\quad + E_{x,i}^{\mu_j^*} \int_{t_j}^{\infty} e^{-\beta t} L(x(t), \theta(t), u^1(t)dt \\
&= E_{x,i}^{\mu_{j-1}^*} \int_0^{t_{j-1}} e^{-\beta t} L(x(t), \theta(t), u^1(t)dt + E_{x,i}^{\mu_j^*} \int_{t_{j-1}}^{t_j} e^{-\beta t} L(x(t), \theta(t), u^1(t)dt \\
&\quad + \min_{\mu | x = x_j} E_{x,i}^{\mu} \int_{t_j}^{\infty} e^{-\beta t} L(x(t), \theta(t), u^1(t)dt \\
&\leq E_{x,i}^{\mu_{j-1}^*} \int_0^{t_{j-1}} e^{-\beta t} L(x(t), \theta(t), u^1(t)dt + E_{x,i}^{\mu_{j-1}^*} \int_{t_{j-1}}^{\infty} e^{-\beta t} L(x(t), \theta(t), u^1(t)dt \\
&= \hat{J}_\beta(x, i; \mu_{j-1}^*).
\end{aligned}$$

A few words are now in order for the steps used above, and one notation introduced: The breaking of the interval $[0, \infty)$ into three subintervals in step one is consistent with the construction of policy

μ_j^*. The first two expressions have been carried in tact to the second step, with the expectation now taken under the policy μ_{j-1}^*, since it agrees with μ_j^* on the subinterval $[0, t_j)$ as per the construction. The last expression in the second step stands for minimization of the cost over the interval $[t_j, \infty)$, over $\mu_1 \in \mathcal{U}_1$ restricted to the same interval and $\mu^2 \in M_2^j$ – with the conditioning with respect to the observed value x_j of $x(t_j)$ as well as that of the Markov chain, and then the entire expression averaged with respect to the probability measure corresponding to μ_j^* over the subinterval $[0, t_j)$. This expression clearly bounds from below the same for an arbitrary control picked over the interval $[t_j, \infty)$, and in particular μ_{j-1}^* – which leads to the inequality and the second expression of the third step. The fourth step simply follows from the definition of μ_j^*. ∎

The significance of the result given above, and the accompanying recursive construction, is that one can obtain a constantly improving sequence of approximations to a possibly non-quadratic value function, which is also nondifferentiable, in terms of piecewise quadratic value functions, whose construction is fairly easy. As a byproduct, we also obtain a sequence of easily computable policies for the team, which improve with increasing j. The limiting properties of this sequence of policies, and *ditto* for the value functions still need to be studied.

6 An Illustrative Example

To motivate the class of hybrid problems studied in this paper, and to illustrate the theory for the scalar case, consider the following communication network, which can be viewed as a modified version of the models studied recently in [1] and [2], with a new component. It is assumed that the network has linearized dynamics (for the control of queue length), and all performance measures (such as throughput, delays, loss probabilities, etc.) are determined essentially by a bottleneck node. Both these assumptions have theoretical as well as experimental justifications; see, [1].

Let q_t denote the queue length at a bottleneck link, and r_t denote the effective service rate available for traffic of the given source at that link at time t. We let r_t be arbitrary, but assume that the controllers have perfect measurements of it. Let ξ_t denote the (controlled) source rate at time t. $u_t^1 := \xi_t - r_t$ is its shifted version, it is determined by controller 1, called the flow controller. Consider the following linearized dynamics for the queue length:

$$\frac{dq}{dt} = u^1 , \tag{71}$$

which is called *linearized* because the end-point effects have been ignored. The objectives of the flow controller are *(i)* to ensure that the bottleneck queue size stays around some desired level $\bar{Q}$, and *(ii)* to achieve good tracking between input and output rates. In particular, the choice of $\bar{Q}$ and the variability around it have direct impact on loss probabilities and throughput. We therefore define a shifted version of q:

$$x_t := q_t - \bar{Q},$$

in view of which (71) now becomes

$$\frac{dx}{dt} = u^1. \tag{72}$$

An appropriate local cost function that is compatible with the objectives stated above would be the one that penalizes variations in x_t and u_t^1 around *zero* — a candidate for which is the weighted quadratic cost function.

Suppose that there are several types (say s) of possible traffic, with different kinds of requirements on the performance measures. Associated with type i traffic are the positive constants $Q(i)$ and $R(i)$ appearing in the immediate cost: $L(x,i,u) = |x|^2_{Q(i)} + |u^1|^2_{R(i)}$. Typically, traffic requiring higher quality of service (QoS) might have a larger $Q(i)$, which reflects the fact that it might require lower loss probabilities and higher throughput. It could be receiving a higher priority from the network in the sense that larger variations in u^1 will be tolerated so as to achieve the required QoS; thus the corresponding $R(i)$ might be smaller. The occurrence of these different types of traffic will be governed by a continuous-time Markov jump process, with transitions controlled by a second controller with a finite action set.

A typical admission control problem is the following: by default the system always accepts traffic of some given type, say 1. Traffic of type 1 is transmitted until a session consisting of another type of traffic of higher priority, say type i ($i = 2, 3, \ldots, s$), is accepted. When it is accepted, the session cannot be interrupted until it ends. Thus at states $\theta = i > 1$, there are no (admission) control actions available.

The controller 2 is thus effective only at state 1, at which the rate of arrival of sessions of high priority traffic is to be determined. To each traffic type there corresponds two admission decisions (that are part of the action to be chosen by controller 2): 0 - corresponding to low admission rate $\underline{\lambda}(j)$, and 1- corresponding to a high admission rate $\bar{\lambda}(j)$. The control action at state 1 is thus of the form $u^2 = (u_1^2, \ldots, u_s^2)$, $u_j^2 \in \{0,1\}$. The controlled transition rates have the form

$$\lambda_{1,u^2,j} \quad = \quad \underline{\lambda}(j) 1_{\{u_j^2=0\}} + \bar{\lambda}(j) 1_{\{u_j^2=1\}},$$

where $1_{\{\cdot\}}$ denotes the set indicator function. Note that, if at state 1 the control action u^2 is fixed, then the next type of session to be accepted will be $j (j > 1)$ with probability

$$[\underline{\lambda}(j) 1_{\{u_j^2=0\}} + \bar{\lambda}(j) 1_{\{u_j^2=1\}}] / \sum_{k=2}^{s} [\underline{\lambda}(k) 1_{\{u_k^2=0\}} + \bar{\lambda}(k) 1_{\{u_k^2=1\}}].$$

The problem then is to minimize the expected discounted or total expected cost (with instantaneous cost being L above) with respect to the multi-strategy $\mu := (\mu^1, \mu^2)$, where μ^1 is the flow controller and μ^2 the admission controller, both having as arguments the current and past values of ξ, r, and θ.

We now apply the results presented in the previous sections to the admission-flow control problem. We simply let $A(i) = 0$ and $B(i) = 1$, for all $i \in \mathcal{S}$. To obtain some explicit results in this context, we took $\beta = 0$ and $s = 3$ (i.e., $\mathcal{S} = \{1, 2, 3\}$). We further took $U_2(1) = \{00, 01, 11, 10\}$, $U_2(2) = \{0\}$, $U_2(3) = \{0\}$; $\lambda_{1002} = \lambda_{1012} = 1$, $\lambda_{1003} = \lambda_{1103} = 2$, $\lambda_{1102} = \lambda_{1112} = 10$, $\lambda_{1013} = \lambda_{1113} = 20$; $Q(1) = 1$, $Q(2) = 10$, $Q(3) = 20$; $R(1) = 10$, $R(2) = 2$, $R(3) = 1$. For each possible choice of μ^2, we computed the corresponding optimal policy for controller 1 (i.e., μ^1), and the associated optimal cost function. By comparing the optimal value functions obtained for all possible (four) admission policies, we found the unique optimal controller 2 to be $\mu^{2*}(1) = 00$, and the solution to (29) to be: $P^*(1) = 4.141$, $P^*(2) = 4.439$, $P^*(3) = 4.352$. The unique optimal flow controller is then

$$\mu^{1*}(q, i) = \begin{cases} -0.4141\, q & i = 1 \\ -2.220\, q & i = 2 \\ -4.352\, q & i = 3 \end{cases}$$

Typical system responses under all four admission control policies are depicted in Figures 1–4. In each of the figures, we have plotted the time history of the queue size, traffic type, the flow control and the integral cost incurred. For illustration purposes, we have taken the initial queue size to be 30 units, which is 10 units larger than the desired queue length of 20 units. Since the high priority traffic types (2 and 3) have stringent QoS specifications, the corresponding flow control is more aggressive to maintain the queue size around its desired value. On the other hand, for the low priority type (1), the control is smoother in order to minimize the jitter (i.e., variability) in the network. These observations are consistent with what we would have expected from the design. We further observe that by admitting the high priority traffic at a higher rate, the queue size reaches its desired value more quickly, at the expense of a larger control jitter. Although they only represent one sample path of the stochastic system in each of the four cases, the simulations corroborate the theory very well. The smallest integral cost is achieved under $\mu^2(1) = 00$, which is actually the optimal admission rule for the total (expected) value of the cost function.

We next considered the case when $R(1)$ was increased to $R(1) = 100$, with all parameter values remaining the same as above. In this case, the optimal admission controller turned out to be $\mu^{2*}(1) = 11$, and the solution to (29) was: $P^*(1) = 4.508$, $P^*(2) = 4.476$, $P^*(3) = 4.485$. The unique optimal flow controller was then

$$\mu^{1*}(q, i) = \begin{cases} -0.0451\, q & i = 1 \\ -2.238\, q & i = 2 \\ -4.485\, q & i = 3 \end{cases}$$

Typical system responses for this case are depicted in Figures 5–8. Qualitative behaviors similar to those in the previous case are also observed here. Because of the increased weighting of $R(1)$, the flow control magnitude for the type 1 traffic is reduced significantly. The integral costs incurred under the four admission control laws are ranked differently from the previous case. Coincidentally, the cost incurred under the optimal admission control $\mu^{2*}(1) = 11$ is the smallest among the four particular sample paths simulated.

Finally, while holding $R(1)$ at 100, we increased $Q(3)$, from 20 to $Q(3) = 30$. With this change, the optimal controller 2 turned out to be $\mu^{2*}(1) = 10$, and the solution to (29) was: $P^*(1) = 4.680$, $P^*(2) = 4.493$, $P^*(3) = 5.223$. The unique optimal flow controller was then

$$\mu^{1*}(q,i) = \begin{cases} -0.0468\,q & i=1 \\ -2.246\,q & i=2 \\ -5.223\,q & i=3 \end{cases}$$

Due to space limitations, we have included here only the figure corresponding to the optimal μ^2 (see Figure 9). In all four simulations we have observed qualitative behavior similar to the previous two cases. We have seen a significant increase in the flow control magnitude for type 3 traffic; this is due to the increased weighting of $Q(3)$, which corresponds to a more stringent QoS for type 3 traffic than in the previous cases. Again coincidentally, the cost incurred under the optimal admission control $\mu^{2*}(1) = 10$ turned out to be the smallest among the four particular sample paths simulated.

7 Conclusions

Several extensions of the results of this paper can be envisioned, both for the general theoretical model and for the special telecommunication network application. For the former, one question that this paper has left unanswered is the structure of the solution to the HJB equation (9), or its infinite-horizon version (10), when the minimization over u^2 in (23) depends on the state x. There is also the issue of developing computational tools for the solutions of (9) and (10) when general structural information is lacking. One extension of the general model of this paper would be the inclusion of an additional additive term in (1), which would represent an unknown disturbance – modeled either as a stochastic process with known statistics (such as a Brownian motion process) or as a completely unknown deterministic process (as in H^∞ control [4]). In the latter case, one chooses as performance index the ratio of J_β introduced here, to the energy of the unknown deterministic process, whose maximum over the unknown input will now have to be minimized over the multi-strategy μ, in the team problem (over the $\mathcal{U}_1$ in the game problem). For discussions on the solution to this problem for the special case when the transition rate matrix is not controlled, but under various types of measurements for the controller (including the noise-perturbed measurement scheme not covered in this paper), see [7].

As far as the specific telecommunication network model of Section 6 is concerned, there is the potential to extend it to the more general case where the effective service rate r_t is not known, but is measured in some additive noise. Furthermore, r_t could be generated by a stochastic ARMA process, or by a deterministic linear model driven by an unknown deterministic process with finite energy. Such models have been considered before in [2], but for a single type traffic (i.e., with $s = 1$), and the extensions to the cases where there are multiple types of traffic (as in this paper) with fixed or controlled transition rates remain today as interesting but challenging research topics to pursue.

References

[1] E. Altman, F. Baccelli, and J.-C. Bolot, "Discrete-time analysis of adaptive rate control mechanisms," in H. G. Perros and Y. Viniotis, editors, *High Speed Networks and their Performance*, pp. 121–140, North Holland, 1994.

[2] E. Altman and T. Başar, "Optimal rate control for high speed telecommunication networks," *Proc. 34th IEEE Conference on Decision and Control*, pp. 1389–1394, New Orleans, LA, December 13-15, 1995.

[3] T. Başar, "Minimax control of switching systems under sampling," *Systems & Control Letters*, 25(5): 315–325, August 1995.

[4] T. Başar and P. Bernhard, *H^∞-Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach*, 2nd edition, Birkhäuser, 1995

[5] W. H. Fleming and H. M. Soner, *Controlled Markov Processes and Viscosity Solutions*, vol. 25, Springer-Verlag, New York, 1993.

[6] Y. Ji and H.J. Chizeck, "Controllability, stabilizability, and continuous-time Markovian jump linear quadratic control," *IEEE Transactions on Automatic Control*, vol. AC-35, pp. 777-788, July 1990.

[7] Z. Pan and T. Başar, "H^∞-control of Markovian jump systems and solutions to associated piecewise-deterministic differential games," in G.J. Olsder, editor, *Annals of Dynamic Games*, vol. 3, pp. 61-94, Birkhäuser, 1995.

[8] R. J. Elliott, *Viscosity Solution and Optimal Control*, "Longman Scientific & Technical, New York, N.Y., 1987

[9] A. Friedman, "The Cauchy Problem for first order differential equations," *Indiana University Journal of Mathematics*, vol. 23, pp. 27–40, 1973.

[10] N. N. Krassovskii and E. A. Lidskii, "Analytical design of controllers in systems with random attributes I, II, III," *Automation Remote Contr.*, vol. 22, pp. 1021–1025, 1141–1146, 1289–1294, 1961.

[11] E. A. Lidskii, "Optimal control of systems with random properties," *Appl. Math. Mech.*, vol. 27, pp. 33–45, 1963.

[12] J. J. Florentin, "Optimal control of continuous-time Markov, stochastic systems," *J. Electron. Control*, vol. 10, 1961.

[13] H. J. Kushner, "On the stochastic maximum principle: Fixed time of control," *J. Math. Anal. Appl.*, vol. 11, pp. 78–92, 1965.

[14] R. Rishel, "Dynamic programming and minimum principles for systems with jump Markov disturbances," *SIAM Journal on Control and Optimization*, vol. 13, pp. 338–371, February 1975.

[15] D. D. Sworder, "Feedback control of a class of linear systems with jump parameters," *IEEE Trans. Automat. Control*, vol. 14, pp. 9–14, 1969.

[16] W. M. Wonham, "Random differential equations in control theory," in *Probabilistic Methods in Applied Mathematics* (A. T. Bharucha-Reid, ed.), pp. 191–199, New York, NY: Academic Press, 1970.

[17] M. Mariton, "On controllability of linear systems with stochastic jump parameters," *IEEE Trans. Automat. Control*, vol. AC-31, pp. 680–683, 1986.

[18] T. Başar and A. Haurie, *Feedback equilibria in differential games with structural and modal uncertainties*, vol. 1 of *Advances in Large Scale Systems*, pp. 163–201. Connecticut: JAI Press Inc., May 1984. (J. B. Cruz, Jr., ed.).

[19] C. E. de Souza and M. Fragoso, "H^∞ control of linear systems with Markovian jumping parameters," *Control Theory and Technology*, vol. 9, no. 2, pp. 457–466, 1993.

[20] Z. Pan and T. Başar, "H^∞ control of large scale jump linear systems via averaging and aggregation," *Proc. 34th IEEE Conference on Decision and Control*, New Orleans, Louisiana, pp. 2574–2579, December (13-15) 1995.

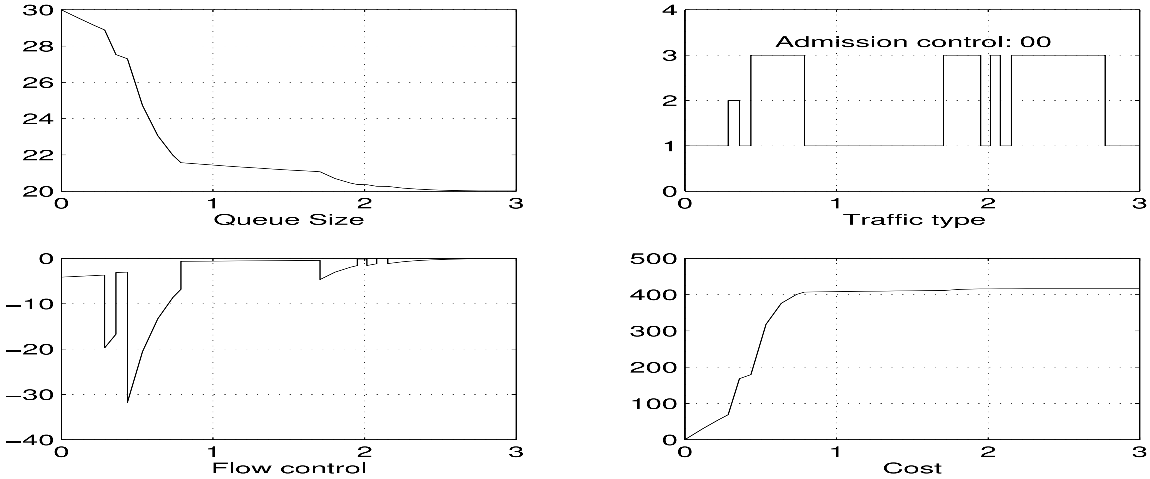

Figure 1: Case 1: Performance under μ^{1*} w.r.t. $\mu^{2*} = 00$.

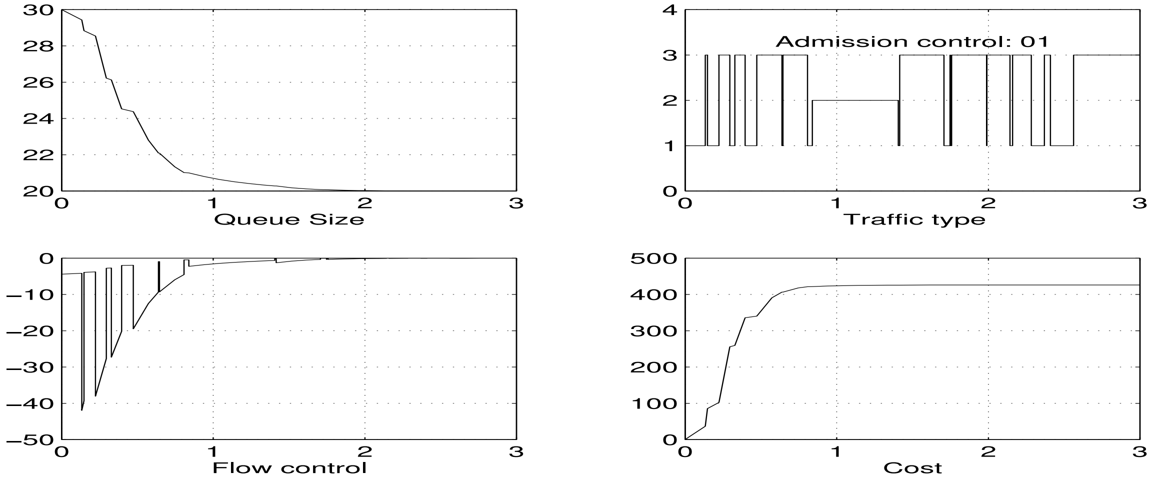

Figure 2: Case 1: Performance under μ^{1*} w.r.t. $\mu^{2*} = 01$.

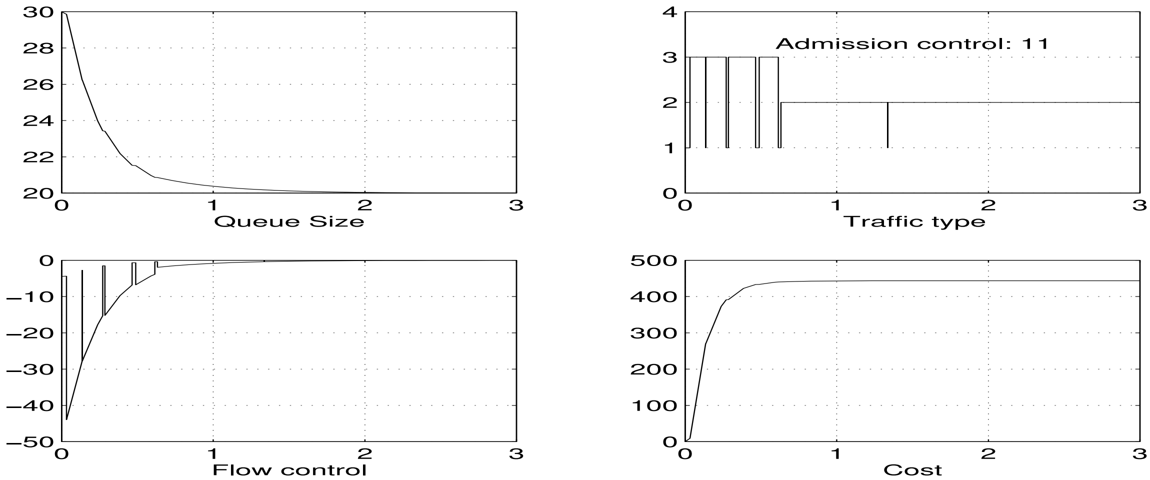

Figure 3: Case 1: Performance under μ^{1*} w.r.t. $\mu^{2*} = 11$.

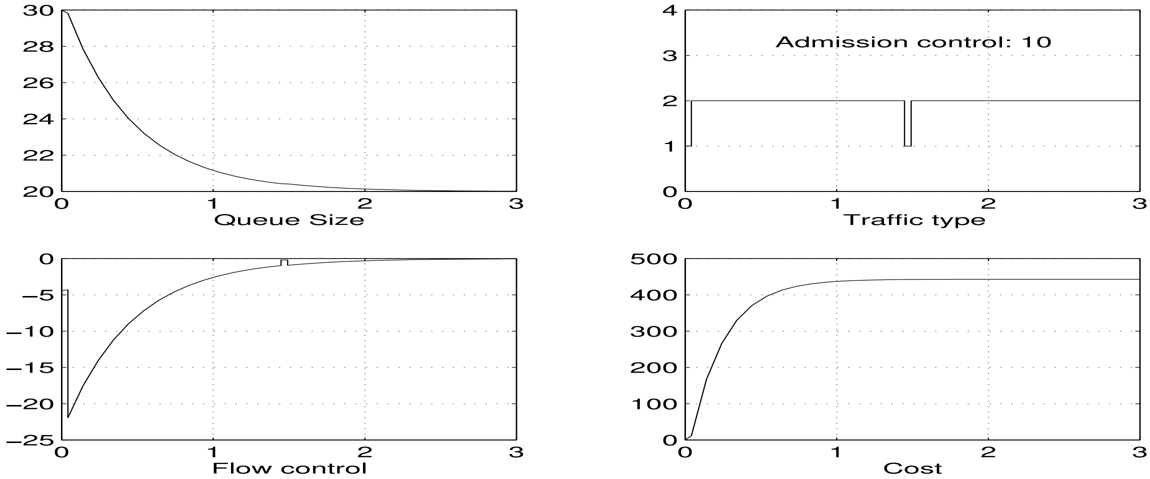

Figure 4: Case 1: Performance under μ^{1*} w.r.t. $\mu^{2*} = 10$.

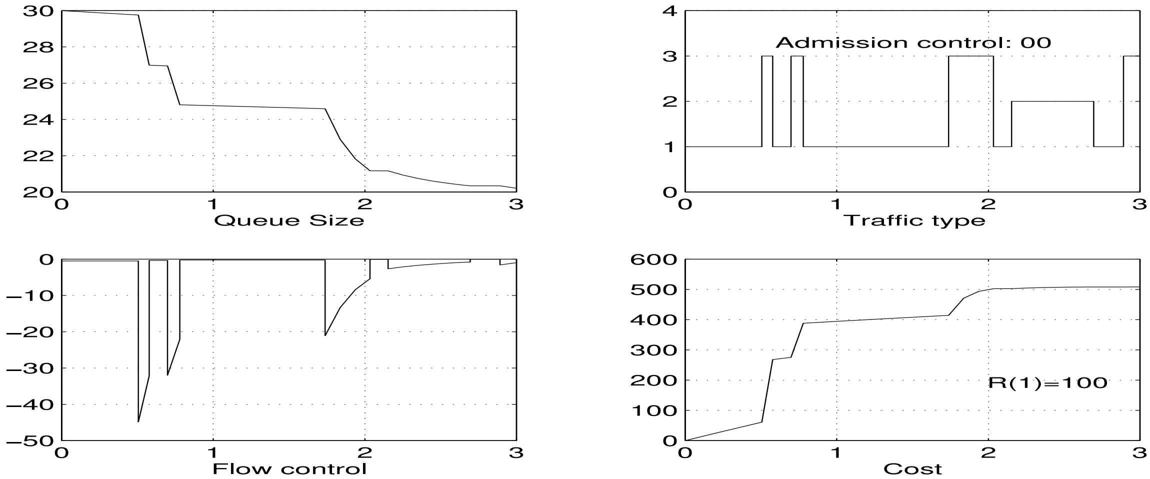

Figure 5: Case 2: Performance under μ^{1*} w.r.t. $\mu^{2*} = 00$.

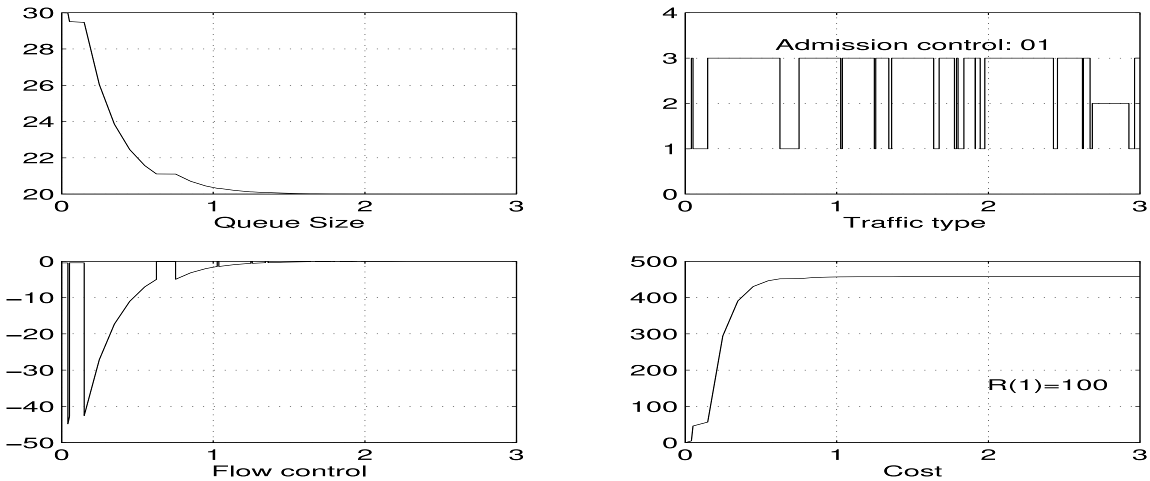

Figure 6: Case 2: Performance under μ^{1*} w.r.t. $\mu^{2*} = 01$.

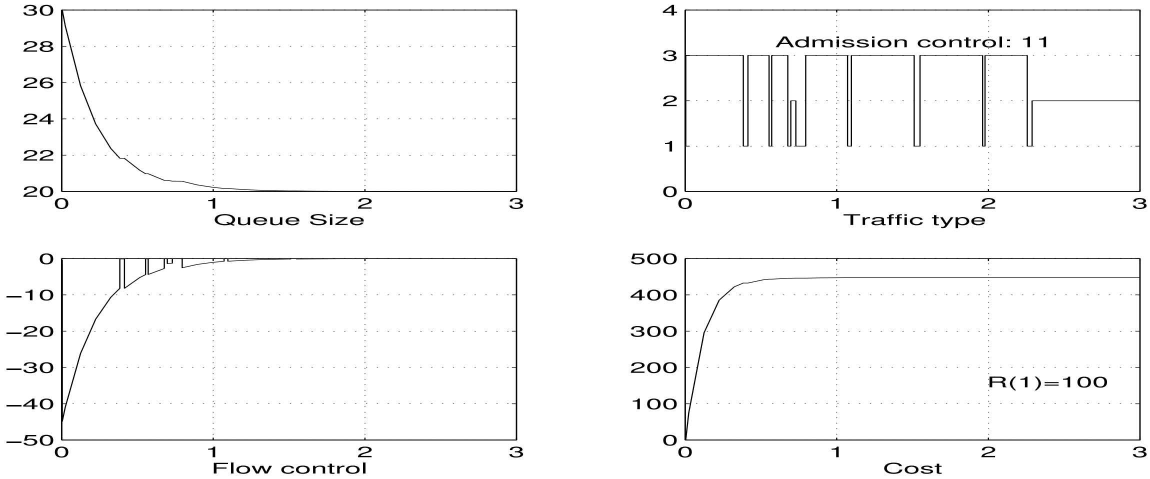

Figure 7: Case 2: Performance under μ^{1*} w.r.t. $\mu^{2*} = 11$.

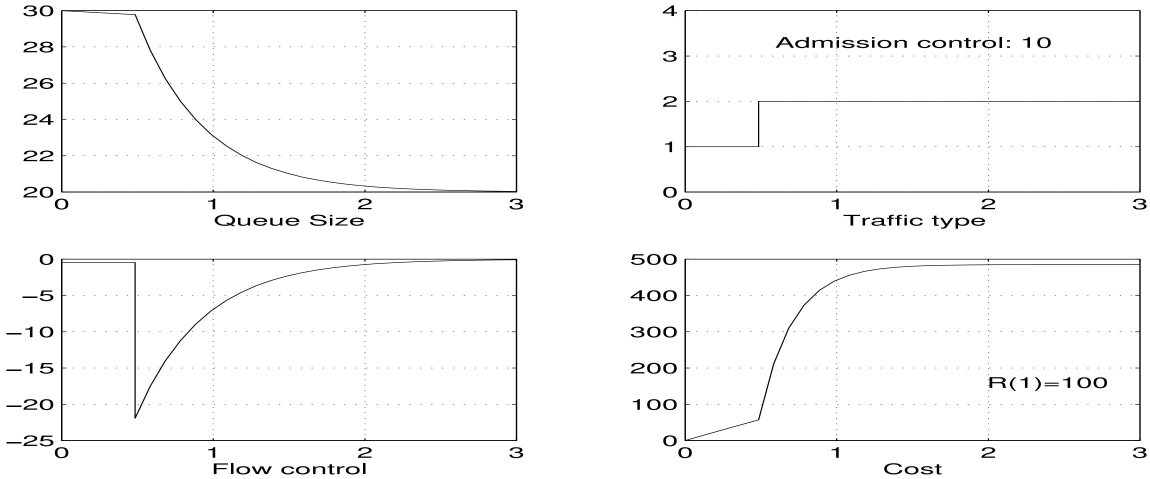

Figure 8: Case 2: Performance under μ^{1*} w.r.t. $\mu^{2*} = 10$.

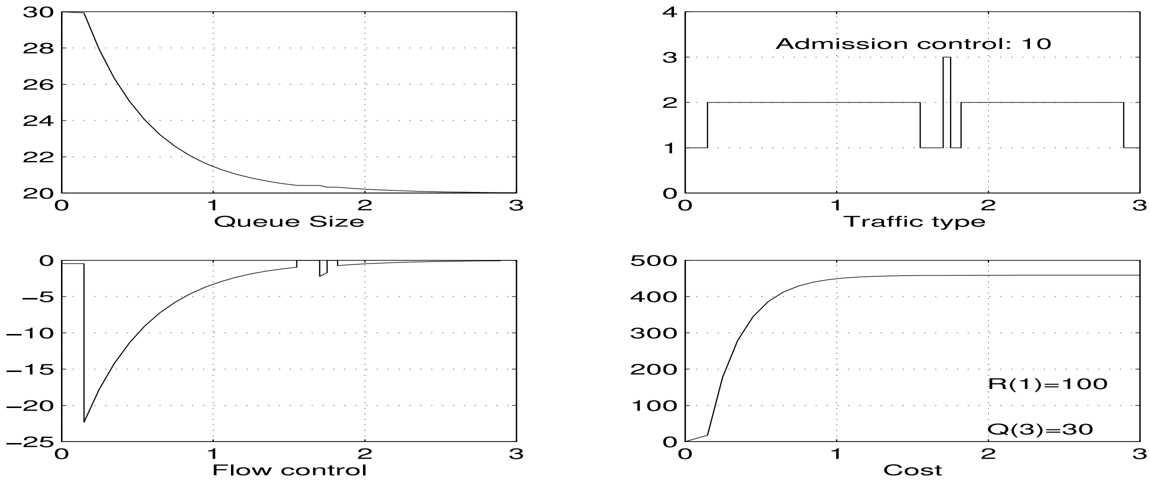

Figure 9: Case 3: Performance under μ^{1*} w.r.t. $\mu^{2*} = 10$.

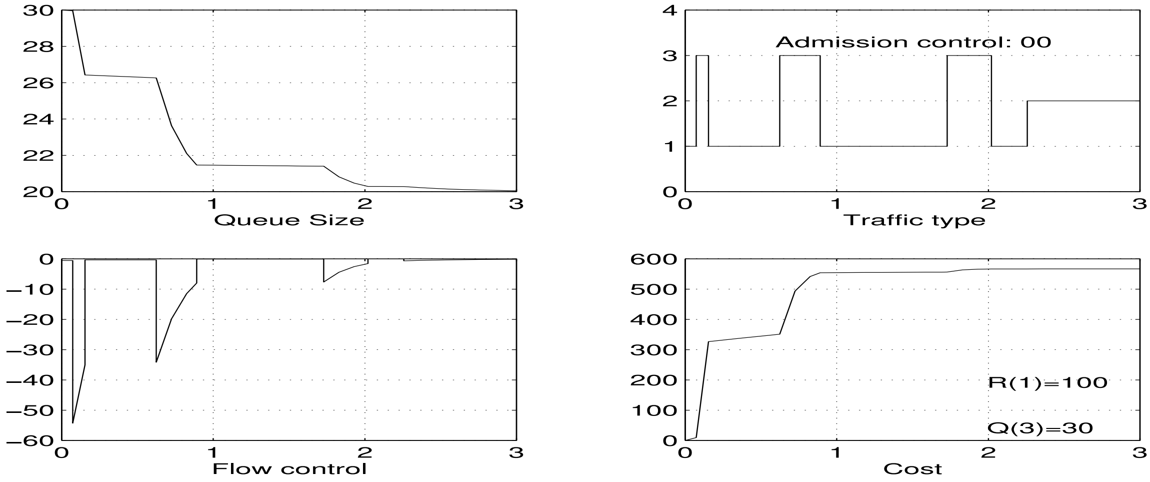

Figure 10: Case 3: Performance under μ^{1*} w.r.t. $\mu^{2*} = 00$.

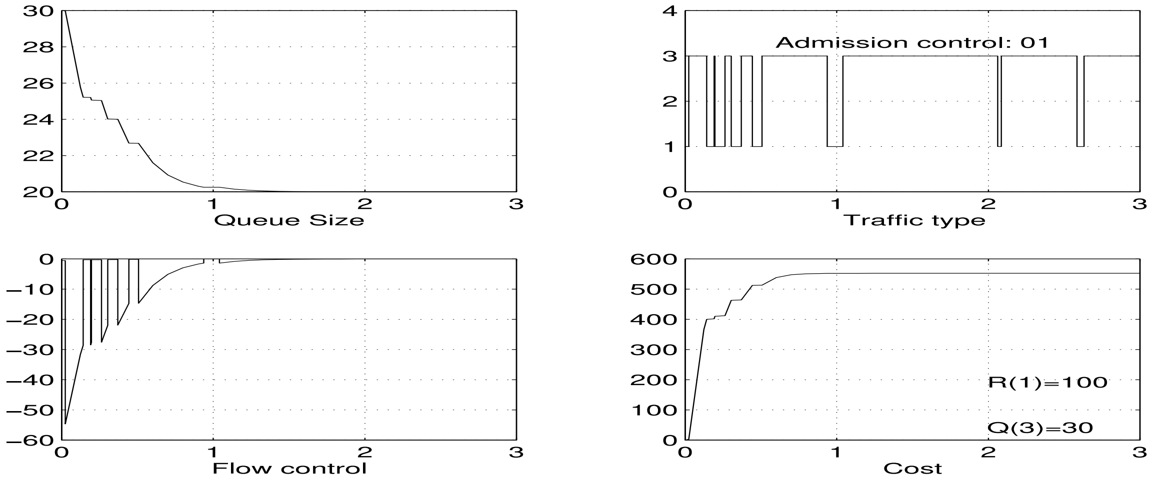

Figure 11: Case 3: Performance under μ^{1*} w.r.t. $\mu^{2*} = 01$.

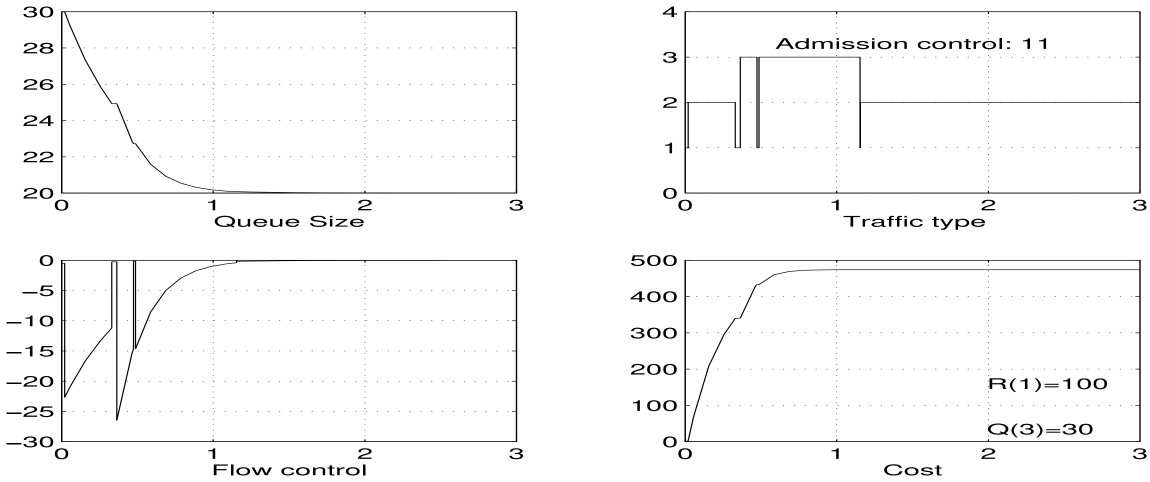

Figure 12: Case 3: Performance under μ^{1*} w.r.t. $\mu^{2*} = 11$.