Convexity of the cost functional in the optimal control problem for a class of positive switched systems ⋆

Patrizio Colaneri [a], Richard H. Middleton [b], Zhiyong Chen [b], Danilo Caporale [a], Franco Blanchini [c],

[a] *DEI, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy*

[b] *School of Electrical and Computer Engineering, University of Newcastle, Callaghan, New South Wales, 2308, Australia*

[c] *Dipartimento di Matematica e Informatica, Università di Udine, 33100 Udine, Italy*

Abstract

This paper deals with the optimal control of a class of positive switched systems. The main feature of this class is that switching alters only the diagonal entries of the dynamic matrix. The control input is represented by the switching signal itself and the optimal control problem is that of minimizing a positive linear combination of the final state variable. First, the switched system is embedded in the class of bilinear systems with control variables living in a simplex, for each time point. The main result is that the cost is convex with respect to the control variables. This ensures that any Pontryagin solution is optimal. Algorithms to find the optimal solution are then presented and an example, taken from a simplified model for HIV mutation mitigation is discussed.

Key words: Optimal control; switched systems;

1 Introduction

This paper is motivated by the recent work [15], where a drug therapy scheduling problem in HIV infection was studied, using simplified switched linear system models of HIV mutation and treatment. An analytic optimal control solution for a particular class of switched systems was provided using necessary conditions based on the Pontryagyn principle. It is well known that a Pontryagin solution does not yield optimality in general. However, optimality can be guaranteed if the cost functional is convex with respect to the control variable.

Optimal control of switched and hybrid systems has been widely studied [7], [23], [9], [10] and the variational approach has been developed by [22], [4] and [17]. However, the authors are unaware of published papers on the optimal control problem for switched positive linear system. In particular, in general, such problems are non-convex.

In this note we prove, for a class of switched positive systems, that the cost functional is indeed convex with respect to the control variables. Thus, for this class of systems, we guarantee that any Pontryagin solution is globally optimal. Such a solution can be found through iterative algorithms that exploit the convexity of the cost. Also, by exploiting the concavity of the cost function with respect to the initial state, a min-max procedure can be implemented to find the worst initial state for the cost function.

Before proceeding we introduce some basic notation. The semiring of nonnegative real numbers is $\mathbb{R}_+$. A square matrix $A = [a_{ij}]$ is said to be *Metzler* if its off-diagonal entries are nonnegative, namely $a_{ij} \geq 0$ for every $i \neq j$. For matrices or vectors, we use $x \gg 0$ (x is strictly positive) to denote that every element of x is positive.

The symbol $\mathbf{1}_n$ denotes the n-dimensional vector with all entries equal to 1. The suffix n will be omitted when the vector size is clear from the context. The unit simplex, that is, the convex polytope of the nonnegative m-tuples,

---

⋆ Patrizio Colaneri is also associated to the institute IEIIT of the Italian National Research Council.

Email addresses: colaneri@elet.polimi.it (Patrizio Colaneri), richard.middleton@newcastle.edu.au (Richard H. Middleton), zhydong.chen@newcastle.edu.au (Zhiyong Chen), caporale@elet.polimi.it (Danilo Caporale), blanchini@uniud.it (Franco Blanchini).

$m \in \mathbb{N}$, that sum to 1 will be denoted by:

$$\mathcal{U} \doteq \left\{ \alpha \in \mathbb{R}^m_+ : \mathbf{1}'\alpha = 1 \right\}.$$

For further details on positive systems the reader is referred to [13].

2 Optimal control problem

Consider the switched positive linear system:

$$\Sigma_A : \dot{x}(t) = A_\sigma x(t), \quad x(0) = x_0 \tag{1}$$

where $\sigma : \mathbb{R}^n_+ \to \mathcal{I} = \{1, \ldots, m\}$ denotes the mode selection, that in general may be a function of $x(t)$, and A_σ is a family of $n \times n$ matrices. The system (1) is positive if the non-negative orthant is positively invariant for any switching signal. Positivity is well known to be equivalent to all matrices A_σ being Metzler.

System (1) may be motivated as a simplified model of treatment of HIV infection dynamics (see for example [15]). In this case, the state x represents the concentrations of various viral mutants in a patient, and σ represents the selection of a suitable therapy. Alternatively, in the widespread SIR (or SI, or SIRS) models of epidemiology over a network, in the initial infection phase (also termed epidemic outbreak, where the concentration of susceptible individuals is approximately constant), the dynamics of infected individuals, are linear (see for example [21,19]).

For simplicity no constraints or penalty terms are imposed on the switching. We consider cost functionals of the following form:

$$J := c' x(t_f) \tag{2}$$

where c is a strictly positive vector, $c \gg 0$ and $t_f > 0$ is fixed.

When dealing with switched systems we can encounter sliding trajectories, i.e. infinite frequency switching of $\sigma(t)$. To include sliding trajectories, we embed the switched system in the larger class described by

$$\dot{x}(t) = A(u(t))x(t) \tag{3}$$

where matrix $A(u)$ is defined as

$$A(u) := \sum_{i=1}^{m} A_i u_i \tag{4}$$

and $u(t) \in \mathcal{U}$, for all $t \in [0, t_f]0$, is the control vector. The passage from system (3) to system (3) has been emphasized in many papers, see e.g. [16] where a different context is however considered.

By construction the system (3) includes the system (1) since $u_i(t) = 1$ (and hence $u_j(t) = 0, j \neq i$) corresponds to $\sigma(t) = i$. If, for some t, $u(t)$ is not a vertex of the simplex, then there is no directly equivalent $\sigma(t)$. Note however, see e.g. [1], that the set of possible trajectories of (1) are dense in the set of trajectories generated by (3). Therefore, extending the concept of valid switching signals to sliding modes based on the appropriate differential inclusions, we consider optimal control of the system (3). For further details of the related *viscous* solutions of differential equations and optimal control of differential inclusions see [2] and [5], respectively. The role of sliding modes (singular control) in optimization problems in terms of finite time convergence to the sliding surface is emphasized in [18].

Remark 1 *Note that the optimal control for system (3)-(4) and cost (2) always exists. Indeed, a sufficient condition for the existence is that the sets of velocities $F(x, u) := \{A(u)x; \quad u \in \mathcal{U}\}$ are convex and that the vector field is bounded by a affine function of the norm of the state variable, i.e. $\|A(u)x\| \le \alpha(1 + \|x\|)$ for some positive scalar α, and for all $x \in \mathbb{R}^n_+$ and $u \in \mathcal{U}$, see e.g. Theorem 5.1.1 in [6]. These conditions are satisfied for our problem and therefore the optimal control exists.*

In the literature on optimal control, a great importance has been given to the analysis of necessary conditions for optimality. In most cases these necessary conditions are the starting point to find the optimal solutions, since direct sufficient conditions (for instance associated with Hamilton-Jacobi-Bellman equations) are often unpractical. For our control problem, necessary conditions can be easily found by writing the Hamiltonian function $H(x, u, \pi) =: \pi' A(u)x$ and using a minor extension of the Pontryagin principle to cope with input-affine form of this function. As a result we now introduce the definition of a Pontragyin solution, namely candidate optimal solutions satisfying the necessary conditions. For further details see e.g. [6].

Definition 1 *A triple $u^o(t) : [0, t_f] \times \mathcal{U}$, $x^o(t)$, $\pi^o(t)$, that satisfies (for almost all t) the system of equations:*

$$\dot{x}^o(t) = \left(\sum_{i=1}^{m} u_i^o(t) A_i \right) x^o(t) \tag{5}$$

$$-\dot{\pi}^o(t) = \left(\sum_{i=1}^{m} u_i^o(t) A_i' \right) \pi^o(t) \tag{6}$$

$$u^o(t) \in \underset{u \in \mathcal{U}}{\operatorname{argmin}} \{ \pi^{o\prime}(t) \left(\sum_{i=1}^{m} u_i A_i \right) x^o(t) \} \tag{7}$$

with the boundary conditions $x^o(0) = x_0$ and $\pi^o(t_f) = c$, is called a Pontryagin solution for the optimal control *problem.*

As noted earlier, in general a Pontryagin solution need not be optimal, since the conditions expressed by Definition 1 are only necessary for optimality. We know that for linear systems and (for instance) quadratic cost, the Pontryagin solution is also optimal and can be found through backward integration of a Riccati differential equation. Two classes of optimal control problems where any Pontryagin solution is necessarily optimal are discussed in the following remark.

Remark 2 *In some cases, the necessary conditions of the Pontryagin solution are also sufficient to guarantee optimality. One is trivially the case when the Pontryagin solution is unique. The second case is convexity of the cost functional with respect to the control variable. The following result can be stated, see [6], Theorem 7.2.1. Assume that the final state $x(t_f)$ is constrained to belong to a target set S and define the set of the admissible controls that steer the system state to the target set, i.e.*

$$\mathcal{U}_m = \{u : [0, t_f] \to \mathcal{U},\ x(t_f) \in S\}$$

To be precise $\mathcal{U}_m$ is the set of measurable and locally integrable *functions taking values in $\mathcal{U}$ and such that $x(t_f) \in S$. If $\mathcal{U}_s$ is convex and the functional $u \to c'x(t_f)$ from $\mathcal{U}_m$ into $\mathbb{R}_+$ is convex, then any Pontryagin solution gives an optimal input trajectory u^o and state trajectory x^o. For our control problem the final state $x(t_f)$ is not constrained (that is, S is the positive orthant) and $\mathcal{U}_m = \{u : [0, t_f] \to \mathcal{U}\}$ is a convex set. Our analysis will focus on the convexity of the functional $u \to c'x(t_f)$ from $\mathcal{U}_m$ to $\mathbb{R}_+$, under an additional assumption on the system matrices A_i, $i = 1, 2, \cdots, m$.*

We are ready to formulate the main theorem of this note, dealing with the optimal control problem for system (3) and cost (2) under the additional assumption below.

Assumption 1 *The off-diagonal entries of the Metzler matrices A_i, $i = 1, 2, \cdots, m$ do not depend on i, i.e.*

$$A_i = D_i + M$$

where M is a suitable Metzler matrix and D_i are diagonal matrices, $i = 1, 2, \cdots, m$.

Remark 3 *The class of switched positive systems that satisfy Assumption 1 is relevant to several applications. At the end of this paper we will write a system, included in this class, that comes when approximating the dynamics of HIV mitigation under therapy switching. It is assumed that the therapy only affects the diagonal elements of the matrices, and this simplifying assumption seems to be good enough to represent the behaviour in a particular transient of the disease grow. Moreover, this class also encompasses some epidemiology models (see for example [21,19]). With some additional assumptions on the interactions, and considering the initial phase of infection, when trying to slow the spread of a disease, the model above is appropriate.*

Theorem 1 *Consider the system (3), the cost (2), and let Assumption 1 be verified. Then the optimal control problem admits at least one Pontryagin solution (u^o, x^o, π^o) and $u^o(t)$ is a global optimal control signal relative to x_0. Moreover, the value of the optimal cost functional is $\pi^{o\prime}(0)x_0$.*

The theorem above will be proved in the next section. Its proof requires a result on convexity that is important in itself and discussed in the following section.

3 Convexity

Given $u \in \mathcal{U}_s$, system (3) under Assumption 1 can be rewritten as

$$\dot{x}(t) = (M + \Lambda(t))x(t) \tag{8}$$

where

$$\Lambda(t) = \sum_{i=1}^{m} D_i u_i(t) = diag\{\lambda_1(t), \lambda_2(t), \ldots, \lambda_n(t)\}$$

Hence $\Lambda(t) \in \mathcal{L}$, $\forall t \geq 0$, where $\mathcal{L}$ is a predefined convex and compact set of diagonal matrices.

Let $\Phi(\Lambda, t, t_0)$ be the transition matrix of $M + \Lambda(t)$, i.e.

$$\frac{d}{dt}\Phi(\Lambda, t, t_0) = (M + \Lambda(t))\Phi(\Lambda, t, t_0), \qquad \Phi(\Lambda, t_0, t_0) = I$$

Given $t_f > 0$, and a positive vector c rewrite the cost as

$$J(\Lambda, x_0) = c'\Phi(\Lambda, t_f, t_0)x_0$$

We now prove that under the above assumptions that

(i) the cost $J(\Lambda, x_0)$ is a convex function of constant diagonal matrix functions, i.e. $\Lambda \in \mathcal{L}$
(ii) the functional $\Lambda \to J(\Lambda, x_0)$ from $\mathcal{L}_p$ into $\mathbb{R}_+$ is convex, where $\mathcal{L}_p = \{\Lambda : [0, t_f] \to \mathcal{L}\}$ is the set of piecewise constant diagonal matrix functions taking values in $\mathcal{L}$.
(iii) the functional $\Lambda \to J(\Lambda, x_0)$ from $\mathcal{L}_m$ into $\mathbb{R}_+$ is convex, where $\mathcal{L}_m = \{\Lambda : [0, t_f] \to \mathcal{L}\}$ is the set of measurable and locally integrable diagonal matrix functions taking values in $\mathcal{L}$.

Let us start by stating two technical lemmas.

Lemma 1 *For any $w \in \mathbb{R}^n$, the function $f : \mathbb{R}^n \to \mathbb{R}_+$*

defined as

$$f(z) = e^{w'z}$$

is convex.

Proof. The gradient of $f(z)$ is $\frac{\partial f(z)}{\partial z} = f(z)w'$ whereas the Hessian matrix is $f(z)ww'$, that is positive semidefinite, so implying convexity of $f(z)$. $\diamond$

Lemma 2 *Let $f_k(z)$, $k = 1, 2, \cdots, p$ be a sequence of convex functions on a convex domain, and assume that the sequence converges to a function $f(z)$. Then $f(z)$ is convex.*

Proof. Assume by contradiction that there exist two points z_1 and z_2 and $0 < \alpha < 1$ such that, denoting by $z = \alpha z_1 + (1-\alpha)z_2$, we have

$$f(z) > \alpha f(z_1) + (1-\alpha) f(z_2) \tag{9}$$

On the other hand, from the convexity assumption,

$$-f_k(z) + \alpha f_k(z_1) + (1-\alpha) f_k(z_2) \geq 0 \tag{10}$$

for all k. Taking the limit, we have

$$f(z) \leq \alpha f(z_1) + (1-\alpha) f(z_2)$$

which contradicts (9).

3.1 Convexity of the cost in $\mathcal{L}$

Let us prove convexity in the case of a constant diagonal matrix functions Λ, i.e. point (i) above.

Lemma 3 *The function $f(\Lambda) : \mathcal{L} \to \mathbb{R}_+$ defined by*

$$f(\Lambda) = c' e^{(M+\Lambda)t_f} x_0 \tag{11}$$

is convex in Λ.

Proof. We first note that $f(\Lambda)$ in (11) is a well defined continuous function of Λ. Then recall a useful formula (Trotter formula) for the exponential of the sum of two matrices ([8]), ([12]):

$$e^{(M+\Lambda)t_f} = \lim_{k\to\infty} \left(e^{\frac{Mt_f}{k}} \; e^{\frac{\Lambda t_f}{k}} \right)^k$$

Therefore, defining the functions

$$f_k(\Lambda) = c' \left(e^{\frac{Mt_f}{k}} \; e^{\frac{\Lambda t_f}{k}} \right)^k \; x_0$$

we have $f_k(\Lambda) \to f(\Lambda)$.

Let us consider the generic function $f_k(\Lambda)$. Since $e^{\frac{Mt_f}{k}}$ is a nonnegative matrix and $e^{\frac{\Lambda t_f}{k}}$ is a diagonal matrix with elements $\xi_i \doteq e^{(\lambda_i T/k)}$ we have that $f_k(\Lambda)$ is a positive polynomial in the variables ξ_i. Formally

$$f_k(\Lambda) = \sum_{k_1+k_2+\ldots k_n = k} \alpha_{k_1,k_2,\ldots,k_N} \xi_1^{k_1} \xi_2^{k_2} \ldots \xi_n^{k_n},$$

with all $\alpha \geq 0$. On the other hand if we substitute $\xi_i = e^{(\lambda_i t_f/k)}$ we get that each monomial can be expressed as

$$\xi_1^{k_1} \xi_2^{k_2} \ldots \xi_n^{k_n} = e^{\lambda_1 k_1 t_f/k + \lambda_2 k_2 t_f/k + \cdots + \lambda_n k_n t_f/k}$$

and therefore it is a convex function of λ_i in view of Lemma 1.

The proof follows from the fact that since $f_k(\Lambda) \to f(\Lambda) = J(\Lambda, x_0)$, and hence $J(\Lambda, x_0)$ is a convex function of Λ. $\diamond$

3.2 Convexity of the cost in $\mathcal{L}_p$

We start by considering piecewise constant diagonal functions Λ.

Lemma 4 *Let $\Lambda(t)$ be piecewise constant function of t*

$$\Lambda(t) = \Lambda_i, \qquad t_{i-1} \leq t < t_i = t_{i-1} + T_i$$

$i = 1, 2, \ldots, K$. Then

$$J(\Lambda, x_0) = c' \prod_i e^{(M+\Lambda_i)T_i} x_0 \tag{12}$$

is convex in the values Λ_i.

Proof. The proof follows similar lines of argument to the proof of Lemma 3. We approximate each exponential

$$e^{(M+\Lambda_i)T_i} \approx \left(e^{\frac{MT_i}{k}} \; e^{\frac{\Lambda_i T_i}{k}} \right)^k$$

as before and we notice that we get a polynomial with positive coefficients in the unknowns $\xi_{i,j} = e^{\lambda_{i,j} T_i/k}$, where $\lambda_{i,j}$, $j = 1, 2, \cdots, n$ are the elements on the diagonal of Λ_i. This polynomial is convex and hence the limit function is convex as well. $\diamond$

Remark 4 *Note that the convexity results have been presented in the case in which the intervals $[t_{k-1}, t_k]$ are common to all functions. We can state the same results in the class in which each function $\Lambda(t)$ has its own interval partition by considering the "intersection". Indeed two piecewise constant functions with switching point $t_k^{(1)}$ and $t_k^{(2)}$ have a common interval partition, precisely that* *achieved by considering all the ordered instants $t_k^{(1)}$ and $t_k^{(2)}$.*

3.3 Convexity of the cost in $\mathcal{L}_s$

We can conclude with the main result, that is achieved by taking suitable approximations of a measurable and locally integrable function. The idea of approximation of an arbitrary function by piecewise constant ones is certainly not new in control system theory, see e.g. the book [24].

Theorem 2 *Let $\Lambda \in \mathcal{L}_s$. The cost*

$$J(\Lambda, x_0) = c'\Phi(\Lambda, t_f, t_0)x_0$$

is convex in $\mathcal{L}_s$.

Proof. For any measurable and locally integrable function $f(t)$ there exists a sequence of continuous functions $f_k(t), k = 1, 2, \cdots$, such that $\int_0^{t_f} \|f(t)-f_k(t)\|dt \leq \epsilon_k \to 0$ holds (density Theorem, see e.g. [11]). Moreover, any continuous function $f_k(t)$ can be approximated (again in the L_1 sense) by a sequence of piecewise constant functions $f_{kq}(t)$, $q = 1, 2, \cdots$. Consider now the system dynamic matrix $\bar{A}(t) = \Lambda(t) + M$ for a certain $\Lambda \in \mathcal{L}_s$ and a sequence $\bar{A}_k(t) = \Lambda_k(t) + M$ with Λ_k continuous diagonal matrix functions such that

$$\int_0^{t_f} \|\bar{A}_k(t) - \bar{A}(t)\|dt \leq \epsilon_k, \quad \lim_{k\to\infty} \epsilon_k = 0$$

Moreover, for each k, take a sequence $\bar{A}_{kq}(t)$ of piecewise constant matrix functions such that

$$\int_0^{t_f} \|\bar{A}_{kq}(t) - \bar{A}_k(t)\|dt \leq \epsilon_{kq}, \quad \lim_{q\to\infty} \epsilon_{kq} = 0, \forall k$$

It follows that

$$\begin{aligned}
&\int_0^{t_f} \|\bar{A}_{kq}(t) - \bar{A}(t)\|dt \leq \\
&\leq \int_0^{t_f} \|\bar{A}_{kq}(t) - \bar{A}_k(t) + \bar{A}_k(t) - \bar{A}(t)\|dt \leq \\
&\leq \int_0^{t_f} \|\bar{A}_{kq}(t) - \bar{A}(t)\|dt + \int_0^{t_f} \|\bar{A}_{kq}(t) - \bar{A}_k(t)\|dt \leq \\
&\leq \epsilon_{kq} + \epsilon_k, \quad \lim_{k\to\infty} = 0, \ \lim_{q\to\infty} \epsilon_{kq} = 0, \forall k
\end{aligned}$$

Denote now by $x(t_f)$ and $x_{kq}(t_f)$ the solution corresponding to $\bar{A}(t)$ and $\bar{A}_{kq}(t)$, with initial condition x_0, namely

$$x(t_f) = x_0 + \int_0^{t_f} \bar{A}(t)x(t)dt$$

$$x_{kq}(t_f) = x_0 + \int_0^{t_f} \bar{A}_{kq}(t)x_{kq}(t)dt$$

Denote by $\mu(t) = \|x_{kq}(t) - x(t)\|$ and note that $x(t)$ is bounded in $[0, t_f]$, i.e. $\|x(t)\| \leq \xi$ and $\bar{A}_{kq}(t)$, $k = 1, 2, \cdots, q = 1, 2, \cdots$ are bounded as well, i.e. $\|\bar{A}_{kq}(t)\| \leq \alpha$, for each k and q. Adding and subtracting $\bar{A}_{kq}(t)x(t)$, we get

$$\begin{aligned}
\mu(t_f) &\leq \int_0^{t_f} \|\bar{A}_{kq}(t) - \bar{A}(t)\|\|x(t)\|dt + \int_0^{t_f} \|\bar{A}_{kq}(t)\|\mu(t)dt \\
&\leq (\epsilon_k + \epsilon_{kq})t_f\xi + \alpha \int_0^{t_f} \mu(t)dt
\end{aligned}$$

Notice that $\mu(0) = 0$. Therefore, in view of Gronwall's lemma

$$\mu(t_f) \leq (\epsilon_k + \epsilon_{kq})\frac{t_f\xi}{\alpha}\ (e^{\alpha t_f} - 1)$$

so that $\mu(t_f)$ tends to zero for both k and q going to infinity. This proves continuity of $x(t_f)$ for $\Lambda \in \mathcal{L}_s$.

As for convexity, assume by contradiction assume that $\Lambda \in \mathcal{L}_s$,

$$\Lambda(t) = \alpha\Lambda_1(t) + (1 - \alpha)\Lambda_2(t)$$

with $\Lambda_i \in \mathcal{L}_s$, $i = 1, 2$, is such that

$$c'\Phi(\Lambda, t, t_0)x_0 > \alpha c'\Phi(\Lambda_1, t, t_0)x_0 + (1-\alpha)\Phi(\Lambda_2, t, t_0)x_0$$

for some $0 < \alpha < 1$. On the other hand each of the terms appearing in this expression can be approximated by sequences of piecewise constant functions, i.e.

$$\Lambda^{k,q}(t) = \alpha\Lambda_1^{kq}(t) + (1 - \alpha)\Lambda_2^{kq}(t)$$

with $\Lambda^{k,q} \to 0$, $\Lambda_i^{k,q} \to 0$, $i = 1, 2$, for k and q going to infinity. Thanks to the convexity result (Lemma 4) it follows

$$c'\Phi(\bar{\Lambda}^{kq}, t, t_0)x_0 \leq \alpha c'\Phi(\bar{\Lambda}_1^{kq}, t, t_0)x_0 + (1-\alpha)\Phi(\bar{\Lambda}_2^{kq}, t, t_0)x_0$$

and this leads to a contradiction. ⋄

3.4 Proof of Theorem 1

The optimal control does exist, as noted in Remark 1. Let an optimal triple be x^o, u^o, π^o. This triple is a Pontryagin solution, as defined in Definition 1. Indeed, the Hamiltonian function associated with system (3) and the linear cost (2) is

$$H(x, u, \pi) = \pi(t)' \sum_{i=1}^{m} u_i A_i x(t)$$

and $\dot{\pi}(t) = -\left(\frac{\partial H}{\partial x}\right)' = -\sum_{i=1}^{m} u_i(t)A_i'\pi(t)$, $\dot{x}(t) = \left(\frac{\partial H}{\partial \pi}\right)' = \sum_{i=1}^{m} u_i(t)A_i x(t)$, with $\pi(t_f) = c$ and

$x(0) = x_0$. The transversal conditions are satisfied and for all $u \in \mathcal{U}$:

$$H(x^o, u^o, \pi^o) \leq H(x^o, u, \pi^o).$$

In view of the Pontryagin principle, the triple (x^o, π^o, u^o) satisfies the necessary conditions for optimality. Theorem 2 states the convexity of the cost with respect to the functions $\lambda_i(t)$, diagonal entries of $\Lambda(t) = \sum_{i=1}^m D_i u_i(t)$. Hence the cost is convex with respect to the control variable $u \in \mathcal{U}_s$. This fact is sufficient, see [6, Theorem 5.1.1], to guarantee optimality. □

Remark 5 *For almost all t, the scalar function $v(x,t) = \pi^o(t)'x$ satisfies:*

$$0 = \frac{\partial v}{\partial t}(x^o(t), t) + \min_u H\left(x^o(t), u, \frac{\partial v}{\partial x}(x^o(t), t)'\right)$$

with the boundary condition

$$v(x^o(t_f), t_f) = \pi^o(t_f)'x^o(t_f) = c'x^o(t_f)$$

Based on the convexity result, this is enough to guarantee that any Pontryagin solution is also optimal

Note that if $u^o(t)$ lies at one of the vertices of $\mathcal{U}$, then an *admissible switching signal* (i.e. signals $\sigma(t) \in \{1, 2...m\}$ for almost all t), $\sigma^o(t)$ can be constructed as follows:

$$\begin{aligned} \dot{x}^o(t) &= A_{\sigma^o(t,x_0)} x^o(t) \\ -\dot{\pi}^o(t) &= A'_{\sigma^o(t,x_0)} \pi^o(t) \\ \sigma^o(t, x_0) &= \arg\min_{i \in \mathcal{I}} \{\pi^{o\prime}(t) A_i x^o(t)\} \end{aligned}$$

with the boundary conditions $x^o(0) = x_0$, $\pi^o(t_f) = c$, and $J(x_0, x^o, \sigma^o) = \pi^{o\prime}(0) x_0$.

3.5 Extensions

The proof of convexity has been carried out by looking at the so-called *Mayer* problem for switched positive systems, i.e. a problem where the cost is a function of the final state only. The same result can be proved for the more general cost function

$$J(x_0) = c'x(t_f) + \int_0^{t_f} d'x(t)dt \qquad (13)$$

where d is a nonnegative vector. Indeed, the optimal control problem for system (3) and cost (13) can be transformed in the optimal control problem for cost (3) and cost (2) by taking the extended system

$$\dot{\xi} = \bar{A}(u)\xi, \quad \bar{A}(u) = \begin{bmatrix} A(u) & 0 \\ d' & 0 \end{bmatrix}, \quad \xi(0) = \begin{bmatrix} x_0 \\ 0 \end{bmatrix}$$

and the extended cost

$$J = \bar{c}'\xi(t_f), \quad \bar{c}' = \begin{bmatrix} c' & 1 \end{bmatrix}$$

Notice that the assumption on diagonal switching, i.e. Assumption 1, is met with for $\bar{A}(u)$ as well, since d does not depends on u.

Following a similar rationale we can also establish an extension of the presented theory for switched systems affected by a constant input, i.e.

$$\dot{x} = A(u)x + b$$

where b is a nonnegative vector. This system can be rewritten as

$$\dot{\xi} = \bar{A}(u)\xi, \quad \bar{A}(u) = \begin{bmatrix} A(u) & b \\ 0 & 0 \end{bmatrix}, \quad \xi(0) = \begin{bmatrix} x_0 \\ 0 \end{bmatrix}$$

and the cost

$$J = \bar{c}'\xi(t_f), \quad \bar{c}' = \begin{bmatrix} c' & 0 \end{bmatrix}$$

Again the new $\bar{A}(u)$ satisfies Assumption (1) if $A(u)$ does.

4 Algorithm and application to optimal therapy scheduling

In this section we provide a computational scheme for the optimal control and consider the application example in [15].

4.1 Algorithm

The convexity property allows the use of different types of algorithms to find the solution of

$$\min_{u \in \mathcal{U}_s} J(u), \quad J(u) = c'x(t_f)$$

Computations can be cast in discrete-time, by taking, as in the previous section, a subdivision of the interval $[0, t_f]$ into N intervals $T_1, T_2, \cdots, T_N$. The control variable may be approximated as piecewise constant, i.e.

$$u(t) = \begin{cases} \bar{u}_1 & t \in [0, T_1) \\ \bar{u}_2 & t \in [T_1, T_1 + T_2) \\ \vdots & \vdots \\ \bar{u}_N & t \in [\sum_{i=1}^{N-1} T_i, t_f) \end{cases}$$

The discretized control is denoted by $\bar{u}$ taking values in the cartesian products of $\mathcal{U}$, denoted by $\bar{\mathcal{U}}$. Hence the problem is to find

$$\min_{u\in\bar{\mathcal{U}}} c' \prod_{i=N}^{1} e^{(M+D_i\bar{u}_i)T_i} x_0$$

Letting

$$J(\bar{u}) = c' \prod_{i=N}^{1} e^{(M+D_i\bar{u}_i)T_i} x_0 \tag{14}$$

the constrained optimization problem can be solved using the standard Matlab function fmincon.m or an it ad hoc algorithm based on a projected (sub)gradient method. Notice indeed that $\bar{\mathcal{U}}$ is a convex set and that $J(\bar{u})$ is a convex function of $\bar{u}$, see Theorem 2. As for the gradient method, the standard scheme is given by the sequence (k indicates the iteration index)

$$\bar{u}^{[k+1]} = \mathrm{Proj}_{\bar{\mathcal{U}}}\left(\bar{u}^{[k]} - \alpha g^{[k]}\right) \tag{15}$$

where α is a speed factor (possibly varying with k), $\mathrm{Proj}_{\bar{\mathcal{U}}}$ is the projection on $\bar{\mathcal{U}}$ and $g^{[k]}$ is the gradient of $J(\bar{u})$ evaluated at $\bar{u} = \bar{u}^{[k]}$. The gradient of $J(\bar{u})$ is a N-dimensional row vector and the j^{th} entry g_j can be computed in a simple way from (14). Indeed, simple calculations show that

$$g_j = c' \prod_{i=N}^{j+1} e^{(M+D_i\bar{u}_i)T_i} \Phi_j(T_j) \prod_{i=j-1}^{1} e^{(M+D_i\bar{u}_i)T_i} x_0$$

where $\Phi_j(t) = \frac{\partial e^{(M+D_j\bar{u}_j)t}}{\partial u_j}$. It follows that

$$\dot{\Phi}_j(t) = (M + D_j\bar{u}_j)\Phi_j(t) + D_j e^{(M+D_j\bar{u}_j)t} \tag{16}$$

and therefore

$$\Phi_j(t) = \int_0^t e^{(M+D_j\bar{u}_j)(t-\tau)} D_j e^{(M+D_j\bar{u}_j)\tau} d\tau$$

Alternatively, it can then be shown that

$$\Phi_j(T_j) = \begin{pmatrix} 0 & I \end{pmatrix} e^{\begin{pmatrix} M + D_j\bar{u}_j & 0 \\ D_j & M + D_j\bar{u}_j \end{pmatrix} T_j} \begin{pmatrix} I \\ 0 \end{pmatrix}$$

from which g_j can be calculated. Notice that numerical algorithms might be further enhanced by explicit computation of the Hessian matrix using similar techniques.

4.2 Saddle point problem

The optimal cost, i.e. $\min_{\Lambda\in\mathcal{L}_s} J(\Lambda, x_0)$, is a concave function of x_0, see [14], [3]. Then taking $x_0 \in \mathcal{A}$ for some predefined set of initial states, it may also be of interest to find a saddle point solution of the min-max problem

$$\min_{u\in\mathcal{U}_s} \max_{x_0\in\mathcal{A}} J(x_0, u)$$

i.e. a solution u^*, x_0^* such that $J(x_0, u^*) \leq J(x_0^*, u^*) \leq J(x_0^*, u)$ for any $x_0 \in \mathcal{A}$ and any $u \in \mathcal{U}_s$. In this regard, taking again the above discretization of u, we are able to write the computational scheme

$$\bar{u}^{[k+1]} = \mathrm{Proj}_{\bar{\mathcal{U}}}\left(\bar{u}^{[k]} - \alpha g^{[k]}\right) \tag{17}$$

$$\bar{x}_0^{[k+1]} = \mathrm{Proj}_{\mathcal{A}}\left(\bar{x}_0^{[k]} + \alpha h^{[k]}\right) \tag{18}$$

where $\mathrm{Proj}_{\mathcal{A}}$ is the projection on $\mathcal{A}$, $h^{[k]}$ is the gradient of $J(x_0, u)$ with respect to x_0 at the k-th iteration. The vector h can be easily computed thanks to the linearity of $J(x_0, u)$ with respect to x_0, as $h = c' \prod_{i=N}^{1} e^{(M+D_i\bar{u}_i)T_i}$.

4.3 HIV mitigation

Under several simplifying assumptions, [15], the drug treatment scheduling problem of HIV infection can be cast as as an optimal control problem for an autonomous positive switched systems of the form

$$\dot{x}(t) = \left(R_{\sigma(t)} - \delta_V I\right) x(t) + \mu M_u x(t) \tag{19}$$

where $M_u := [m_{ij}]$ and $R_{\sigma(t)} := diag\{\rho_{i,\sigma(t)}\}$. $\rho_{i,\sigma(t)}$ is the replication rate for viral genotype (i) and therapy combination σ; μ represents the mutation rate, δ_V is the viral clearance and $m_{i,j} \in \{0, 1\}$ represents the genetic connections between genotypes, that is, $m_{i,j} = 1$ if and only if it is possible for genotype j to mutate into genotype i.

As a specific example, we consider a system with 4 states, and two treatment options, $m = 2$, of the following structure

$$A_\sigma = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_{2\sigma} & 0 & 0 \\ 0 & 0 & \lambda_{3\sigma} & 0 \\ 0 & 0 & 0 & \lambda_4 \end{bmatrix} + \mu \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \tag{20}$$

With the additional assumption

$$\lambda_{21} > 0, \lambda_{22} < 0, \lambda_{31} < 0, \lambda_{32} > 0, \ \lambda_{21} - \lambda_{22} + \lambda_{31} - \lambda_{32} = 0 \tag{21}$$

we have previously been able to give an explicit form for the optimal control of this specific problem [15].

Notice that this 4 variant model is a significant simplification, more amenable to analysis, of complex mutation dynamics in real HIV behaviour.

Define $k_1 = \text{argmin}\{x_2(0), x_3(0)\}$, $k_2 = \text{argmin}\{c_2, c_3\}$, $T_1 = \text{argmin}_{t \geq 0} : [0\ 1\ \ -1\ 0]e^{A_{k_1}t}x(0) = 0$, $T_2 = \text{argmax}_{t \leq t_f} : [0\ 1\ \ -1\ 0]e^{-A_{k_2}(t-t_f)}c = 0$, and $\alpha = (\lambda_{32} - \lambda_{22})(\lambda_{32} - \lambda_{22} + \lambda_{21} - \lambda_{31})^{-1}$.

Notice that, thanks to the definition of k_1, k_2 and the monotonicity conditions of $x_2(t) - x_3(t)$, $\pi_2(t) - \pi_3(t)$, solutions of (5), (6), the time instants T_1 and T_2 are well defined and unique. Clearly, by definition $x_2(T_1) = x_3(T_1)$ and $\pi_2(T_2) = \pi_3(T_2)$. Moreover, from (21), $\alpha \in (0, 1)$.

Recall the main result for the four variant model for the case $T_1 \leq T_2$ (Long horizon case), (see [15]):

Theorem 3 *Consider the system (3), (20) with (21) and cost (2) and assume that $T_1 \leq T_2$. Then, the optimal control is given by $u_1(t) = 2 - k_1$, $\quad t \in [0, T_1]$ and $u_1(t) = 2 - k_2$, $\quad t \in [T_2, t_f]$. For $t \in [T_1, T_2]$, the optimal control is given by $u_1(t) = \alpha$, i.e. the trajectory along the plane $x_2 = x_3$, with dynamical matrix $A_\alpha = \alpha A_1 + (1 - \alpha)A_2$.*

In [15], the above theorem (as well as other results relative to all other cases) has been proven with the additional assumption that $\mu > 0$, by showing that the necessary (Pontryagin) conditions are satisfied and that only one Pontragyn solution can be found for the problem. From Theorem 2 we know that, given the convexity of the cost, the necessary conditions are also sufficient to guarantee optimality. This optimality applies even in cases of non-strict convexity (e.g. the pure diagonal case, $\mu = 0$). In this case there are infinitely many Pontryagin solutions, all of which achieve the same cost.

4.4 Simulation results

Here we show some simulation results for the optimal control of (3), (20) with cost (2). The parameters, taken from [15] to which the reader is referred to for additional details, are as follows: $\mu = 10^{-4}$, $\lambda_1 = -0.19$, $\lambda_{21} = \lambda_{32} = 0.03$, $\lambda_{31} = \lambda_{22} = -0.19$, $\lambda_4 = 0.03$ so that $\alpha = 0.5$. In all simulations the final time is $t_f = 50$ and we take an asymmetric initial state, $x_0 = [10^3\, 5\, 0\, 10^{-5}]$ so that $k_1 = 2$ and $T_1 = 12.29$.

We have compared the interior point method in the Matlab routine fmincon.m and an *ad hoc* developed algorithm based on the projected gradient method (15), with optimally varying speed, see [20]. In the first simulation the linear cost is chosen with $c = [1\,1\,1\,1]$, and hence k_2 is undefined and $T_2 = t_f$. In the second simulation $c = [1\,5\,1\,1]$. It follows that $k_2 = 2$ and $T_2 = 42.69$. In the third simulation $c = [1\,1\,5\,1]$ so that $k_2 = 1$ and $T_2 = 42.69$. As expected, the results using both algorithms are identical apart from minor differences due to numerical issues. In Figure 4.4 the optimal control u is shown for the three cases of different terminal cost vector. These results are of course in accordance with Theorem 3.

As for the min-max optimal control problem, algorithm (17), (18) has been applied for several different choices of c. As a result, the saddle point solution is found to be given by (x_0^*, u^*) where $x_0^* = [0\,0\,0\,1]'$ and u^* coinciding with the optimal control function associated with x_0^*, that means $u_1(t) = u_2(t) = 0.5$ for $t \in [0, T_2]$ and $u_1(t) = 2 - k_2$, $u_2(t) = |1 - k_2|$, $\quad t \in [T_2, t_f]$.

Finally, we have slightly perturbed one parameter of the system in order to violate (21), i.e. $\lambda_{21} - \lambda_{22} + \lambda_{31} - \lambda_{32} = 0$. In particular we have set $\lambda_{21} = 0.07$. In this case Theorem 3 is no longer valid and the optimal solution, even in the long horizon case, cannot be easily computed in closed-form. Figure 4.4 shows the results using the numerical algorithms of Section 4.1. Although there appears to remain a sliding mode optimal solution, in this case, the equivalent control is no longer constant in time. Note however, that the results here do not prove that in general the asymmetric case also exhibits sliding mode solutions.

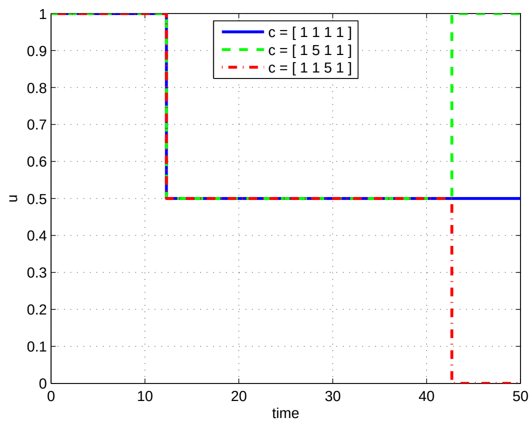

Figure 1. Optimal control variable - symmetric case

5 Conclusions

In this paper we have examined the optimal control problem for a class of linear switched positive systems. In this class, only the diagonal entries of the dynamical matrices associated with the modes are permitted to vary as a function of the switching signal. This assumption,

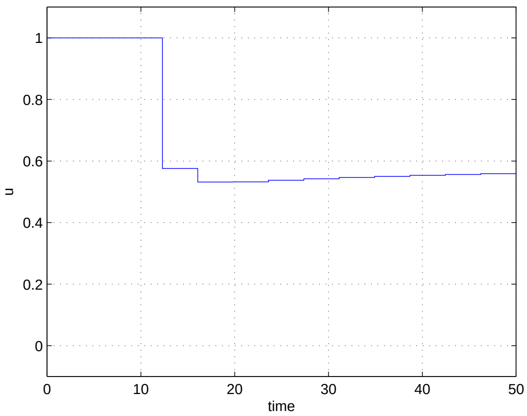

Figure 2. Optimal control variable - asymmetric case

relevant in some biological and epidemiological applications, is shown to guarantee the important property of convexity of the cost function with respect to the diagonal entries of the dynamical matrices. In turn, this also guarantees that any Pontryagin solution is also optimal. This fact opens the way to the use efficient computational methods to find the optimal control. Concavity of the optimal cost with respect to the initial state also allows efficient computation of a saddle point solution of the relevant min-max optimal control problem. The results are verified through simulations in a 4-dimensional switched system representing in a simplified way the effect of a therapy scheduling on HIV mutation.

Acknowledgments: The authors thank Prof. Fabio Zanolin for helpful suggestions.

References

[1] X. Bai and X.S. Yang. A new proof of a theorem on optimal control of switched systems. *Journal of Mathematical Analysis and Applications*, 331(2):895–901, 2007.

[2] M. Bardi and I. Capuzzo-Dolcetta. *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations.* Birkhäuser Boston, 2008.

[3] F. Blanchini, P. Colaneri, and M.E. Valcher. Co-positive lyapunov functions for the stabilization of positive switched systems. *Automatic Control, IEEE Transactions on*, 57(12):3038–3050, 2012.

[4] U. Boscain. Stability of planar switched systems: the linear single input case. *SIAM journal on control and optimization*, 41(1):89–112, 2002.

[5] P. Brandi and A. Salvadori. On measure differential inclusions in optimal control theory. *Universitae Politecnico di Torino*, page 69, 1998.

[6] A. Bressan and B. Piccoli. *Introduction to Mathematical Control Theory.* 2007. AIMS Series on Applied Mathematics, Philadelphia, 2007.

[7] C.G. Cassandras, D.L. Pepyne, and Y. Wardi. Optimal control of a class of hybrid systems. *Automatic Control, IEEE Transactions on*, 46(3):398–415, 2001.

[8] J.E. Cohen. Convexity of the dominant eigenvalue of an essentially nonnegative matrix. *Proceedings of the American Mathematical Society*, 81(4):657–658, 1981.

[9] A.V. Dmitruk and A.M. Kaganovich. The hybrid maximum principle is a consequence of pontryagin maximum principle. *Systems & Control Letters*, 57(11):964–970, 2008.

[10] A.V. Dmitruk and A.M. Kaganovich. Maximum principle for optimal control problems with intermediate constraints. *Computational Mathematics and Modeling*, 22(2):180–215, 2011.

[11] B. K Driver. *Analysis Tools with Applications.* Springer, 2003.

[12] D Elliott. *Bilinear Control Systems: Matrices in Action.* Springer, 2009.

[13] L. Farina and S. Rinaldi. *Positive linear systems: Theory and applications.* Wiley New York, 2000.

[14] E. Hernandez-Vargas, P. Colaneri, R. Middleton, and F. Blanchini. Discrete-time control for switched positive systems with application to mitigating viral escape. *International Journal of Robust and Nonlinear Control*, 21(10):1093–1111, 2011.

[15] Esteban A Hernandez-Vargas, Patrizio Colaneri, and Richard H Middleton. Optimal therapy scheduling for a simplified hiv infection model. *Automatica*, 2013.

[16] J. L. Mancilla-Aguilar and R. F. Garcia. On measure differential inclusions in optimal control theory. *Automatica*, 49(2):441–447, 2013.

[17] M. Margaliot. Stability analysis of switched systems using variational principles: An introduction. *Automatica*, 42(12):2059–2077, 2006.

[18] D.B. McDonald. Lyapunov optimizing sliding mode control for linear systems with bounded disturbance. *Applied Mathematical Sciences*, 2(19):901–918, 2008.

[19] Y. Moreno, R. Pastor-Satorras, and A. Vespignani. Epidemic outbreaks in complex heterogeneous networks. *The European Physical Journal B-Condensed Matter and Complex Systems*, 26(4):521–529, 2002.

[20] Y Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course.* Kluwer Academic Publishers, 2004.

[21] M.A. Rami, V.S. Bokharaie, O. Mason, and F.R. Wirth. Stability criteria for sis epidemiological models under switching policies. *arXiv preprint arXiv:1306.0135*, 2013.

[22] L. Rapoport. Asymptotic stability and periodic motions of selector-linear differential inclusions. *Robust control via variable structure and Lyapunov techniques*, pages 269–285, 1996.

[23] W. Spinelli, P. Bolzern, and P. Colaneri. A note on optimal control of autonomous switched systems on a finite time interval. In *American Control Conference, IEEE*, pages 6–pp, 2006.

[24] Jurdjevic V. *Geometric Control Theory.* Cambridge press, 1997.