UNIVERSITÀ DEGLI STUDI DI PADOVA

SCUOLA DI DOTTORATO IN INGEGNERIA DELL'INFORMAZIONE

Indirizzo "Automatica e Ricerca Operativa" – XX Ciclo

Tesi di dottorato / Ph.D. Thesis

Topics on Switched Systems

Ph.D. Advisor: Prof.ssa Maria Elena Valcher

Ph.D. Co-Advisor: Prof. Joao Pedro Hespanha

Ph.D. Candidate: Paolo Santesso

Padova, January 2008

Contents

Introduction

Thesis Outline

This thesis presents several results pertaining two rather distinct research topics within the broader area of the so-called “Switched Systems”.

The first part of the work features a deep investigation of the structural properties, namely reachability and zero-controllability, of “Positive Switched Systems”, both for the discrete-time and the continuous-time case.

All the notation relative to this contribution is defined in Chapter 1.

Together with considerations on the motivational aspect, in Chapter 2, the familiar concepts of *reachability* and *zero-controllability* are properly defined within the context of positive switched systems.

Then, several results are presented, first dealing with the discrete-time case, and subsequently addressing the continuous-time one.

More specifically, in Chapter 3, the zero-controllability of a discrete-time positive switched system is proved to be equivalent to the mortality property of the set of system matrices; some sufficient conditions for this property to hold are then provided, together with an algorithm designed to find the correct switching sequence, if any, which is needed to drive any positive state vector to the zero vector.

In Chapter 4 the reachability issue for discrete-time positive switched systems is addressed. First, the problem is restated into a geometric form, then the property of *monomial reachability*, known to be equivalent to the reachability for standard (meaning non-switched) positive systems, but only necessary in our setting, is fully explored and characterized.

All the chapters from 5 to 8 tackle with the continuous-time case.

In particular, in Chapter 5 the possibility for a continuous-time positive switched system to be zero-controllable is ruled out.

The reachability issue is first addressed in Chapter 6 where, similarly to the discrete-time case, we investigate the *monomial* as well as the *pattern reachability* property, which represent two necessary conditions for the general reachability of the system. Then, in Chapter 7, a useful sufficient condition for the reachability is provided; a geometric equivalent description of a reachable system is also introduced. Finally, further contributions to the problem of finding conditions ensuring the reachability of a continuous-time positive switched system are presented in Chapter 8, where the useful concept of *asymptotic exponential cone* of a Metzler matrix (an ordered set of Metzler matrices) is first defined and then fully characterized.

Results pertaining to a different stream of research [1] are included in the chapters 9 and 10.

More specifically, in Chapter 9 the case when a traditional Linear Time Invariant plant is controlled by a switching *multicontroller* whose transfer function may commute among different ones, each of them stabilizing the system, is considered. In particular, we focus our interest on the design of the function, called *Reset Map*, ruling the update of the multicontroller state vector at every switching time. It turns out that a proper choice of it may deeply improve the controlled system transient behaviour.

The application of the same principles is then suggested in Chapter 10 in the context of non-switching reset controllers. The result presented within this chapter represents a substantial enhancement with respect to the traditional approach which is known in the literature under the name of *Reset Control Strategy*.

The Appendix, besides a series of technical results which are preliminary to those presented in this thesis, features an extensive contribution to the study of the exponential of a Metzler matrix. This topic has been initially addressed as a mathematical mean for solving certain specific problems within the setting of positive switched systems. Indeed, the analysis of reachability property for continuous-time positive switched systems requires a deep knowledge of the behaviour of these exponential matrices. For this reason, we decided to include the results in the Appendix. However, we believe that they deserve some interest by themselves, as their significance and extension exceed by far what we needed for their initial application.

[1] The material included in these chapters comes from a joint work with Prof. J.P. Hespanha, University of California at Santa Barbara, USA.

Introduzione

Contenuto della tesi

Questa tesi contiene alcuni risultati riguardanti due distinti argomenti di ricerca, entrambi collocabili all'interno della vasta area di studio inerente ai cosiddetti "sistemi a commutazione".

La prima parte di questo lavoro rappresenta un'approfondito studio delle proprietà strutturali, ovvero raggiungibilità e contrallabilità a zero, dei "sistemi positivi a commutazione", sia nel contesto di sistemi a tempo discreto sia in quello di sistemi a tempo continuo.

La notazione utilizzata in questo contributo è esaurientemente definita nel Capitolo 1.

Assieme a considerazioni di natura motivazionale, nel Capitolo 2 vengono definiti i familiari concetti di *raggiungibilità* e *controllabilità a zero* nel contesto dei sistemi positivi a commutazione.

Nel proseguo della tesi sono raccolti vari risultati, dapprima relativi al caso dei sistemi a tempo discreto, e quindi di quelli a tempo continuo.

Più specificatamente, nel Capitolo 3, viene provata l'equivalenza tra la controllabilità a zero dei sistemi positivi a commutazione a tempo discreto e la nota proprietà di mortalità dell'insieme delle matrici di sistema; vengono poi fornite alcune condizioni sufficenti per la controlabilità di tali modelli matematici, assieme ad un algoritmo in grado di individuare la corretta sequenza di commutazione, qualora essa esista, che è necessaria per condurre ogni vettore di stato positivo allo stato finale nullo.

Nel Capitolo 4 viene affrontata la questione relativa alla raggiungibilità dei sistemi positivi a commutazione a tempo discreto. Inizialmente, il problema viene riformulato in un contesto puramente geometrico, quindi viene dettagliatamente esplorata la proprietà della cosiddetta "raggiugilità monomia", equivalente alla "raggiungiblità" semplice per i sistemi positivi standard (ovvero non a commutazione), condizione semplicemente necessaria nel nostro contesto.

I Capitoli dal 5 al 8 sono dedicati allo studio dei sistemi a tempo continuo.

In particolare, nel Capitolo 5 si dimostra come un sistema positivo a commutazione a tempo continuo non possa mai risultare controllabile a zero.

La questione della raggiungibilità è inizialmente affrontata nel Capitolo 6 dove, come nel caso discreto, vengono approfondite le proprietà di raggiungiblità *monomia* e di *pattern*, che assieme rappresentano due condizioni necessarie per la raggiungibilità generale del sistema. Successivamente, nel Capitolo 7, viene ricavata una utile condizione sufficiente per la raggiungibilità, ed infine viene fornita in termini puramente geometrici una caratterizzazione equivalente di un sistema raggiungibile. In ultima analisi, ulterori contributi alla soluzione del problema di trovare condizioni che assicurino la raggiungibilità dei sistemi positivi a commutazione a tempo continuo vengono illustrati nel Capitolo 8, dove vengono definiti e completamente caratterizzati gli utili concetti di *cono esponenziale asintotico* di una matrice di Metzler (di un insieme ordinato di matrici di Metzler).

Nei capitoli 9 e 10 vengono invece presentati i risultati ottenuti in un diverso contesto di ricerca [2]

Con più precisione, nel Capitolo 9 viene contemplato il caso di un tradizionale sistema Lineare Tempo-Invariantethe che viene controllato da un *multicontrollore* a commutazione, la cui funzione di trasferimento può variare tra quella di diversi modelli, ciascuno dei quali risulti stabilizzante per il dato sistema. In particolare, ci siamo concentrati sulla scelta della funzione, detta *funzione di aggiornamento*, che controlla l'aggiornamento del vettore di stato del multicontrollore ad ogni istante di commutazione. La nostra ricerca mostra chiaramente come un'accurata scelta di tale oggetto matematico possa influenzare in maniera importante il comportamento in regime transitorio del sistema controllato.

Gli stessi principi vengono poi applicati nel Capitolo 10 al contesto dei sistemi non a commutazione. I risultati presentati costituiscono un sostanziale progresso rispetto all'approccio tradizionalmente conosciuto in letteratura come *Controllo a reset*.

L'Appendice, oltre a contenere una serie di risultati tecnici preliminari

[2] Il materiale incluso in questi capitoli è frutto di un lavoro condotto assieme al Prof. J.P. Hespanha, presso l'Università della California, Santa Barbara, USA.

a quelli illustrati nella tesi, è anche arricchita da un esauriente contributo relativo allo studio della forma esponenziale di una matrice di Metzler. L'argomento è stato inizialmente affrontato in quanto strumento matematico atto a risolvere specifici problemi nel contesto dei sistemi positivi a commutazione a tempo continuo, particolarmente nello studio della raggiungibilità. Per tale ragione, abbiamo deciso di includere questi risultati nell'Appendice. Va notato d'altronde come la generalità dei contributi presentati vada ben oltre la semplice applicazione al caso in esame nel presente lavoro.

Chapter 1

Notation

Before starting, we introduce here some useful notation.

For every $k \in \mathbb{N}$, we set $\langle k \rangle := \{1, 2, \ldots, k\}$. In the sequel, the (i, j)th entry of some matrix A is denoted by $[A]_{i,j}$. If A is block partitioned, we denote its (i, j)th block either by $\mathrm{block}_{(i,j)}[A]$ or by A_{ij}. In the special case of a vector $\mathbf{v}$, its ith entry is $[\mathbf{v}]_i$ and its ith block is $\mathrm{block}_i[\mathbf{v}]$. Given a block partitioned matrix A with blocks A_{ij}, $i, j \in \langle \ell \rangle$, for every i and j with $i \leq j$, we will denote by $A_{\{i,j\}}$ the following submatrix of A:

$$A_{\{i,j\}} := \begin{bmatrix} A_{ii} & A_{i,i+1} & \ldots & A_{ij} \\ A_{i+1,i} & A_{i+1,i+1,} & \ldots & A_{i+1,j} \\ \vdots & & \ddots & \vdots \\ A_{ji} & \ldots & \ldots & A_{jj} \end{bmatrix}. \tag{1.1}$$

Given an ordered set of real numbers $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$, we define

$$\mathrm{diag}(\lambda_1, \ldots, \lambda_n) := \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & & & \vdots \\ 0 & \ldots & 0 & \lambda_n \end{bmatrix}.$$

Given a matrix $A \in \mathbb{R}^{q \times r}$, by the *zero pattern* of A we mean the set of index pairs corresponding to its zero entries, namely

$$\mathrm{ZP}(A) := \{(i, j) : [A]_{i,j} = 0\}.$$

For a column vector $\mathbf{v}$ (corresponding to $r = 1$), the zero pattern is accordingly defined as

$$\mathrm{ZP}(\mathbf{v}) := \{i : [\mathbf{v}]_i = 0\}.$$

Conversely, the *nonzero pattern* is the set of indices corresponding to the nonzero entries of a matrix A (a vector $\mathbf{v}$) and it is denoted by $\overline{\mathrm{ZP}}(A)$ ($\overline{\mathrm{ZP}}(\mathbf{v})$).

If $\mathbf{v}$ is an n-dimensional vector and $\mathcal{S} = \{i_1, i_2, \dots, i_r\}$, with $i_1 < i_2 < \dots < i_r$, is a subset of $\{1, 2, \dots, n\}$, we denote by $\mathbf{v}_\mathcal{S}$ the r-dimensional vector such that $[\mathbf{v}_\mathcal{S}]_k = [\mathbf{v}]_{i_k}, k = 1, 2, \dots, r$, namely the restriction of $\mathbf{v}$ to the entries with indices in $\mathcal{S}$. Moreover, $P_\mathcal{S} \in \mathbb{R}^{n \times r}$ is the selection matrix which singles out the columns indexed on $\mathcal{S}$, such that $\mathbf{v}_\mathcal{S} = P_\mathcal{S}^T \mathbf{v}$.

We let $\mathbf{e}_i$ denote the ith vector of the canonical basis in $\mathbb{R}^n$ (where n is always clear from the context), whose entries are all zero except for the ith one which is unitary. We say that a vector $\mathbf{v} \in \mathbb{R}^n$ is an *i-monomial vector* if it has the same nonzero pattern as $\mathbf{e}_i$, namely $\overline{\mathrm{ZP}}(\mathbf{v}) = \overline{\mathrm{ZP}}(\mathbf{e}_i) = \{i\}$. To every subset $\mathcal{S}$ of $\{1, 2, \dots, n\}$ we may associate the nonnegative vector

$$\mathbf{e}_\mathcal{S} := \sum_{i \in \mathcal{S}} \mathbf{e}_i,$$

which is the only vector having non-zero pattern equal to $\mathcal{S}$ and whose nonzero entries are all unitary.

The symbol $\mathbb{R}_+$ denotes the semiring of nonnegative real numbers. By a *nonnegative matrix* we mean a matrix whose entries are all nonnegative and hence in $\mathbb{R}_+$. A nonnegative matrix A is typically denoted, for the sake of compactness, by means of the notation $A \geq 0$. If $A \geq 0$ and at least one entry is positive, A is a *positive matrix*[1] ($A > 0$), while if all its entries are positive it is a *strictly positive matrix* ($A \gg 0$). The same notation is adopted for nonnegative, positive and strictly positive vectors.

In our context, a *monomial matrix* is a nonsingular square nonnegative matrix whose columns are (of course, distinct) monomial vectors. A monomial matrix can always be seen as the product of a permutation matrix and a diagonal matrix with positive diagonal entries. Also, we denote by $\bar{1}$ the vector (whose size is clear from the context) with all entries equal to 1. The *spectral radius* of a positive matrix A is defined as the modulus of its largest eigenvalue, and denoted by $\rho(A)$. The Perron-Frobenius Theorem [4, 10, 37] ensures that $\rho(A)$ is always an eigenvalue of A_+, corresponding to a positive eigenvector.

[1] In the literature, see, e.g., [26, 47], positive matrices have also been referred to as semi-positive matrices, while strictly positive matrices as positive matrices.

Definition 1.1 *Given a matrix function* $M(t), t \in \mathbb{R}_+$, *(in particular, a vector function) taking values in* $\mathbb{R}_+^{k\times p}$, *a real number* λ *and a nonnegative integer* m, *we say that* $M(t)$ *has the* pseudo-exponential growth rate (λ, m) *if there exists a strictly positive matrix* $M_\infty \in \mathbb{R}_+^{k\times p}$ *such that*

$$\lim_{t\to+\infty} \frac{M(t)}{e^{\lambda t}\frac{t^m}{m!}} = M_\infty.$$

When so, we write $M(t) \sim e^{\lambda t}\frac{t^m}{m!}$.

A *Metzler matrix* is a real square matrix, whose off-diagonal entries are nonnegative. If A is an $n \times n$ Metzler matrix, then there exist a nonnegative matrix $A_+ \in \mathbb{R}_+^{n\times n}$ and a nonnegative number α such that $A = A_+ - \alpha I_n$. As a consequence, the spectrum of A, $\sigma(A)$, is obtained from the spectrum of A_+ by simple translation. This ensures, in particular, that [48]:

- $\lambda_{\max}(A) = \rho(A_+) - \alpha \in \sigma(A)$ is a real dominant eigenvalue, by this meaning that $\lambda_{\max}(A) > \mathrm{Re}(\lambda), \forall\, \lambda \in \sigma(A), \lambda \neq \lambda_{\max}(A)$;
- there exists a nonnegative eigenvector $\mathbf{v}_1$ corresponding to $\lambda_{\max}(A)$.

To every $n\times n$ Metzler matrix A we associate [10, 47] a *directed graph* $\mathcal{G}(A)$ of *order* n, with vertices indexed by $1, 2, \dots, n$. There is an arc (j, i) from j to i if and only if $[A]_{ij} \neq 0$. We say that vertex i is *accessible* from j if there exists a path (i.e., a sequence of adjacent arcs $(j, i_1), (i_1, i_2), \dots, (i_{k-1}, i)$) in $\mathcal{G}(A)$ from j to i (equivalently, $\exists\, k \in \mathbb{N}$ such that $[A^k]_{ij} \neq 0$). Two distinct vertices i and j are said to *communicate* if each of them is accessible from the other. Each vertex is assumed to communicate with itself. The concept of communicating vertices allows to partition the set of vertices $\langle n\rangle$ into *communicating classes*, say $\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_\ell$. Given an index $i \in \{1, \dots, n\}$, we let $\mathcal{C}(i)$ be the index of the irreducibility class the vertex i belongs to (w.r.t. the directed graph $\mathcal{G}(A)$). A directed graph $\mathcal{G}(A)$ is *strongly connected* if it consists of a single communicating class.

The *reduced graph* $\mathcal{R}(A)$ [47] associated with A (with $\mathcal{G}(A)$) is the (acyclic) graph having the classes $\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_\ell$ as vertices. There is an arc (j, i) from $\mathcal{C}_j$ to $\mathcal{C}_i$ if and only if $\mathrm{block}_{(i,j)}[A] \neq 0$. With any class $\mathcal{C}_i$ we associate two index sets:

$$\mathcal{A}(\mathcal{C}_i) := \{j :\ \text{the class } \mathcal{C}_j \text{ has access to the class } \mathcal{C}_i\}$$
$$\mathcal{D}(\mathcal{C}_i) := \{j :\ \text{the class } \mathcal{C}_j \text{ is accessible from the class } \mathcal{C}_i\}.$$

Each class $\mathcal{C}_i$ is assumed to have access to itself. Any (acyclic) path (i_1, i_2), $(i_2,\ i_3), \dots, (i_{k-1}, i_k)$ in $\mathcal{R}(A)$ identifies a *chain of classes* $(\mathcal{C}_{i_1}, \mathcal{C}_{i_2}, \dots, \mathcal{C}_{i_k})$, having $\mathcal{C}_{i_1}$ as *initial class* and $\mathcal{C}_{i_k}$ as *final class.*

An $n \times n$ Metzler matrix A is *reducible* if there exists a permutation matrix P such that

$$P^T A P = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix},$$

where A_{11} and A_{22} are square (nonvacuous) matrices, otherwise it is *irreducible*. It follows that 1×1 matrices are always irreducible. In general, given a square Metzler matrix A, a permutation matrix P can be found such that

$$P^T A P = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1\ell} \\ & A_{22} & \dots & A_{2\ell} \\ & & \ddots & \vdots \\ & & & A_{\ell\ell} \end{bmatrix}, \tag{1.2}$$

where each A_{ii} is irreducible. (1.2) is usually known as *Frobenius normal form* of A [37]. Clearly, the directed graphs $\mathcal{G}(A)$ and $\mathcal{G}(P^T A P)$ are isomorphic and the irreducible matrices $A_{11}, A_{22}, \dots, A_{\ell\ell}$ correspond to the communicating classes $\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_\ell$ of $\mathcal{G}(P^T A P)$ (coinciding with those of $\mathcal{G}(A)$, after a suitable relabelling).

When dealing with the graph of a matrix in Frobenius normal form (1.2), we will let $\mathcal{C}_i = \{(n_1+n_2+\dots+n_{i-1})+1,\ (n_1+\dots+n_{i-1})+2,\ \dots,\ (n_1+\dots+n_{i-1})+n_i\}$ denote[2] the ith communicating class of $\mathcal{G}(A)$, associated with A_{ii}. For every $i \in \langle \ell \rangle$, $\mathcal{A}(\mathcal{C}_i) \subseteq \{i, i+1, \dots, \ell\}$, while $\mathcal{D}(\mathcal{C}_i) \subseteq \{1, 2, \dots, i\} = \langle i \rangle$, so that $\mathcal{A}(\mathcal{C}_i) \cap \mathcal{D}(\mathcal{C}_i) = \{i\}$. On the other hand, if $i > j$ then $\mathcal{A}(\mathcal{C}_i) \cap \mathcal{D}(\mathcal{C}_j) = \emptyset$, while if $i < j$ then $\mathcal{A}(\mathcal{C}_i) \cap \mathcal{D}(\mathcal{C}_j) \neq \emptyset \Leftrightarrow i \in \mathcal{D}(\mathcal{C}_j) \Leftrightarrow j \in \mathcal{A}(\mathcal{C}_i)$. Also, a class $\mathcal{C}_i$ is *initial* if $\mathcal{A}(\mathcal{C}_i) = \{i\}$, and it is *distinguished* [47] if $\lambda_{\max}(A_{ii}) > \lambda_{\max}(A_{jj})$ for every $j \in \mathcal{D}(\mathcal{C}_i), j \neq i$. If A is irreducible ($\mathcal{G}(A)$ has a single communicating class), then $\lambda_{\max}(A)$ is a simple eigenvalue and the corresponding nonnegative eigenvector $\mathbf{v}_1$ is strictly positive. Moreover, the only nonnegative eigenvector or generalized eigenvector of A is $\mathbf{v}_1$ (and its positive mutiples). The *cyclicity index* $c(A)$ [4] of an irreducible matrix A is the greatest common divisor of the lengths of the cycles in $\mathcal{G}(A)$. If $c(A) = 1$, A is *primitive.*

We introduce a basis of eigenvectors and generalized eigenvectors of a Metzler matrix A, whose nonzero patterns are related to the block-triangular

[2] We assume, by definition, $n_0 := 0$.

structure of the Frobenius normal form of A. Such a vector basis will be rather useful for describing and investigating the asymptotic behavior of the exponential matrix e^{At}.

Definition 1.2 *Let A be an $n \times n$ Metzler matrix in Frobenius normal form (1.2). An ordered family $\mathcal{B}_e = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ of linearly independent eigenvectors and generalized eigenvectors of A is said to be an* echelon basis *for A if*

- $\mathbf{v}_j, j \in \mathcal{C}_1$, *are n_1 (possibly generalized) eigenvectors with $\overline{\mathrm{ZP}}(\mathbf{v}_j) \subseteq \mathcal{C}_1$;*
- $\mathbf{v}_j, j \in \mathcal{C}_2$, *are n_2 (possibly generalized) eigenvectors with $\overline{\mathrm{ZP}}(\mathbf{v}_j) \subseteq \mathcal{C}_1 \cup \mathcal{C}_2$ and $\overline{\mathrm{ZP}}(\mathbf{v}_j) \cap \mathcal{C}_2 \neq \emptyset$;*
- $\dots$
- $\mathbf{v}_j, j \in \mathcal{C}_\ell$, *are n_ℓ (possibly generalized) eigenvectors with $\overline{\mathrm{ZP}}(\mathbf{v}_j) \subseteq \mathcal{C}_1 \cup \mathcal{C}_2 \cup \dots \cup \mathcal{C}_\ell$ and $\overline{\mathrm{ZP}}(\mathbf{v}_j) \cap \mathcal{C}_\ell \neq \emptyset$.*

When so, we say that the vector $\mathbf{v}_j \in \mathcal{B}_e$ corresponds to the class $\mathcal{C}_i$ *if $j \in \mathcal{C}_i$, or, equivalently,* $\mathrm{block}_i[\mathbf{v}_j] \neq 0$ *and* $\mathrm{block}_h[\mathbf{v}_j] = 0$ *for every $h > i$.*

Notice that if $\mathbf{v}_j \in \mathcal{B}_e$ is a generalized eigenvector of A of order (also called "height" [22, 23]) k corresponding to the class $\mathcal{C}_i$ and to the eigenvalue $\lambda \in \sigma(A)$, then $\mathrm{block}_i[\mathbf{v}_j]$ is a generalized eigenvector of A_{ii} (of course, corresponding to the same eigenvalue λ, which thus must be in $\sigma(A_{ii})$ and possibly in the spectrum of some other diagonal block) of order not greater than k.

Given a set of n-dimensional matrices $\mathcal{A} := \{A_1, A_2, \dots, A_p\}$, positive integers $l, k \in \mathbb{Z}_+$, with $l \leq k$, and a sequence $\{\sigma(t)\}_{t \in \mathbb{N}}$ with $\sigma(t) \in \langle p \rangle$, we set

$$A_\sigma\big|_l^{k-1} := \begin{cases} A_{\sigma(k-1)} A_{\sigma(k-2)} \dots A_{\sigma(l)}, & \text{if } l < k; \\ I_n, & \text{if } l = k. \end{cases}$$

Clearly, for every nonnegative integer h such that $l \leq h \leq k-1$, we have $A_\sigma\big|_l^{k-1} = A_\sigma\big|_{h+1}^{k-1} A_\sigma\big|_l^h$.

If the matrices in $\mathcal{A}$ are $n \times n$ Metzler matrices, to every $\mathcal{S} \subseteq \{1, \dots, n\}$ we associate the set $\mathcal{I}_\mathcal{S} := \{i \in \langle p \rangle : \overline{\mathrm{ZP}}(e^{A_i} \mathbf{e}_\mathcal{S}) = \mathcal{S}\}$.

Basic definitions and results about cones may be found, for instance, in [4]. We recall here only those facts that will be used within this work. A set $\mathcal{K} \subset \mathbb{R}^n$ is said to be a *cone* if $\alpha\mathcal{K} \subset \mathcal{K}$ for all $\alpha \geq 0$; a cone is *convex* if it contains, with any two points, the line segment between them. A convex cone $\mathcal{K}$ is *solid* if the interior of $\mathcal{K}$ is nonempty, and it is *pointed* if $\mathcal{K} \cap \{-\mathcal{K}\} = \{0\}$. A closed, pointed, solid convex cone is called a *proper cone*. A cone $\mathcal{K}$ is said to be *polyhedral* if it can be expressed as the set of nonnegative linear combinations of a finite set of *generating vectors*. This amounts to saying that a positive integer k and an $n \times k$ matrix C can be found, such that $\mathcal{K}$ coincides with the set of nonnegative combinations of the columns of C. In this case, we adopt the notation $\mathcal{K} := \text{Cone}(C)$. A proper polyhedral cone $\mathcal{K}$ in $\mathbb{R}^n$ is said to be *simplicial* if it admits n linearly independent generating vectors. In other words, $\mathcal{K} := \text{Cone}(C)$ for some nonsingular square matrix C. When so, a vector $\mathbf{v}$ belongs to the boundary of the simplicial cone $\mathcal{K} = \text{Cone}(C)$ if and only if $\mathbf{v} = C\mathbf{u}$ for some nonnegative vector $\mathbf{u}$, with $\text{ZP}(\mathbf{u}) \neq \emptyset$.

To efficiently introduce our results, we also need some definitions borrowed from the algebra of non-commutative polynomials [44]. Given the *alphabet* $\Xi = \{\xi_1, \xi_2, \ldots, \xi_p\}$, the free monoid Ξ^* with base Ξ is the set of all *words* $w = \xi_{i_1}\xi_{i_2}\cdots\xi_{i_k}, k \in \mathbb{N}, \xi_{i_h} \in \Xi$. The integer k is called the *length* of w and is denoted by $|w|$, while $|w|_i$ represents the number of occurencies of ξ_i in w. If $\tilde{w} = \xi_{j_1}\xi_{j_2}\cdots\xi_{j_p}$ is another element of Ξ^*, the product is defined by concatenation $w\tilde{w} = \xi_{i_1}\xi_{i_2}\cdots\xi_{i_m}\xi_{j_1}\xi_{j_2}\cdots\xi_{j_p}$. This produces a monoid with $\varepsilon = \emptyset$, the empty word, as unit element. Clearly, $|w\tilde{w}| = |w| + |\tilde{w}|$ and $|\varepsilon| = 0$.

$\mathbb{C}\langle\xi_1, \xi_2, \ldots, \xi_p\rangle$ is the algebra of polynomials in the noncommuting indeterminates $\xi_1, \xi_2, \ldots, \xi_p$. For every family of p matrices in $\mathbb{C}^{n\times n}$, $\mathcal{A} := \{A_1, A_2, \ldots, A_p\}$, the map ψ defined on $\{\varepsilon, \xi_1, \xi_2, \ldots, \xi_p\}$ by the assignments $\psi(\varepsilon) = I_n$ and $\psi(\xi_i) = A_i$, $i = 1, 2, \ldots, p$, uniquely extends to an algebra morphism of $\mathbb{C}\langle\xi_1, \xi_2, \ldots, \xi_p\rangle$ into $\mathbb{C}^{n\times n}$ (as an example, $\psi(\xi_1\xi_2) = A_1A_2 \in \mathbb{C}^{n\times n}$). If w is a word in Ξ^* (i.e. a monic monomial in $\mathbb{C}\langle\xi_1, \xi_2, \ldots, \xi_p\rangle$), the ψ-image of w is denoted by $w(A_1, A_2, \ldots, A_p)$. Finally, $w \in \Xi^*$ is an *annihilating word* for the matrix family $\mathcal{A}$ if $w(A_1, A_2, \ldots, A_p) = 0$.

Chapter 2

Switched Positive Systems

Switched linear systems have attracted the interest of several scientists, in the last ten years. Initially treated as special cases of the broader class of hybrid systems, they have later gained complete autonomy and have been the object of an in-depth analysis. While the first contributions were almost exclusively concentrated on the stability and stabilizability properties [34, 55], nowadays several other issues have been investigated and, in particular, structural properties, like reachability, controllability and observability have been explored [20, 51, 54, 56].

Despite of the numerous research efforts, these issues still offer a quite interesting set of open problems. Indeed, structural properties have found a rather complete characterization for the class of continuous-time switched systems and for the class of reversible discrete-time switched systems (by this meaning that the system matrices of all the subsystems among which the system commutes are nonsingular). The non-reversible discrete-time case, however, still deserves investigation, since necessary and sufficient conditions for reachability (and observability) have been provided only under certain structural constraints (see, e.g., [16]). However, it must be pointed out that some interesting properties of the controllable sets for (both reversible and non-reversible) discrete-time switched systems have been investigated in the pioneering works of Conner and Stanford [30, 31, 49].

Positive systems, on the other hand, are linear systems in which the state variables are always positive, or at least nonnegative, in value. These systems have received considerable attention in the literature, as they naturally arise in various fields such as bioengineering (compartmental models), economic modelling, behavioral science, and stochastic processes (Markov chains), where the state variables represent quantities that may not have meaning unless nonnegative. The theory of positive systems [17] is deep and elegant, and firmly built upon the classical positive matrix theory, which has its cornerstone in the celebrated Perron-Frobenius theorem [4]. While in the past the positivity constraint has often been ignored or accomodated in order to take advantage of the well-developed theory of linear systems, in the last two decades system issues have been addressed specifically for positive systems, by taking advantage of the powerful tools coming out of positive matrix theory and, even more, of graph theory. In particular, the analysis of controllability and reachability properties of positive discrete-time systems has been the object of a noteworthy interest [9, 13, 53].

Switched positive systems deserve investigation both for theoretical reasons and for practical applications. Indeed, switching among different system models naturally arises as a way to mathematically formalize the fact that the system laws change under different operating conditions. This is true, e.g., when resorting to compartmental models; each of them may satisfactorily capture the behavior of a physiological system only under specific working conditions, so when these conditions change the system model has to switch to a different structure. For instance, the insulin-sugar metabolism [11, 28] is captured by two different compartmental models: one for the steady-state and the other for describing the evolution under perturbed conditions (following an oral assumption or an intravenous injection).

Similarly, different positive systems, which arise when discretizing linear differential equations describing processes whose state variables are temperatures, pressures, population levels, etc., may undergo different working conditions and, consequently, switch among different mathematical models.

2.1 Continuous-time Positive Switched Systems

A *continuous-time switched positive system* is described by the following equation

$$\dot{x}(t) = A_{\sigma(t)}x(t) + B_{\sigma(t)}u(t), \qquad t \in \mathbb{R}_+, \tag{2.1}$$

where $x(t)$ and $u(t)$ denote the n-dimensional state variable and the m-dimensional input[1], respectively, at the time instant t, σ is a switching se-

[1]The extension of the current analysis to the case when the input size varies as $\sigma(t)$ varies within $\mathcal{P}$ is rather straightforward and does not affect at all the results presented

quence, taking values in a finite set $\mathcal{P} = \{1, 2, \dots, p\}$.

We assume that the switching sequence is piece-wise constant, and hence in every time interval $[0, t]$ there is a finite number of discontinuities, which corresponds to a finite number of switching instants $0 = t_0 < t_1 < \dots < t_k < t$. Also, we assume that, at the switching time t_ℓ, σ is right continuous. For each $i \in \mathcal{P}$, the pair (A_i, B_i) represents a continuous-time positive system, which means that A_i is an $n \times n$ Metzler matrix and B_i is an $n \times m$ nonnegative matrix.

Given a time interval $[0, t[$ and a switching sequence $\sigma : [0, t]$, corresponding to a set of switching instants $\{t_0, t_1, \dots, t_k\}$ satisfying $0 = t_0 < t_1 < \dots < t_k < t$, we first observe that the state at the time instant t, starting from the initial condition $x(0)$ and under the action of the soliciting input $u(\tau), \tau \in [0, t[$, can be expressed as follows:

$$\begin{aligned} x(t) &= e^{A_{i_k}(t-t_k)} \dots e^{A_{i_0}(t_1-t_0)} x(0) + \\ &+ e^{A_{i_k}(t-t_k)} \dots e^{A_{i_1}(t_2-t_1)} \int_{t_0}^{t_1} e^{A_{i_0}(t_1-\tau)} B_{i_0} u(\tau) \mathrm{d}\tau + \\ &+ e^{A_{i_k}(t-t_k)} \dots e^{A_{i_2}(t_3-t_2)} \int_{t_1}^{t_2} e^{A_{i_1}(t_2-\tau)} B_{i_1} u(\tau) \mathrm{d}\tau + \\ &+ \dots + \int_{t_k}^{t} e^{A_{i_k}(t-\tau)} B_{i_k} u(\tau) \mathrm{d}\tau, \end{aligned} \tag{2.2}$$

where $i_\ell = \sigma(t_\ell), \ell = 0, 1, \dots, k$.

The definitions of controllability and reachability for switched positive systems may be given by suitably adjusting the analogous definitions given in [20, 51], in order to introduce the nonnegativity constraint on the state and input variables.

Definition 2.1 *A state $x_f \in \mathbb{R}^n_+$ is said to be (positively) reachable if there exist some time instant $t_f > 0$, a switching sequence $\sigma : [0, t_f] \to \mathcal{P}$ and an input $u : [0, t_f] \to \mathbb{R}^m_+$ that lead the state trajectory from $x(0) = 0$ to $x(t_f) = x_f$.*

A switched positive system is said to be (positively) reachable if every state $x_f \in \mathbb{R}^n_+$ is (positively) reachable.

In the sequel, we will omit the specification "positively", since it is clear that we will steadily work under this assumption.

in this paper.

Definition 2.2 *A state $x_0 \in \mathbb{R}^n_+$ is said to be zero controllable if there exist some time instant $t_f > 0$, a switching sequence $\sigma : [0, t_f] \to \mathcal{P}$ and an input $u : [0, t_f] \to \mathbb{R}^m_+$ that lead the state trajectory from $x(0) = x_0$ to $x(t_f) = 0$.*

A switched positive system is said to be zero controllable if every state $x_0 \in \mathbb{R}^n_+$ is zero controllable.

2.2 Discrete-time Positive Switched Systems

A *discrete-time positive switched system* is described by a first-order difference equation of the following type:

$$x(t+1) = A_{\sigma(t)}x(t) + B_{\sigma(t)}u(t), \qquad t \in \mathbb{Z}_+, \tag{2.3}$$

where $x(t)$ and $u(t)$ denote the n-dimensional state variable and the m-dimensional input variable, respectively, at the time instant t, while σ is a switching sequence, defined on $\mathbb{Z}_+$ and taking values in a finite set $\mathcal{P} = \{1, 2, \dots, p\}$. For each $i \in \mathcal{P}$, the pair (A_i, B_i) represents a discrete-time positive system, which means that $A_i \in \mathbb{R}^{n\times n}_+$ and $B_i \in \mathbb{R}^{n\times m}_+$.

The definitions of controllability and reachability for discrete-time positive switched systems may be given by suitably adjusting the analogous definitions given in [20, 56], in order to introduce the nonnegativity constraint on the state and input variables. As for continuous-time systems, in the sequel, the specification "positively" will be omitted.

Definition 2.3 *A state $x_f \in \mathbb{R}^n_+$ is said to be* reachable at time $k \in \mathbb{N}$ *if there exist a switching sequence $\sigma : [0, k-1] \to \mathcal{P}$ and an input sequence $u : [0, k-1] \to \mathbb{R}^m_+$ that lead the state trajectory from $x(0) = 0$ to $x(k) = x_f$. System (2.3) is said to be* reachable *if every state $x_f \in \mathbb{R}^n_+$ is reachable at some time instant k.*

Definition 2.4 *A state $x_0 \in \mathbb{R}^n_+$ is said to be* zero controllable at time $k \in \mathbb{N}$ *if there exist a switching sequence $\sigma : [0, k-1] \to \mathcal{P}$ and an input sequence $u : [0, k-1] \to \mathbb{R}^m_+$ that lead the state from $x(0) = x_0$ to $x(k) = 0$. System (2.3) is said to be* zero controllable *if every state is zero controllable at some time instant k.*

Since reachability and zero controllability properties always refer to a finite time interval, focusing on the value of the state at the final instant k, only the values of the switching sequence σ within $[0, k-1]$ are relevant. So, we refer to the cardinality of the discrete time interval $[0, k-1]$ as to the *length* of the switching sequence σ and we denote it by $|\sigma|$ (in this case, $|\sigma| = k$).

When reachability (zero controllability) property is ensured, a natural goal one may want to pursue is that of determining the minimum number of steps required to reach (to control to zero) every nonnegative state. This leads to the definition of reachability (controllability) index.

Definition 2.5 *Given a reachable (zero controllable) switched system (2.3), we define its* reachability index *(*controllability index*) as*

$$I_R := \sup_{x \in \mathbb{R}^n_+} \min\{k : x \textit{ is reachable at time } k\} \tag{2.4}$$

$$(I_C := \sup_{x \in \mathbb{R}^n_+} \min\{k : x \textit{ is zero controllable at time } k\}). \tag{2.5}$$

As we will see, while for zero controllable systems the index I_C always takes finite values, reachable systems can be found endowed with an infinite[2] I_R. This fact represents a significant difference with respect to both standard switched systems and positive systems.

It is first convenient to provide the explicit expression of the state at any time instant $k \in \mathbb{N}$, starting from the initial condition $x(0)$, under the effect of the input sequence $u(0), u(1), \dots, u(k-1)$, and of the switching sequence $\sigma(0), \sigma(1), \dots, \sigma(k-1)$. It turns out (see, for instance, [20]) that

$$x(k) = A_\sigma\big|_0^{k-1} x(0) + A_\sigma\big|_1^{k-1} B_{\sigma(0)} u(0) + A_\sigma\big|_2^{k-1} B_{\sigma(1)} u(1) + \cdots + B_{\sigma(k-1)} u(k-1). \tag{2.6}$$

[2]It is worthwhile to underline that even when I_R is infinite, each single nonnegative state can be reached in a finite number of steps. However, such a number of steps may be arbitrarily high. This concept must not be confused with the weak reachability property of positive systems [53], which allows to reach certain states only asymptotically.

Chapter 3

Zero Controllability of Discrete-time Positive Switched Systems

As for standard positive systems, the nonnegativity of the soliciting input constrains the forced component of the state evolution in (2.6) to be nonnegative. As a consequence, the goal of forcing to zero the nonnegative initial state $x(0)$ is by no means simplified by the use of a nonnegative input, and either the free state evolution goes to zero in a finite number of steps, or there is no way to control to zero the initial state.

3.1 Existence of an annihilating word

Differently from the standard positive case, where zero controllability is simply equivalent to the nilpotency of the system matrix [13, 53], when dealing with switched systems we have some spare degree of freedom to exploit: the switching sequence σ. Of course, if the system is zero controllable, each nonnegative initial state may be controlled to zero in a minimum number of steps, by choosing a switching sequence σ tailored to the specific state or, more precisely, to its nonzero pattern. However, as shown in Proposition 3.1, below, zero controllability property is equivalent to the existence of a single switching sequence σ of length k such that $A_\sigma\big|_0^{k-1} = 0$. This amounts to saying that some $w \in \Xi^*$ exists, with $|w| = k$, such that w is an annihilating word for $\{A_1, A_2, \ldots, A_p\}$, i.e. $w(A_1, A_2, \ldots, A_p) = 0$.

In the literature, the existence of such an annihilating word is known as *mortality* of the set of matrices [6]. The problem of deciding whether a certain set of matrices is mortal or not is known to be NP-complete, and an approach for building an algorithm for the mortality problem for matrices with nonnegative entries has been suggested in [5] (p. 286), where it is also remarked that the problem can be reduced to the analogous one with boolean entries.

Proposition 3.1 *For the switched system (2.3), the following facts are equivalent:*

i) the system is zero controllable;

ii) there exists $w \in \Xi^*$ such that $w(A_1, A_2, \dots, A_p) = 0_{n\times n}$;

iii) there exists $\tilde{w} \in \Xi^*$ such that $\tilde{w}(A_1, A_2, \dots, A_p)$ is nilpotent.

Proof: i) $\Rightarrow$ ii) If the system is zero controllable, then, in particular, $x(0) = \bar{1}$ is zero controllable. Hence there exists a switching sequence σ such that $A_\sigma\big|_0^{k-1}\bar{1} = 0$. But then, $A_\sigma\big|_0^{k-1} = A_{\sigma(k-1)} \dots A_{\sigma(1)} A_{\sigma(0)} = 0$, thus proving ii).

ii) $\Rightarrow$ i) and ii) $\Leftrightarrow$ iii) are obvious. ■

It is worthwhile noticing that if $w = \xi_{i_1}\xi_{i_2}\dots\xi_{i_m}$ corresponds to a nilpotent matrix then every word obtained from w by means of circular permutation of its symbols, namely $\xi_{i_k}\xi_{i_{k+1}}\dots\ \dots\xi_{i_m}\xi_{i_1}\dots\xi_{i_{k-1}}$, with $1 < k \leq m$, corresponds to a nilpotent matrix, too. This is due to the fact that if $(A_{i_1}A_{i_2}\dots A_{i_m})^n = 0$ then $\left(A_{i_k}A_{i_{k+1}}\dots A_{i_m}A_{i_1}\dots A_{i_{k-1}}\right)^n = 0$.

A set of necessary conditions for the existence of an annihilating word is presented.

Proposition 3.2 *If there exists some annihilating word $w \in \Xi^*$ and we let $I = \{i_1, i_2, \dots, i_r\}$ be the set of distinct indices in $\mathcal{P}$ such that $|w|_i > 0$, then*

i) $A_{i_1} * A_{i_2} * \dots * A_{i_r}$ is nilpotent, where $*$ represents the Hadamard product (entry by entry) of the r matrices;

ii) there is at least one index $i \in I$ such that the matrix A_i has at least one null column (one null row).

Proof: Assume w.l.o.g. that all matrices A_i's explicitly appear in $w(A_1, A_2, \ldots, A_p)$, and hence $I = \mathcal{P}$.

i) Let P be a permutation matrix that reduces the nonnegative matrix $A := A_1 * A_2 * \cdots * A_p$ to Frobenius normal form, i.e.,

$$P^T AP = (P^T A_1 P) * (P^T A_2 P) * \ldots * (P^T A_p P) = \begin{bmatrix} Q_1 & \star & \star \\ & \ddots & \star \\ & & Q_r \end{bmatrix},$$

where every diagonal block Q_i is either $0_{1\times 1}$ or irreducible. If A is not nilpotent, then $\exists\ i$ such that Q_i is nonzero and irreducible [53]. Let $\mathcal{C}_i$ be the communicating class corresponding to Q_i. Consider, now, the digraph associated with $P^T AP$, and let h be an arbitrary vertex in $\mathcal{C}_i$. By the irreducibility of Q_i (the strong connectedness of $\mathcal{C}_i$), for every $k > 0$, there is a path of length k in $D(P^T AP)$, starting from vertex h and reaching some other vertex $\tilde{h}$ in $\mathcal{C}_i$. Since the arcs of this path belong to $D(P^T A_i P)$ for every index i, this means that for every $h \in \mathcal{C}_i$ and every $k > 0$ we may find a path of length k starting from h, consisting of arcs arbitrarily selected in the various $D(P^T A_i P)$, and leading to some other vertex in $\mathcal{C}_i$. But then, $\forall\, w \in \Xi^*$, the vector $w(A_1, \ldots, A_p)\mathbf{e}_h$ has (at least) one nonzero entry corresponding to some index $\tilde{h}$ corresponding to Q_i, thus contradicting condition $w(A_1, A_2, \ldots, A_p) = 0$. As a consequence, A must be nilpotent.

ii) Notice, first, that if the ith column of BC, with $B, C \in \mathbb{R}_+^{n\times n}$, is zero, then either the ith column of C is zero, or B has at least one zero column. So, if $w(A_1, \ldots, A_p)$ is the zero matrix, and $w = \xi_{i_1}\xi_{i_2}\ldots\xi_{i_k}$, then either A_{i_1} or $A_{i_2}\ldots A_{i_k}$ has a zero column. So, by recursively applying this reasoning, we show that there exists $i_l \in I$ such that A_{i_l} has a zero column.

On the other hand, if $BC = 0$ then $C^T B^T = 0$, and the result applies also to the rows. ∎

3.2 Zero controllability algorithm

In this section, a "branch and bound" procedure, which allows to test whether an annihilating word exists, is presented. This procedure proves to stop within 2^n steps. So, not only is the procedure computationally useful, but it provides an upper bound on the index I_C.

In order to sketch a compact algorithm, we introduce the following definition.

Definition 3.3 *Given $A \in \mathbb{R}_+^{n\times n}$, the* positive kernel *of A is*

$$\ker_+(A) := \big\{v \in \mathbb{R}_+^n : \ A\,v = 0\big\}.$$

$\ker_+(A)$ is a polyhedral (convex) cone generated by a subset of the canonical basis. Indeed, upon setting $\mathcal{I} := \{i \in \{1,2,\dots,n\} : \mathbf{e}_i \in \ker_+(A)\}$, it is easily seen that $v = \sum_{i=1}^n v_i\,\mathbf{e}_i \in \mathbb{R}_+^n$ belongs to $\ker_+(A)$ if and only if $v = \sum_{i\in\mathcal{I}} v_i\,\mathbf{e}_i$. So, $\ker_+(A)$ may be uniquely associated to the index set $\mathcal{I}$ of the canonical vectors which generate it. In the following, given a word $w \in \Xi^*$, we denote by $\mathcal{I}_w$ the index set corresponding to $\ker_+(w(A_1, A_2, \dots, A_p))$. Clearly, w is an annihilating word if and only if $\mathcal{I}_w = \{1,2,\dots,n\}$. The following lemma will be useful in the sequel.

Lemma 3.4 *Let w_1 and w_2 be in Ξ^*. If $\mathcal{I}_{w_1} \subseteq \mathcal{I}_{w_2}$, then $\mathcal{I}_{w_1 w} \subseteq \mathcal{I}_{w_2 w}$ for each $w \in \Xi^*$.*

Proof: If $\mathcal{I}_{w_1} \subseteq \mathcal{I}_{w_2}$ then $\ker_+(w_1(A_1,\dots,A_p)) \subseteq \ker_+(w_2(A_1,\dots,A_p))$, and hence, for every matrix product $w(A_1,\dots,A_p)$ and every $x \in \mathbb{R}_+^n$, condition

$$w_1(A_1,\dots,A_p)w(A_1,\dots,A_p)x = 0$$

implies

$$w_2(A_1,\dots,A_p)w(A_1,\dots,A_p)x = 0.$$

This equivalently means that $\mathcal{I}_{w_1 w} \subseteq \mathcal{I}_{w_2 w}$. ∎

Algorithm: The variables we deal with are the integer `Length`, the set `ActiveWords` and, for each "active word" $w \in \Xi^*$ (namely, every element of `ActiveWords`), the Index Set $\mathcal{I}_w$ of the associated positive kernel.

Step 1: Initialization. Set:

- `Length`$:= 0$.
- `ActiveWords`$:=\{\varepsilon\}$, ε being the empty word, corresponding to the identity matrix I_n.
- The active word ε is associated with the index set $\mathcal{I}_1 = \emptyset$.

Step 2: Analysis of the active words of maximum length. We define:

$$\mathtt{AW}_{\mathtt{Length}} := \mathtt{ActiveWords} \cap \{w \in \Xi^* : |w| = \mathtt{Length}\}.$$

If $\mathtt{AW}_{\mathtt{Length}} = \emptyset$, then the algorithm stops and no annihilating word exists. Otherwise, $\forall w \in \mathtt{AW}_{\mathtt{Length}}$ introduce the p new words $w\xi_i, i = 1, \dots, p$, and their associated index set $I_{w\xi_i}$.
Step 3: Update of `ActiveWords`. Set $\mathtt{Length} := \mathtt{Length} + 1$, and ordinately consider every word $\tilde{w} = w\xi_i$ with $w \in \mathtt{AW}_{\mathtt{Length-1}}$. If $\mathcal{I}_{\tilde{w}}$ coincides with $\{1, \dots, n\}$, the algorithm stops and $\tilde{w}$ is an annihilating word of length `Length`. Otherwise, compare $\mathcal{I}_{\tilde{w}}$ with the index sets of all active words. If there exists an active word w such that $\mathcal{I}_w \supseteq \mathcal{I}_{\tilde{w}}$, then neglect $\tilde{w}$ and move to the following word. If not, set $\mathtt{ActiveWords} := \mathtt{ActiveWords} \cup \{\tilde{w}\}$. When all words $\tilde{w} = w\xi_i$, with $w \in \mathtt{AW}_{\mathrm{Length}-1}$, have been considered, go back to step 2.

We want to show that the algorithm stops, providing an annihilating word of minimum length, if any. To this end, observe that two conditions cause the algorithm to stop:
Case 1: A word $\tilde{w}$ is found, associated with $\mathcal{I}_{\tilde{w}} = \{1, \dots, n\}$. If so, $\tilde{w}$ is clearly an annihilating word of minimal length, as all words of smaller length were not associated with that index set.
Case 2: There exists $k \in \mathbb{Z}_+$ such that all $w \in \mathtt{AW}_k$ produce new words $\tilde{w} = w\xi_i$ with $\mathcal{I}_{\tilde{w}} \subseteq \mathcal{I}_{\hat{w}}, \exists\ \hat{w} \in \mathtt{ActiveWords}$. In this case, it is not worth further exploring the words of higher length, as, indeed, by Lemma 3.4, they would be associated with positive kernels which could never be greater than those already obtained, and hence could never coincide with $\mathbb{R}^n_+$.

Finally, since the index sets $\mathcal{I}_w$ are all subsets of $\{1, 2, \dots, n\}$, the maximum number of distinct index sets we may encounter in this algorithm is exactly 2^n and the maximum value that the variable `Length` may achieve is exactly 2^n. This case occurs if and only if for every value of `Length` there is only one active word, and the index set $\{1, \dots, n\}$ is obtained for $\mathtt{Length} = 2^n - 1$.

We provide, below, a simple example where the bound on I_C is reached.

Example 3.5 *Consider the positive system, switching among the following subsystems:*

$$(A_1, B_1) = \left(\begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right),\ (A_2, B_2) = \left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right),$$

$$(A_3, B_3) = \left(\begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right).$$

It is a matter of simple computation to show that $A_3 A_2 A_1 = 0$*; more specifically in this case our algorithm produce* $\mathcal{I}_\varepsilon = \emptyset,\ \mathcal{I}_3 = \{1\},\ \mathcal{I}_{32} = \{2\}\ \mathcal{I}_{321} = \{1, 2\}$*. However every product involving only two matrices turns out to be different from the null matrix, hence in this case* $I_C = 3 = 2^2 - 1$*.*

Chapter 4

Reachability of Discrete-time Positive Switched Systems

4.1 Preliminary remarks and a sufficient condition

It is immediately seen from eq. (2.6) that, when the initial condition $x(0)$ is zero and the input sequence $u(\cdot)$ is nonnegative, the state at the time instant k belongs to the polyhedral cone generated by the (columns of the) matrices $A_\sigma\big|_l^{k-1} B_{\sigma(l-1)}$, as l ranges from 1 to k, namely to the cone generated by the columns of the *discrete-time reachability matrix associated with the switching sequence* σ *of length* k:

$$\mathcal{R}_k(\sigma) := \left[B_{\sigma(k-1)}\ A_\sigma\big|_{k-1}^{k-1} B_{\sigma(k-2)}\ \dots\ A_\sigma\big|_2^{k-1} B_{\sigma(1)}\ A_\sigma\big|_1^{k-1} B_{\sigma(0)}\right].$$

When dealing with standard discrete-time switched systems, it has been proved [20] that the system is reachable if and only if there exists a single switching sequence σ (of length say k) such that $\mathrm{Im}\,(\mathcal{R}_k(\sigma)) = \mathbb{R}^n$. For positive switched systems, instead, this represents an obvious sufficient condition for reachability, but not a necessary one (see Example 4.2, below). Even the weaker condition that there exists a finite number of switching sequences of finite lengths, such that the union of the cones generated by the columns of their reachability matrices covers the positive orthant, is only sufficient for the system reachability (see Example 4.3, below).

Proposition 4.1 *If there exist switching sequences* $\sigma_1, \sigma_2, \dots, \sigma_\ell$ *(of lengths*

$k_1, k_2, \dots, k_\ell$, respectively) such that $\bigcup_{i=1}^{\ell} \mathrm{Cone}\,(\mathcal{R}_{k_i}(\sigma_i)) = \mathbb{R}^n_+$, the switched system (2.3) is reachable.

Example 4.2 *Consider the positive system, switching among the following subsystems:*

$$(A_1, B_1) = \left(\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right), \qquad (A_2, B_2) = \left(\begin{bmatrix} 0 & 2 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right).$$

It is clearly seen that, upon setting $\sigma_1(0) = \sigma_1(1) = 1$, and $\sigma_2(0) = \sigma_2(1) = 2$, we get

$$\mathrm{Cone}\,(\mathcal{R}_2(\sigma_1)) \cup \mathrm{Cone}\,(\mathcal{R}_2(\sigma_2)) = \mathrm{Cone}\left(\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \right) \cup \mathrm{Cone}\left(\begin{bmatrix} 0 & 2 \\ 1 & 1 \end{bmatrix} \right) = \mathbb{R}^2_+.$$

Consequently, the system is reachable. However, every single switching sequence σ (of length k) corresponds to a reachability matrix $\mathcal{R}_k(\sigma)$ having only one monomial column. Consequently, either $\mathbf{e}_1$ or $\mathbf{e}_2$ does not belong to $\mathrm{Cone}\,(\mathcal{R}_k(\sigma))$, which thus can never coincide with $\mathbb{R}^2_+$.

Example 4.3 *Consider the positive system, switching among the following subsystems:*

$$(A_1, B_1) = \left(\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right), \qquad (A_2, B_2) = \left(\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right).$$

It is clearly seen that every 1st monomial vector $x_f = \begin{bmatrix} x_1 & 0 \end{bmatrix}^T$, $x_1 > 0$, can be reached in a single step, by setting $\sigma_1(0) = 1$ (and $u(0) = x_1$). On the other hand, for every $x_f = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^T \geq 0$, with $x_2 > 0$, there exists a sufficiently large $k \in \mathbb{Z}_+, k \geq 2$, such that

$$x_f \in \mathrm{Cone}\left(\begin{bmatrix} 0 & 1 & 2 & \dots & k-1 \\ 1 & 1 & 1 & \dots & 1 \end{bmatrix} \right) = \mathrm{Cone}(\mathcal{R}_k(\sigma_2)),$$

where $\sigma_2(i) = 2$, for every $i \in [0, k-1]$. In particular, from equation (2.6) together with the expression of $\mathcal{R}_k(\sigma_2)$ we may deduce that $x_1(k) \leq (k-1)x_2(k)$. Thus, x_f can be reached in a minimum of $k = \left\lceil \frac{x_1}{x_2} \right\rceil + 1$ steps. As *a particular case, when* $x_1 = 0$ *(hence* $k = 1$*) it is sufficient to set* $u_0 = x_2$*; if* $k > 1$*, then* x_f *can be reached by setting*

$$\begin{aligned} u_0 &= \tfrac{x_1}{k-1}, \\ u_i &= 0, \qquad i = 1, 2, \dots, k-2, \\ u_{k-1} &= x_2 - u_0, \end{aligned}$$

where the nonnegativity of u_{k-1} *is ensured by the definition of* k*. So,* x_f *can be reached in* k *steps. This ensures that*

$$\mathbb{R}^2_+ = \mathrm{Cone}(\mathcal{R}_1(\sigma_1)) \bigcup \left(\cup_{k=0}^{+\infty} \mathrm{Cone}(\mathcal{R}_k(\sigma_2))\right),$$

and hence the system is reachable. However, since every nonnegative vector in $\mathbb{R}^2_+$ *which is not a* 1*-monomial vector can only be reached by steadily setting the switching sequence to the value* 2*, we may deduce that:*
1) for every finite $k \in \mathbb{Z}_+$, $\mathrm{Cone}(\mathcal{R}_1(\sigma_1)) \cup \mathrm{Cone}(\mathcal{R}_k(\sigma_2)) \neq \mathbb{R}^2_+$*, thus proving that Proposition 4.1 gives only a sufficient condition;*
2) there is no upper bound on $\sup_{x \in \mathbb{R}^2_+} \min\{k : \exists\, \sigma \text{ with } |\sigma| = k \text{ s.t. } x \in \mathrm{Cone}(\mathcal{R}_k(\sigma))\} = I_R$*. Consequently, the system is reachable, but* I_R *is not finite.*

4.2 Monomial Reachability

We introduce here the following definition (extending the classical one given for positive system, see for instance [13]).

Definition 4.4 *The switched system (2.3) is said to be* monomially reachable *if every monomial vector* $\alpha \mathbf{e}_i \in \mathbb{R}^n_+, \alpha \in \mathbb{R}_+, \alpha \neq 0, i = 1, 2, \dots, n,$ *is reachable.*

While the reachability property for positive systems is equivalent to the *monomial reachability*, for positive switched systems, instead, monomial reachability is an obvious necessary condition, but it is not sufficient. This is due to the fact that different monomial vectors can be reached by means of different switching sequences. Monomial reachability is easily captured.

Proposition 4.5 *The switched system (2.3) is monomially reachable if and only if there exists some positive integer N such that the* discrete-time reachability matrix in N steps

$$\mathcal{R}_N = \big[w(A_1,\dots,A_p)B_1 \; w(A_1,\dots,A_p)B_2 \; \dots \; w(A_1,\dots,A_p)B_p\big]_{\substack{w\in\Xi^* \\ 0\le|w|\le N-1}}$$

includes an $n\times n$ monomial submatrix.

Proof: Monomial reachability is equivalent to the possibility of reaching every canonical vector $\mathbf{e}_i, i = 1,2,\dots,n$. However, $\mathbf{e}_i$ is reachable if and only if there exist $k \in \mathbb{Z}_+$, a switching sequence $\sigma : [0,k-1] \to \mathcal{P}$ and a nonnegative input sequence $u(0),u(1),\dots,u(k-1)$, such that (2.6) holds for $x(0) = 0$ and $x(k) = \mathbf{e}_i$. Since each vector appearing in the right-hand side of (2.6) is nonnegative, such an identity holds if and only if there exists ℓ such that $A_\sigma\big|_\ell^{k-1} B_{\sigma(\ell-1)}u(\ell-1)$ is an i-monomial vector. This, in turn, is possible if and only if one of the columns of $A_\sigma\big|_\ell^{k-1} B_{\sigma(\ell-1)}$ is an i-monomial. So, there must be some $w_i \in \Xi^*$ and $j_i \in \mathcal{P}$ such that $w_i(A_1,A_2,\dots,A_p)B_{j_i}$ has an i-monomial column. Since this applies to each $i \in \langle n\rangle$, the proposition statement holds for $N := \max_i |w_i| + 1$. ■

As a further consequence of monomial reachability, we can obtain a new necessary condition for the reachability of discrete-time systems.

Corollary 4.6 *If the system (2.3) is monomially reachable (and hence, a fortiori, if it is reachable), then the matrix $[A_1 \; A_2 \; \dots \; A_p \mid B_1 \; B_2 \; \dots \; B_p]$ has an $n\times n$ monomial submatrix.*

Proof: If the system is reachable, for all i there exist $w_i \in \Xi^*$ and $j_i \in \mathcal{P}$ such that $w_i(A_1,A_2,\dots,A_p)B_{j_i}$ has an i-monomial column. So, if $|w_i| = 0$ then B_{j_i} includes an i-monomial vector, otherwise if $w_i = \xi_{i_k}\dots\xi_{i_1}$, with $k \ge 1$, then A_{i_k} has an i-monomial column. ■

At this point a natural question arises: if the system is monomially reachable and we let N denote the minimum nonnegative integer such that $\mathcal{R}_N$ includes an $n\times n$ monomial matrix, what is the maximum value that N may reach?

If the system is monomially reachable, for every $i \in \langle n \rangle$, there exists a matrix B_{j_i}, a specific column b in B_{j_i}, and some word $w \in \Xi^*$, such that $w(A_1, \dots, A_p)b$ is an i-monomial vector. So, the problem can be equivalently stated as: *given an n-dimensional vector $b \geq 0$ and $n \times n$ nonnegative matrices, $A_1, A_2, \dots, A_p$, find an upper bound on the minimum length of the word $w \in \Xi^*$ such that $\overline{\mathrm{ZP}}(w(A_1, \dots, A_p)b) = \{i\}$, provided that such a word exists.*

For the sake of compactness, in the sequel we adopt the following shorthand notation: given a switching sequence σ, we set $b_k := A_\sigma\big|_1^k b$, for every $k \geq 0$.

Proposition 4.7 *Let i be an element of $\langle n \rangle$. If for every word $w \in \Xi^*$, with $|w| \leq 2^n - 2$, $\overline{\mathrm{ZP}}(w(A_1, \dots, A_p)b) \neq \{i\}$, then $\overline{\mathrm{ZP}}(w(A_1, \dots, A_p)b) \neq \{i\}$ for every word $w \in \Xi^*$. Hence, if system (2.3) is monomially reachable, then $\mathcal{R}_{2^n-1}$ contains an $n \times n$ monomial submatrix.*

Proof: Suppose, by contradiction, that

$$\mathcal{W}_i := \{w \in \Xi^* : \ \overline{\mathrm{ZP}}(w(A_1, \dots, A_p)b) = \{i\}\} \neq \emptyset,$$

but the word of minimum length in $\mathcal{W}_i$, say $\bar{w}$, has length $k := |\bar{w}| > 2^n - 2$. Let σ be the switching sequence corresponding to $\bar{w}$, meaning that $A_\sigma\big|_1^k = \bar{w}(A_1, \dots, A_p)$. Consider, now, the finite sequence $b_j = A_\sigma\big|_1^j b$, for $j = 0, \dots, k$. By the assumption, $\overline{\mathrm{ZP}}(b_k) = \{i\}$. Clearly, $\overline{\mathrm{ZP}}(b_j) \neq \emptyset$ for every j (otherwise it would be $b_k = 0$). Since the family of the b_j's consists of at least 2^n vectors, and n-dimensional vectors may exhibit only $2^n - 1$ distinct nonzero patterns (excluding the empty one), two indices $\ell_1 < \ell_2 \leq k$ may be found such that $\overline{\mathrm{ZP}}(A_\sigma\big|_1^{\ell_1} b) = \overline{\mathrm{ZP}}(b_{\ell_1}) = \overline{\mathrm{ZP}}(b_{\ell_2}) = \overline{\mathrm{ZP}}(A_\sigma\big|_1^{\ell_2} b)$. But then, it is easily seen that a switching sequence σ' of length $k' = k - (l_2 - l_1) < k$ exists such that $\overline{\mathrm{ZP}}(A_{\sigma'}\big|_1^{k'} b) = \overline{\mathrm{ZP}}(A_\sigma\big|_{l_2+1}^{k} A_\sigma\big|_1^{l_1} b) = \{i\}$, a contradiction. ∎

In order to show that the bound provided by Proposition 4.7 is tight, we will provide an example of a reachable (and hence monomially reachable) system for which the discrete-time reachability matrix $\mathcal{R}_{2^n-1}$ contains an $n \times n$ monomial submatrix, but, for every $k < 2^n - 1$, $\mathcal{R}_k$ does not.

For the sake of simplicity, in the remaining part of the section, we replace each vector (matrix) with the vector (matrix) with entries in the boolean algebra $\mathcal{B} = \{0, 1\}$, endowed with the same zero pattern. We briefly recall here that the rules for addition and multiplication in $\mathcal{B}$ are:

$+$	0	1
0	0	1
1	1	1

$\cdot$	0	1
0	0	0
1	0	1

Example 4.8 *Consider all distinct proper subsets of* $\{1, 2, \ldots, n\}$*, namely all* $\mathcal{S}_j \in \mathscr{P}(\{1, 2, \ldots, n\})$ *with* $0 \le |\mathcal{S}_j| \le n-1$*, and suppose that they are ordered according to their cardinality. This means that* $\mathcal{S}_0 = \emptyset$*, all sets* $\mathcal{S}_j$ *for* $j = 1, \ldots, n$*, have* $|\mathcal{S}_j| = 1$*, and so on. Of course many different orderings satisfy this condition, simply consider one among them. We now accordingly define the vector sequence* $\{\tilde{b}_j\}_{j=0,1,\ldots,2^n-2}$*, where* $\tilde{b}_j$ *is the (only) vector in* $\mathcal{B}^n$ *with* $\mathrm{ZP}(\tilde{b}_j) = \mathcal{S}_j$ *(notice that we are constraining the zero pattern to coincide with* $\mathcal{S}_j$*, and hence the nonzero pattern* $\overline{\mathrm{ZP}}(\tilde{b}_j)$ *to coincide with* $\{1, 2, \ldots, n\} \setminus \mathcal{S}_j$*). It follows that* $\tilde{b}_j \neq 0$ *for every* j*,* $\tilde{b}_0 = \bar{1}$ *and* $\tilde{b}_{2^n-2} = \mathbf{e}_i$ *for some* i*.*

Let $A_j \in \mathcal{B}^{n\times n}$*,* $j \in \{1, 2, \ldots, 2^n-2\}$*, be defined according to the following rule:*

$$[A_j]_{rc} := \begin{cases} 0, & \text{if } \big(r \in \mathrm{ZP}(\tilde{b}_j)\big) \text{ and } \big(c \notin \mathrm{ZP}(\tilde{b}_{j-1})\big); \\ 1, & \text{otherwise.} \end{cases} \tag{4.1}$$

Consider the discrete-time system (2.3), switching among $(A_j, \tilde{b}_0)$*, with* $j = 1, \ldots, 2^n-2$*. It is just a matter of simple computation to check that by assuming* $\sigma(j) = j$ *for* $j = 1, 2, \ldots, 2^n-2$*, we obtain* $A_\sigma\big|_1^j \tilde{b}_0 = A_j A_{j-1} \ldots A_1 \tilde{b}_0 = A_j \tilde{b}_{j-1} = \tilde{b}_j$ *for* $j = 0, 1, \ldots, 2^n-2$*, and hence* $\mathbf{e}_i$ *is reached at time* $k = 2^n-1$ *(by means of a switching sequence* σ *defined on* $[1, 2^n-2]$*, and hence of length* 2^n-2*). It remains to show that no sequence of shorter length reaching* $\mathbf{e}_i$ *can be built for this system. To this end, consider the following lemma.*

Lemma 4.9 *By referring to the set of matrices* A_i*'s,* $i = 1, 2, \ldots, 2^n-2$*, previously defined, for every* $j \in \{1, 2, \ldots, 2^n-2\}$*, if* $q \neq j$*, then* $A_q \tilde{b}_{j-1} = \tilde{b}_k$ *with* $k \le j-1$*.*

Proof: *Suppose that* $j \in \{1, 2, \ldots, 2^n-2\}$ *is assigned. Then two situations may arise:*

Case $q < j$*: suppose that an index* $c \in \mathrm{ZP}(\tilde{b}_{q-1})$ *exists such that* $c \in \overline{\mathrm{ZP}}(\tilde{b}_{j-1})$*. Since, by definition (4.1), if* $[\tilde{b}_{q-1}]_c = 0$*, then the cth column of* A_q *coincides with* $\bar{1}$*, it follows that* $A_q \tilde{b}_{j-1} = \bar{1} = \tilde{b}_0$ *and the statement holds.*

Suppose, now, that $q < j$*, but for every* $h \in \mathrm{ZP}(\tilde{b}_{q-1})$ *we have* $h \in \mathrm{ZP}(\tilde{b}_{j-1})$ *(namely,* $\mathrm{ZP}(\tilde{b}_{q-1}) \subsetneq \mathrm{ZP}(\tilde{b}_{j-1})$*). We want to show that the vectors*

$\tilde{b}_q$ and $A_q\tilde{b}_{j-1}$ exhibit the same zero pattern. Since the zero patterns uniquely identify the (binary) vectors in the sequence $\{\tilde{b}_j\}$, this will imply that $\tilde{b}_q = A_q\tilde{b}_{j-1}$.

From (4.1) we deduce that:

a) If $r \in \overline{\mathrm{ZP}}(\tilde{b}_q)$, then $[A_q]_{rc} = 1$ for every c. Since $\tilde{b}_{j-1} \neq 0$, it follows that $[A_q\tilde{b}_{j-1}]_r = 1$ and $r \in \overline{\mathrm{ZP}}(A_q\tilde{b}_{j-1})$.

b) If $r \in \mathrm{ZP}(\tilde{b}_q)$, then $[A_q]_{rc} = 1$ implies $c \in \mathrm{ZP}(\tilde{b}_{q-1}) \subsetneq \mathrm{ZP}(\tilde{b}_{j-1})$. So, for every index c, $[A_q]_{rc} = 1$ implies $[\tilde{b}_{j-1}]_c = 0$ and hence $[A_q\tilde{b}_{j-1}]_r = 0$ or, equivalently $r \in \mathrm{ZP}(A_q\tilde{b}_{j-1})$.

This means that $\mathrm{ZP}(A_q\tilde{b}_{j-1}) = \mathrm{ZP}(\tilde{b}_q)$ and hence $A_q\tilde{b}_{j-1} = \tilde{b}_q$. So, in this case, $k = q \leq j-1$.

Case $q > j$: certainly there exists $c \in \overline{\mathrm{ZP}}(\tilde{b}_{j-1})$ such that $c \in \mathrm{ZP}(\tilde{b}_{q-1})$. But this implies the cth column of A_q is $\bar{1}$ and, as a consequence, $A_q\tilde{b}_{j-1} = \bar{1} = b_0$. Thus, again, $k = 0 \leq j-1$. ■

By the previous lemma, if we try to change the sequence $\{\tilde{b}_j\}$, by applying to the vector $\tilde{b}_{j-1}$, for some $j \in \{1, 2, \ldots, 2^n - 2\}$, a matrix A_q with $q \neq j$, we obtain a vector $\tilde{b}_k$ that we have already encountered in the sequence. So, even if another switching sequence may exist, eventually leading to the vector $\mathbf{e}_i$, its length is necessarily greater than $2^n - 2$.

So, the switching sequence σ we have provided in Example 4.8 is the shortest which allows to reach $\mathbf{e}_i$, and, consequently, the bound given in Theorem 4.7 is tight.

Chapter 5

Controllability of Continuous-time Positive Switched Systems

The nonnegativity constraint on the initial state and on the input signal rules out the possibility of controlling to zero the state trajectory in the continuous-time case. So, switched positive systems cannot be zero-controllable.

Proposition 5.1 *The continuous-time switched positive system (2.1) is never zero controllable.*

Proof: Consider equation (2.2) in Section 2. If system (2.1) were zero controllable, then for every $x(0) \geq 0$ there would be a time instant $t_f > 0$, a switching function σ on $[0, t_f]$ (defining an index sequence $\{i_0, \dots, i_k\}$) and a nonnegative input signal $u(\cdot)$ such that $x(t_f) = 0$. This would require every term of the sum in (2.2) to be the null vector. However, the matrix product $e^{A_{i_k}(t_f - t_k)} \dots e^{A_{i_0}(t_1 - t_0)}$is always nonsingular, so $e^{A_{i_k}(t_f - t_k)} \dots e^{A_{i_0}(t_1 - t_0)} x(0) = 0 \ \Leftrightarrow \ x(0) = 0$. Consequently, the only state that can be controlled to zero is the zero state and the claim is proved. ■

Remark 5.2 *Proposition 5.1 shows that a switched positive system (2.1) is never zero controllable in the usual sense, namely within a finite time interval. If we try to weaken Definition 2.2, by allowing t_f to take an infinite value, we actually deal with the stabilizability problem. Such an issue goes beyond the purposes of the present thesis.*

Chapter 6

Necessary Conditions for Reachability

Accordingly to how we dealt with the reachability issue in the context of Discrete-time positive switched systems, we first introduce some preliminary necessary condition for the reachability of Continuous-time Positive Switched Systems.

6.1 Monomial Reachability

A first necessary condition for reachability naturally descends from the fact that, among all nonnegative vectors, monomial vectors in particular have to be reachable. We refer to this condition as to "monomial reachability". As in the discrete-time case, and as shown in Example 6.8 below, monomial reachability does not ensure reachability.

Definition 6.1 *A continuous-time positive switched system (2.1) is said to be* monomially reachable *if every monomial vector* $\alpha \mathbf{e}_i \in \mathbb{R}^n_+, \alpha \in \mathbb{R}_+, \alpha \neq 0, i = 1, 2, \ldots, n$, *is reachable.*

Monomial reachability admits a family of interesting equivalent conditions.

Proposition 6.2 *For the continuous-time positive switched system (2.1) the following equivalent conditions hold:*

i) the system is monomially reachable;

ii) $\forall\, i \in \{1, \dots, n\}$ *there exist* $w \in \Xi^*$ *and some index* $j(i) \in \{1, 2, \dots, p\}$, *such that* $w(e^{A_1}, e^{A_2}, \dots, e^{A_p})e^{A_{j(i)}}B_{j(i)}$ *has an* i*th monomial column;*

iii) $\forall\, i \in \{1, \dots, n\}$ *there exists an index* $j(i)$ *such that* $\mathbf{e}_i$ *is an eigenvector of* $A_{j(i)}$ *(i.e.* $A_{j(i)}\mathbf{e}_i = \alpha_i \mathbf{e}_i$ *for some* $\alpha_i \geq 0$*, and hence the* i*th column of* $A_{j(i)}$ *is either an* i*th monomial vector or the zero vector) and one of the columns of* $B_{j(i)}$ *is, in turn, an* i*th monomial vector;*

iv) $\exists\, N \in \mathbb{N}$ *such that the* continuous-time reachability matrix in N steps

$$\mathcal{R}_N = \Big[w(e^{A_1}, \dots, e^{A_p})e^{A_1}B_1 | \dots | w(e^{A_1}, \dots, e^{A_p})e^{A_p}B_p \Big]_{\substack{w \in \Xi^* \\ 0 \leq |w| \leq N-1}}$$

has an $n \times n$ *monomial submatrix.*

Proof: If system (2.1) is monomially reachable, then $\forall\, i \in \{1, 2, \dots, n\}$ there exist a time instant $\bar{t}_i$, a switching sequence $\sigma_i : [0, \bar{t}_i] \to \mathcal{P}$, and a nonnegative input sequence $u_i(\tau), \tau \in [0, \bar{t}_i]$, that steer the system state from $x(0) = 0$ to $x(\bar{t}_i) = \mathbf{e}_i$. By resorting to equation (2.2), with $x(0) = 0$, $t = \bar{t}_i$, $u = u_i$ and $x(\bar{t}_i) = \mathbf{e}_i$, it is clear that in order to reach $\mathbf{e}_i$ it is necessary and sufficient that at least one of the matrix products

$$e^{A_{i_k}(\bar{t}_i - t_k)} e^{A_{i_{k-1}}(t_k - t_{k-1})} \dots e^{A_{i_l}(t_{l+1} - t_l)} \int_{t_{l-1}}^{t_l} e^{A_{i_{l-1}}(t_l - \tau)} B_{i_{l-1}} u_i(\tau) \mathrm{d}\tau \quad (6.1)$$

is an ith monomial vector. Note, first of all, that, by the properties of the exponential of a Metzler matrix given in Lemma A.2, if the above matrix product is an ith monomial vector, then the first exponential matrix in (6.1), $e^{A_{i_k}(\bar{t}_i - t_k)}$, must have the ith column which is an ith monomial vector and, in turn, $e^{A_{i_{k-1}}(t_k - t_{k-1})} \dots e^{A_{i_l}(t_{l+1} - t_l)} \int_{t_{l-1}}^{t_l} e^{A_{i_{l-1}}(t_l - \tau)} B_{i_{l-1}} u_i(\tau) \mathrm{d}\tau$ must be an ith monomial vector. So, by proceeding in this way, we show that all exponential matrices $e^{A_{i_{r-1}}(t_r - t_{r-1})}$ have the ith column which is an ith monomial vector and $\int_{t_{l-1}}^{t_l} e^{A_{i_{l-1}}(t_l - \tau)} B_{i_{l-1}} u_i(\tau) \mathrm{d}\tau$ is an ith monomial vector. This integral expression, however, is obtained by integrating a vector of continuous and nonnegative functions. So, it can be an ith monomial vector only if $e^{A_{i_{l-1}}(t_l - \tau)} B_{i_{l-1}} u_i(\tau)$ is an ith monomial column for every $\tau < t_l$. This proves that $e^{A_{i_k}(\bar{t}_i - t_k)} e^{A_{i_{k-1}}(t_k - t_{k-1})} \dots e^{A_{i_l}(t_{l+1} - t_l)} e^{A_{i_{l-1}}(t_l - \tau)} B_{i_{l-1}}$ has a column which is an ith monomial vector, namely, by Lemma A.2, $e^{A_{i_k}} e^{A_{i_{k-1}}} \dots e^{A_{i_l}} e^{A_{i_{l-1}}} B_{i_{l-1}}$ has a column which is an ith monomial vector. This way we have shown that monomial reachability implies ii).

In order to prove that ii) implies iii), recall that when

$$e^{A_{i_k}} \dots e^{A_{i_l}} e^{A_{i_{l-1}}} B_{i_{l-1}}$$

has a column which is an ith monomial vector, then the same holds true for $e^{A_{i_{l-1}}} B_{i_{l-1}}$, as it was shown in the first part of the proof. This, in turn, implies that the ith column of $e^{A_{i_{l-1}}}$ is an ith monomial vector and at least one of the columns of $B_{i_{l-1}}$ is an ith monomial vector. But the ith column of $e^{A_{i_{l-1}}}$ is an ith monomial vector if and only if the ith column of $A_{i_{l-1}}$ is either zero or an ith monomial vector. This proves iii). Conversely, if iii) holds, then the matrix $e^{A_{j(i)}} B_{j(i)}$ has an ith monomial column, and condition ii) holds.

Suppose now that iii) holds. But then iv) is verified for $N = 1$. Indeed, observe that the only word $w \in \Xi^* : \ |w| = 0$ is the empty word ε and, by iii), we certainly know that the matrix

$$\mathcal{R}_1 = \left[e^{A_1} B_1 | \dots | e^{A_p} B_p \right]$$

has an $n \times n$ monomial submatrix.

Finally, if iv) holds, than for every $i \in \{1, \dots, n\}$ a matrix product $w(e^{A_1}, \dots, e^{A_p}) e^{A_j} B_j$, for some $j \in \mathcal{P}$, can be found having among its columns an ith monomial vector. Note that, according to the previous reasonings, the matrix product $e^{A_j} B_j$ itself must exhibit an ith monomial column. Let k be the index corresponding to this ith monomial column, such that $e^{A_j} B_j \mathbf{e}_k = \beta \mathbf{e}_i$, for some $\beta > 0$ (note that $\mathbf{e}_k \in R_+^m$, while $\mathbf{e}_i \in R_+^n$). But then, in order to design a switching sequence and a nonnegative input such that any $n$dimensional ith monomial vector $\alpha \mathbf{e}_i$, $\alpha > 0$ can be reached it suffices to set $\sigma(t) = j$ and $u(t) = \dfrac{\alpha \mathbf{e}_k}{\int_0^{t_f} e^{A_{\sigma(\tau)}(t_f - \tau)} B_{\sigma(\tau)} d\tau}$ (see equation (2.1)).

■

The following necessary condition for the monomial reachability of positive switched systems is now straightforward (its proof follows the same line of reasoning of the proof of Corollary 4.6 in Section 4.2).

Corollary 6.3 *If the continuous-time positive switched system (2.1) is monomially reachable, then the matrix* $\left[e^{A_1} B_1 \ e^{A_2} B_2 \ \dots \ e^{A_p} B_p \right]$ *has a monomial submatrix.*

Remark 6.4 *Clearly, the result of Proposition 6.2 applies also to the special case when $|\mathcal{P}| = 1$, namely when we are dealing with (non-switched) continuous-time positive systems. So, if a continuous-time positive system (A, B) is reachable, then $A = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$, with $\lambda_i \geq 0$, and B contains an $n \times n$ monomial submatrix. On the other hand, it is immediately seen that a continuous-time positive system endowed with such a special structure is reachable. So, we have obtained a characterization of reachable positive systems.*

Proposition 6.5 *A continuous-time positive system $\dot{x}(t) = Ax(t) + Bu(t)$ is reachable if and only if $A = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$, for some $\lambda_i \geq 0$, and B contains an $n \times n$ monomial submatrix.*

Remark 6.6 *It is worthwhile noticing that condition ii) in Proposition 6.2 necessarily constrains the matrix $A_{j(i)}$ to be reducible. Consequently, all subsystems (A_i, B_i) with A_i irreducible play no role in the monomial reachability.*

Proposition 6.2 is immediately stated for single-input systems in the following simpler form.

Corollary 6.7 *If a single-input n-dimensional switched continuous-time positive system (2.1) is reachable, then there exists a relabeling of the p subsystems $(A_i, b_i), i \in \mathcal{P}$, such that that*

$$A_i \mathbf{e}_i = \alpha_i \mathbf{e}_i \text{ and } b_i = \beta_i \mathbf{e}_i, \qquad \text{for } i = 1, \dots, n, \tag{6.2}$$

where $\alpha_i \geq 0$ and $\beta_i > 0$. So, in particular, p must be greater than or equal to n.

Unfortunately, monomial reachability is not sufficient for reachability. Indeed, consider the following example.

Example 6.8 *Consider the continuous-time positive switched system (2.1), switching among the following $p = 3$ subsystems*

$$(A_1, B_1) = \left(\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \right)$$

$$(A_2, B_2) = \left(\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \right)$$

$$(A_3, B_3) = \left(\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \right).$$

Clearly, the necessary condition of Proposition 6.2 is satisfied. We aim to show, however, that the switched system is not reachable. To this end, preliminarily observe that $\mathrm{ZP}(e^{A_i t}) = \mathrm{ZP}(e^{A_i}) = \mathrm{ZP}(A_i)$*, for each* $i \in \{1, 2, 3\}$ *and every* $t > 0$*. So, it is clear that if* $i \neq k$ *then* $\mathrm{ZP}(e^{A_i \tau_i} e^{A_k \tau_k}) = \emptyset$, $\forall \, \tau_i, \, \tau_k > 0$*. Consequently, each nonnegative vector* x_f *with* $\mathrm{ZP}(x_f) \neq \emptyset$*, cannot be reached by switching the system structure between two different subsystems. For both subsystems, however, the only reachable states with nontrivial zero pattern are monomial vectors. So the system is not reachable.*

What may be regarded as somewhat surprising is that Proposition 6.2 can be "reversed" when the system size is $n = 2$. Indeed, suppose that condition ii) of Proposition 6.2 holds. Then either one of the following two situations arises:

Case 1. There exists an index $i \in \mathcal{P}$ such that A_i is diagonal and the corresponding input matrix B_i has an $n \times n$ monomial submatrix. By Proposition 6.5, the ith subsystem is reachable and hence the whole switched system is trivially reachable.

Case 2. There exist two indices, say 1 and 2, for the sake of simplicity, and two monomial vectors $\mathbf{e}_{j_1}$ and $\mathbf{e}_{j_2}$ such that

$$(A_1, \, B_1 \mathbf{e}_{j_1}) = \left(\begin{bmatrix} a_1 & \star \\ 0 & \star \end{bmatrix}, \begin{bmatrix} b_1 \\ 0 \end{bmatrix} \right), \quad (A_2, \, B_2 \mathbf{e}_{j_2}) = \left(\begin{bmatrix} \star & 0 \\ \star & a_2 \end{bmatrix}, \begin{bmatrix} 0 \\ b_2 \end{bmatrix} \right),$$

where $a_1, \, a_2 \geq 0, \; b_1, \, b_2 > 0$, while $\star$ denotes a nonnegative entry. Since in the following we will use only the specific columns $B_1 \mathbf{e}_{j_1}$ and $B_2 \mathbf{e}_{j_2}$, we will assume that each matrix B_i consists of that single column. Monomial reachability is trivially verified. Consider now any vector $x_f \gg 0$, and assume that $u(\cdot)$ is piece-wise constant and described as follows:

$$u(\tau) = \begin{cases} u_1, & \text{for } \tau \in [0, 1[, \\ u_2, & \text{for } \tau \in [1, 1+t], \end{cases} \tag{6.3}$$

where the positive values of u_1, u_2 and t will be chosen later. Notice that the expression of the forced component of the state evolution at the time instant $1+t$, after two switching instants $0 = t_0 < t_1 = 1 < 1+t$, by assuming that $\sigma(t_0) = 1$ and $\sigma(t_1) = 2$ is:

$$\begin{aligned} x(1+t) &= e^{A_2 t} \int_0^1 e^{A_1(1-\tau)} B_1 u_1 \mathrm{d}\tau + \int_1^{1+t} e^{A_2(1+t-\tau)} B_2 u_2 \mathrm{d}\tau \\ &= e^{A_2 t} \int_0^1 e^{a_1 \tau} \mathrm{d}\tau \cdot b_1 u_1 \mathbf{e}_1 + \int_0^t e^{a_2 \tau} \mathrm{d}\tau \cdot b_2 u_2 \mathbf{e}_2. \end{aligned} \tag{6.4}$$

Assume that $t > 0$ has been fixed, and notice that $\int_0^1 e^{a_1\tau}\mathrm{d}\tau \cdot b_1 u_1 \mathbf{e}_1$ and $\int_0^t e^{a_2\tau}\mathrm{d}\tau \cdot b_2 u_2 \mathbf{e}_2$ are a 1-monomial vector and a 2-monomial vector, respectively, whose nonzero entries may be arbitrarily assigned by suitably choosing u_1 and u_2. On the other hand, due to Lemma A.24, it is possible to choose t such that the first column of $e^{A_2 t}$ is a nonnegative vector arbitrarily close to the monomial vector $\mathbf{e}_1$. As a result, for every choice of $x_f \gg 0$ it is possible to choose $t > 0$ arbitrarily small so that x_f belongs to the cone generated by $e^{A_2 t}\mathbf{e}_1$ and $\mathbf{e}_2$. This implies that

$$x_f = e^{A_2 t} \int_0^1 e^{A_1(1-\tau)} B_1 u_1 \mathrm{d}\tau + \int_1^{1+t} e^{A_2(1+t-\tau)} B_2 u_2 \mathrm{d}\tau$$

for suitable u_1, u_2 and t.

So, we have proved the following result.

Proposition 6.9 *A continuous-time positive switched system (2.1) of size $n = 2$ is reachable if and only if $\forall\, i \in \{1, 2\}$ there exists an index $j \in \mathcal{P}$ such that $\mathbf{e}_i$ is an eigenvector of A_j and one of the columns of B_j is, in turn, an i-monomial vector (namely (6.2) holds for suitable $\alpha_i \geq 0$ and $\beta_i > 0$, $i = 1, 2$).*

The following proposition provides a sufficient condition for reachability.

Proposition 6.10 *If there exists a permutation π of the set $\{1, 2, \ldots, n\}$, such that $\forall\ k \in \{1, \ldots, n\}$ there exists an index $i_k \in \mathcal{P}$ such that all the $\pi(1)$th, $\pi(2)$th, ..., $\pi(k)$th columns of A_{i_k} are ordinately equal to $\alpha_1 \mathbf{e}_{\pi(1)}$, $\alpha_2 \mathbf{e}_{\pi(2)}$, ..., $\alpha_k \mathbf{e}_{\pi(k)}$, for suitable $\alpha_1, \ldots, \alpha_k \geq 0$, and one of the columns of B_{i_k} is, in turn, a $\pi(k)$-monomial vector then the system is reachable.*

Proof: For the sake of simplicity, we assume that π is the identical permutation, that leaves all the elements of $\{1, 2, \ldots, n\}$ invariant. Notice that $\forall k \in \{1, \ldots, n\}$ the matrix product $e^{A_{i_n}} e^{A_{i_{n-1}}} \ldots e^{A_{i_k}} B_{i_k}$ contains a k-monomial column. Consequently,

$$\left[e^{A_{i_n}} B_{i_n}\ \ e^{A_{i_n}} e^{A_{i_{n-1}}} B_{i_{n-1}}\ \ldots\ e^{A_{i_n}} \ldots e^{A_{i_1}} B_{i_1} \right]$$

has an $n \times n$ monomial submatrix. We now show that this fact implies reachability. Actually, let x_f be an arbitrary vector in $\mathbb{R}^n_+$. Consider equation (2.2) and assume that $x(0) = 0$, the switching instants are $t_\ell = \ell$ for $\ell = 0, 1, \ldots, n-1$, and the final time instant is $t_f = n$. Set $\sigma(t) := i_{\ell+1}$ when $\ell \leq t < \ell + 1$. Finally, assume that the input signal $u(\tau)$ is defined as follows: in every time interval $[\ell, \ell + 1[$ the jth entry of $u(\tau)$ is constant and positive if $e^{A_{i_n}} \ldots e^{A_{i_\ell}} B_{i_\ell} \mathbf{e}_j$ is an ℓ-monomial vector and $\ell \in \overline{\text{ZP}}(x_f)$, otherwise assume it is zero. By suitably choosing the positive values of the u_j's in every interval $[\ell-1, \ell[$, $\ell = 1, \ldots, n$, we can obtain x_f through equation (2.2). ■

It is worth to observe that the sufficient condition given in Proposition 6.10 guarantees something more than simple reachability. Indeed, as the proof shows, it ensures that there exists a single switching path along which every vetor in $\mathbb{R}^n_+$ can be reached. In fact, in the sequel, we will refer to the matrix

$$\left[e^{A_{i_n}} B_{i_n}\ \ e^{A_{i_n}} e^{A_{i_{n-1}}} B_{i_{n-1}}\ \ldots\ e^{A_{i_n}} \ldots e^{A_{i_1}} B_{i_1} \right] =: \mathcal{R}(i_1, i_2, \ldots, i_n)$$

as to the *continuous-time reachability matrix associated with the switching sequence* $(i_1, i_2, \ldots, i_n)$ (the switching times are necessarily $t_0 = 0 < t_1 = 1 < \cdots < t_{n-1} = n - 1$).

6.2 Pattern reachability

The concept of monomial reachability may be extended to the broader concept of "pattern reachability". Indeed, one may wonder which types of nonzero patterns may be reached (starting from zero initial conditions), by resorting to suitable nonnegative inputs and switching sequences, independently of the specific values taken by the positive entries associated with the nonzero patterns. This, of course, represents a necessary preliminary step toward the investigation of the more challenging reachability property.

Definition 6.11 *An n-dimensional continuous-time positive switched system (2.1) is said to be* pattern reachable *if for every (non-empty) subset $\mathcal{S}$ of $\langle n \rangle$ there exists a positive vector $x_f \in \mathbb{R}^n_+$, with $\overline{\mathrm{ZP}}(x_f) = \mathcal{S}$, such that x_f is reachable.*

Proposition 6.12 *An n-dimensional continuous-time positive switched system (2.1) is pattern reachable if and only if for every $\mathcal{S} \subseteq \{1, 2, \dots, n\}$ there exist $k \in \mathbb{N}$ and indices $i_0, i_1, \dots, i_k \in \mathcal{P}$ such that the cone generated by the columns of* the continuous-time reachability matrix associated with the switching sequence $(i_0, i_1, \dots, i_k)$, *i.e.*

$$\mathcal{R}(i_0, \dots, i_k) = \left[e^{A_{i_k}} B_{i_k} \,\middle|\, \dots \,\middle|\, e^{A_{i_k}} \dots e^{A_{i_0}} B_{i_0} \right],$$

contains a vector v with $\overline{\mathrm{ZP}}(v) = \mathcal{S}$.

Proof: Let $\mathcal{S}$ be a subset of $\{1, 2, \dots, n\}$ and let x_f be any nonnegative vector with $\overline{\mathrm{ZP}}(x_f) = \mathcal{S}$. The vector x_f is reachable if and only if there exist a time instant $t_f > 0$, a switching sequence $\sigma : [0, t_f) \mapsto \mathcal{P}$, and a nonnegative input sequence $u(\tau), \tau \in [0, t_f)$, that steers the system state from $x(0) = 0$ to $x(t_f) = x_f$. By resorting to equation (2.2), with $x(0) = 0$ and $x(t_f) = x_f$, it is clear that in order to reach x_f it is necessary and sufficient that there exists a finite number of nonzero matrix products in (2.2) of the following type

$$e^{A_{i_k}(t_f - t_k)} \dots e^{A_{i_l}(t_{l+1} - t_l)} \int_{t_{l-1}}^{t_l} e^{A_{i_{l-1}}(t_l - \tau)} B_{i_{l-1}} u(\tau) \mathrm{d}\tau \qquad (6.5)$$

which sum up to x_f. We may easily observe that, when our interest is only in nonzero patterns, the role of the nonnegative input $u(t)$ in every time interval $[t_{l-1}, t_l)$ is just that of "selecting" the columns of $B_{i_{l-1}}$. So, it can always be assumed positive and constant (say u_{l-1}), with either zero or unitary entries. On the other hand, due to the fact that the integral of $e^{A_{i_{l-1}}(t_l - \tau)} B_{i_{l-1}} u_{l-1}$ has the same zero pattern as $e^{A_{i_{l-1}}(t_l - t_{l-1})} B_{i_{l-1}} u_{l-1}$, and that the zero pattern of the exponential matrix at any positive time instant coincides with its zero pattern at the time instant $t = 1$ (see Lemma A.2), it follows that a nonnegative vector v with $\overline{\mathrm{ZP}}(v) = \overline{\mathrm{ZP}}(x_f) = \mathcal{S}$ is reachable if and only if there exists a finite number of nonzero matrix products of the following type

$$e^{A_{i_k}} e^{A_{i_{k-1}}} \dots e^{A_{i_l}} e^{A_{i_{l-1}}} B_{i_{l-1}} u_{l-1} \tag{6.6}$$

which sum up to v. But this amounts to saying that v belongs to the (polyhedral) cone, $\mathrm{Cone}(\mathcal{R}(i_0, i_1, \dots, i_k))$, generated by the columns of $\mathcal{R}(i_0, \dots, i_k)$. ■

An interesting fact about pattern reachability is that, in order to reach any nonzero pattern $\mathcal{S} \subseteq \{1, 2, \dots, n\}$, the system needs to switch no more than $|\mathcal{S}|$ times (including the initial configuration).

Proposition 6.13 *If an n-dimensional continuous-time positive switched system (2.1) is pattern reachable, then for every $\mathcal{S} \subseteq \{1, 2, \dots, n\}$ there exist $k < |\mathcal{S}|$ and indices $i_0, i_1, \dots, i_k \in \mathcal{P}$ such that* $\mathrm{Cone}(\mathcal{R}(i_0, i_1, \dots, i_k))$ *contains a nonnegative vector v with* $\overline{\mathrm{ZP}}(v) = \mathcal{S}$.

Proof: We prove this result by induction on the cardinality r of the set $\mathcal{S}$. If $r = |\mathcal{S}| = 1$, then $\mathcal{S} = \{i\}$, for some index $i \in \{1, 2, \dots, n\}$, and we are dealing with ith monomial reachability. As we have seen in Proposition 6.2, ith monomial reachability is equivalent to the existence of some index $j \in \mathcal{P}$ such that $A_j \mathbf{e}_i = \alpha \mathbf{e}_i$, for some $\alpha \geq 0$ and there exists $k \in \{1, 2, \dots, m\}$ such that $B_j \mathbf{e}_k$ is an ith monomial vector. Consequently, $\mathbf{e}_i = e^{A_j} B_j \mathbf{e}_k u_k$, for some suitable $u_k > 0$. This ensures that we need a switching sequence of length $r = 1$ in order to reach vectors with a single positive entry.

We assume, now, by induction, that given any subset $\mathcal{S}'$ of $\{1, 2, \dots, n\}$, with $|\mathcal{S}'| < r$, there exists a vector $v' \geq 0$ with $\overline{\mathrm{ZP}}(v') = \mathcal{S}'$ that can be reached by means of a switching sequence of length not greater than $|\mathcal{S}'|$. We aim to prove that the result extends to all subsets $\mathcal{S}$ of $\{1, 2, \dots, n\}$, with $|\mathcal{S}| = r$. Indeed, let v be a nonnegative vector, with $\overline{\mathrm{ZP}}(v) = \mathcal{S}$, which is reachable by means of a switching sequence $(i_0, i_1, \dots, i_k)$, and hence

$$v = e^{A_{i_k}} e^{A_{i_{k-1}}} \dots e^{A_{i_1}} e^{A_{i_0}} B_{i_0} \bar{u}_0 + \dots + e^{A_{i_k}} e^{A_{i_{k-1}}} B_{i_{k-1}} \bar{u}_{k-1} + e^{A_{i_k}} B_{i_k} \bar{u}_k,$$

for suitable $\bar{u}_i \geq 0$, with $\bar{u}_0 \neq 0$ (if $\bar{u}_0 = 0$ the switching sequence can surely be shortened). Since each of these terms is left multiplied by $e^{A_{i_k}}$, it follows the sum can be expressed as $e^{A_{i_k}}\mathcal{B}_k$ with

$$\mathcal{B}_k := e^{A_{i_{k-1}}}\dots e^{A_{i_0}}B_{i_0}\bar{u}_0 + \dots + e^{A_{i_{k-1}}}B_{i_{k-1}}\bar{u}_{k-1} + B_{i_k}\bar{u}_k.$$

By Lemma A.7, then

$$\mathcal{S} = \overline{\mathrm{ZP}}(v) = \overline{\mathrm{ZP}}(e^{A_{i_k}}\mathcal{B}_k) \;\Rightarrow\; \mathcal{S} = \overline{\mathrm{ZP}}(e^{A_{i_k}}\mathbf{e}_{\mathcal{S}})$$

and $\mathcal{S} \supseteq \overline{\mathrm{ZP}}(\mathcal{B}_k)$. Set $\mathcal{S}_k := \mathcal{S}$ and $\mathcal{S}_{k-1} := \overline{\mathrm{ZP}}(\mathcal{B}_k)$. As we have seen, $\mathcal{S}_k \supseteq \mathcal{S}_{k-1}$. We distinguish the following three situations:

a) $\overline{\mathrm{ZP}}(B_{i_k}\bar{u}_k) \subseteq \overline{\mathrm{ZP}}(e^{A_{i_{k-1}}}\dots e^{A_{i_1}}e^{A_{i_0}}B_{i_0}\bar{u}_0 + \dots + e^{A_{i_{k-1}}}B_{i_{k-1}}\bar{u}_{k-1})$ and $\mathcal{S}_k = \mathcal{S}_{k-1}$;

b) $\overline{\mathrm{ZP}}(B_{i_k}\bar{u}_k) \subseteq \overline{\mathrm{ZP}}(e^{A_{i_{k-1}}}\dots e^{A_{i_1}}e^{A_{i_0}}B_{i_0}\bar{u}_0 + \dots + e^{A_{i_{k-1}}}B_{i_{k-1}}\bar{u}_{k-1})$ and $\mathcal{S}_k \supsetneq \mathcal{S}_{k-1}$;

c) $\overline{\mathrm{ZP}}(B_{i_k}\bar{u}_k) \not\subseteq \overline{\mathrm{ZP}}(e^{A_{i_{k-1}}}\dots e^{A_{i_1}}e^{A_{i_0}}B_{i_0}\bar{u}_0 + \dots + e^{A_{i_{k-1}}}B_{i_{k-1}}\bar{u}_{k-1})$.

In the case a), $\overline{\mathrm{ZP}}(e^{A_{i_{k-1}}}\dots e^{A_{i_1}}e^{A_{i_0}}B_{i_0}\bar{u}_0+\dots+e^{A_{i_{k-1}}}B_{i_{k-1}}\bar{u}_{k-1}) = \overline{\mathrm{ZP}}(\mathcal{B}_k) = \mathcal{S}_{k-1} = \mathcal{S}_k = \mathcal{S}$. Consequently, we have found a shorter switching sequence $(i_0, i_1, \dots, i_{k-1})$ allowing to reach a vector $\tilde{v}$ with $\overline{\mathrm{ZP}}(\tilde{v}) = \mathcal{S}$ and we may consider, again, this shorter switching the sequence at the light of the possible three cases.

In the case b), we have

$$\overline{\mathrm{ZP}}(e^{A_{i_{k-1}}}\dots e^{A_{i_1}}e^{A_{i_0}}B_{i_0}\bar{u}_0 + \dots + e^{A_{i_{k-1}}}B_{i_{k-1}}\bar{u}_{k-1}) = \overline{\mathrm{ZP}}(\mathcal{B}_k) = \mathcal{S}_{k-1} \subsetneq \mathcal{S}.$$

So, on the one hand, the nonnegative vector

$$\tilde{v} := e^{A_{i_k}}\left(e^{A_{i_{k-1}}}\dots e^{A_{i_0}}B_{i_0}\bar{u}_0 + \dots + e^{A_{i_{k-1}}}B_{i_{k-1}}\bar{u}_{k-1}\right)$$

satisfies $\overline{\mathrm{ZP}}(\tilde{v}) = \mathcal{S}_k = \mathcal{S}$. On the other hand, condition $|\mathcal{S}_{k-1}| < r$ allows to apply the inductive hypothesis and hence to find a vector w, with $\overline{\mathrm{ZP}}(w) = \mathcal{S}_{k-1} = \overline{\mathrm{ZP}}(e^{A_{i_{k-1}}}\dots e^{A_{i_1}}e^{A_{i_0}}B_{i_0}\bar{u}_0 + \dots + e^{A_{i_{k-1}}}B_{i_{k-1}}\bar{u}_{k-1})$, and a switching sequence $(j_0, j_1, \dots, j_l)$, with $l + 1 \leq |\mathcal{S}_{k-1}| \leq r - 1$, such that

$$w = e^{A_{j_l}}\dots e^{A_{j_1}}e^{A_{j_0}}B_{j_0}u_0 + \dots + e^{A_{j_l}}B_{j_l}u_l$$

for suitable $u_i \geq 0$. Since

$$\mathcal{S} = \overline{\mathrm{ZP}}(e^{A_{i_k}}\mathcal{B}_k) = \overline{\mathrm{ZP}}(e^{A_{i_k}}(w + B_{i_k}0)),$$

we have found a switching sequence $(j_0, j_1, \dots, j_l, i_k)$ of length not greater than r that allows to reach the pattern $\mathcal{S}$.

Finally, if we are in the case c), there exists some index $\ell \in \overline{\mathrm{ZP}}(B_{i_k}\bar{u}_k) \subseteq \mathcal{S}$ such that $\ell \notin \overline{\mathrm{ZP}}(e^{A_{i_{k-1}}} \dots e^{A_{i_1}} e^{A_{i_0}} B_{i_0}\bar{u}_0 + \dots + e^{A_{i_{k-1}}} B_{i_{k-1}}\bar{u}_{k-1})$. Consequently,

$$\overline{\mathrm{ZP}}(e^{A_{i_{k-1}}} \dots e^{A_{i_1}} e^{A_{i_0}} B_{i_0}\bar{u}_0 + \dots + e^{A_{i_{k-1}}} B_{i_{k-1}}\bar{u}_{k-1}) \subsetneq \mathcal{S}.$$

So, we may apply the inductive hypothesis and find a vector w with

$$\overline{\mathrm{ZP}}(w) = \overline{\mathrm{ZP}}(e^{A_{i_{k-1}}} \dots e^{A_{i_0}} B_{i_0}\bar{u}_0 + \dots + e^{A_{i_{k-1}}} B_{i_{k-1}}\bar{u}_{k-1})$$

and a switching sequence $(j_0, j_1, \dots, j_l)$, with $l+1 \le |\overline{\mathrm{ZP}}(e^{A_{i_{k-1}}} \dots e^{A_{i_0}} B_{i_0}\bar{u}_0 + \dots + e^{A_{i_{k-1}}} B_{i_{k-1}}\bar{u}_{k-1})| \le r-1$, such that

$$w = e^{A_{j_l}} \dots e^{A_{j_1}} e^{A_{j_0}} B_{j_0} u_0 + \dots + e^{A_{j_l}} B_{j_l} u_l$$

for suitable $u_i \ge 0$. Since

$$\mathcal{S} = \overline{\mathrm{ZP}}(e^{A_{i_k}} \mathcal{B}_k) = \overline{\mathrm{ZP}}(e^{A_{i_k}}(w + B_{i_k}\bar{u}_k)),$$

we have found a switching sequence $(j_0, j_1, \dots, j_l, i_k)$ of length not greater than r that allows to reach the pattern $\mathcal{S}$. ∎

We are, now, in a position to provide the final characterization of pattern reachability. Even though the result could be easily given for multiple input systems, for the sake of simplicity we state it for single input systems.

Proposition 6.14 *A single-input positive switched system (2.1) is pattern reachable if and only if for every set $\mathcal{S} \subseteq \{1, \dots, n\}$ there exist an integer $\ell \le |\mathcal{S}|$, indices $j_1, j_2, \dots, j_\ell$, and a subset sequence $\mathcal{S}_0 \subseteq \mathcal{S}_1 \subset \mathcal{S}_2 \subset \dots \subset \mathcal{S}_\ell = \mathcal{S}$, such that*

$$\overline{\mathrm{ZP}}(e^{A_{j_h}} \mathbf{e}_{\mathcal{S}_{h-1}}) = \mathcal{S}_h, \qquad \forall\, h \in \langle \ell \rangle \tag{6.7}$$

$$\emptyset \ne \overline{\mathrm{ZP}}(B_{j_1}) \subseteq \mathcal{S}_1. \tag{6.8}$$

Proof: If system (2.1) is pattern reachable, then, by Proposition 6.13, for every $\mathcal{S} \subseteq \{1, \dots, n\}$ there exist $k < |\mathcal{S}|$, indices $i_0, i_1, \dots, i_k \in \mathcal{P}$, and a positive vector v, with $\overline{\mathrm{ZP}}(v) = \mathcal{S}$, such that

$$v = e^{A_{i_k}} e^{A_{i_{k-1}}} \dots e^{A_{i_1}} e^{A_{i_0}} B_{i_0} u_0 + \dots + e^{A_{i_k}} e^{A_{i_{k-1}}} B_{i_{k-1}} u_{k-1} + e^{A_{i_k}} B_{i_k} u_k,$$

where, w.l.o.g., the scalars $u_0, u_1, \dots, u_k$ take values in $\{0,1\}$, and k is the smallest such index. Set $\ell := \min\{d \geq 1 : u_{k-d+1} = 1\}$, and $j_h := i_{k-\ell+h}$ for $h = 1, 2, \dots, \ell$. Then:

$$v = e^{A_{j_\ell}} e^{A_{j_{\ell-1}}} \dots e^{A_{j_1}} \Big[e^{A_{i_{k-\ell}}} \dots e^{A_{i_0}} B_{i_0} u_0 + \dots + e^{A_{i_{k-\ell}}} B_{i_{k-\ell}} u_{k-\ell} + B_{j_1} u_{k-\ell+1} \Big].$$

Set $\mathcal{B}_0 := e^{A_{i_{k-\ell}}} \dots e^{A_{i_1}} e^{A_{i_0}} B_{i_0} u_0 + \dots + e^{A_{i_{k-\ell}}} B_{i_{k-\ell}} u_{k-\ell} + B_{j_1} u_{k-\ell+1}$ and $\mathcal{B}_h := e^{A_{j_h}} \mathcal{B}_{h-1}, h = 1, 2, \dots, \ell$. Notice that $\mathcal{B}_\ell = v$. Set, finally, $\mathcal{S}_h := \overline{\mathrm{ZP}}(\mathcal{B}_h)$. By recursively applying Lemma A.7, we can prove that

$$\mathcal{S} = \overline{\mathrm{ZP}}(v) = \overline{\mathrm{ZP}}(\mathcal{B}_\ell) \supseteq \overline{\mathrm{ZP}}(\mathcal{B}_{\ell-1}) \cdots \supseteq \overline{\mathrm{ZP}}(\mathcal{B}_1) \supseteq \overline{\mathrm{ZP}}(B_{j_1}).$$

On the other hand, all the inequalities $\mathcal{S}_h \supseteq \mathcal{S}_{h-1}, h = 2, 3, \dots, \ell$, must be strict, otherwise the sequence could be shortened. Therefore $\ell \leq |\mathcal{S}|$ and (6.7) holds. Finally, condition $\mathcal{S}_1 \supseteq \mathcal{S}_0 = \overline{\mathrm{ZP}}(\mathcal{B}_0) \supseteq \overline{\mathrm{ZP}}(B_{j_1})$ ensures that (6.8) holds.

Assume, now, that (6.7)-(6.8) hold. We prove that the system is pattern reachable by induction on $s := |\mathcal{S}|$. To this end, consider, first the case $s = 1$, namely $\mathcal{S} = \{i\}$ for some $i \in \{1, \dots, n\}$. If so, $\ell = 1$ and there exists an index j_1 and sets $\mathcal{S}_0 = \mathcal{S}_1 = \mathcal{S}$ such that $\overline{\mathrm{ZP}}(e^{A_{j_1}} \mathbf{e}_i) = \{i\}$, and $\emptyset \neq \overline{\mathrm{ZP}}(e^{B_{j_1}}) = \{i\}$. So, by Proposition 6.2, the system is monomially reachable.

Suppose, now, that for every set $\mathcal{S}'$ of cardinality smaller than s, there exists a reachable vector v' with $\overline{\mathrm{ZP}}(v') = \mathcal{S}'$. Consider an arbitrary set $\mathcal{S}$ of cardinality s and let $\ell, j_1, \dots, j_\ell, \mathcal{S}_0, \mathcal{S}_1, \dots, \mathcal{S}_\ell$ be the corresponding indices and sets as they appear in the proposition's statement. Consider the (possibly empty) set $\mathcal{S}' = \mathcal{S}_0 \setminus \overline{\mathrm{ZP}}(B_{j_1})$ whose cardinality is smaller than s. By the inductive assumption, there exist indices $i_0, i_1, \dots, i_k$ in $\mathcal{P}$ such that the cone generated by the columns of the reachability matrix $\mathcal{R}(i_0, i_1, \dots, i_k)$ includes a vector v' with $\overline{\mathrm{ZP}}(v') = \mathcal{S}'$ (if $\mathcal{S}' = \emptyset$, simply choose $v' = 0$). Let u be a binary vector such that $\overline{\mathrm{ZP}}(\mathcal{R}(i_0, i_1, \dots, i_k)u) = \mathcal{S}'$. Then, the vector

$$v = e^{A_{j_\ell}} e^{A_{j_{\ell-1}}} \dots e^{A_{j_1}} \left[\mathcal{R}(i_0, i_1, \dots, i_k)u + B_{j_1} \right]$$

satisfies $\overline{\mathrm{ZP}}(v) = \mathcal{S}$. This ensures that a vector with nonzero pattern $\mathcal{S}$ is reachable through the switching sequence $(i_0, i_1, \dots, i_k, j_1, j_2, \dots, j_\ell)$. ■

Chapter 7

Reachability of Continuous-time Single-Input Positive Switched Systems

As a result of the pattern reachability analysis, given a single-input switched system (2.1), switching among p positive subsystems $(A_i, b_i), i \in \mathcal{P}$, a positive vector v, with $\overline{\mathrm{ZP}}(v) = \mathcal{S}$, is reachable only if there is an index $j = j(\mathcal{S}) \in \mathcal{P}$ such that $\overline{\mathrm{ZP}}(e^{A_{j(\mathcal{S})}}\mathbf{e}_{\mathcal{S}}) = \mathcal{S}$. Consequently, a necessary condition for reachability is that, for every $\mathcal{S} \subseteq \langle n \rangle$, the set $\mathcal{I}_{\mathcal{S}} := \{i \in \langle p \rangle : \overline{\mathrm{ZP}}(e^{A_i}\mathbf{e}_{\mathcal{S}}) = \mathcal{S}\} \neq \emptyset$. Note that, if $\mathcal{S} = \langle n \rangle$, $\mathcal{I}_{\mathcal{S}} = \mathcal{P}$, and the previous condition is trivially satisfied.

In this section we focus on the derivation of necessary and/or sufficient conditions for reachability, by restricting our attention to single-input systems and, occasionally, on single-input systems of size n which commute among $p = n$ subsystems. As we have seen, this represents the minimum number of subsystems among which a single-input positive switched system has to commute in order to be reachable.

7.1 A sufficient condition

The first result of the section is a sufficient condition for reachability.

Proposition 7.1 *Consider a positive switched system (2.1), commuting among p single-input subsystems $(A_i, b_i), i \in \mathcal{P}$. If $\forall\, \mathcal{S} \subseteq \langle n \rangle$, $\exists\, j(\mathcal{S}) \in \mathcal{I}_{\mathcal{S}}$*

such that

$$\overline{\mathrm{ZP}}(b_{j(\mathcal{S})}) \subseteq \mathcal{S}, \text{ with } |\overline{\mathrm{ZP}}(b_{j(\mathcal{S})})| = 1,$$

then the switched system is reachable.

Proof: Given any positive vector $v \in \mathbb{R}^n_+$, set $r := |\overline{\mathrm{ZP}}(v)|$. Set, now, $\mathcal{S}_r := \overline{\mathrm{ZP}}(v)$, and let $j(\mathcal{S}_r)$ be an index which makes the Proposition assumption satisfied, and hence $\overline{\mathrm{ZP}}(e^{A_{j(\mathcal{S}_r)}}\mathbf{e}_{\mathcal{S}_r}) = \mathcal{S}_r$, and $\{i_r\} := \overline{\mathrm{ZP}}(b_{j(\mathcal{S}_r)}) \subseteq \mathcal{S}_r$. For each $h \in \langle r-1 \rangle$, we may recursively define sets $\mathcal{S}_h$ and indices i_h, as

$$\mathcal{S}_h := \mathcal{S}_{h+1} \setminus \{i_{h+1}\}, \qquad \{i_h\} := \overline{\mathrm{ZP}}(b_{j(\mathcal{S}_h)}).$$

Notice that, by the way the sets $\mathcal{S}_h$ are defined, $|\mathcal{S}_h| = h$. Moreover, when $h \neq q$ we have $j(\mathcal{S}_h) \neq j(\mathcal{S}_q)$. Now, we show that by suitably choosing a final time instant $t_r > 0$, the values of the switching instants $t_i, i = 0, 1, \ldots, r-1$, with $0 = t_0 < \ldots < t_{r-1} < t_r$, and positive input values $\bar{u}_i$ in every time interval $[t_{i-1}, t_i)$, we may ensure that

$$\begin{aligned} v = {} & e^{A_{j(\mathcal{S}_r)}(t_r - t_{r-1})} e^{A_{j(\mathcal{S}_{r-1})}(t_{r-1}-t_{r-2})} \ldots e^{A_{j(\mathcal{S}_2)}(t_2-t_1)} \int_{t_0}^{t_1} e^{A_{j(\mathcal{S}_1)}(t_1-\tau)} d\tau \, b_{j(\mathcal{S}_1)} \bar{u}_1 \\ & + \ldots + \int_{t_{r-1}}^{t_r} e^{A_{j(\mathcal{S}_r)}(t_r-\tau)} d\tau \, b_{j(\mathcal{S}_r)} \bar{u}_r \end{aligned} \tag{7.1}$$

By the previous considerations, every term in (7.1) has a nonzero pattern included in $\mathcal{S}$. Moreover, by Lemma A.24, it is easy to conclude that, since every exponential matrix can be made as close as we want to the identity matrix and since $b_{j(\mathcal{S}_\ell)}$ is an i_ℓ-monomial vector, then each positive term

$$e^{A_{j(\mathcal{S}_r)}(t_r - t_{r-1})} e^{A_{j(\mathcal{S}_{r-1})}(t_{r-1}-t_{r-2})} \ldots e^{A_{j(\mathcal{S}_{\ell+1})}(t_{\ell+1}-t_\ell)} \int_{t_{\ell-1}}^{t_\ell} e^{A_{j(\mathcal{S}_\ell)}(t_\ell-\tau)} d\tau \, b_{j(\mathcal{S}_\ell)} \tag{7.2}$$

can be made as close as we want to the monomial vector $\mathbf{e}_{i_\ell}$ (and, of course, its nonzero pattern is included in $\mathcal{S}$), by suitably choosing the time intervals between two consecutive switching instants sufficiently small. If we assume that the switching time instants are given, in order to ensure that the aforementioned terms are desired approximations of selected monomial vectors, the only values we have to choose are the constant values $\bar{u}_\ell$, and the problem we have to solve can be seen as that of solving an algebraic equation of the following type: $A\bar{u} = \bar{v}$, where $\bar{v} \in \mathbb{R}^r_+$ is the (strictly positive) vector consisting of the nonzero entries of v, $A \in \mathbb{R}^{r\times r}_+$ is the positive matrix whose columns are those terms (7.2) (approximating the monomial vectors) which pertain to the indices in $\mathcal{S}$, and $\bar{u}$ is the vector containing the associated input vectors $\bar{u}_\ell$. By Lemma A.23, this linear equation admits a positive solution, and hence the vector $v \in \mathbb{R}^n_+$ is reachable. ■

Remark. It is worthwhile noticing that, when the previous sufficient condition holds, all states in $\mathbb{R}^n_+$ are reached by resorting to a suitable switching sequence $(i_1, i_2, \dots, i_k)$ and by applying a nonnegative input which is surely nonzero during the last switching interval (when the system has commuted to the i_kth subsystem). Of course, this is not the general case, and a state may be reached even by eventually leaving the system freely evolve (meaning that no soliciting input is applied during the last part of the time interval), meanwhile commuting from one subsystem to another. Consequently, the above condition is only sufficient for reachability, as shown in the following example.

Example 7.2 *Consider the positive switched system (2.1), switching among the following three subsystems*

$$(A_1, B_1) = \left(\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right) \qquad (A_2, B_2) = \left(\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right)$$

$$(A_3, B_3) = \left(\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right).$$

Note that the hypothesis of Proposition 7.1 is fulfilled $\forall\, \mathcal{S} \neq \{1, 2\}$*. Therefore, in order to show that the switched system is reachable, we only need to prove that every vector* v *with* $\overline{\mathrm{ZP}}(v) = \{1, 2\}$ *is reachable. Observe now that*

$$e^{A_3 t} = \begin{bmatrix} e^t & t\,e^t & 0 \\ 0 & e^t & 0 \\ 0 & 0 & e^t \end{bmatrix}.$$

Hence, given $v = \begin{bmatrix} v_1 \\ v_2 \\ 0 \end{bmatrix}$*, with* $v_1,\, v_2 \neq 0$*, set* $t = \frac{v_1}{v_2} + 1,\ t_1 = 1,\ t_0 = 0.$ *Introduce the piece-wise constant input function and the switching sequence:*

$$u(t) = \begin{cases} \frac{v_2}{(e-1)e^{\frac{v_1}{v_2}}}, & \text{for } 0 \leq t < t_1; \\ 0, & \text{for } t_1 \leq t < t; \end{cases} \qquad \sigma(t) = \begin{cases} 2, & \text{for } 0 \leq t < t_1; \\ 3, & \text{for } t_1 \leq t < t. \end{cases}$$

By referring to equation (2.2), we get

$$x(t) = e^{A_3(t-t_1)} \int_{t_0}^{t_1} e^{A_2(t_1-\tau)} B_2 u(\tau) d\tau + \int_{t_1}^{t} e^{A_3(t-\tau)} B_3 u(\tau) d\tau$$

$$= e^{A_3 \frac{v_1}{v_2}} \int_0^1 \begin{bmatrix} 0 \\ e^{1-\tau} \\ 0 \end{bmatrix} d\tau \, \frac{v_2}{(e-1)e^{\frac{v_1}{v_2}}} + 0 = \frac{v_2\,(e-1)}{(e-1)e^{\frac{v_1}{v_2}}} \begin{bmatrix} \frac{v_1}{v_2} e^{\frac{v_1}{v_2}} \\ e^{\frac{v_1}{v_2}} \\ 0 \end{bmatrix} = v.$$

As a consequence, the switched system is reachable.

7.2 A geometric characterization of reachability

Aiming to provide an equivalent condition for reachability, we first introduce a technical lemma which allows us to use, when dealing with single-input systems, only piece-wise constant input signals.

Lemma 7.3 *Consider an n-dimensional monomially reachable positive switched system (2.1), switching among n single-input subsystems (A_i, b_i), $i \in \langle n \rangle$, with*[1]

$$A_i \mathbf{e}_i = \alpha_i \mathbf{e}_i, \qquad b_i = \beta_i \mathbf{e}_i, \qquad \exists\ \alpha_i \geq 0 \text{ and } \beta_i > 0. \tag{7.3}$$

Given $t > 0$, $v \in \mathbb{R}^n_+$, $k \in \mathbb{N}$, time instants $0 = t_0 < t_1 < \ldots < t_k < t$ and indices $i_0, i_1, \ldots, i_k \in \langle n \rangle$, if there exists a nonnegative input $u(\cdot)$ such that:

$$\begin{aligned} v = {} & e^{A_{i_k}(t-t_k)} e^{A_{i_{k-1}}(t_k - t_{k-1})} \ldots e^{A_{i_1}(t_2-t_1)} \int_{t_0}^{t_1} e^{A_{i_0}(t_1-\tau)} b_{i_0} u(\tau) d\tau \\ & + \ldots + \int_{t_k}^{t} e^{A_{i_k}(t-\tau)} b_{i_k} u(\tau) d\tau, \end{aligned} \tag{7.4}$$

then there exists a piece-wise constant input $u(\cdot)$, taking some suitable constant value $u_i \geq 0$ in every time interval $[t_i, t_{i+1})$, such that

$$\begin{aligned} v = {} & e^{A_{i_k}(t-t_k)} e^{A_{i_{k-1}}(t_k - t_{k-1})} \ldots e^{A_{i_1}(t_2-t_1)} \int_{t_0}^{t_1} e^{A_{i_0}(t_1-\tau)} b_{i_0} d\tau \cdot u_0 \\ & + \ldots + \int_{t_k}^{t} e^{A_{i_k}(t-\tau)} b_{i_k} d\tau \cdot u_k. \end{aligned} \tag{7.5}$$

[1]Notice that this assumption is by no means restrictive, since, by Proposition 6.2, we can always reduce ourselves to this case by means of a simple relabeling.

Proof: By the assumption (7.3), $e^{A_i t} b_i = e^{\alpha_i t} \beta_i \mathbf{e}_i, \forall t \in \mathbb{R}_+$. Consequently,

$$\begin{aligned}\int_{t_i}^{t_{i+1}} e^{A_i(t_{i+1}-\tau)} b_i u(\tau) d\tau &= \int_{t_i}^{t_{i+1}} e^{\alpha_i(t_{i+1}-\tau)} \beta_i \mathbf{e}_i u(\tau) d\tau \\ &= \left[\int_{t_i}^{t_{i+1}} e^{\alpha_i(t_{i+1}-\tau)} u(\tau) d\tau \right] \cdot \beta_i \mathbf{e}_i, \qquad (7.6)\end{aligned}$$

where the term inside the square brackets is a nonnegative number. But then, a nonnegative coefficient u_i can always be found such that

$$\int_{t_i}^{t_{i+1}} e^{\alpha_i(t_{i+1}-\tau)} u(\tau) d\tau = \int_{t_i}^{t_{i+1}} e^{\alpha_i(t_{i+1}-\tau)} d\tau \cdot u_i. \qquad (7.7)$$

This immediately implies the lemma statement. ■

From the previous lemma, we get the following Proposition.

Proposition 7.4 *Consider an n-dimensional positive switched system (2.1), switching among n single-input systems $(A_i,\ b_i), i \in \langle n \rangle$, and suppose that for every index $i \in \langle n \rangle$ the pair (A_i, b_i) satisfies (7.3). The system is reachable if and only if for every positive vector $v \in \mathbb{R}^n_+$ there exist $k \in \mathbb{N}$, strictly positive intervals $\tau_1, \ldots, \tau_k$ and switching values $i_0, i_1, \ldots, i_k \in \langle n \rangle$, such that*

$$\begin{aligned} v &\in \mathrm{Cone}[e^{A_{i_k}\tau_k} b_{i_k} | e^{A_{i_k}\tau_k} e^{A_{i_{k-1}}\tau_{k-1}} b_{i_{k-1}} | \ldots | e^{A_{i_k}\tau_k} \ldots e^{A_{i_1}\tau_1} e^{A_{i_0}\tau_0} b_{i_0}] \\ &= \mathrm{Cone}[\mathbf{e}_{i_k} | e^{A_{i_k}\tau_k} \mathbf{e}_{i_{k-1}} | \ldots | e^{A_{i_k}\tau_k} \ldots e^{A_{i_1}\tau_1} \mathbf{e}_{i_0}]. \end{aligned}$$

Proof: By the assumption on the n subsystems (A_i, b_i), the identity

$$\begin{aligned} &\mathrm{Cone}[e^{A_{i_k}\tau_k} b_{i_k} | e^{A_{i_k}\tau_k} e^{A_{i_{k-1}}\tau_{k-1}} b_{i_{k-1}} | \ldots | e^{A_{i_k}\tau_k} \ldots e^{A_{i_1}\tau_1} e^{A_{i_0}\tau_0} b_{i_0}] = \\ &\mathrm{Cone}[\mathbf{e}_{i_k} | e^{A_{i_k}\tau_k} \mathbf{e}_{i_{k-1}} | \ldots | e^{A_{i_k}\tau_k} \ldots e^{A_{i_1}\tau_1} \mathbf{e}_{i_0}] \end{aligned}$$

immediately follows. So, in the sequel, we only refer to the latter expression.

[Necessity] If the system is reachable, then $\forall\, v \in \mathbb{R}^n_+$ there exist parameters t, t_j, i_j (endowed with suitable properties) and an input $u(\cdot) \in \mathbb{R}_+$ such that:

$$\begin{aligned} v = {}& e^{A_{i_k}(t-t_k)} e^{A_{i_{k-1}}(t_k - t_{k-1})} \ldots e^{A_{i_1}(t_2-t_1)} \int_{t_0}^{t_1} e^{A_{i_0}(t_1-\tau)} b_{i_0} u(\tau) d\tau \\ & + \ldots + \int_{t_k}^{t} e^{A_{i_k}(t-\tau)} b_{i_k} u(\tau) d\tau. \qquad (7.8) \end{aligned}$$

But then, by Lemma 7.3, this means that there exist suitable $u_j \geq 0$ such that

$$v = e^{A_{i_k}(t-t_k)} e^{A_{i_{k-1}}(t_k-t_{k-1})} \dots e^{A_{i_1}(t_2-t_1)} \int_{t_0}^{t_1} e^{A_{i_0}(t_1-\tau)} b_{i_0} d\tau \cdot u_0$$
$$+ \dots + \int_{t_k}^{t} e^{A_{i_k}(t-\tau)} b_{i_k} d\tau \cdot u_k = e^{A_{i_k}\tau_k} e^{A_{i_{k-1}}\tau_{k-1}} \dots e^{A_{i_1}\tau_1} \mathbf{e}_{i_0} c_{i_0} + \dots + \mathbf{e}_{i_k} c_{i_k},$$

where $t_{k+1} := t$, $\tau_j := t_{j+1} - t_j$ and $c_{i_j} = \int_{t_j}^{t_{j+1}} e^{\alpha_{i_j}(t_{j+1}-\tau)} \beta_{i_j} d\tau \cdot u_j$. Hence,

$$v \in \text{Cone}[\mathbf{e}_{i_k} | e^{A_{i_k}\tau_k} \mathbf{e}_{i_{k-1}} | \dots | e^{A_{i_k}\tau_k} \dots e^{A_{i_1}\tau_1} \mathbf{e}_{i_0}]. \tag{7.9}$$

[Sufficiency] Conversely, suppose that for every positive vector v we can find $k \in \mathbb{N}$, intervals $\tau_1, \dots, \tau_k > 0$ and switching values $i_0, i_1, \dots, i_k \in \langle n \rangle$, such that (7.9) holds. Let $c_{i_j}, j = 0, 1, \dots, k$, be nonnegative coefficients such that $v = e^{A_{i_k}\tau_k} e^{A_{i_{k-1}}\tau_{k-1}} \dots e^{A_{i_1}\tau_1} \mathbf{e}_{i_0} c_{i_0} + \dots + \mathbf{e}_{i_k} c_{i_k}$. Set, now, $t_0 := 0, t_1 := 1$ and $t_{j+1} := t_j + \tau_j$ for every $j \in \langle k \rangle$. Then, by assuming

$$u_j := \frac{c_{i_j}}{\displaystyle\int_{t_j}^{t_{j+1}} e^{\alpha_{i_j}(t_{j+1}-\tau)} d\tau \beta_{i_j}}, \text{ we get}$$

$$e^{A_{i_k}(t_{k+1}-t_k)} e^{A_{i_{k-1}}(t_k-t_{k-1})} \dots e^{A_{i_1}(t_2-t_1)} \int_{t_0}^{t_1} e^{A_{i_0}(t_1-\tau)} b_{i_0}\ u_0\ d\tau + \dots +$$
$$+ \int_{t_k}^{t_{k+1}} e^{A_{i_k}(t-\tau)} b_{i_k}\ u_k\ d\tau = e^{A_{i_k}\tau_k} e^{A_{i_{k-1}}\tau_{k-1}} \dots e^{A_{i_1}\tau_1} \mathbf{e}_{i_0} c_{i_0} + \dots + \mathbf{e}_{i_k} c_{i_k} = v$$

thus proving that v is reachable. ∎

We are now ready to provide an algebraic equivalent characterization of the reachability property for n-dimensional positive switched systems (2.1), commuting among n single-input subsystems.

Proposition 7.5 *Given an n-dimensional positive switched system (2.1), commuting among n single-input subsystems $(A_i, b_i), i = 1, 2, \dots, n$, the following facts are equivalent:*

i) the switched system (2.1) is reachable;

ii) for every proper subset $\mathcal{S} \subset \langle n \rangle$ we have:

iia) *if* $|\mathcal{S}| = 1$*, then* $\exists\, j(\mathcal{S}) \in \mathcal{I}_{\mathcal{S}}$ *such that* ($\overline{\mathrm{ZP}}(e^{A_{j(\mathcal{S})}}\mathbf{e}_{\mathcal{S}}) = \mathcal{S}$ *and)* $\overline{\mathrm{ZP}}(b_{j(\mathcal{S})}) = \mathcal{S}$*;*

iib) *if* $|\mathcal{S}| > 1$*, then* $\mathcal{I}_{\mathcal{S}} \neq \emptyset$*, and either*

1. $\exists\, j(\mathcal{S}) \in \mathcal{I}_{\mathcal{S}}$ *such that*[2] $\overline{\mathrm{ZP}}(b_{j(\mathcal{S})}) \subset \mathcal{S}$*,* or
2. $\forall\, v \in \mathbb{R}^n_+$*, with* $\overline{\mathrm{ZP}}(v) = \mathcal{S}$*, there exist* $m \in \mathbb{N}$*,* $\tau_1, \dots, \tau_m > 0$ *and* $i_1, \dots, i_m \in \mathcal{I}_{\mathcal{S}}$*, such that* v *can be obtained as the nonnegative combination of no more than* $|\mathcal{S}| - 1$ *columns of* $e^{A_{i_m}\tau_m} \dots e^{A_{i_1}\tau_1} P_{\mathcal{S}}$*, where* $P_{\mathcal{S}}$ *is the selection matrix which selects all the columns corresponding to the indices*[3] *appearing in* $\mathcal{S}$*.*

Proof: i) $\Rightarrow$ ii) Suppose, first, that system (2.1) is reachable. Since condition iia) is equivalent to monomial reachability, its necessity has already been proved, and we may assume, as usual, that each pair (A_i, b_i) satisfies (7.3).

Now, let $\mathcal{S}$ be any subset of $\langle n \rangle$ with cardinality $|\mathcal{S}| > 1$, and let v be any positive vector with nonzero pattern $\overline{\mathrm{ZP}}(v) = \mathcal{S}$. By the reachability assumption and by Proposition 7.4, there exist $k \in \mathbb{N}$, positive intervals $\tau_0, \tau_1, \dots, \tau_k$, switching values $i_0, i_1, \dots, i_k \in \langle n \rangle$ (with $i_j \neq i_{j+1}$ w.l.o.g.), and nonnegative coefficients c_{i_j}, $j = 0, 1, \dots, k$, such that

$$\begin{aligned} v &= e^{A_{i_k}\tau_k} \dots e^{A_{i_1}\tau_1} e^{A_{i_0}\tau_0} b_{i_0} c_{i_0} + .. + e^{A_{i_k}\tau_k} e^{A_{i_{k-1}}\tau_{k-1}} b_{i_{k-1}} c_{i_{k-1}} + e^{A_{i_k}\tau_k} b_{i_k} c_{i_k} \\ &= e^{A_{i_k}\tau_k} \left[e^{A_{i_{k-1}}\tau_{k-1}} \dots e^{A_{i_1}\tau_1} e^{A_{i_0}\tau_0} b_{i_0} c_{i_0} + \dots + b_{i_k} c_{i_k} \right]. \end{aligned} \qquad (7.10)$$

Clearly, by Lemma A.7, $\overline{\mathrm{ZP}}(e^{A_{i_k}}\mathbf{e}_{\mathcal{S}}) = \mathcal{S}$. So, the set $\mathcal{I}_{\mathcal{S}}$ is nonempty. If there exist $j(\mathcal{S}) \in \mathcal{I}_{\mathcal{S}}$ such that $\overline{\mathrm{ZP}}(b_{j(\mathcal{S})}) \subset \mathcal{S}$ we fall in case 1. of iib). Suppose, now, that $\forall\, j(\mathcal{S}) \in \mathcal{I}_{\mathcal{S}}, \overline{\mathrm{ZP}}(b_{j(\mathcal{S})}) \not\subset \mathcal{S}$. Consequently, in (7.10), $c_{i_k} = 0$, and hence (7.10) becomes $v = e^{A_{i_k}\tau_k}\mathcal{B}_k$, with

$$\mathcal{B}_k := e^{A_{i_{k-1}}\tau_{k-1}} \dots e^{A_{i_1}\tau_1} e^{A_{i_0}\tau_0} b_{i_0} c_{i_0} + e^{A_{i_{k-1}}\tau_{k-1}} b_{i_{k-1}} c_{i_{k-1}}. \qquad (7.11)$$

From Lemma A.7, $\mathcal{S}_k := \overline{\mathrm{ZP}}(\mathcal{B}_k) \subseteq \mathcal{S}$. Now, either $\mathcal{S}_k \subsetneq \mathcal{S}$ or $\mathcal{S}_k = \mathcal{S}$.

(i) If $\mathcal{S}_k \subsetneq \mathcal{S}$, then v lies on a face of $\mathrm{Cone}(e^{A_{j_i}\tau_k} P_{\mathcal{S}}), \exists\, \tau_k > 0$, namely it can be obtained by combining no more than $|\mathcal{S}| - 1$ columns of $e^{A_{j_i}\tau_k} P_{\mathcal{S}}$.

[2]Note that condition iia) together with the fact that we are switching among n subsystems ensure that $\overline{\mathrm{ZP}}(b_{j(\mathcal{S})}) \neq \emptyset$.

[3]Notice that since $\overline{\mathrm{ZP}}(e^{A_{i_h}}\mathbf{e}_{\mathcal{S}}) = \mathcal{S}$ for $h = 1, 2, \dots, m$, the polyhedral cone $\mathrm{Cone}(e^{A_{i_m}\tau_m} \dots e^{A_{i_1}\tau_1} P_{\mathcal{S}})$ is generated by $|\mathcal{S}|$ linearly independent vectors whose nonzero pattern is included in $\mathcal{S}$. Indeed, $\mathrm{Cone}(P_{\mathcal{S}}^T e^{A_{i_m}\tau_m} \dots e^{A_{i_1}\tau_1} P_{\mathcal{S}})$ is a simplicial cone in $\mathbb{R}^{|\mathcal{S}|}_+$ and it coincides with $\mathrm{Cone}(e^{\tilde{A}_{i_m}\tau_m} \dots e^{\tilde{A}_{i_1}\tau_1})$, $\tilde{A}_{i_h} = P_{\mathcal{S}}^T A_{i_h} P_{\mathcal{S}}$.

(ii) If $\mathcal{S}_k = \mathcal{S}$ then $\overline{\mathrm{ZP}}(e^{A_{i_{k-1}}}\mathbf{e}_{\mathcal{S}}) = \mathcal{S}$, which, in turn, implies that $i_{k-1} \in \mathcal{I}_{\mathcal{S}}$. But since $\overline{\mathrm{ZP}}(b_{i_{k-1}}) \not\subset \mathcal{S}$ (and hence $c_{i_{k-1}} = 0$), it follows that we can iterate this reasoning until we find some index ℓ such that $i_k, i_{k-1}, \dots, i_\ell \in \mathcal{I}_{\mathcal{S}}$, $i_{\ell-1} \notin \mathcal{I}_{\mathcal{S}}$ and $v = e^{A_{i_k}\tau_k} \dots e^{A_{i_\ell}\tau_\ell}\mathcal{B}_\ell$, for some suitable $\mathcal{B}_\ell$ with $\overline{\mathrm{ZP}}(\mathcal{B}_\ell) = \mathcal{S}_\ell \subsetneq \mathcal{S}$. Consequently, again, v can be obtained by combining no more than $|\mathcal{S}| - 1$ columns of $e^{A_{i_k}\tau_k} \dots e^{A_{i_\ell}\tau_\ell} P_{\mathcal{S}}$.

In both cases we fall in case 2. of iib).

ii) $\Rightarrow$ i) Let us see, now, whether condition iia) and iib) are also sufficient for reachability. We prove this fact by induction on the cardinality of the nonzero pattern $|\mathcal{S}| = |\overline{\mathrm{ZP}}(v)|$ of any vector $v \in \mathbb{R}^n_+$. If $|\mathcal{S}| = 1$, condition iia), corresponding to monomial reachability, ensures that v is reachable.

Suppose now that, under the assumptions ii), every positive vector w, with $|\overline{\mathrm{ZP}}(w)| < s$, is reachable. Let v be a positive vector with $|\mathcal{S}| = |\overline{\mathrm{ZP}}(v)| = s$. If for the set $\mathcal{S}$ the case 1. applies, it has been already proved in Proposition 7.1 that v is reachable. Suppose now that only case 2. holds. Then $\exists\, m > 0$, $\exists\ i_1, \dots, i_m \in \mathcal{I}_{\mathcal{S}}$, $\exists\ \tau_1, \dots, \tau_m > 0$ such that v is obtained by combining no more than $r - 1$ columns of $e^{A_{i_m}\tau_m} \dots e^{A_{i_1}\tau_1} P_{\mathcal{S}}$ and hence $\exists\, w \geq 0$, with $\overline{\mathrm{ZP}}(w) \subsetneq \mathcal{S}$ (and therefore $|\overline{\mathrm{ZP}}(w)| < r$), such that $v = e^{A_{i_m}\tau_m} \dots e^{A_{i_1}\tau_1} w$. Since vector w is reachable for hypothesis, also v is. Indeed, upon reaching w, we switch ordinately to the subsystems $i_1, i_2, \dots, i_m$ and leave the system freely evolve at each stage for a lapse of time equal to τ_i. ∎

Chapter 8

Further results on the Reachability of Continuous-Time Single-Input Positive Switched Systems

Not every condition provided in Proposition 7.5 can be easily verified. Specifically, there is no obvious way of testing whether indices $i_1, \dots, i_m$ and positive time intervals $\tau_1, \dots, \tau_m$ can be found, such that a given vector $\mathbf{v} > 0$, with $\overline{\mathrm{ZP}}(\mathbf{v}) = \mathcal{S}$, can be obtained by combining less than $|\mathcal{S}|$ columns of $e^{A_{i_m}\tau_m} \dots e^{A_{i_1}\tau_1} P_{\mathcal{S}}$.

It is important to note that this situation corresponds to the case when some vector $\mathbf{v}$ can be reached only by means of some nonnegative input function $\mathbf{u}(t)$ which must be set equal to zero during the time interval $\tau_1, \dots, \tau_m$ corresponding to the last m commutations of the system (2.1), thus letting the system freely evolve during this lapse of time.

We start in Section 8.1 by analyzing in depth the case when only during the last commutation no soliciting input acts on the switched system (2.1). In Section 8.2 we will deal with the general case when multiple switchings take place, meanwhile no input acts on the system. Finally, in Section 8.3 the case of systems presenting special properties is investigated.

8.1 The asymptotic exponential cone: the single matrix case

As a first step toward the general problem solution, in this section we explore the restrictive case when $m = 1$. In other words, we are interested in investigating when a positive vector $\mathbf{v}$, with $\mathcal{S} := \overline{\mathrm{ZP}}(\mathbf{v})$ of cardinality s, can be expressed as the positive combination of at most $s-1$ columns of $e^{A_{j(\mathcal{S})}\tau}P_{\mathcal{S}}$, for some suitable $j(\mathcal{S}) \in \mathcal{I}_{\mathcal{S}}$ and $\tau > 0$ (as usual, $P_{\mathcal{S}}$ is the selection matrix that singles out the columns indexed on $\mathcal{S}$). Notice, though, that this is equivalent to investigating when the restriction of $\mathbf{v}$ to its positive entries (which is a strictly positive vector, say $\mathbf{v}_{\mathcal{S}}$, of size s) belongs to the boundary of the simplicial cone, $\mathrm{Cone}[P_{\mathcal{S}}^T e^{A_{j(\mathcal{S})}\tau}P_{\mathcal{S}}]$. Thus, our problem may be restated in a just apparently restrictive, but in fact absolutely general, formulation, by assuming $\mathcal{S} = \langle n \rangle$ and $\mathcal{I}_{\mathcal{S}} = \mathcal{P}$ (and, consequently, $\mathbf{v}_{\mathcal{S}} = \mathbf{v}$).

Problem Statement: Given an $n \times n$ Metzler matrix A, search for conditions ensuring that every strictly positive vector $\mathbf{v} \in \mathbb{R}^n_+$ can be obtained as

$$\mathbf{v} = e^{A\tau}\mathbf{u}, \qquad \exists\, \tau > 0, \text{ and } \mathbf{u} \in \mathbb{R}^n_+ \text{ with } \mathrm{ZP}(\mathbf{u}) \neq \emptyset. \tag{8.1}$$

To solve this problem, we introduce a new concept which turns out to be very meaningful for our investigation.

Definition 8.1 *Given an $n \times n$ Metzler matrix A, we define its* asymptotic exponential cone, $\mathrm{Cone}_\infty(e^{At})$, *as the polyhedral cone generated by the vectors $\mathbf{v}_i^\infty$, which represent the asymptotic directions of the columns of e^{At}, i.e.*

$$\mathbf{v}_i^\infty := \lim_{t\to\infty} \frac{e^{At}\mathbf{e_i}}{\|e^{At}\mathbf{e_i}\|}, \qquad i = 1, 2, \dots, n. \tag{8.2}$$

It is not hard to prove that $\mathrm{Cone}_\infty(e^{At})$ always exists, it is a polyhedral convex cone in $\mathbb{R}^n_+$, and it is never the empty set. Moreover, except for the case of a diagonal matrix A (in which case $\mathrm{Cone}(e^{At}) = \mathrm{Cone}_\infty(e^{At}) = \mathbb{R}^n_+$ for every $t \geq 0$), we have for every $0 < t_1 < t_2 < +\infty$:

$$\mathbb{R}^n_+ = \mathrm{Cone}(e^{A\cdot 0}) \supsetneq \mathrm{Cone}(e^{At_1}) \supsetneq \mathrm{Cone}(e^{At_2}) \supsetneq \mathrm{Cone}_\infty(e^{At}).$$

Notice, also, that while $\mathrm{Cone}(e^{At})$ is a simplicial cone for every $t \geq 0$, $\mathrm{Cone}_\infty(e^{At})$ is typically not, since it is not generally solid.

In this section we investigate the relationship between the asymptotic exponential cone and the boundary of the cone generated by a single exponential matrix. By making use of this characterization and of the fundamental result of Proposition 7.5, we will be able to provide a family of sufficient conditions for reachability.

Lemma 8.2 *Given an $n \times n$ Metzler matrix A and a strictly positive vector $\mathbf{v} \in \mathbb{R}^n_+$, the following facts are equivalent:*

i) there exists $\tau > 0$ such that $\mathbf{v}$ belongs to $\partial\mathrm{Cone}(e^{A\tau})$;

ii) $\mathbf{v} \notin \mathrm{Cone}_\infty(e^{At})$.

Even more, if any of the above equivalent conditions holds, there exists a unique $\tau > 0$ such that $\mathbf{v}$ belongs to $\partial\mathrm{Cone}(e^{A\tau})$.

Proof: *i)* $\Rightarrow$ *ii)* If there exists $\tau > 0$ such that $\mathbf{v}$ belongs to $\partial\mathrm{Cone}(e^{A\tau})$, then $\mathbf{v} = e^{A\tau}\mathbf{u}$, for some $\mathbf{u} \geq 0$ with $\overline{\mathrm{ZP}}(\mathbf{u}) = \mathcal{S} \subsetneq \langle n \rangle$. We want to prove that for every $\delta > 0$ the vector $\mathbf{v}$ does not belong to $\mathrm{Cone}(e^{A(\tau+\delta)})$ and hence, a fortiori, it does not belong to $\mathrm{Cone}_\infty(e^{At})$. If this were the case, then

$$\mathbf{v} = e^{A\tau}\mathbf{u} = e^{A\tau}[e^{A\delta}\mathbf{w}],$$

for some nonnegative $\mathbf{w}$. By the invertibility of $e^{A\tau}$, this would mean $\mathbf{u} = e^{A\delta}\mathbf{w}$. Since $\overline{\mathrm{ZP}}(\mathbf{u}) = \mathcal{S}$, by Lemma A.7, it must be $\overline{\mathrm{ZP}}(e^{A\tau}\mathbf{e}_{\mathcal{S}}) = \mathcal{S}$. But then $\overline{\mathrm{ZP}}(e^{A\tau}\mathbf{u})$ should be $\mathcal{S}$, too, thus contradicting the strict positivity assumption on $\mathbf{v}$.

ii) $\Rightarrow$ *i)* Conversely, suppose that $\mathbf{v} \gg 0$ and $\mathbf{v} \notin \mathrm{Cone}_\infty(e^{At})$. Notice, then, that $\mathbf{v}$ is an internal point of $\mathrm{Cone}(e^{At})\big|_{t=0} = \mathbb{R}^n_+$. By the continuity of the exponential matrix and the fact that $\mathrm{Cone}(e^{At})$ is monotonically decreasing with t (in the sense of the inclusion chain mentioned before), it follows that there exists $\bar{t} > 0$ such that $\mathbf{v} \notin \mathrm{Cone}(e^{A\bar{t}})$. Define a distance function $d(t)$ between the vector $\mathbf{v}$ and $\mathrm{Cone}(e^{At})$ as:

$$d(t) := \inf\{\|\mathbf{v} - e^{At}\mathbf{x}\| : \mathbf{x} \geq 0\}.$$

Clearly, $d(0) = 0$ and $d(\bar{t}) > 0$, moreover $d(t)$ is a continuous function. So, once we define $\tau := \sup\{t \geq 0 : d(t) = 0\}$, it is easily seen that $\mathbf{v} \in \mathrm{Cone}(e^{A\tau})$ (as polyhedral cones are closed sets) and it must lie on the boundary of the cone, namely on some "face", otherwise it would contradict the definition of τ. This further proves that $\tau = \max\{t \geq 0 : d(t) = 0\}$.

Suppose now, by contradiction, that there exist $\tau_1, \tau_2 > 0$, with $\tau_1 \neq \tau_2$, such that $\mathbf{v} = e^{A\tau_1}\mathbf{u}_1 = e^{A\tau_2}\mathbf{u}_2$, for some positive vectors $\mathbf{u}_1, \mathbf{u}_2$ with nontrivial zero patterns. If we assume, w.l.o.g., $\tau_2 > \tau_1$, then from the previous identity one gets $\mathbf{u}_1 = e^{A(\tau_2-\tau_1)}\mathbf{u}_2$, which ensures (see Lemma A.2 in the Appendix)

$$\overline{\mathrm{ZP}}(\mathbf{u}_1) = \overline{\mathrm{ZP}}(e^{A(\tau_2-\tau_1)}\mathbf{u}_2) = \overline{\mathrm{ZP}}(e^{A\tau_2}\mathbf{u}_2) = \overline{\mathrm{ZP}}(\mathbf{v}),$$

a contradiction. ∎

As an immediate corollary of Lemma 8.2, we get.

Corollary 8.3 *Given an $n \times n$ Metzler matrix A, the following are equivalent:*

i) $\forall \mathbf{v} \gg 0$ *there exists* $\tau > 0$ *such that* $\mathbf{v}$ *belongs to* $\partial\mathrm{Cone}(e^{A\tau})$*;*

ii) $\mathrm{Cone}_\infty(e^{At}) \subseteq \partial\mathbb{R}^n_+$*;*

iii) there exists some index $r \in \langle n \rangle$ *such that* $r \in \mathrm{ZP}(\mathbf{v}_j^\infty)$ *for every* $j \in \langle n \rangle$ *(with* $\mathbf{v}_j^\infty$ *as in eq.(8.2)).*

Note that, if A is an irreducible matrix, it admits only one positive eigenvector of unitary norm, which is strictly positive and corresponds to the dominant eigenvalue [4]. Therefore $\mathrm{Cone}_\infty(e^{At})$ collapses into a one dimensional cone (a ray) which lies in the interior of the positive orthant. So, condition ii) in Corollary 8.3 cannot be fulfilled, unless A is a reducible matrix.

At this point, we want to analyze when either one of the equivalent conditions in Corollary 8.3 is verified. Equivalently, we want to derive a characterization of the condition $\mathrm{Cone}_\infty(e^{At}) \not\subseteq \partial\mathbb{R}^n_+$.

Proposition 8.4 *Let A be an $n\times n$ Metzler matrix in Frobenius normal form (1.2).* $\mathrm{Cone}_\infty(e^{At}) \not\subseteq \partial\mathbb{R}^n_+$ *if and only if every initial class is distinguished.*

Proof: [Sufficiency] Suppose that for every class $\mathcal{C}_j, j \in \langle \ell \rangle$, which is initial $\lambda_{\max}(A_{jj}) > \lambda_{\max}(A_{kk})$ for every $k \in \mathcal{D}(\mathcal{C}_j)$. Let j be an arbitrary index in $\langle \ell \rangle$. If $\mathcal{C}_j$ is an initial class, then for every index i such that $\mathcal{C}(i) = \mathcal{C}_j$ (namely, $i \in \mathcal{C}_j$), $\mathrm{block}_j[\mathbf{v}_i^\infty] \gg 0$ (see Proposition A.10). On the other hand, when $\mathcal{C}_j$ is not an initial class, and we let $\mathcal{C}_h$ be an initial class accessing $\mathcal{C}_j$, then for every index i such that $\mathcal{C}(i) = \mathcal{C}_h$, $\text{block}_j[\mathbf{v}_i^\infty] \gg 0$. This proves that for every $j \in \langle \ell \rangle$ there is at least one vector $\mathbf{v}_i^\infty$ with $\text{block}_j[\mathbf{v}_i^\infty] \gg 0$, and this ensures that $\text{Cone}_\infty(e^{At}) \not\subseteq \partial\mathbb{R}_+^n$.

[Necessity] Assume, by contradiction, that there is one initial class $\mathcal{C}_j$, $j \in \langle \ell \rangle$, such that $\lambda_{\max}(A_{jj}) \leq \lambda_{\max}(A_{kk})$ for some $k \in \mathcal{D}(\mathcal{C}_j)$. Let i be an arbitrary index in $\langle n \rangle$. If $i \notin \mathcal{C}_j$ then $\text{block}_j[e^{At}\mathbf{e}_i] = 0$ and hence $\text{block}_j[\mathbf{v}_i^\infty] = 0$ (see Theorem A.16). On the other hand, if $i \in \mathcal{C}_j$ then there exists $h < i$ such that $\text{block}_h[e^{At}\mathbf{e}_i]$ strictly dominates $\text{block}_j[e^{At}\mathbf{e}_i]$. Consequently, $\text{block}_j[\mathbf{v}_i^\infty] = 0$. This ensures that all vectors $\mathbf{v}_i^\infty$ have the jth block identically zero, and this implies that $\text{Cone}_\infty(e^{At}) \subseteq \partial\mathbb{R}_+^n$. ■

By making use of Proposition A.17 and of Lemma A.21, in the Appendix, the derivation of the following characterization is straightforward.

Lemma 8.5 *Given an $n \times n$ Metzler matrix A,*

i) for every $i \in \langle n \rangle$, the ith generating vector of $\text{Cone}_\infty(e^{At})$, $\mathbf{v}_i^\infty$, *is a positive eigenvector (of unitary norm) of A, corresponding to the dominant eigenvalue of some distinguished class; as a consequence,* $\text{Cone}_\infty(e^{At})$ *is A-invariant and therefore e^{At}-invariant $\forall\, t \geq 0$.*

ii) A positive eigenvector $\mathbf{v}$ of A, corresponding to some eigenvalue $\lambda \in \sigma(A)$, can be expressed as the nonnegative combination of all those eigenvectors $\mathbf{v}_i^\infty$ which correspond to the eigenvalue λ, and hence $\mathbf{v}$ belongs to $\text{Cone}_\infty(e^{At})$.

iii) $\text{Cone}_\infty(e^{At})$ *coincides with the (polyhedral convex) cone in $\mathbb{R}_+^n$ generated by the set of positive eigenvectors of A. Even more,* $\text{Cone}_\infty(e^{At})$ *is the polyhedral cone generated by a full column rank positive matrix.*

Proof: i) Follows immediately from Proposition A.17. Note that the vectors $\tilde{\mathbf{v}}_i$, $i = 1, \ldots, \ell$, of eq. (A.10) in Proposition A.17 are the asymptotic directions defined in eq. (8.2)[1].

[1]As a matter of fact, while the vector $\mathbf{v}_i^\infty$ represents the asymptotic directions the ith column aligns to, the vector $\tilde{\mathbf{v}}_i$ represents the asymptotic direction all the columns of the ith class align to.

ii) Suppose w.l.o.g. that $\|\mathbf{v}\| = 1$ and that A is in Frobenius normal form (1.2). Since $e^{At}\mathbf{v} = e^{\lambda t}\mathbf{v}$, it is easily seen that

$$\lim_{t\to+\infty} \frac{e^{At}\mathbf{v}}{\|e^{At}\mathbf{v}\|} = \lim_{t\to+\infty} \frac{e^{\lambda t}\mathbf{v}}{\|e^{\lambda t}\mathbf{v}\|} = \lim_{t\to+\infty} \mathbf{v} = \mathbf{v}.$$

On the other hand, by resorting to Proposition A.19, we may say that, when t tends to $+\infty$, then

$$e^{At}\mathbf{v} \approx \sum_{i\in I} \mathbf{v}_i^{\infty}[\mathbf{v}]_i \; m(t),$$

where

- $m(t)$ is the dominant mode within the set $\left\{ e^{\lambda_j t}\frac{t^{m_j}}{m_j!} : j \in \overline{\mathrm{ZP}}(\mathbf{v})\right\}$, with $\lambda_j = \max\{\, \lambda_{\max}(A_{kk}) : k \in \mathcal{D}(\mathcal{C}(j))\}$ and m_j+1 the maximum number of classes $\mathcal{C}_k$ with $\lambda_{\max}(A_{kk}) = \lambda_j$ that lie in a single chain starting from $\mathcal{C}(j)$ in $\mathcal{R}(A)$;
- $I := \{i \in \overline{\mathrm{ZP}}(\mathbf{v}) : m_i(t) = m(t)\}$.

Consequently, $\lim_{t\to+\infty} \frac{e^{At}\mathbf{v}}{\|e^{At}\mathbf{v}\|} = \frac{\sum_{i\in I} \mathbf{v}_i^{\infty}[\mathbf{v}]_i}{\|\sum_{i\in I} \mathbf{v}_i^{\infty}[\mathbf{v}]_i\|}$. So, it must be

$$\mathbf{v} = \frac{\sum_{i\in I} \mathbf{v}_i^{\infty}[\mathbf{v}]_i}{\|\sum_{i\in I} \mathbf{v}_i^{\infty}[\mathbf{v}]_i\|},$$

which concludes the proof of ii).

iii) Let V be the set of all positive eigenvectors of A. By the previous point ii), $\mathrm{Cone}(V) \subseteq \mathrm{Cone}_{\infty}(e^{At})$.
On the other hand, $\mathrm{Cone}_{\infty}(e^{At}) = \mathrm{Cone}(\mathbf{v}_1^{\infty}, \mathbf{v}_2^{\infty}, \dots, \mathbf{v}_n^{\infty}) \subseteq \mathrm{Cone}(V)$, and hence $\mathrm{Cone}(V) = \mathrm{Cone}_{\infty}(e^{At})$. ∎

By resorting to Proposition 7.5 and Corollary 8.3, we get the following sufficient condition for reachability.

Proposition 8.6 *Consider an n-dimensional positive switched system (2.1), commuting among n single-input subsystems (A_i, b_i), $i = 1, 2, \dots, n$, and suppose that, for every proper subset $\mathcal{S} \subset \langle n\rangle$, $|\mathcal{I}_{\mathcal{S}}| = 1$, namely there exists a unique index $j(\mathcal{S}) \in \langle n\rangle$ such that $\overline{\mathrm{ZP}}(e^{A_{j(\mathcal{S})}}\mathbf{e}_{\mathcal{S}}) = \mathcal{S}$. Then the system is reachable if and only if the following two conditions hold:*

a) the system is monomially reachable;

b) for every $\mathcal{S}$, with $r := |\mathcal{S}| > 1$, either

- $\overline{\mathrm{ZP}}(b_{j(\mathcal{S})}) \subseteq \mathcal{S}$ *or*
- $\mathrm{Cone}_\infty\big(P_\mathcal{S}^T\, e^{A_{j(\mathcal{S})}t} P_\mathcal{S}\big) \subseteq \partial\mathbb{R}_+^r$,

where $P_\mathcal{S}$ is the selection matrix which selects all the columns corresponding to the indices belonging to $\mathcal{S}$.

Proof: [Sufficiency] Notice, first, that if $\mathrm{Cone}_\infty\big(P_\mathcal{S}^T\, e^{A_{j(\mathcal{S})}t} P_\mathcal{S}\big) \subseteq \partial\mathbb{R}_+^r$, then, by Corollary 8.3, for every strictly positive vector $\mathbf{v}_\mathcal{S} \in \mathbb{R}_+^r$ there exists $\tau > 0$ such that $\mathbf{v}_\mathcal{S} \in \partial\mathrm{Cone}\big(P_\mathcal{S}^T\, e^{A_{j(\mathcal{S})}\tau} P_\mathcal{S}\big)$. So, as a consequence of condition $\overline{\mathrm{ZP}}(e^{A_{j(\mathcal{S})}}\mathbf{e}_\mathcal{S}) = \mathcal{S}$, for every positive vector $\mathbf{v} \in \mathbb{R}_+^n$, with $\overline{\mathrm{ZP}}(\mathbf{v}) = \mathcal{S}$, there exists $\tau > 0$ such that $\mathbf{v} = e^{A_{j(\mathcal{S})}\tau} P_\mathcal{S}\mathbf{u}_\mathcal{S}$, with $\mathrm{ZP}(\mathbf{u}_\mathcal{S}) \neq \emptyset$. Consequently, assumptions a) and b) imply conditions iia) and iib) of Proposition 7.5, and reachability follows.

[Necessity] By comparing the proposition statement with the result of Proposition 7.5, it remains to prove that if the system is reachable and for every $\mathcal{S} \subset \langle n\rangle$ there is a single index $j(\mathcal{S})$ such that $\overline{\mathrm{ZP}}(e^{A_{j(\mathcal{S})}}\mathbf{e}_\mathcal{S}) = \mathcal{S}$, then condition $\emptyset \neq \overline{\mathrm{ZP}}(b_{j(\mathcal{S})}) \not\subseteq \mathcal{S}$ implies $\mathrm{Cone}_\infty\big(P_\mathcal{S}^T\, e^{A_{j(\mathcal{S})}t} P_\mathcal{S}\big) \subseteq \partial\mathbb{R}_+^r$. Indeed, let $\mathbf{v} > 0$ with $\overline{\mathrm{ZP}}(\mathbf{v}) = \mathcal{S}$. By referring to the same notation employed in the proof of Proposition 7.5, we have that $\mathbf{v} = e^{A_{j(\mathcal{S})}\tau_k}\mathcal{B}_k$, with $\tau_k > 0$ and

$$\mathcal{B}_k := e^{A_{i_{k-1}}\tau_{k-1}} \dots e^{A_{i_1}\tau_1} e^{A_{i_0}\tau_0} b_{i_0} c_{i_0} + \dots + e^{A_{i_{k-1}}\tau_{k-1}} b_{i_{k-1}} c_{i_{k-1}}, \tag{8.3}$$

for suitable indices i_ℓ (with $i_\ell \neq i_{\ell+1}$), positive time intervals τ_ℓ and nonnegative coefficients c_ℓ. From Lemma A.7, it follows that $\mathcal{S}_k := \overline{\mathrm{ZP}}(\mathcal{B}_k) \subseteq \mathcal{S}$, and the uniqueness of $j(\mathcal{S})$ ensures that $\mathcal{S}_k \subsetneq \mathcal{S}$. So, $\mathbf{v} = e^{A_{j(\mathcal{S})}\tau_k} P_\mathcal{S}\mathbf{u}_\mathcal{S}$, $\exists\ \mathbf{u}_\mathcal{S} \geq 0$, with $\mathrm{ZP}(\mathbf{u}_\mathcal{S}) \neq \emptyset$. But since this must be true for every vector $\mathbf{v} \in V_\mathcal{S} := \{v : \overline{\mathrm{ZP}}(\mathbf{v}) = \mathcal{S}\}$, then every $\mathbf{v}_\mathcal{S} \in \mathbb{R}_+^r$, with $\mathbf{v}_\mathcal{S} \gg 0$, must lie on the boundary of $\mathrm{Cone}(P_\mathcal{S}^T e^{A_{j(\mathcal{S})}\tau} P_\mathcal{S})$ for some $\tau = \tau(v_\mathcal{S}) > 0$. By Corollary 8.3, then, it must be $\mathrm{Cone}_\infty\big(P_\mathcal{S}^T\, e^{A_{j(\mathcal{S})}t} P_\mathcal{S}\big) \subseteq \partial\mathbb{R}_+^n$. ∎

We are ready to derive, at this point, a sufficient condition for reachability which is based on the structure of the cones $\mathrm{Cone}_\infty\big(P_\mathcal{S}^T\, e^{A_{j(\mathcal{S})}t} P_\mathcal{S}\big)$.

Proposition 8.7 *Consider an n-dimensional positive switched system (2.1), commuting among n single-input subsystems (A_i, b_i), $i = 1, 2, \dots, n$, and suppose that the system is monomially reachable. If for every proper subset*

$\mathcal{S} \subset \langle n \rangle$, *with* $|\mathcal{S}| \geq 2$,

$$\cap_{i\in\mathcal{I}_\mathcal{S}} \mathrm{Cone}_\infty(e^{P_\mathcal{S}^T A_i P_\mathcal{S} t}) \subseteq \partial\mathbb{R}_+^{|\mathcal{S}|}, \tag{8.4}$$

then the system is reachable.

Proof: Monomial reachability ensures that all monomial vectors are reachable. On the other hand, consider the case of any vector $\mathbf{v}$ with $\mathcal{S} = \overline{\mathrm{ZP}}(\mathbf{v})$ of cardinality greater than 1 and let $\mathbf{v}_\mathcal{S}$ be the restriction of $\mathbf{v}$ to the indices corresponding to $\mathcal{S}$. If (8.4) holds, than, by the strict positivity of $\mathbf{v}_\mathcal{S}$, there exists at least one index $j = j(\mathbf{v}_\mathcal{S}) \in \mathcal{I}_\mathcal{S}$ such that $\mathbf{v}_\mathcal{S} \notin \mathrm{Cone}_\infty(e^{P_\mathcal{S}^T A_j P_\mathcal{S} t})$. Thus, by Lemma 8.2, there exists $\tau > 0$ such that $\mathbf{v}_\mathcal{S} \in \partial\mathrm{Cone}(e^{P_\mathcal{S}^T A_j P_\mathcal{S} \tau})$ and $\mathbf{v}$ is reachable. ■

8.2 Asymptotic exponential cone: the multiple exponential case

Analogously to the single matrix case, we introduce here the following definition

Definition 8.8 *Given an ordered set of* $n \times n$ *Metzler matrices* $A_{i_1}, \dots, A_{i_m}$ *and a positive vector* $\bar{\alpha} = (\alpha_1, \dots, \alpha_m)$*, we define their* asymptotic exponential cone *along* $\bar{\alpha}$

$$\mathrm{Cone}_\infty^{\bar{\alpha}}\left(e^{A_{i_1} t} \dots e^{A_{i_m} t}\right)$$

as the polyhedral cone generated by the (normalized) vectors $\mathbf{v}_i^\infty$ *which represent the asymptotic directions of the columns of* $e^{A_{i_1}\alpha_1 t} \dots e^{A_{i_m}\alpha_m t}$*, i.e.*

$$\mathbf{v}_i^\infty := \lim_{t\to\infty} \frac{e^{A_{i_1}\alpha_1 t} \dots e^{A_{i_m}\alpha_m t}\mathbf{e_i}}{\|e^{A_{i_1}\alpha_1 t} \dots e^{A_{i_m}\alpha_m t}\mathbf{e_i}\|}, \qquad i = 1, 2, \dots, n. \tag{8.5}$$

Again, $\mathrm{Cone}_\infty^{\bar{\alpha}}\left(e^{A_{i_1} t} \dots e^{A_{i_m} t}\right)$ is a polyhedral convex cone in $\mathbb{R}_+^n$, and it is never the empty set. However, no monotonicity property can be generally guaranteed, as it happens for a single matrix exponential.

One may wonder why there is the need for introducing a whole family of asymptotic cones corresponding to a certain index family $\{i_1, i_2, \dots, i_m\}$.

The reason is that, unfortunately, different directions $\bar{\alpha}$ may lead to different asymptotic cones. So, while in the single exponential case we are dealing with a single cone, when considering m exponentials we are typically dealing with a family of cones. This simple example clarifies this point.

Example 8.9 *Consider the two Metzler matrices*

$$A_1 = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}, \qquad A_2 = \begin{bmatrix} 6 & 1 \\ 0 & 4 \end{bmatrix}.$$

It is a matter of simple computation to show that, for any $\bar{\alpha} = (\alpha_1, \alpha_2) \in \mathbb{R}^2_+$, *we get*

$$e^{A_1 \alpha_1 t} e^{A_2 \alpha_2 t} = \begin{bmatrix} e^{(\alpha_1 + 6\alpha_2)t} & e^{(\alpha_1 + 6\alpha_2)t} + e^{(2\alpha_1 + 4\alpha_2)t} + l.t. \\ 0 & e^{(2\alpha_1 + 4\alpha_2)t} \end{bmatrix}$$

where "l.t." ("lower terms") denotes terms which are surely dominated by the two terms appearing in the $(1,2)$*-entry. Consequently, we distinguish the following three cases:*

(i) $\alpha_1 + 6\alpha_2 > 2\alpha_1 + 4\alpha_2$, *namely* $\alpha_1 < 2\alpha_2$*: if so,*

$$\mathrm{Cone}^{\bar{\alpha}}_{\infty}(e^{A_1 t} e^{A_2 t}) = \mathrm{Cone}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} \right);$$

(ii) $\alpha_1 + 6\alpha_2 = 2\alpha_1 + 4\alpha_2$, *namely* $\alpha_1 = 2\alpha_2$, *in which case*

$$\mathrm{Cone}^{\bar{\alpha}}_{\infty}(e^{A_1 t} e^{A_2 t}) = \mathrm{Cone}\left(\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \right);$$

(iii) $\alpha_1 + 6\alpha_2 < 2\alpha_1 + 4\alpha_2$, *namely* $\alpha_1 > 2\alpha_2$, *for which*

$$\mathrm{Cone}^{\bar{\alpha}}_{\infty}(e^{A_1 t} e^{A_2 t}) = \mathrm{Cone}\left(\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \right).$$

It is easily seen, though, that even if $\bar{\alpha}$ may vary continuously in $\mathbb{R}^m_+$, the number of asymptotic cones is necessarily finite, as it depends on which mode dominates each column in the matrix product $e^{A_{i_1}\alpha_1 t} e^{A_{i_2}\alpha_2 t} \dots e^{A_{i_m}\alpha_m t}$. Since the dominant modes are obtained by multiplying the dominant modes of each single entry of the various factors $e^{A_{i_h}\alpha_h t}$, the number of possible combinations as $\bar{\alpha}$ varies in $\mathbb{R}^m_+$ is necessarily finite.

In this section we explore the properties of the cones generated by an ordered family of exponential matrices, along certain directions. As illustrated in Example 8.9, once the indices $i_1, i_2, \dots, i_m \in \mathcal{P}$ have been chosen, we are dealing with a family of asymptotic exponential cones, and not a single one. Nonetheless, the order of the indices constrains the asymptotic cone to lie within the asymptotic cone related to the first index i_1.

Proposition 8.10 *Given a set $\mathcal{A} = \{A_1, \dots, A_p\}$ of Metzler matrices and indices $i_1, i_2, \dots, i_m \in \mathcal{P}$, we have that, for every $\bar{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_m)$*

$$\mathrm{Cone}_\infty^{\bar{\alpha}}(e^{A_{i_1}t} \dots e^{A_{i_m}t}) \subseteq \mathrm{Cone}_\infty(e^{A_{i_1}t}).$$

Proof: Assume, initially, $\bar{\alpha} = \bar{1}$. Notice, first, that for every $i \in \langle n \rangle$ the jth entry of the vector $\hat{\mathbf{v}}_i(t) := e^{A_{i_2}t} \dots e^{A_{i_m}t}\mathbf{e}_i$, if nonzero, takes the following form

$$[\hat{\mathbf{v}}_i(t)]_j = c_{ji}\; e^{\lambda_{ji}t} \frac{t^{m_{ji}}}{m_{ji}!} + [\hat{\mathbf{v}}_i]_{j,\mathrm{rem}}(t), \tag{8.6}$$

with $c_{ji} > 0$ and $\lim_{t\to+\infty} \frac{[\hat{\mathbf{v}}_i]_{j,\mathrm{rem}}(t)}{e^{\lambda_{ji}t}\frac{t^{m_{ji}}}{m_{ji}!}} = 0$.

So,

$$\lim_{t\to+\infty} \frac{e^{A_{i_1}t} \dots e^{A_{i_m}t}\mathbf{e}_i}{\|e^{A_{i_1}t} \dots e^{A_{i_m}t}\mathbf{e}_i\|} = \lim_{t\to+\infty} \frac{e^{A_{i_1}t}\hat{\mathbf{v}}_i(t)}{\|e^{A_{i_1}t}\hat{\mathbf{v}}_i(t)\|}$$

for some vector $\hat{\mathbf{v}}_i(t)$ whose nonzero entries take the form (8.6). If we assume that $\mathrm{Cone}_\infty(e^{A_{i_1}t}) = \mathrm{Cone}(\mathbf{v}_1^\infty, \dots, \mathbf{v}_n^\infty)$, by making use of the expression of $e^{A_{i_1}t}$ given in Proposition A.17, we may say that, when t that tends to $+\infty$,

$$e^{A_{i_1}t}\hat{\mathbf{v}}_i(t) \approx \sum_{h\in H} \mathbf{v}_h^\infty c_{hi}\; m(t),$$

where

- $m(t)$ is the dominant mode among $\left\{m_j(t)e^{\lambda_{ji}t}\frac{t^{m_{ji}}}{m_{ji}!} : j \in \overline{\mathrm{ZP}}(\hat{\mathbf{v}}_i(t))\right\}$;
- $H := \{h \in \overline{\mathrm{ZP}}(\mathbf{v}) : m_h(t)e^{\lambda_{hi}t}\frac{t^{m_{hi}}}{m_{hi}!} = m(t)\}$.

Consequently,

$$\lim_{t\to+\infty} \frac{e^{A_{i_1}t}\hat{\mathbf{v}}_i(t)}{\|e^{A_{i_1}t}\hat{\mathbf{v}}_i(t)\|} = \frac{\sum_{h\in H} \mathbf{v}_h^\infty c_{hi}}{\|\sum_{h\in H} \mathbf{v}_h^\infty c_{hi}\|}.$$

So, it must be

$$\lim_{t\to+\infty} \frac{e^{A_{i_1}t} \dots e^{A_{i_m}t}\mathbf{e}_i}{\|e^{A_{i_1}t} \dots e^{A_{i_m}t}\mathbf{e}_i\|} = \frac{\sum_{h\in H} \mathbf{v}_h^\infty c_{hi}}{\|\sum_{h\in H} \mathbf{v}_h^\infty c_{hi}\|}.$$

Consequently, all generators of $\text{Cone}_\infty^{(1,1,\dots,1)}(e^{A_{i_1}t}\dots e^{A_{i_m}t})$ are nonnegative combinations of the generators of $\text{Cone}_\infty(e^{A_{i_1}t})$, which concludes the first part of the proof.

Now, setting $\bar{A}_{i_h} := A_{i_h}\alpha_h$, we can follow the same proof as before, upon noticing that $\text{Cone}_\infty(e^{\bar{A}_{i_1}t}) = \text{Cone}_\infty(e^{A_{i_1}\alpha_1 t}) = \text{Cone}_\infty(e^{A_{i_1}t})$. This will allow to prove the general statement. ■

Unfortunately, the result of Lemma 8.2 for the asymptotic cone of a single exponential matrix can be only partially extended, thus getting the following proposition.

Proposition 8.11 *Given a set $\mathcal{A} = \{A_1, \dots, A_p\}$ of Metzler matrices and a strictly positive vector $\mathbf{v} \in \mathbb{R}^n_+$, let m be in $\mathbb{N}$ and let $i_1, i_2, \dots, i_m$ be indices in $\mathcal{P}$. If $\mathbf{v} \notin \text{Cone}_\infty^{\bar{\alpha}}(e^{A_{i_1}t}\dots e^{A_{i_m}t})$, for some $\bar{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_m) \in \mathbb{R}^m_+$, then $\exists\ \bar{\tau}_1, \dots, \bar{\tau}_m > 0$ such that $\mathbf{v} \in \partial\,\text{Cone}(e^{A_{i_1}\bar{\tau}_1}\dots e^{A_{i_m}\bar{\tau}_m})$.*

Proof: As in the previous proof, consider initially $\bar{\alpha} = \bar{1}$. Suppose that $\mathbf{v} \notin \text{Cone}_\infty^{\bar{1}}(e^{A_{i_1}t}\dots e^{A_{i_m}t})$. Surely though, $\mathbf{v}$ is an internal point of $\text{Cone}(e^{A_{i_1}0}\dots e^{A_{i_m}0})$. By the continuity of the exponential matrices, the boundary surface of $\text{Cone}(e^{A_{i_1}t}\dots e^{A_{i_m}t})$ evolves continuously with $t \geq 0$. So, if we define a distance $d(t)$ between the vector $\mathbf{v}$ and $\text{Cone}(e^{A_{i_1}t}\dots e^{A_{i_m}t})$ as:

$$d(t) := \inf\{\|v - e^{A_{i_1}t}\dots e^{A_{i_m}t}x\| : x \geq 0\}.$$

Clearly, $d(0) = 0$ and $d(+\infty) > 0$, moreover $d(t)$ is a continuous function. So, once we define $\tau := \sup\{t \geq 0 : d(t) = 0\}$, it is easily seen that $\mathbf{v} \in \text{Cone}(e^{A_{i_1}\tau}\dots e^{A_{i_m}\tau})$ (as polyhedral cones are closed sets) and it must lie on the boundary of the cone, namely on some face, otherwise it would contradict the definition of τ. This further proves that $\tau = \max\{t \geq 0 : d(t) = 0\}$. The extension to the more general statement proceeds along the same lines, by assuming $\bar{A}_{i_h} := A_{i_h}\alpha_h$ and $\tau_h = \alpha_h\bar{\tau}_h$. ■

This immediately brings, as a corollary, a sufficient condition for the PROBLEM solution.

Corollary 8.12 *Consider a set $\mathcal{A} = \{A_1, \dots, A_p\}$ of $n \times n$ Metzler matrices and a nonnegative vector $\mathbf{v} \in \mathbb{R}^n_+$. Set $\mathcal{S} := \overline{\text{ZP}}(\mathbf{v})$, and let $P_\mathcal{S}$ denote the* *(column) selection matrix corresponding to the indices in* $\mathcal{S}$*, and* $\mathbf{v}_\mathcal{S} = P_\mathcal{S}^T \mathbf{v}$ *the subvector obtained by restricting* $\mathbf{v}$ *to the entries corresponding to* $\mathcal{S}$*. If*

$$\mathbf{v}_\mathcal{S} \notin \bigcap_{m\geq 1} \bigcap_{i_1,\dots,i_m \in \mathcal{I}_\mathcal{S}} \bigcap_{\bar{\alpha}\in\mathbb{R}_+^m} \mathrm{Cone}_\infty^{\bar{\alpha}}(P_\mathcal{S}^T e^{A_{i_1}t} \dots e^{A_{i_m}t} P_\mathcal{S}),$$

then $\exists\, i_1, i_2, \dots, i_m \in \mathcal{I}_\mathcal{S}$ *and* $\tau_1, \dots, \tau_m > 0$ *such that* $\mathbf{v} = e^{A_{i_1}\tau_1} \dots e^{A_{i_m}\tau_m}\mathbf{u}$*, with* $\overline{\mathrm{ZP}}(\mathbf{u}) \subsetneq \mathcal{S}$*.*

Unfortunately, up to now, we have not been able to reverse the statement of Proposition 8.11.

8.3 Special Cases

There are some special cases, though, when we are able to forecast that each strictly positive vector lies in the boundary of some cone generated by the product of some exponential matrices.

Proposition 8.13 *Consider a set* $\mathcal{A} = \{A_1, \dots, A_p\}$ *of* $n \times n$ *pairwise commuting Metzler matrices and a nonnegative vector* $\mathbf{v} \in \mathbb{R}_+^n$*. Set* $\mathcal{S} := \overline{\mathrm{ZP}}(\mathbf{v})$*, and let* $P_\mathcal{S}$*,* $\mathbf{v}_\mathcal{S}$ *and* $\mathcal{I}_\mathcal{S}$ *be as in Corollary 8.12. Then, the following facts are equivalent:*

i) $\exists\, i_1, i_2, \dots, i_m \in \mathcal{I}_\mathcal{S}$ *and* $\tau_1, \dots, \tau_m > 0$ *such that* $\mathbf{v} = e^{A_{i_1}\tau_1} \dots e^{A_{i_m}\tau_m}\mathbf{u}$*, with* $\overline{\mathrm{ZP}}(\mathbf{u}) \subsetneq \mathcal{S}$*;*

ii) $\exists\, i_1, i_2, \dots, i_m \in \mathcal{I}_\mathcal{S}$ *such that* $\mathbf{v}_\mathcal{S} \notin \mathrm{Cone}_\infty(P_\mathcal{S}^T e^{A_{i_1}t} \dots e^{A_{i_m}t} P_\mathcal{S})$*;*

iii) $\mathbf{v}_\mathcal{S} \notin \bigcap_{m\geq 1} \bigcap_{i_1,i_2,\dots,i_m \in \mathcal{I}_\mathcal{S}} \mathrm{Cone}_\infty(P_\mathcal{S}^T e^{A_{i_1}t} \dots e^{A_{i_m}t} P_\mathcal{S})$*.*

Proof: i) $\Rightarrow$ ii) Consider the subvector $\mathbf{v}_\mathcal{S}$. By recursively applying Lemma A.7, we can say that there exists a permutation matrix P such that

$$P^T\mathbf{v} = \begin{bmatrix} \mathbf{v}_\mathcal{S} \\ 0 \end{bmatrix}, \qquad P^T A_{i_k} P = \begin{bmatrix} A_{\mathcal{S},k} & * \\ 0 & * \end{bmatrix}, k = 1, 2, \dots, m, \qquad P^T\mathbf{u} = \begin{bmatrix} \mathbf{u}_\mathcal{S} \\ 0 \end{bmatrix}, \tag{8.7}$$

with $A_{\mathcal{S},k}$ of size $|\mathcal{S}| \times |\mathcal{S}|$ and $\mathbf{u}_\mathcal{S} \in \mathbb{R}_+^{|\mathcal{S}|}$, with $\mathcal{R} := \overline{\mathrm{ZP}}(\mathbf{u}_\mathcal{S}) \subsetneq \mathcal{S}$. Moreover,

$$\mathbf{v}_\mathcal{S} = e^{A_{\mathcal{S},1}\tau_1} \dots e^{A_{\mathcal{S},m}\tau_m}\mathbf{u}_\mathcal{S} \gg 0.$$

Note that the commutativity of the matrices A_i's ensures the commutativity of the matrices $A_{\mathcal{S},i}$ and hence of the corresponding exponential matrices. Also, $\mathrm{Cone}_\infty(P_{\mathcal{S}}^T e^{A_{i_1}t} \dots e^{A_{i_m}t} P_{\mathcal{S}}) = \mathrm{Cone}_\infty(e^{A_{\mathcal{S},1}t} \dots e^{A_{\mathcal{S},m}t})$.

We want to show that, for every $\delta_1, \delta_2, \dots, \delta_m > 0$, the vector $\mathbf{v}_{\mathcal{S}}$ does not belong to

$$\mathrm{Cone}\Big(e^{A_{\mathcal{S},1}(\tau_1+\delta_1)} \dots e^{A_{\mathcal{S},m}(\tau_m+\delta_m)}\Big) = \mathrm{Cone}(e^{A_{\mathcal{S},1}\tau_1} \dots e^{A_{\mathcal{S},m}\tau_m} e^{A_{\mathcal{S},1}\delta_1} \dots e^{A_{\mathcal{S},m}\delta_m}).$$

If this were the case, and hence there would be some $\mathbf{w}_{\mathcal{S}} \geq 0$ such that

$$\mathbf{v}_{\mathcal{S}} = e^{A_{\mathcal{S},1}\tau_1} \dots e^{A_{\mathcal{S},m}\tau_m} \Big(e^{A_{\mathcal{S},1}\delta_1} \dots e^{A_{\mathcal{S},m}\delta_m}\Big) \mathbf{w}_{\mathcal{S}},$$

then, by the invertibility of the exponential matrices, it would be

$$\mathbf{u}_{\mathcal{S}} = \Big(e^{A_{\mathcal{S},1}\delta_1} \dots e^{A_{\mathcal{S},m}\delta_m}\Big) \mathbf{w}_{\mathcal{S}}.$$

By Lemma A.2, it must be $\overline{\mathrm{ZP}}\Big(e^{A_{\mathcal{S},1}\delta_1} \dots e^{A_{\mathcal{S},m}\delta_m} \mathbf{e}_{\mathcal{R}}\Big) = \mathcal{R}$.
But then $\overline{\mathrm{ZP}}(e^{A_{\mathcal{S},1}\tau_1} \dots e^{A_{\mathcal{S},m}\tau_m}\ \mathbf{u}_{\mathcal{S}}) = \overline{\mathrm{ZP}}(\mathbf{v}_{\mathcal{S}})$ should be $\mathcal{R}$, too, thus contradicting the strict positivity assumption on $\mathbf{v}_{\mathcal{S}}$.
In particular, the previous result holds if we consider $m$tuples $(\delta_1, \dots, \delta_m)$ such that

$$\delta_i = \Big(\sum_{j \neq i} \tau_j\Big) + t, \qquad t \in \mathbb{R}_+, i \in \{1, 2, \dots, m\}.$$

So, once we set $\tau := \sum_{j=1}^m \tau_j$, we have proved that

$$\mathbf{v}_{\mathcal{S}} \notin \mathrm{Cone}(e^{A_{\mathcal{S},1}(\tau+t)} \dots e^{A_{\mathcal{S},m}(\tau+t)})$$

for every $t \in \mathbb{R}_+$. As a result, $\mathbf{v}_{\mathcal{S}} \notin \mathrm{Cone}_\infty(e^{A_{\mathcal{S},1}(\tau+t)} \dots e^{A_{\mathcal{S},m}(\tau+t)}) = \mathrm{Cone}_\infty(e^{A_{\mathcal{S},1}t} \dots e^{A_{\mathcal{S},m}t})$, since

$$\lim_{t \to +\infty} \frac{e^{A_{\mathcal{S},1}(\tau+t)} \dots e^{A_{\mathcal{S},m}(\tau+t)} \cdot \mathbf{e}_j}{\|e^{A_{\mathcal{S},1}(\tau+t)} \dots e^{A_{\mathcal{S},m}(\tau+t)} \cdot \mathbf{e}_j\|} = \lim_{t \to +\infty} \frac{e^{A_{\mathcal{S},1}t} \dots e^{A_{\mathcal{S},m}t} \cdot \mathbf{e}_j}{\|e^{A_{\mathcal{S},1}t} \dots e^{A_{\mathcal{S},m}t} \cdot \mathbf{e}_j\|} = \mathbf{v}_j^\infty.$$

So, being $\mathbf{v}_{\mathcal{S}}$ strictly positive, $\mathbf{v}_{\mathcal{S}} \notin \mathrm{Cone}_\infty(e^{A_{\mathcal{S},1}t} \dots e^{A_{\mathcal{S},m}t})$.

ii) $\Rightarrow$ i) Has been proved in Corollary 8.12.

ii) $\Leftrightarrow$ iii) is obvious. ∎

Lemma 8.14 *Let $A_1, A_2 \in \mathbb{R}^{n\times n}$ be two Metzler matrices and let $\mathbf{v}$ be a strictly positive vector in $\mathbb{R}^n_+$. If* [2] $\text{Cone}(\mathbf{v}) = \text{Cone}_\infty(e^{A_2 t})$, *and* $\mathbf{v} \in \text{Cone}_\infty(e^{A_1 t})$, *but it is not an eigenvector of A_1, then for every $\tau_1 > 0$ there exists $\tau_2 > 0$ such that* $\mathbf{v} \in \partial\text{Cone}(e^{A_1\tau_1}e^{A_2\tau_2})$.

Proof: This amounts to proving that for every $\tau_1 > 0$ there exists $\tau_2 > 0$ such that

$$\mathbf{v} = e^{A_1\tau_1}e^{A_2\tau_2}\mathbf{u} \ \Leftrightarrow \ e^{-A_1\tau_1}\mathbf{v} = e^{A_2\tau_2}\mathbf{u},$$

for some $\mathbf{u} > 0$ with $\text{ZP}(\mathbf{u}) \neq \emptyset$. We first observe that for every $\tau_1 > 0$, $\mathbf{w} := e^{-A_1\tau_1}\mathbf{v}$ is not a multiple of $\mathbf{v}$ and hence it does not belong to $\text{Cone}_\infty(e^{A_2 t})$. On the other hand, since $\mathbf{v} \in \text{Cone}_\infty(e^{A_1 t})$, then $\mathbf{v}$ is an internal point of $\text{Cone}(e^{A_1 t})$, for every $t \geq 0$. So, in particular, $\mathbf{v}$ is an internal point of $\text{Cone}(e^{A_1\tau_1})$, which amounts to saying that $\mathbf{v} = e^{A_1\tau_1}\mathbf{u}_1$ for some $\mathbf{u}_1 \gg 0$. Clearly, by the invertibility of the exponential matrix, $\mathbf{w} = \mathbf{u}_1 \gg 0$. So, we have shown that $\mathbf{w}$ is a strictly positive vector which does not belong to $\text{Cone}_\infty(e^{A_2 t})$. This implies that $\mathbf{w} \in \partial\text{Cone}(e^{A_2\tau_2})$ for some $\tau_2 > 0$, and hence $\mathbf{w} = e^{A_2\tau_2}\mathbf{u}$, for some $\mathbf{u} > 0$ with $\text{ZP}(\mathbf{u}) \neq \emptyset$. This completes the proof. ∎

The previous technical result leads to the following sufficient condition for reachability

Proposition 8.15 *An n-dimensional continuous-time positive switched system (2.1), commuting among n single-input subsystems $(A_i, b_i), i = 1, \dots, n$, is reachable if for every proper subset $\mathcal{S} \subset \langle n\rangle$ we have:*

a) if $|\mathcal{S}| = 1$, then $\exists\, j(\mathcal{S}) \in \mathcal{I}_\mathcal{S}$ such that $\overline{\text{ZP}}(b_{j(\mathcal{S})}) = \mathcal{S}$;

b) if $|\mathcal{S}| > 1$, then either

1. $\exists\, j(\mathcal{S}) \in \mathcal{I}_\mathcal{S}$ such that $\overline{\text{ZP}}(b_{j(\mathcal{S})}) \subset \mathcal{S}$,

or

2. $\exists\, j_i(\mathcal{S}), j_k(\mathcal{S}) \in \mathcal{I}_\mathcal{S}$ such that $P_\mathcal{S}^T A_{j_i(\mathcal{S})} P_\mathcal{S}$ is irreducible and its strictly positive eigenvector (of unitary modulus) is not an eigenvector of $P_\mathcal{S}^T A_{j_k(\mathcal{S})} P_\mathcal{S}$.

[2]Note that this happens if and only if A_2 is irreducible, and when so $\mathbf{v}$ is a (strictly positive) dominant eigenvector.

Proof: We only need to show that condition b) - 2. implies condition iib - 2) in Proposition 7.5.

To this end, let $\mathbf{v}$ be a positive vector with $\mathcal{S} = \overline{\mathrm{ZP}}(\mathbf{v})$ of cardinality greater than 1, and notice that, under assumption b)-2., there exists $j_i \in \mathcal{I}_{\mathcal{S}}$ such that $\mathrm{Cone}_\infty(P_{\mathcal{S}}^T A_{j_i} P_{\mathcal{S}})$ coincides with the cone generated by a single strictly positive vector $\mathbf{w}$.

Let $\mathbf{v}_{\mathcal{S}}$ be the restriction of $\mathbf{v}$ to the indices corresponding to $\mathcal{S}$. If $\mathbf{v}_{\mathcal{S}} \neq \mathbf{w}$, then $\mathbf{v}_{\mathcal{S}} \notin \mathrm{Cone}_\infty(e^{P_{\mathcal{S}}^T A_{j_i} P_{\mathcal{S}} t})$, and hence there exists $\tau > 0$ such that $\mathbf{v}_{\mathcal{S}} \in \partial\mathrm{Cone}(e^{P_{\mathcal{S}}^T A_{j_i} P_{\mathcal{S}} \tau})$. This ensures that $\mathbf{v}$ is reachable. If $\mathbf{v}_{\mathcal{S}} = \mathbf{w}$, then either $\mathbf{v}_{\mathcal{S}} \notin \mathrm{Cone}_\infty(e^{P_{\mathcal{S}}^T A_j P_{\mathcal{S}} t})$ for some other $j \in \mathcal{I}_{\mathcal{S}}$ (and if so, by repeating the previous argument, we may say that $\mathbf{v}_{\mathcal{S}} \in \partial\mathrm{Cone}(e^{P_{\mathcal{S}}^T A_j P_{\mathcal{S}} \tau}), \exists\, \tau > 0$, and hence $\mathbf{v}$ is reachable), or for every $j \neq i, j \in \mathcal{I}_{\mathcal{S}}$, we have $\mathbf{v}_{\mathcal{S}} \in \mathrm{Cone}_\infty(e^{P_{\mathcal{S}}^T A_j P_{\mathcal{S}} t})$. For one such index $j_k \in \mathcal{I}_{\mathcal{S}}$, though, $\mathbf{v}_{\mathcal{S}}$ is not an eigenvector of $P_{\mathcal{S}}^T A_{j_k} P_{\mathcal{S}}$. So, by applying Lemma 8.14, we may say that there exist $\tau_k, \tau_i > 0$ such that

$$\mathbf{v}_{\mathcal{S}} = e^{P_{\mathcal{S}}^T A_{j_k} P_{\mathcal{S}} \tau_k} e^{P_{\mathcal{S}}^T A_{j_i} P_{\mathcal{S}} \tau_i} \mathbf{u}_{\mathcal{S}},$$

for some positive vector $\mathbf{u}_{\mathcal{S}}$, with $\mathrm{ZP}(\mathbf{u}_{\mathcal{S}}) \neq \emptyset$. This ensures, again, that $\mathbf{v}$ is reachable. ■

Lemma 8.16 *Let A be an $n \times n$ Metzler matrix in Frobenius normal form (1.2), and let $\mathcal{C}_{j_1}, \dots, \mathcal{C}_{j_r}$ be the distinguished classes*[3] *in $\mathcal{G}(A)$. We know that if $\tilde{\mathbf{v}}_i$, $i \in \langle r \rangle$, is the positive eigenvector corresponding to the dominant eigenvalue $\lambda_{\max}(A_{j_i j_i})$ of the distinguished class $\mathcal{C}_{j_i}$, then $\mathrm{Cone}_\infty(e^{At})$ is the cone generated by the full column rank positive matrix $V_\infty = [\tilde{\mathbf{v}}_1 \ \dots \ \tilde{\mathbf{v}}_r]$.*

Suppose that the $s \leq r$ distinct eigenvalues the previous eigenvectors correspond to are ordered as $\lambda_1 < \lambda_2 < \cdots < \lambda_s$. Define, for every $k = 1, \dots, s$, the following sets:

- $I_k := \{i \in \langle r \rangle : \tilde{\mathbf{v}}_i$ *is an eigenvector corresponding to* $\lambda_k\}$;
- $\mathcal{D}_k := \bigcup_{i \in I_k} \mathcal{D}(\mathcal{C}_{j_i})$;
- $\mathcal{V} := \{k \in \langle s \rangle : \bigcup_{j \geq k} \mathcal{D}_j \neq \langle \ell \rangle\}$.

Then, for any $k \in \langle s \rangle$, there exists a positive vector $\mathbf{c} \in \mathbb{R}_+^r$ satisfying

[3]Note that we do not introduce any specific ordering within the set of indices $\{j_1, \dots, j_r\}$.

- $V_\infty \mathbf{c}$ *is strictly positive;*
- $k := \min\{i \in \langle s\rangle : \overline{\mathrm{ZP}}(\mathbf{c}) \cap I_i \neq \emptyset\}$.

if and only if $k \notin \mathcal{V}$.

Proof: Notice, first, that since $\tilde{\mathbf{v}}_i$, $i \in \langle r\rangle$, is the eigenvector corresponding to the dominant eigenvalue of the distinguished class $\mathcal{C}_{j_i}$, its nonzero pattern obeys the following rules (see Proposition A.10):

$$\mathrm{block}_k[\tilde{\mathbf{v}}_i] = \begin{cases} \gg 0, & \text{if } k \in \mathcal{D}(\mathcal{C}_{j_i}); \\ 0, & \text{otherwise.} \end{cases} \tag{8.8}$$

For any index $k \in \langle s\rangle$, the set $\mathcal{D}_k$ represents the set of indices of those classes that are reached by (at least) one distinguished class corresponding to λ_k. Clearly, as I_k is the set of indices in $\{1, 2, \dots, r\}$ such that $\tilde{\mathbf{v}}_i$ is an eigenvector corresponding to λ_k, then, by (8.8), $\overline{\mathrm{ZP}}(\sum_{i\in I_k} \tilde{\mathbf{v}}_i) = \mathcal{D}_k$. Finally, $\mathcal{V}$ represents the set of all indices $k \in \langle s\rangle$ for which

$$\overline{\mathrm{ZP}}\left(\sum_{i\in I_k\cup I_{k+1}\cup\cdots\cup I_r} \tilde{\mathbf{v}}_i\right) \neq \langle n\rangle, \text{ namely } \sum_{i\in I_k\cup I_{k+1}\cup\cdots\cup I_r} \tilde{\mathbf{v}}_i \text{ is not strictly positive.}$$

As a consequence, if $k \in \mathcal{V}$, then there is no way of finding some $\mathbf{c} \in \mathbb{R}^r_+$ such that $V_\infty \mathbf{c} \gg 0$ and $k := \min\{i : \overline{\mathrm{ZP}}(\mathbf{c}) \cap I_i \neq \emptyset\}$.

Conversely, if $k \notin \mathcal{V}$, there exists $\mathbf{c} \in \mathbb{R}^r_+$ such that $k := \min\{i : \overline{\mathrm{ZP}}(\mathbf{c}) \cap I_i \neq \emptyset\}$ and $V_\infty \mathbf{c} \gg 0$. ∎

Remark 8.17 *Note that the hypothesis requiring the matrix A to be in Frobenius normal form is not necessary at all to conclude the result, even though it greatly simplifies the notation of the proof.*

Proposition 8.18 *Let A_1 and A_2 be two $n \times n$ Metzler matrices, and adopt the same notation as in Lemma 8.16, where all the symbols $\tilde{\mathbf{v}}_1, \dots, \tilde{\mathbf{v}}_r, \lambda_1, \dots, \lambda_s, I_k, \mathcal{V}$ now refer to the matrix A_1, and we assume $\lambda_1 < \cdots < \lambda_s$. If each positive eigenvector of A_1 belonging to $\cup_{k\notin\mathcal{V}} \mathcal{K}_k$, with*

$$\mathcal{K}_k := \mathrm{Cone}(\{\tilde{\mathbf{v}}_i, i \in I_k\}),$$

does not belong to $\mathrm{Cone}_\infty(e^{A_2 t})$, *then for every strictly positive vector $\mathbf{v} \in \mathbb{R}^n_+$ there exists $\tau_1, \tau_2 \geq 0$ such that $v \in \partial\mathrm{Cone}(e^{A_1\tau_1} e^{A_2\tau_2})$.*

Proof: We already know from Lemma 8.2 that the result is true for every $\mathbf{v} \notin \text{Cone}_\infty(e^{A_1 t}) \cap \text{Cone}_\infty(e^{A_2 t})$, by setting either τ_1 or τ_2 equal to zero. Pick now $\mathbf{v} \in \text{Cone}_\infty(e^{A_1 t}) \cap \text{Cone}_\infty(e^{A_2 t})$. For two positive time instants τ_1, τ_2 to exist such that $\mathbf{v} = e^{A_1\tau_1} e^{A_2\tau_2}\mathbf{u}$ for some $\mathbf{u} \in \partial\mathbb{R}^n_+$, a time instant τ_1 must exist such that the vector $\mathbf{w}(\tau_1) := e^{-A_1\tau_1}\mathbf{v}$ does not belong to $\text{Cone}_\infty(e^{A_2 t})$.

Since $\mathbf{v} \in \text{Cone}_\infty(e^{A_1 t}) = \text{Cone}(V_\infty)$, it can be expressed as $\mathbf{v} = \sum_{i=1}^r c_i \tilde{\mathbf{v}}_i$. On the other hand, by the strict positivity of $\mathbf{v}$ it also follows that $k := \min\{i : \overline{\text{ZP}}(\mathbf{c}) \cap I_i \neq \emptyset\}$ does not belong to $\mathcal{V}$ (this comes directly from the definition of the set $\mathcal{V}$ itself). As a consequence, if we express $\mathbf{v}$ as

$$\mathbf{v} = \sum_{i\in I_k} c_i\tilde{\mathbf{v}}_i + \sum_{i\in I_{k+1}} c_i\tilde{\mathbf{v}}_i + \cdots + \sum_{i\in I_s} c_i\tilde{\mathbf{v}}_i,$$

then

$$\mathbf{w}(\tau_1) = \left(\sum_{i\in I_k} c_i\tilde{\mathbf{v}}_i\right) e^{-\lambda_k\tau_1} + \left(\sum_{i\in I_{k+1}} c_i\tilde{\mathbf{v}}_i\right) e^{-\lambda_{k+1}\tau_1} + \cdots + \left(\sum_{i\in I_s} c_i\tilde{\mathbf{v}}_i\right) e^{-\lambda_s\tau_1}.$$

As τ_1 goes to $+\infty$, $\mathbf{w}(\tau_1)$ will converge to the eigenvector $\mathbf{w}(+\infty) := \sum_{i\in I_k} c_i$ $\tilde{\mathbf{v}}_i$, corresponding to the eigenvalue $\lambda_k, k \notin \mathcal{V}$.

But then, by the proposition's assumptions, $\mathbf{w}(+\infty) \notin \text{Cone}_\infty(e^{A_2 t})$, and since $\text{Cone}_\infty(e^{A_2 t})$ is a closed set, it is possible to find some $0 < \bar{\tau}_1 < +\infty$ such that for every $\tau_1 > \bar{\tau}_1$, $\mathbf{w}(\tau_1) \notin \text{Cone}_\infty(e^{A_2 t})$. So, by Lemma 8.2, for every such τ_1 it will be possible to find some $\tau_2 > 0$ such that $\mathbf{w}(\tau_1) \in \partial\text{Cone}(e^{A_2\tau_2})$, and hence $\mathbf{v} \in \partial\text{Cone}(e^{A_1\tau_1} e^{A_2\tau_2})$. ∎

Chapter 9

Optimal Reset Map for Switched Systems

When dealing with the control of complex systems, multiple conflicting requirements on the closed-loop system often make a single linear time-invariant (LTI) controller unsuitable [7]. In this context, a convenient solution consists of designing several LTI controllers with transfer functions $\{k_p(s) : p \in \mathcal{P}\}$, each one of them designed to meet only some specifications, and then switching between them in order to achieve the best overall performance [24, 25, 50, 33]. In this chapter we do not address the problem of how to select a "suitable" switching sequence, but instead focus on how to guarantee stability of the switched closed-loop and how to obtain the best possible performance for an externally given desired sequence of controllers.

9.1 Background and problem formulation

Consider a linear time-invariant process Σ, with transfer function $g(s)$ from the input $u(t)$ to the output $y(t)$, and let $\{k_p(s)\}_{p\in\mathcal{P}}$ be a finite family of controller transfer functions from the tracking error $e_T(t) := r(t) - y(t)$ to the control input $u(t)$, where $r(t)$ denotes a piecewise constant reference signal, all of them stabilizing the plant Σ. The switched system considered here arises from the feedback interconnection of the plant Σ to be controlled with a *multicontroller* $\mathbf{C}(\sigma)$ whose inputs are the usual tracking error $e_T(t)$ as well as a piecewise constant *switching signal* $\sigma : [0, +\infty) \to \mathcal{P}$ that roughly determines which should be the controller transfer function at time t (cf. Figure 9.1). In particular, given an n-dimensional state space realization

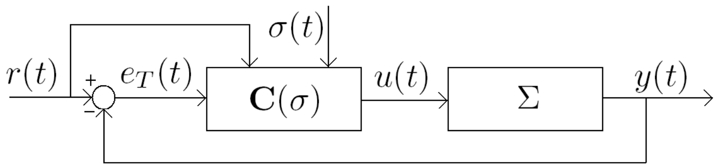


Figure 9.1: Controller architecture

(E_p, F_p, G_p, H_p) for each $k_p(s)$, $p \in \mathcal{P}$, the state $x_{\text{mult}}(t)$ of the multicontroller $\mathbf{C}(\sigma)$ evolves according to

$$\begin{cases} \dot{x}_{\text{mult}}(t) = E_{\sigma(t)} x_{\text{mult}}(t) + F_{\sigma(t)} e_T(t) \\ u(t) = G_{\sigma(t)} x_{\text{mult}}(t) + H_{\sigma(t)} e_T(t) \end{cases} \tag{9.1}$$

on any time interval on which the switching signal $\sigma(t)$ remains constant, and according to

$$x_{\text{mult}}(t) = F(x_{\text{mult}}(t^-), \sigma(t^-), \sigma(t), r(t)), \tag{9.2}$$

at every time t, called a *switching time*, at which $\sigma(t)$ is discontinuous. The function $F(\cdot)$ is called the *reset map* and, given a signal $z(\cdot)$, we denote by $z(t^-)$ its limit from the left at time t, $\lim_{\tau \uparrow t} z(\tau)$. All signals are assumed to be continuous from the right. One should emphasize that even if each $k_p(s)$ stabilizes the process, as the value of $\sigma(t)$ switches within $\mathcal{P}$ the stability of the closed loop may be lost [25, 33].

A wide body of literature is available on conditions for the uniform stability of switched systems [14, 35, 34, 33] and on the optimal control of such systems [52, 58, 39, 2, 57, 3]. However, the special structure given by the feedback interconnection of a multicontroller with a non-switching plant provides a special structure that is usually not taken into account in the previously mentioned works.

Besides the choice of the controller transfer functions $k_p(s)$ and the selection of the switching signal $\sigma(t)$, there are two additional degrees of freedom available to the designer of the multi-controller: the selection of state space realizations for each controller and the construction of the reset map. A study of how these choices affect system stability appeared in [25], which addressed the problem of finding realizations and reset maps for a given family of stabilizing controller transfer functions such that the closed-loop system remains uniformly stable for every switching signal $\sigma(t)$. However, [25] does not explore the performance implications of the controller realizations and reset maps. In addition, this reference only considers very special types of reset maps.

The more recent paper [50] considers a similar setup and suggests the creation of several "candidate" control signals $v_p(t)$, $p \in \mathcal{P}$, one for each controller, by letting each individual controller evolve continuously without resets. Then the switching signal $\sigma(t)$ selects which one among the $v_p(t)$'s should be actually employed for control purposes. In order to get a smoother transient response, the piecewise continuous signal $v(t) := v_{\sigma(t)}(t)$ thus obtained is filtered in order to generate a continuous control input $u(t)$.

Inspired by the control scheme proposed in [25], we deal with the issue of appropriately designing the reset map, so that the closed-loop switched system produces transients that minimize a given cost function, while preserving the (input-to-state) stability of the closed-loop switched system. Simulation results compare the performance of our switching controller with those in [25] and [50].

For simplicity we restrict our attention to asymptotically stable single input single output (SISO) processes Σ. However, all the results presented here could be generalized to not necessarily stable multiple input multiple output processes (MIMO) processes, following the approach in [25].

The construction of the multicontroller follows [25] and is inspired by the Youla-Kucera parametrization of all the stabilizing controllers. This parameterization motivates us to express each controller transfer function as

$$k_p(s) = \frac{q_p(s)}{1 - q_p(s)g(s)}, \tag{9.3}$$

which can be viewed as a *positive* feedback interconnection between a system with transfer function $g(s)$ and an asymptotically stable system with transfer function

$$q_p(s) = \frac{k_p(s)}{1 + g(s)k_p(s)}.$$

Note that $q_p(s)$ is indeed asymptotically stable because $k_p(s)$ stabilizes $g(s)$ (see Figure 9.2).

Let (A, B, C) denote a stabilizable and detectable n_{pl}-dimensional realization for the process transfer function $g(s)$ and select stabilizable and detectable realizations (A_p, B_p, C_p, D_p) for each $q_p(s)$, $p \in \mathcal{P}$. The (not necessarily minimal) realizations for all the $q_p(s)$ should have the same dimension n_{cn} and all the Hurwitz matrices A_p should admit the squared-norm of the state as a Lyapunov function, which is to say that

$$A_p + A_p^T < 0, \quad \forall p \in \mathcal{P}. \tag{9.4}$$

This is always possible because of [25, Lemma 7]. In view of the diagram in Figure 9.2, each transfer function $k_p(s)$ has a realization (E_p, F_p, G_p, H_p) with

$$E_p := \begin{bmatrix} A_p & B_pC \\ BC_p & A + BD_pC \end{bmatrix}, \qquad F_p := \begin{bmatrix} B_p \\ BD_p \end{bmatrix}, \tag{9.5}$$

$$G_p := \begin{bmatrix} C_p & D_pC \end{bmatrix}, \qquad H_p := D_p. \tag{9.6}$$

The proposed multicontroller $\mathbf{C}_\sigma$ has state $x_{\mathrm{mult}}(t) := [x_{\mathrm{cn}}(t)^T \, x_{\mathrm{copy}}(t)^T]^T$, which evolves according to (9.1) on any time interval on which $\sigma(t)$ remains constant and

$$x_{\mathrm{mult}}(t) = \begin{bmatrix} x_{\mathrm{cn}}(t) \\ x_{\mathrm{copy}}(t) \end{bmatrix} = \begin{bmatrix} F(x_{\mathrm{mult}}(t^-), \sigma(t^-), \sigma(t), r(t)) \\ x_{\mathrm{copy}}(t^-) \end{bmatrix} \tag{9.7}$$

at every switching time t. Note that (9.7) slightly differs from (9.2), since in the former the reset map $F(\cdot)$ only affects the component x_{cn} of the state of $\mathbf{C}(\sigma)$. In (9.7), the component x_{copy} of the state of $\mathbf{C}(\sigma)$ remains continuous and it will actually converge to the process state $x_{\mathrm{pl}}(t)$. In fact, it follows from (9.1) and (9.7) that

$$\dot{x}_{\mathrm{copy}}(t) = Ax_{\mathrm{copy}}(t) + Bu(t) \tag{9.8}$$

for all times and, because we are assuming that A is asymptotically stable, we indeed have that $x_{\mathrm{copy}}(t)$ converges to $x_{\mathrm{pl}}(t)$, regardless of the control signal $u(t)$.

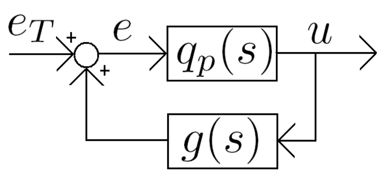


Figure 9.2: Controller $k_p(s)$

Connecting the plant Σ with realization (A, B, C) with the multicontroller $\mathbf{C}(\sigma)$, through the negative feedback interconnection in Figure 9.1, results in a switched system with a state $x(t) := [x_{\text{pl}}(t)^T \, x_{\text{cn}}(t)^T \, x_{\text{copy}}(t)^T]^T$ that evolves according to

$$\dot{x}(t) = \hat{A}_{\sigma(t)} x(t) + \hat{B}_{\sigma(t)} r(t), \qquad y(t) = \hat{C}_{\sigma(t)} x(t) \tag{9.9}$$

on any time interval on which $\sigma(t)$ remains constant and

$$x(t) = \begin{bmatrix} x_{\text{pl}}(t) \\ x_{\text{cn}}(t) \\ x_{\text{copy}}(t) \end{bmatrix} = \begin{bmatrix} x_{\text{pl}}(t^-) \\ F(x(t^-), \sigma(t^-), \sigma(t), r(t)) \\ x_{\text{copy}}(t^-) \end{bmatrix} \tag{9.10}$$

at every switching time t. The matrices in (9.9) are defined by

$$\hat{A}_p := \begin{bmatrix} A & BC_p & BD_pC \\ -B_pC & A_p & B_pC \\ -BD_pC & BC_p & A + BD_pC \end{bmatrix}, \qquad \hat{B}_p := \begin{bmatrix} BD_p \\ B_p \\ BD_p \end{bmatrix}$$
$$\hat{C}_p := \begin{bmatrix} C & 0 & 0 \end{bmatrix}, \quad \forall p \in \mathcal{P}.$$

with all the $\hat{A}_p$ Hurwitz, since $k_p(s)$ stabilizes $g(s)$.

The remainder of this chapter is focused on the goal of selecting an appropriate reset map $F(\cdot)$ that achieves optimal transient performance at switching times, while maintaining (9.9)–(9.10) stable. The issue of optimizing transient performance is the subject of the next section.

9.2 An optimal definition of the reset map

Suppose that at time $t = t_0$ the switching signal jumps from $\sigma(t_0^-) = p$ to $\sigma(t_0) = q$. Our goal is to select the post-switching state defined by the reset map

$$x_{\text{cn}}(t_0) = F(x_{\text{mult}}(t_0^-), p, q, r(t_0)) \tag{9.11}$$

so as to optimize the resulting transient performance, as measured by a quadratic cost of the following form

$$J = \int_{t_0}^{t_1} \Big(e_T(t)^T R e_T(t) + \dot{y}(t)^T W \dot{y}(t) + u(t)^T K u(t) \Big) dt + \big(x(t_1) - x_\infty\big)^T T \big(x(t_1) - x_\infty\big) \tag{9.12}$$

where R, W, K, T are appropriately selected symmetric positive semi-definite matrices and

$$x_\infty := -\hat{A}_q^{-1}\hat{B}_q r(t_0). \tag{9.13}$$

The choice of the matrices R, W, K, T allows one to penalize the tracking error magnitude, output rate of change, control effort, and final state magnitude, respectively. In performing this optimization, it will be assumed that the switching signal $\sigma(t)$ and the reference $r(t)$ remain constant and equal to q and $r(t_0)$, respectively, over the optimization horizon $[t_0, t_1]$. If $\sigma(t)$ turns out to switch again before t_1, the value to which x_{cn} was reset at time t_0, will generally not be optimal. However, we will later make sure that even in this case, stability is guaranteed. Note that the vector x_∞ that appears in the terminal term in (9.12) is the steady-state value to which $x(t)$ would converge as $t \to \infty$ if both $\sigma(t)$ and $r(t)$ were to remain constant.

9.2.1 Optimization of transient performance

To find the value of $x_{\mathrm{cn}}(t_0)$ that minimizes (9.12) we need to introduce some notation. Let Q_q denote the symmetric solution to the following Lyapunov equation

$$Q_q\hat{A}_q + \hat{A}_q^T Q_q = -P_q, \tag{9.14}$$

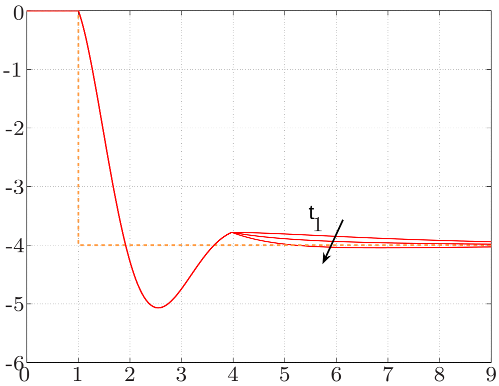
where $P_q := \hat{C}_q^T R\hat{C}_q + \hat{A}_q^T\hat{C}_q^T W\hat{C}_q\hat{A}_q + \tilde{C}_q^T K\tilde{C}_q \geq 0$ and $\tilde{C}_q := \begin{bmatrix}-D_qC & C_q & D_qC\end{bmatrix}$. Such solution exists and is at least positive semi-definite because $\hat{A}_q$ is a Hurwitz matrix.

Set $\Delta := t_1 - t_0$ and introduce the positive semi-definite matrix $M_q := Q_q - e^{\hat{A}_q^T\Delta}(Q_q + T)e^{\hat{A}_q\Delta}$ and the vector

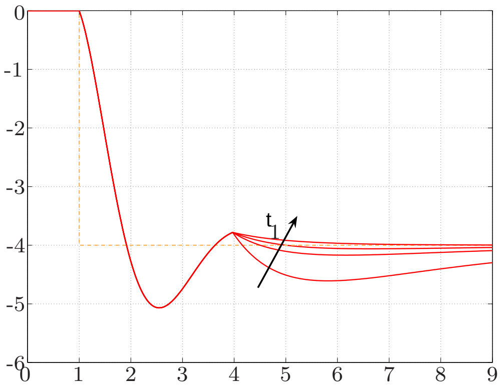
$$\begin{aligned} g_q^T := 2r(t_0)^T\Big[\big(-R\hat{C}_q + \hat{B}_q^T\hat{C}_q^T W\hat{C}_q\hat{A}_q + D_q^T K\tilde{C}_q + {}\\ + Q_q\hat{B}_q^T Q_q\big)\big(I - e^{\hat{A}_q\Delta}\big)\hat{A}_q^{-1} + {}\\ + \hat{B}_q^T\big(e^{\hat{A}_q^T\Delta}(\hat{A}_q^T)^{-1}(Q_q - T) - (\hat{A}_q^T)^{-1}Q_q\big)e^{\hat{A}_q\Delta}\Big]. \end{aligned}$$

We will further need to block-partition the symmetric matrices M_q and the vectors g_q according to the partition in (9.10) of the state vector:

$$M_q = \begin{bmatrix} M_{11}^q & M_{12}^q & M_{13}^q \\ M_{21}^q & M_{22}^q & M_{23}^q \\ M_{31}^q & M_{32}^q & M_{33}^q \end{bmatrix}, \qquad g_q = \begin{bmatrix} g_1^q \\ g_2^q \\ g_3^q \end{bmatrix}$$



(a) $R = I$, $W \neq 0$, $K, T = 0$



(b) $R = I$, $W, K, T = 0$

Figure 9.3: Transient responses for different weighting matrices and different durations for the optimization interval $[t_0, t_1]$. Details on the process and controllers being switched can be found in Section 9.5. In both plots there is a single controller switching at time $t_0 = 4$ sec.

and perform a singular value decomposition of

$$M_{22}^q = \begin{bmatrix} U_1^q \; U_2^q \end{bmatrix} \begin{bmatrix} \Lambda_q & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} (V_1^q)^T \\ (V_2^q)^T \end{bmatrix}$$

(with Λ_q nonsingular). We are now ready to provide the solution to the minimization problem considered above.

Theorem 9.1 *Assuming that* $\sigma(t) = q$ *and* $r(t) = r(t_0)$, $\forall t \in [t_0, t_1]$, *the global minimum to* (9.12) *with smallest norm is given by*

$$x_{\mathrm{cn}}^*(t_0) = V_1^q \Lambda_q^{-1} (U_1^q)^T \cdot \left[\frac{1}{2} g_2^q - \left((M_{12}^q)^T x_{\mathrm{pl}}(t_0) + M_{23}^q x_{\mathrm{copy}}(t_0) \right) \right]. \qquad (9.15)$$

Figure 9.3 depicts the result of numerical simulations, illustrating how varying the length of the optimization interval may influence the system's behavior. It generally happens that the transient response improves as we increase the optimization interval. It should be noted that all the results in this section hold for an infinite horizon, as we make $t_1 \to \infty$ in (9.12), in which case all the matrix exponentials $e^{\hat{A}_q \Delta}$ above become the zero matrix.

Proof: [Theorem 9.1] We start by computing the criterion J in (9.12) along a solution

$$x(t) = e^{\hat{A}_q(t-t_0)}x(t_0) + \int_{t_0}^{t} e^{\hat{A}_q(t-\tau)}\hat{B}_q \cdot r(t_0)\, d\tau,$$
$$u(t) = \tilde{C}_q x(t) + D_q r(t),$$

$\forall t \in [t_0, t_1]$. In what follows, $*$ stands for additive terms that do not depend on the value of $x_{\mathrm{cn}}(t_0)$. Straightforward algebra shows that the terminal term in (9.12) is given by

$$\begin{aligned}\big(x(t_1) - x_\infty\big)^T T\big(x(t_1) - x_\infty\big) &= x(t_0)^T e^{\hat{A}_q^T\Delta} T e^{\hat{A}_q\Delta} x(t_0) + \\ &+ 2r(t_0)^T \hat{B}_q^T e^{\hat{A}_q^T\Delta}(\hat{A}_q^T)^{-1} T e^{\hat{A}_q\Delta} x(t_0) + * \end{aligned} \tag{9.16}$$

and that the integral term in (9.12) is given by

$$\begin{aligned}&\int_{t_0}^{t_1} \Big(e_T(t)^T R e_T(t) + \dot{y}(t)^T W \dot{y}(t) + u(t)^T K u(t)\Big) dt \\ &\quad = \int_{t_0}^{t_1} \Big(r(t_0)^T \big(R + \hat{B}_q^T \hat{C}_q^T W \hat{C}_q \hat{B}_q + D_q^T K D_q\big) r(t_0) \\ &\qquad + x(t)^T P_q x(t) + c_q^T x(t)\Big) dt + * \\ &\quad = \int_{t_0}^{t_1} \Big(x(t)^T P_q x(t) + c_q^T x(t)\Big) dt + *, \end{aligned} \tag{9.17}$$

where P_q has been already defined and $c_q^T = 2r(t_0)^T(-R\hat{C}_q + \hat{B}_q^T\hat{C}_q^T W \hat{C}_q \hat{A}_q + D_q^T K \tilde{C}_q)$. The computational of the integral in (9.17) is fairly standard:

$$\begin{aligned}&\int_{t_0}^{t_1} \Big(x(t)^T P_q x(t) + c_q^T x(t)\Big) dt \\ &\quad = \int_{t_0}^{t_1} \Big(2\, r(t_0)^T \hat{B}_q^T Q_q x(t) - \frac{d}{dt}(x^T Q_q x) + c_q^T x(t)\Big) dt \\ &\quad = \int_{t_0}^{t_1} (2\, r(t_0)^T \hat{B}_q^T Q_q + c_q^T) x(t)\, dt - x^T Q_q x\Big|_{t_0}^{t_1} \\ &\quad = \int_{t_0}^{t_1} (2\, r(t_0)^T \hat{B}_q^T Q_q + c_q^T) e^{\hat{A}_q(t-t_0)} x(t_0) dt - x^T Q_q x\Big|_{t_0}^{t_1} + * \\ &\quad = x(t_0)^T \Big(Q_q - e^{\hat{A}_q^T\Delta} Q_q e^{\hat{A}_q\Delta}\Big) x(t_0) - \\ &\quad - \Big(\big(c_q^T + 2r(t_0)^T Q_q \hat{B}_q^T Q_q\big)\big(I - e^{\hat{A}_q\Delta}\big)\hat{A}_q^{-1} \\ &\qquad + 2r(t_0)^T \hat{B}_q^T (e^{\hat{A}_q^T\Delta} - I)(\hat{A}_q^T)^{-1} Q_q e^{\hat{A}_q\Delta}\Big) x(t_0) + * \end{aligned} \tag{9.18}$$

Combining (9.16) and (9.18), we conclude that

$$J = x(t_0)^T M_q x(t_0) - x(t_0)^T g_q + *,$$

but since our optimization is only performed on the component $x_{\mathrm{cn}}(t_0)$ of $x(t_0)$, we further re-write

$$J = x_{\mathrm{cn}}(t_0)^T M_{22}^q x_{\mathrm{cn}}(t_0) + x_{\mathrm{cn}}(t_0)^T \\ \big(2((M_{12}^q)^T x_{\mathrm{pl}}(t_0) + M_{23}^q x_{\mathrm{copy}}(t_0) - g_2^q\big) + *. \quad (9.19)$$

Since M_{22}^q is positive semi-definite, (9.19) is convex on $x_{\mathrm{cn}}(t_0)$ and any vector $x_{\mathrm{cn}}^*(t_0)$ satisfying the first order condition

$$M_{22}^q x_{\mathrm{cn}}^*(t_0) = \frac{1}{2} g_2^q - (M_{12}^q)^T x_{\mathrm{pl}}(t_0) - M_{23}^q x_{\mathrm{copy}}(t_0) \quad (9.20)$$

provides a global minimum to J [8]. In general, (9.20) may not be solvable, but in our specific problem it can be proved that it always is (see Lemma A.25) and the minimum norm solution to (9.20) is given by (9.15). ■

9.2.2 Choice of the reset map

Since the optimal value for $x_{\mathrm{cn}}^*(t_0)$ in (9.15) depends on the process state $x_{\mathrm{pl}}(t_0)$ that is generally not accessible, we cannot directly use the expression in (9.15) to define the optimal reset map in (9.11). However, as mentioned in Section 9.1, the component $x_{\mathrm{copy}}(t)$ of the multicontroller state converges exponentially fast to the process state $x_{\mathrm{pl}}(t)$, for every control input $u(t)$. If we then replace $x_{\mathrm{pl}}(t_0)$ by $x_{\mathrm{copy}}(t_0)$ in (9.15), we obtain an "asymptotically correct" minimum to (9.12), which justifies the following reset map

$$F(x_{\mathrm{mult}}(t_0^-), p, q, r(t_0)) := V_1^q \Lambda_q^{-1} (U_1^q)^T \cdot \\ \Big(\frac{1}{2} g_2^q - (M_{12}^q)^T x_{\mathrm{copy}}(t_0^-) - M_{23}^q x_{\mathrm{copy}}(t_0^-)\Big). \quad (9.21)$$

Since $F(\cdot)$ is a function of the "past" value of x_{mult}, the right-hand side of (9.21) must only depend on the "past" value of x_{copy}. However, this is not a problem because $x_{\mathrm{copy}}(t_0) = x_{\mathrm{copy}}(t_0^-)$, in view of (9.10).

9.3 Stable reset map

Although the reset map (9.21) minimizes the criteria (9.12), it may not necessarily result in a stable switched system. This is mainly because the optimization assumed that no further switching would occur in the interval $[t_0, t_1]$.

If one were willing to exclude the possibility of consecutive switching times separated by less than a given positive constant τ_D — a condition often referred to as *dwell-time switching* — then it would be possible to compute a sufficiently large τ_D for which the switched system (9.9)–(9.10) would be stable. However, here we do not want to make such an assumption on $\sigma(t)$. Instead, we want to appropriately modify the reset-map so as to guarantee stability *for every* piecewise continuous switching signal $\sigma(t)$, without significantly compromising the optimality of the reset map.

Suppose, as in Section 9.2, that at a time $t = t_0$ the switching signal jumps from $\sigma(t_0^-) = p$ to $\sigma(t_0) = q$. We will show that stability under arbitrary switching can be achieved if we require that, at the switching time t_0, there should be no increase in the distance between x_{cn} and its steady-state value obtained from (9.13). Formally, this can be expressed as follows

$$\|x_{\mathrm{cn}}(t_0) - K_q r(t_0)\|^2 \leq \|x_{\mathrm{cn}}(t_0^-) - K_q r(t_0)\|^2, \tag{9.22}$$

where $K_q := -\begin{bmatrix} 0_{n_{\mathrm{cn}}\times n_{\mathrm{pl}}} & I_{n_{\mathrm{cn}}\times n_{\mathrm{cn}}} & 0_{n_{\mathrm{cn}}\times n_{\mathrm{pl}}} \end{bmatrix} \hat{A}_q^{-1}\hat{B}_q$. Often this constraint will be satisfied by choosing the global minimum to (9.12) with norm closest to $K_q r(t_0)$, which is given by

$$\begin{aligned} x_{\mathrm{cn}}^*(t_0) = V_1^q \Lambda_q^{-1} (U_1^q)^T \cdot \\ \cdot \Big(\frac{1}{2} g_2^q - (M_{12}^q)^T x_{\mathrm{pl}}(t^-) - M_{23}^q x_{\mathrm{copy}}(t^-) \Big) + \\ + V_2^q (V_2^q)^T K_q r(t_0). \end{aligned} \tag{9.23}$$

When compared with (9.15), (9.23) includes the term $V_2^q (V_2^q)^T K_q r(t_0)$, which is the projection of $K_q r(t_0)$ into the kernel of M_{22}^q. Therefore (9.23) still satisfies the first-order condition (9.20).

When (9.23) does not satisfy (9.22), we will need to find the minimum of (9.12) subject to the quadratic inequality (9.22). In view of (9.19), this is a convex quadratic optimization, subject to a convex quadratic constraint, which can be solved numerically very efficiently. As opposed to when (9.23) already satisfies (9.22), in this case the constraint (9.22) may lead to some increase in the value of the criteria (9.12).

9.4 Stability

The switched system (9.9)–(9.10) is said to be *uniformly input-to-state stable (ISS) for a smooth input* $r(t)$ if there exist constants $\lambda, c_0, c_1, c_2 > 0$ such that, for every differentiable $r(t)$ and every piecewise constant switching signal $\sigma(t)$, the following inequality holds along solutions to (9.9)–(9.10):

$$\|x(t)\| \le c_0 e^{-\lambda t}\|x(0)\| + c_1 \sup_{\tau\in[0,t)} \|r(\tau)\| + c_2 \sup_{\tau\in[0,t)} \|\dot{r}(\tau)\|. \tag{9.24}$$

The following result confirms that the constraint (9.22) does guarantee closed-loop stability.

Theorem 9.2 *The switched system* (9.9)–(9.10) *is uniformly ISS for a smooth input* $r(t)$ *if the reset map satisfies the following inequality for every* $p, q \in \mathcal{P}$, *every* $x_{\text{mult}} := [x_{\text{cn}}^T\ x_{\text{copy}}^T]^T$, *and every* r_0:

$$\|F(x_{\text{mult}}, p, q, r_0) - K_q r_0\|^2 \le \|x_{\text{cn}} - K_q r_0\|^2. \tag{9.25}$$

Proof: 9.2 Consider the signal $e(t)$ defined by

$$e(t) = r(t) - y(t) + Cx_{\text{copy}}(t), \quad \forall t \ge 0. \tag{9.26}$$

Because of (9.8) and the fact that the process is asymptotically stable, we conclude that $x_{\text{copy}}(t)$ converges exponentially fast to the process state $x_{\text{pl}}(t)$. Consequently $Cx_{\text{copy}}(t)$ converges to the process output $y(t)$ and we have that

$$\|e(t) - r(t)\| \le d e^{-\gamma t}\|x(0)\|, \tag{9.27}$$

for appropriate positive constants $d, \gamma > 0$. One should emphasize that (9.27) holds regardless of the control input $u(t)$, the switching signal $\sigma(t)$, and any resets to $x_{\text{cn}}(t)$.

The importance of the signal $e(t)$ defined in (9.26) stems from the fact that, along solutions to (9.9)–(9.10), the state $x_{\text{cn}}(t)$ evolves according to

$$\dot{x}_{\text{cn}}(t) = A_{\sigma(t)} x_{\text{cn}}(t) + B_{\sigma(t)} e(t)$$

on any time interval on which $\sigma(t)$ remains constant and

$$x_{\text{cn}}(t) = F(x_{\text{mult}}(t^-), \sigma(t^-), \sigma(t), r(t)),$$

at every switching time t. Suppose now that we define

$$v(t) := \|x_{\text{cn}}(t) - K_{\sigma(t)} r(t)\|^2.$$

On any interval on which $\sigma(t)$ remains constant and equal to some $p \in \mathcal{P}$, we have that

$$\begin{aligned}\dot{v} &= x_{\mathrm{cn}}(t)^T(A_p + A_p^T)x_{\mathrm{cn}}(t)\\ &\quad + 2x_{\mathrm{cn}}(t)^T(B_p e(t) - K_p\dot{r}(t) - A_p^T K_p r(t))\\ &\quad - 2r(t)^T K_p^T(B_p e(t) - K_p\dot{r}(t)).\end{aligned}$$

Because of (9.4) and using fairly standard square completion arguments, it is possible to find a sufficiently small constant μ and sufficiently large constants d_1, d_2, d_3 (independent of the value of p) such that

$$\dot{v}(t) \leq -\mu v(t) + d_1\|e(t)\|^2 + d_2\|r(t)\|^2 + d_3\|\dot{r}(t)\|^2.$$

Since $v(t)$ does not increase at switching times because of (9.25), we conclude that

$$\begin{aligned}v(t) \leq e^{-\mu t}v(0) + d_1\int_0^t e^{-\mu(t-\tau)}\|e(\tau)\|^2 d\tau\\ + \frac{d_2}{\mu}\sup_{\tau\in[0,\tau)}\|r(\tau)\|^2 + \frac{d_3}{\mu}\sup_{\tau\in[0,\tau)}\|\dot{r}(\tau)\|^2.\end{aligned}$$

From this and (9.27), we conclude that $\|x_{\mathrm{cn}}(t)\|$ satisfies an inequality like (9.24). It remains to show that the remaining components $x_{\mathrm{pl}}(t)$ and $x_{\mathrm{copy}}(t)$ of the overall state $x(t)$ also satisfy such inequalities. To this effect note that the input $u(t)$ to the process and to the system (9.8) can be written as

$$u(t) = C_{\sigma(t)}x_{\mathrm{cn}}(t) + D_{\sigma(t)}e(t)$$

[cf. (9.1), (9.5)–(9.6), and (9.26)] and therefore $u(t)$ also satisfies an inequality like (9.24). Finally, since $x_{\mathrm{pl}}(t)$ and $x_{\mathrm{copy}}(t)$ are the states of asymptotically stable LTI systems driven by $u(t)$, these states also satisfy an inequality like (9.24). ∎

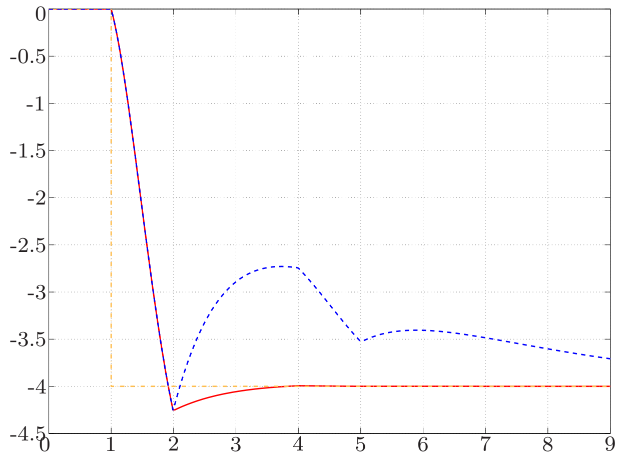
9.5 Simulation results

In this section we compare the transients due to switching for the multicontrollers proposed here, in [50], and in [25].

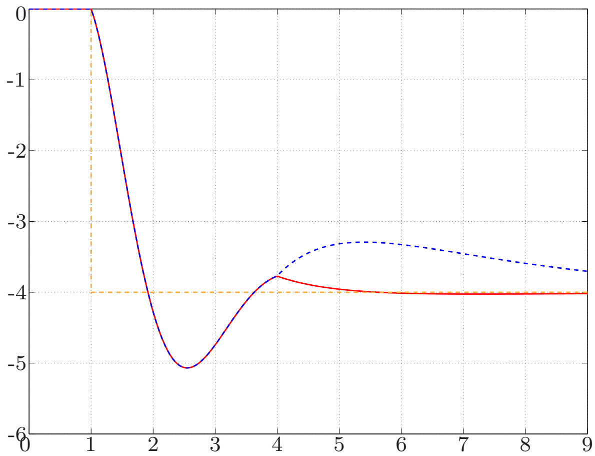
Consider a process with transfer function $g(s) = \frac{10}{s+1}$ and two controllers with transfer functions given by (9.3) with

$$q_1(s) = \frac{2s^2 + 5s + 3}{100s^2 + 120s + 30}, \qquad q_2(s) = \frac{6s^2 + 46s + 40}{100s^2 + 160s + 400}$$

among which one would like to switch. This process and controller transfer functions appeared in [50]. The subsequent figures show the output and reference signals of the resulting closed-loop switched system. In Figure 9.4 the blue dotted lines show the output $y(t)$ of the system for the controller switching strategy suggested in [50], whereas the red solid lines show $y(t)$ for the multicontroller proposed here. Figure 9.5 compares the multicontroller



(a) Switching times: 2, 4, and 5 sec



(b) Switching time: 4 sec

Figure 9.4: Transient responses for the multicontroller proposed here (red solid line) and for the multicontroller proposed in [50] (blue dotted line).

proposed here with the two solutions proposed in [25], which correspond to (i) a reset to zero of $x_{\mathrm{cn}}(t)$ at every switching and (ii) maintaining $x_{\mathrm{cn}}(t)$ continuous at every switching instant. In all simulations the use of the multicontroller proposed here resulted in a significant performance improvement.

It is worth to notice that in all the simulations presented here, the global optimal reset values given by (9.23) automatically satisfied the constraint (9.22). In fact, this was the case in every simulation that we encountered.

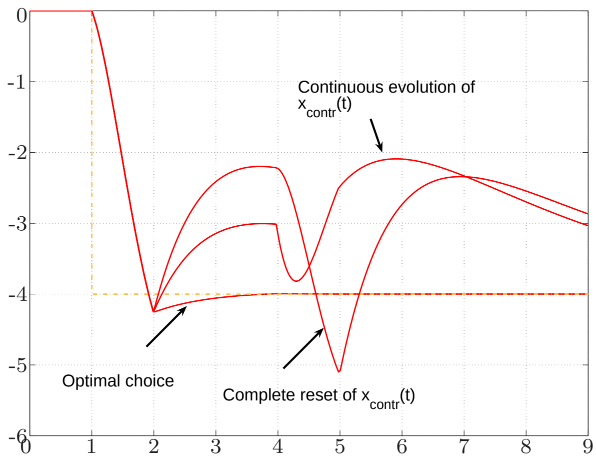


Figure 9.5: Transient responses for the multicontroller proposed here and for the two alternative multicontrollers proposed in [25]. The plot shows the transients due to three control switchings at times 2, 4, and 5 sec.

Chapter 10

Optimal Reset Map: an extension to the non-switched case

The tools developed in Chapter 9 find a natural application also in the context of non-switched system. Indeed, even when dealing with the control of LTI plant, multiple conflicting requirements on the closed-loop system often make a single LTI controller unsuitable [7]-[36]. Several nonlinear approaches have been suggested therefore in the literature in order to overcome the limitation of these simple controllers; among these, an important class to be considered is that of the hybrid controllers (see for instance [1]). A particularly simple yet effective hybrid approach to this problem is known as the *reset control* strategy. The basic structure of a reset control system can be described as a feedback interconnection of the LTI plant Σ to be controlled together with an hybrid controller $\mathbf{C}$ evolving according to the linear dynamics of a selected LTI controller, yet whose state undergoes a reset whenever its output and input satisfy certain conditions. The time instants at which the controller state resets occur are called *reset times*. First introduced in [12] (Clegg integrator), and subsequently developed by various authors (see for instance [27]-[32]), only in more recent years the stability issues for reset controllers has been addressed (see [40]-[41]).

The approach here suggested extends the original ideas of the reset control theory by introducing a function, called *reset map*, that defines at every reset time how the controller state vector must be updated, depending on its previous value and on the measure of the reference signal. According to this, not only resets to zero of the state variable are allowed, but a much wider range of value is now feasible. It is worthwhile to observe that, the asymptotic performance of the closed loop being utterly determined by the choice of the LTI controller (once the system Σ is given), only the transient behaviour of the controlled system may be conceivably improved.

In this chapter, the problem of suitably choosing the reset map function for a given plant-controller pair is addressed. An integral cost function is first defined whose goal is to capture, in an analytical sense, the concept of "good transient behaviour". Then, the choice of the reset map is formulated as an optimization problem with respect to the aforementioned cost function. A procedure leading to a convenient selection of the reset times is then also presented in the same form of a minimization problem, again by referring to the same cost function. Together with the analytical tools required for an efficient numerical solution of this latter problem, a sub-optimal strategy is also provided, greatly reducing the computational effort involved in the optimization procedure. For an overview on the existing results on optimization problems in the context of hybrid systems, the reader is referred to [57, 2, 39, 58] and to references therein.

Simulation results are presented, clearly illustrating the advantages of the proposed solution if compared both to the classic LTI approach and to the standard reset control theory.

10.1 Problem

Consider a linear lime invariant process Σ, with transfer function $g(s)$ from the input $u(t)$ to the output $y(t)$, and let $\mathbf{C}$ be a stabilizing LTI controller of transfer function $k(s)$ from the tracking error $e_T(t) := r(t) - y(t)$ to the control input $u(t)$, where $r(t)$ denotes a piecewise constant reference signal. The controlled system $\mathbf{F}$ arising from the negative feedback interconnection of Σ with $\mathbf{C}$ (see fig. 10.1) can be obviously still described by means of an LTI model. In this section it will be carefully specified how, starting from the equations of the processes Σ and $\mathbf{C}$, a new system of hybrid nature can be constructed, significantly improving the performance of the simple feedback interconnection characterizing $\mathbf{F}$.

For simplicity, we restrict our attention to asymptotically stable SISO processes Σ. However, all the results presented here could be generalized to not necessarily stable MIMO processes [1].

[1] One can resort, for instance, to an approach close to that presented in [25]

From the Youla-Kucera parametrization of all the stabilizing controllers (see for instance [15]), it is well known that the controller transfer function can be expressed as

$$k(s) = \frac{q(s)}{1 - q(s)g(s)}; \tag{10.1}$$

which can be viewed as a *positive* feedback interconnection between a system with the same transfer function $g(s)$ as the process (called *internal model* of the process) and an asymptotically stable system with transfer function

$$q(s) = \frac{k(s)}{1 + g(s)k(s)}. \tag{10.2}$$

Note that $q(s)$ is asymptotically stable because $k(s)$ stabilizes $g(s)$ (see Fig 10.2). Pick now some minimal realizations (A, B, C) for $g(s)$ and (A_q, B_q, C_q, D_q) for $q(s)$. Note that the realization of the transfer function $g(s)$ is supposed to be the same both in the real plant Σ and in its internal copy within the controller $\mathbf{C}$. Let $A \in \mathbb{R}^{n_{\mathrm{pl}} \times n_{\mathrm{pl}}}$ and $A_q \in \mathbb{R}^{n_{\mathrm{cn}} \times n_{\mathrm{cn}}}$; let also $x_{\mathrm{pl}}(t)$, $x_{\mathrm{cn}}(t)$ and $x_{\mathrm{copy}}(t)$ represent the state variables respectively of the plant Σ, of the process of transfer function $q(s)$ and of the internal model of the plant. Once we set $x_{\mathbf{C}} := \begin{bmatrix} x_{\mathrm{cn}}(t) \\ x_{\mathrm{copy}}(t) \end{bmatrix}$, the controller $\mathbf{C}$ can then be described by the following system of equations:

$$\dot{x}_{\mathbf{C}}(t) = \begin{bmatrix} A_q & B_q C \\ B C_q & A + B D_q C \end{bmatrix} x_{\mathbf{C}}(t) + \begin{bmatrix} B_q \\ B D_q \end{bmatrix} e_T(t); \tag{10.3}$$

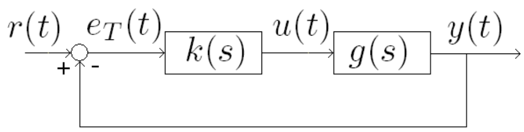


Figure 10.1: Controlled system **F** layout

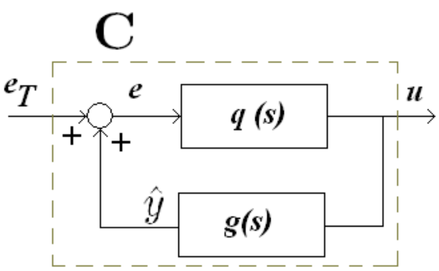


Figure 10.2: Implementation of the controller **C**

$$\text{with} \quad u(t) = \begin{bmatrix} C_q & D_q C \end{bmatrix} x_{\mathbf{C}}(t) + D_q e_T(t). \tag{10.4}$$

Introduce an ordered (possibly infinite) set $\mathcal{S} := \{t_1, t_2, \ldots, t_n\}$ of positive time instants, called *reset times*. We are now ready to define a new controller $\mathbf{C}_{reset}$, whose dynamics is described by eqs. (10.3)-(10.4) for every time instant $t \notin \mathcal{S}$, and whose state is supposed to be updated at every reset time $t \in \mathcal{S}$ according to:

$$x_{\mathbf{C}}(t) = \begin{bmatrix} x_{\text{cn}}(t) \\ x_{\text{copy}}(t) \end{bmatrix} = \begin{bmatrix} F(x_{\mathbf{C}}(t^-), r(t)) \\ x_{\text{copy}}(t^-) \end{bmatrix}. \tag{10.5}$$

The function $F(\cdot, \cdot)$ is called the *reset map* of our system. It is worth to observe that only a part of the state of the controller actually undergoes the action of the reset map; this is because $x_{\text{copy}}(t)$ is supposed to track at every instant the value of $x_{\text{pl}}(t)$, which is a continuous function of time. In (10.5), the component x_{copy} of the state of $\mathbf{C}(\sigma)$ remains continuous and it will actually converge to the process state $x_{\text{pl}}(t)$. In fact, it follows from (10.3) and (10.5) that

$$\dot{x}_{\text{copy}}(t) = A x_{\text{copy}}(t) + B u(t) \tag{10.6}$$

for all times and, because we are assuming that A is asymptotically stable, we indeed have that $x_{\text{copy}}(t)$ converges to $x_{\text{pl}}(t)$, regardless of the control signal $u(t)$. Clearly the process $\mathbf{C}_{reset}$ exhibits an hybrid nature, and we will refer to it in the remainder as to the *linear-based reset controller* for our plant Σ. The stabilizing properties of such a controller will be investigated in section 10.4. By connecting the controller $\mathbf{C}_{reset}$ to the plant Σ according to the usual negative feedback layout of Fig. 10.1, we get another hybrid system which will be denoted as the *reset control system*.

Finally, upon setting $x(t) := \begin{bmatrix} x_{\text{pl}}(t)^T \, x_{\text{cn}}(t)^T \, x_{\text{copy}}(t)^T \end{bmatrix}^T$, the reset control system can be described by the following state space realization:

$$\begin{aligned} \dot{x}(t) &= \begin{bmatrix} A - BD_qC & BC_q & BD_qC \\ -B_qC & A_q & B_qC \\ -BD_qC & BC_q & A{+}BD_qC \end{bmatrix} x(t) + \begin{bmatrix} BD_q \\ B_q \\ BD_q \end{bmatrix} r(t) \\ &=: \hat{A}x(t) + \hat{B}r(t); \end{aligned} \tag{10.7}$$

together with

$$y(t) = \begin{bmatrix} C & 0 & 0 \end{bmatrix} x(t) =: \hat{C}x(t). \tag{10.8}$$

at every time instant $t \notin \mathcal{S}$, while at every reset time $t \in \mathcal{S}$:

$$x(t) = \begin{bmatrix} x_{\text{pl}}(t) \\ x_{\text{cn}}(t) \\ x_{\text{copy}}(t) \end{bmatrix} = \begin{bmatrix} x_{\text{pl}}(t^-) \\ F(x_{\text{C}}(t^-), r(t)) \\ x_{\text{copy}}(t^-) \end{bmatrix}. \tag{10.9}$$

In order to precisely define our goal, we introduce the following:

Definition 10.1 *The system described by (10.7)-(10.8)-(10.9) is said to be* asymptotically stable *if, for every bounded piecewise constant reference signal $r(t)$, the state variable $x(t)$ remains bounded for every possible choice of the set $\mathcal{S}$. We also require the state $x(t)$ to decay to zero when $r(t) \equiv 0$, for all the admissible $\mathcal{S}$.*

The problem here addressed is that of suggesting a convenient choice for the reset map $F(\cdot, \cdot)$ and for the reset times $\mathcal{S}$ in order to:

(i) Guarantee the asymptotic stability of (10.7)-(10.8)-(10.9).

(ii) Achieve an optimal transient behaviour with respect to a criteria to be specified shortly.

10.2 Optimal definition of the reset map

Observe that, referring to (10.7)-(10.8)-(10.9), the following vector:

$$x^{\infty} := -\hat{A}^{-1}\hat{B}\, r \tag{10.10}$$

represents the steady-state value to which $x(t)$ would converge as $t \to +\infty$ if both $r(t)$ was to remain constant (and equal to r) and no resets occur.

Indeed, because of the stability of the closed loop system, asymptotically $\dot{x}(t) \to 0$ and since $\dot{x} = \hat{A}x + \hat{B}r$, the result easily follows. Note that $\hat{A}$ is Hurwitz, hence invertible. In the remainder of the chapter, we will refer to (10.10) as to the *steady state* value of the state vector. Consider the vector x^{∞} to be partitioned accordingly to the partition already introduced for $x(t)$. Acordingly, the steady state value of the input variable $u(t)$ is equal to

$$u^{\infty} := C_q x_{\mathrm{cn}}^{\infty} + D_q r. \tag{10.11}$$

Choose now $t_0 \in \mathcal{S}$ and let $t_1 > t_0$ be given; assume, then, that $t \notin \mathcal{S}$ and $r(t) = r$ for every $t \in]t_0, t_1[$.

We define now a cost function $J = J\big(F(\cdot, \cdot)\big)$ to be minimized. Hence, with the usual meaning of the symbols, introduce

$$J = \int_{t_0}^{t_1} e_T(t)^T R e_T(t) + \dot{y}(t)^T W \dot{y}(t) + (u(t) - u^{\infty})^T K (u(t) - u^{\infty}) dt, \tag{10.12}$$

where R, W, K, T are symmetric positive definite matrices and all the signals involved are subject to (10.7)-(10.8)-(10.9). The choice of the matrices R, W, K, allows to individually penalize the contributions due respectively to the tracking error magnitude, to the output signal oscillations and to the control effort. Note that in the interval $]t_0, t_1[$ the reset control system behaves actually as a regular LTI system described by eq. (10.7)-(10.8). From now on, the explicit indication of the time-dependency of the signals involved in the definition of J will be omitted.

Our goal hence can be conveniently formalized by means of the following:

Problem 10.2 *Select the post-reset state defined by the reset map:*

$$x_{\mathrm{cn}}(t_0) = F(x_{\mathrm{C}}(t_0^-), r(t_0)) \tag{10.13}$$

so as to minimize the cost function J in (10.12).

In order to state the next result, we need to introduce some notation. Let Q be the solution of the Lyapunov equation

$$Q\hat{A} + \hat{A}^T Q = -P, \tag{10.14}$$

where $P = \hat{C}^T R\,\hat{C} + \hat{A}^T\hat{C}^T W\,\hat{C}\hat{A} + \tilde{C}^T K\,\tilde{C} = P^T \geq 0$ and $\tilde{C} := [-D_q C\ C_q\ D_q C]$. Observe that such solution always exists and it is at least positive semi-definite because $\hat{A}$ is a Hurwitz matrix.

Introduce now the symmetric positive semi-definite matrix

$$G = R + \tilde{D}^T K \tilde{D} + \hat{B}^T \hat{C}^T W \hat{C} \hat{B}$$

where

$$\tilde{D} = C_q[0_{n_{\mathrm{cn}}\times n_{\mathrm{pl}}}\ I\ 0_{n_{\mathrm{cn}}\times n_{\mathrm{pl}}}]\hat{A}^{-1}\hat{B},$$

and the following vector

$$g = \left(2r^T(-R\,\hat{C} + \hat{B}^T\hat{C}^T W\,\hat{C}\hat{A} + \tilde{D}^T K\,\tilde{C} + \hat{B}^T Q)\hat{A}^{-1}\right)^T. \tag{10.15}$$

Set $\Delta = t_1 - t_0$. Easy but tedious computations lead to the following:

$$\begin{aligned} J = {} & \Delta \cdot r^T G r + g^T\Big((e^{\hat{A}\Delta} - I)x(t_0) + \big(\hat{A}^{-1}(e^{\hat{A}\Delta} - I) - \Delta I\big)\hat{B}r\Big) \\ & - x(\Delta, x(t_0))^T Q x(\Delta, x(t_0)) + x(t_0)^T Q x(t_0) \end{aligned} \tag{10.16}$$

where $x(\Delta, x(t_0))$ represents the state variable computed according to eqs. (10.7)-(10.8) at time $t = \Delta$, starting from the initial condition $x(0) = x(t_0)$, namely

$$x(\Delta, x(t_0)) = e^{\hat{A}\Delta} x(t_0) + \int_0^{\Delta} e^{\hat{A}(\Delta-\tau)} \hat{B} \cdot r \, d\tau. \tag{10.17}$$

In what follows, $*$ stands for additive terms that do not depend on the value of $x_{\mathrm{cn}}(t_0)$. Notice that the choice of $F(\cdot,\cdot)$ only affects the value of $x_{\mathrm{cn}}(t_0)$.

$$\begin{aligned} J &= g^T(e^{\hat{A}\Delta} - I)x(t_0) - x(\Delta, x(t_0))^T Q x(\Delta, x(t_0)) + x(t_0)^T Q x(t_0) + * \\ &= \tilde{g}^T x(t_0) + x(t_0)^T M x(t_0) + * \end{aligned} \tag{10.18}$$

where

$$\tilde{g} := \left(g^T(e^{\hat{A}\Delta} - I) - 2r^T \hat{B}^T (e^{\hat{A}^T\Delta} - I)\hat{A}^{-T} Q e^{\hat{A}\Delta}\right)^T$$

and

$$M = Q - e^{\hat{A}^T\Delta} Q e^{\hat{A}\Delta}.$$

We will further need to block-partition the symmetric matrices M and the vector $\tilde{g}$ according to the partition in (10.7) of the state vector:

$$M = \begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{bmatrix}, \qquad \tilde{g} = \begin{bmatrix} \tilde{g}_1 \\ \tilde{g}_2 \\ \tilde{g}_3 \end{bmatrix}$$

and perform a singular value decomposition of

$$M_{22} = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} (V_1)^T \\ (V_2)^T \end{bmatrix}$$

(with Λ diagonal and nonsingular). Introduce the pseudo inverse of M_{22}, namely $M_{22}^{\dagger} = V_1 \Lambda^{-1} U_1^T$. We claim now the following.

Theorem 10.3 *Assuming that $r(t) = r(t_0)$, $\forall t \in [t_0, t_1]$, the global minimum to* (10.12) *with smallest norm is given by*

$$x_{\mathrm{cn}}^*(t_0) = M_{22}^{\dagger}\left(\frac{\tilde{g}_2}{2} - [M_{12}^T \; 0 \; M_{23}] \, x(t_0^-)\right). \tag{10.19}$$

Proof: From eq. (10.18), by eliminating all the terms not depending on $x_{\rm cn}(t_0)$ we get

$$J = x_{\rm cn}(t_0)^T M_{22} x_{\rm cn}(t_0) + \\ + [2(x_{\rm pl}(t_0)^T M_{12} + x_{\rm copy}(t_0)^T M_{23}^T) - \tilde{g}_2^T] x_{\rm cn}(t_0) + *. \quad (10.20)$$

Observe that, since $M_{22} = M_{22}^T \geq 0$, the function in (10.20) turns out to be convex; as a consequence, any vector $x_{\rm cn}^\star(t_0)$ satisfying the first order necessary condition

$$2M_{22}x_{\rm cn}^\star(t_0) = g_2 - 2\left(M_{12}^T x_{\rm pl}(t_0) + M_{23}x_{\rm copy}(t_0)\right) \quad (10.21)$$

represents a global minimum for J (see, for instance, [8]). In general, of course, eq. (10.21) may not be solvable, but in our specific context it can be proved that it always is (see Lemma A.26 in the Appendix), and the minimum norm solution is given by (10.19). ■

10.2.1 Choice of the reset map

Since the optimal value for $x_{\rm cn}^*(t_0)$ in (10.19) depends on the process state $x_{\rm pl}(t_0)$, we cannot directly use the expression in (10.19) to define the optimal reset map in (10.13). However, as mentioned in Section 10.1, the component $x_{\rm copy}(t)$ of the multicontroller state converges exponentially fast to the process state $x_{\rm pl}(t)$, for every control input $u(t)$. If we then replace $x_{\rm pl}(t_0)$ by $x_{\rm copy}(t_0)$ in (10.19), we obtain an "asymptotically correct" minimum to (10.12), which justifies the following reset map

$$F(x_{\rm C}(t_0^-), r(t_0)) := M_{22}^\dagger \cdot \left(\frac{\tilde{g}_2}{2} - (M_{12}^T + M_{23})x_{\rm copy}(t_0^-)\right). \quad (10.22)$$

Since $F(\cdot)$ is a function of the "past" value of $x_{\rm C}$, the right-hand side of (10.22) must only depend on the "past" value of $x_{\rm copy}$. However, this is not a problem because $x_{\rm copy}(t_0) = x_{\rm copy}(t_0^-)$, in view of (10.5). Note that the dependence on $r(t_0)$ is actually hidden in the construction of the matrices M and $\tilde{g}$.

10.2.2 Infinite time horizon optimization

Observe that, as a particular case, we have that, as t_1 approaches $+\infty$, $M = Q$ and $\tilde{g} = -g$. Up to this semplification, all the aforementioned results remain of course still valid.

Figures 10.3 and 10.4 depict the results of numerical simulations, illustrating how varying the length of the optimization interval may influence the system's behavior. The dotted line represents the reference signal $r(t)$, while the continuous lines plots the output variable $y(t)$ corresponding to various choices for t_1. The captions specify under what kind of choice for the parameters R, W, K, T the simulations have been runned. It generally happens that the transient response improves as we increase the optimization interval.

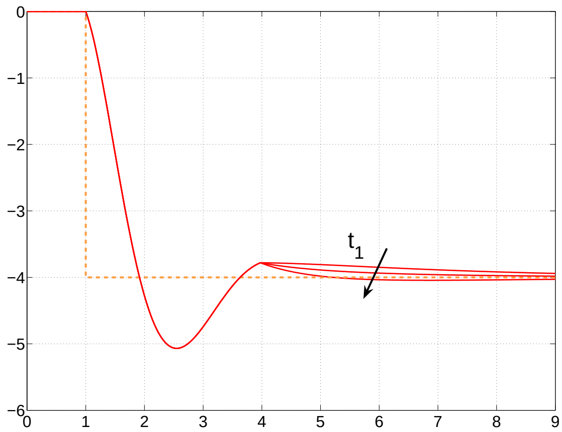


Figure 10.3: Using $R = I$, $W \neq 0$, $K, T = 0$ and different values for t_1

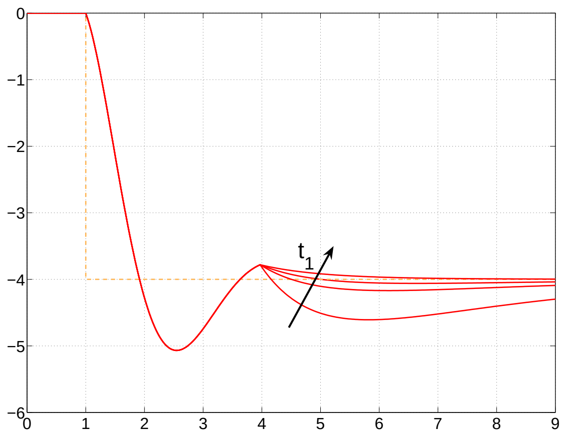


Figure 10.4: Using $R = I$, $W, K, T = 0$ and different values for t_1

10.2.3 Non-minimum norm solution

As we previously mentioned, there might exist more than one optimal solution to Problem 10.2, depending on whether the matrix M_{22} is strictly positive definite or simply positive semi-definite. In the case when the uniqueness of the solution is not guaranteed, it is more reasonable to consider, instead of the minimum norm solution (10.19), the one closest (for instance in the sense of the square norm of the distance) to the asymptotic vector $x_{\text{cn}}^{\infty}(t_0)$. It can be easily verified that, with the usual meaning of the symbols, this latter is given by

$$F(x_{\text{C}}(t_0^-), r(t_0)) := \\ = M_{22}^{\dagger}\Big(\frac{\tilde{g}_2}{2} - (M_{12}^T + M_{23})x_{\text{copy}}(t_0^-)\Big) + V_2 V_2^T x_{\text{cn}}^{\infty}(t_0), \quad (10.23)$$

where the matrix $V_2 V_2^T$ is the projection matrix of $\mathbb{R}^{n_{\text{cn}}}$ on $\text{Ker}(M_{22})$.

10.3 Choice of the reset times

10.3.1 Optimal choice

Suppose that a finite reset time set $\mathcal{S} = \{t_1, t_2, \ldots, t_\ell\}$ is given, together with the initial condition $x(0) = x_0$ and a constant value for the reference signal $r(t) \equiv r$. It is therefore possible to explicitly compute the value of the cost function J in eq. (10.12) when the optimal reset map provided in Theorem 10.3 is used at every reset time.

Let $x(\tau, \tilde{x})$ be defined as in eq. (10.17) and let $x^{\text{hyb}}(\tau)$ be the value of the state variable of the hybrid system (10.7)-(10.8)-(10.9) computed at time $t = \tau$ starting from the initial condition $x(0) = x_0$ when the optimal reset map of Theorem 10.3 is used at every reset time in $\mathcal{S}$.

If follows that:

$$J = \sum_{i=1}^{\ell+1} \Big(\Delta_i r^T G r + g^T \Big((e^{\hat{A}\Delta_i} - I)x^{\text{hyb}}(t_{i-1}) + \big(\hat{A}^{-1}(e^{\hat{A}\Delta_i} - I) - \Delta_i \cdot I\big)\hat{B}r\Big) - \\ - x(\Delta_i, x^{\text{hyb}}(t_{i-1}))^T Q x(\Delta_i, x^{\text{hyb}}(t_{i-1})) + x^{\text{hyb}}(t_{i-1})^T Q x^{\text{hyb}}(t_{i-1})\Big) \quad (10.24)$$

where $t_0 := 0$, $t_{\ell+1} := \infty$ and $\Delta_i = t_i - t_{i-1}$.

But then, it is possible to choose the reset times $t_1, \ldots, t_\ell$ so as to minimize the cost J. A numerical approach is clearly required for this optimization problem. The Newton descent method may be useful for this purpose, and an explicit expression for the partial derivative of the cost function J with respect to any reset time t_i, $i = 1, \ldots, \ell$ can be easily computed.

Indeed the following hold:

$$\frac{\partial x^{\mathrm{hyb}}(t_i)}{\partial t_j} = \begin{cases} 0, & j > i \\ A^* \hat{A}\, e^{\hat{A}\Delta_i}(x^{\mathrm{hyb}}(t_{i-1}) - x^\infty), & j = i \\ (\prod_{k=i}^{j+1} A^* e^{\hat{A}\Delta_k})(\hat{A}A^* + A^*\hat{A})\cdot & \\ \quad \cdot(x^{\mathrm{hyb}}(t_{j-1}) - x^\infty), & j < i \end{cases} \tag{10.25}$$

together with

$$\frac{\partial x(\Delta_i,\, x^{\mathrm{hyb}}(t_i))}{\partial t_j} = \begin{cases} 0, & j > i \\ \hat{A}\, e^{\hat{A}\Delta_i}(x^{\mathrm{hyb}}(t_{i-1}) - x^\infty), & j = i \\ -\frac{\partial x(\Delta_i, x^{\mathrm{hyb}}(t_i))}{\partial t_i} + & \\ +e^{\hat{A}\Delta_i}\frac{\partial x^{\mathrm{hyb}}(t_{i-1})}{\partial t_j}, & j = i-1 \\ e^{\hat{A}\Delta_i}\frac{\partial x^{\mathrm{hyb}}(t_{i-1})}{\partial t_j}, & j < i-1. \end{cases} \tag{10.26}$$

where

$$A^* = \begin{bmatrix} I & 0 & 0 \\ -M_{22}^\dagger M_{12}^T & 0 & -M_{22}^\dagger M_{23} \\ 0 & 0 & I \end{bmatrix}.$$

10.3.2 Suboptimal strategy

However, the solution of a multi-variable optimization problem might become quite demanding from a computational point of view; besides, an a-priori knowledge of the number ℓ of reset times required in order to obtain the desired behavior may be an unrealistic assumption. Another feasible strategy then is to consider, instead of a single multi-variable optimization process, the suboptimal multi-step single-variable procedure described below.

The first reset time t_1 can be chosen by solving the minimization problem of the function in eq. (10.24) with $\ell = 1$. Then, any other reset time t_i, $i = 2, 3, \ldots$ can be computed recursively by again considering the one dimensional minimization of the same function where the initial condition x_0 is chosen to be the value of the hybrid state $x(t)$ evaluated at $t = t_{i-1}$.

Even if, in general, this approach prevents us from actually minimizing the value of the cost function J, given a certain number of allowed reset times ℓ, nonetheless our simulations show that the performances achieved with the multi-step method do not significantly differ from the optimal ones. See section 10.5 for some examples.

10.4 Stability

Consider the control scheme depicted in Fig. 10.1 and Fig. 10.2. What we aim to prove here is that the approach suggested at the end of section 10.3, namely the multi-step single-variable optimization strategy, allows the reset control sysem to be asymptotically stable. According to Definition 10.1, there are two different conditions that need to be verified.

Suppose first $r(t) = r \neq 0$ and note that once r is assigned, then the value of u^∞ is also uniquely determined. Since already with the original non-resetting LTI controller the cost function J turns out to exhibit a finite value (when $t_0 = 0$ and $t_1 = +\infty$), it follows that also with the multi-step strategy of section 10.3 the same property still holds. But this implies $e_T(t)$, $u(t) - u^\infty \in \mathcal{L}_2$[2]. Observe that $y(t)$ is a continuous function. It should be also noted that even if $u(t)$ may fail to be continuous at the reset times, still it is at least a piecewise continuous signal; moreover, the expression of $u(t)$ betwen two consecutive reset times only involves linear combinations of elementary modes of the kind $t^k e^{\lambda t}$, for some $k \in \mathbb{N}$ and $\lambda \in \mathbb{C}$, and also, as the structure of eq.(10.22) clearly reveals, to every finite pre-reset state x_{copy} corresponds a finite post-reset state x_{cn}. We can then claim that finite escape time phenomena are prohibited for $u(t)$; equivalently, $u(t)$ is a bounded signal in every finite interval $[\tau_0, \tau_1] \subset [0, +\infty[$. Now, being u^∞ and $r(t)$ constant, from $e_T(t)$, $u(t) - u^\infty \in \mathcal{L}_2$ it follows that $u(t)$ and $y(t)$ are bounded signals on $[0, +\infty[$. Hence, being $g(s)$ a stable transfer function and (A, B, C) a minimal realization of it, $\hat{y}(t)$ is bounded too, as well as the signals $x_{\text{pl}}(t)$ and $x_{\text{copy}}(t)$. As a further consequence, also $e(t)$ is a bounded signal. It is eventually possible to prove (see Proposition 10.4) that, under this assumption, when the multi-step strategy of section 10.3 is employed also the state $x_{\text{cn}}(t)$ of the hybrid reset controller remains bounded, thus ensuring the first property required by Definition 10.1.

[2]A causal signal $y(t)$ belongs to the vector space $\mathcal{L}_2$ if $(\int_0^{+\infty} y^2(t)dt)^{\frac{1}{2}} < +\infty$

Proposition 10.4 *Let $\mathcal{S} = \{t_0, t_1, \dots\}$ be a discrete (possibly infinite) ordered set of time instants such that there exists $\delta > 0$ having the property that for every $i \in \{1, 2, \dots\}$, $t_i - t_{i-1} > \delta$. Given an hybrid system described by the following equations*

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) & t \notin \mathcal{S} \\ x(t) = Rx(t^-) & t \in \mathcal{S} \\ y(t) = Cx(t) \end{cases} \tag{10.27}$$

where A is a Hurwitz matrix and the pair (A, C) is observable, then

(i) $y(t), u(t)$ bounded imply $x(t)$ bounded.

(ii) $y(t), u(t) \in \mathcal{L}_2$ imply $x(t)$ converging to the zero vector.

Proof: This result is essentially a consequence of the Squashing Lemma in [38]. Note first that system (10.27) is equivalent to the following one

$$\begin{cases} \dot{x}(t) = (A - LC)x(t) + Bu(t) + Ly(t), & t \notin \mathcal{S} \\ x(t) = Rx(t^-), & t \in \mathcal{S} \end{cases} \tag{10.28}$$

where L is a matrix of appropriate dimensions and $y(t)$ is now interpreted as a second input. Because of the observability of the pair (C, A), the Squashing Lemma in [38] guarantees that, for every choice of $\lambda > 0$, it is always possible to find an output-injection matrix L such that

$$\|e^{(A-LC)t}\| \le e^{-\lambda t}. \tag{10.29}$$

Consider first the case when $\|R\| \le 1$. It is then sufficient to consider any matrix L making $A - LC$ Hurwitz in order to verify the correctness of the statement.

Assume now $\|R\| > 1$. Choose the matrix L such that eq. (10.29) is satisfied with $\lambda = \frac{ln\|R\|}{\delta}$. Then it follows that

$$\|e^{(A-LC)t}\| \le \frac{1}{\|R\|^{\frac{t}{\delta}}} \le \frac{1}{\|R\|^{\lfloor \frac{t}{\delta} \rfloor}}. \tag{10.30}$$

This implies that while at every reset time $t \in \mathcal{S}$, when $u, y \equiv 0$, the norm of the state vector may increase of a factor not greater than $\|R\|$, during the reset-free interval between two consecutive reset times (which, by hypothesis, must have length greater than δ) the same norm should decrease at least of the same factor. But then point (i) is immediately proved. To conclude (ii), it is sufficient to observe that in order for $y(t)$ and $u(t)$ to be $\mathcal{L}_2$, they must also be asymptotically converging to zero, since they are continuous in every finite interval between any two consecutive reset times. ■

Proposition 10.4 also allows us to conclude that when $r(t) \equiv 0$, and as a consequence $y(t)$, $u(t)$, $e_T(t)$ and $e(t)$ are signals in $\mathcal{L}_2$, then all the state sub-vectors $x_{\text{pl}}(t)$, $x_{\text{copy}}(t)$ and $x_{\text{cn}}(t)$ are converging to the zero vector.

All the previous considerations lead to the following.

Theorem 10.5 *The system described by equations (10.7)-(10.8)-(10.9) where the reset map in (10.9) is the optimal one suggested in (10.19) is asympotically stable in the sense of Definition 10.1.*

10.5 Simulation results

Consider a process with transfer function $g(s) = \frac{10}{s+1}$ and a controller with transfer function given by (10.1) with

$$q(s) = 10^{-1}\frac{3s^2 + 23s + 20}{5s^2 + 8s + 20}. \tag{10.31}$$

The subsequent Figs. 10.5-10.6 show the comparison between the outputs produced by applying the real optimal procedure described in section 10.3.1 in order to determine the optimal value (red line) for the reset times, and the suboptimal one suggested in section 10.3.2 (blue line). The two plots refer to different choices of the parameters R, W, K.

Referring to the same process and controller transfer functions, in Fig. 10.7 three different outputs are displayed, referring to the case when no reset control strategy is applied, when only one reset time is allowed (whose value has been computed following the approach in section 10.3.1 or 10.3.1, which are identical in this case) and finally, when two reset times are allowed (whose values have been computed following the approach in 10.3.1). It is apparent that, even with a very limited number of reset times, the performance immediately shows a drastic improvement.

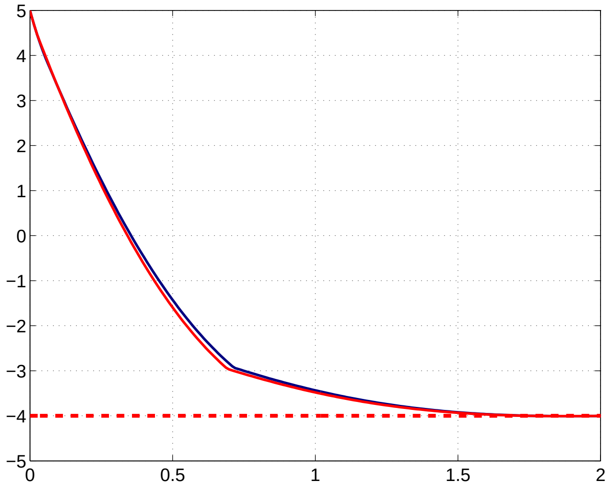


Figure 10.5: $R = 10,\ W = 10,\ K = 0$

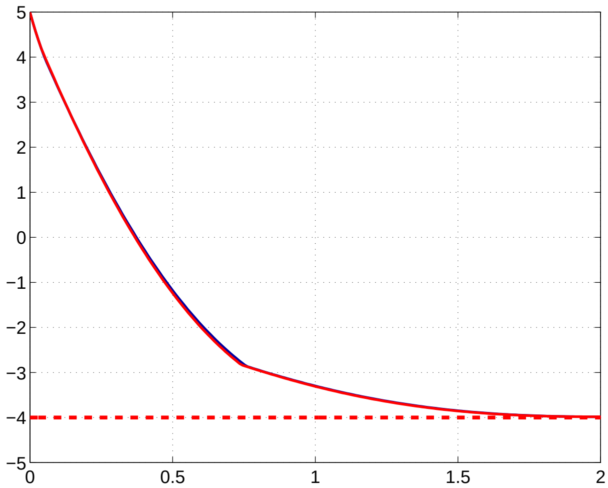


Figure 10.6: $R = 15,\ W = 2,\ K = 1$

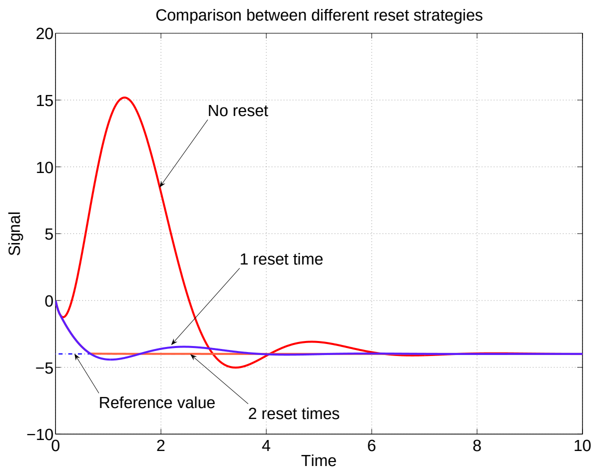


Figure 10.7: Different number of reset times, non-zero initial conditions

Appendix A

Technical results

In order to obtain necessary and sufficient conditions for the reachability of continuous-time positive switched systems, it seems mandatory to preliminarily clarify the zero pattern and dominant modes of $e^{At}, t \in \mathbb{R}_+$, when A is Metzler, and to investigate how the boundary of the cone generated by the columns of e^{At} evolves, as t goes from 0 to $+\infty$.

To achieve this goal, we need to introduce some new tools and to derive some new results, within the broad research area of nonnegative matrix theory, which enable use to explore the zero pattern and the elementary modes of the exponential of a Metzler matrix. In doing this, we can resort to a series of significant results obtained by Hershkowitz, Rothblum, Schneider and others, and pertaining M-matrices (occasionally, Z-matrices) [18, 21, 23, 42, 43, 46, 47, 29] or block triangular matrices [22]. All these results are collected in this Appendix, together with the others which are preliminary to the ones presented in this thesis.

The first result is rather standard in linear system theory, and hence we omit the proof.

Lemma A.1 *Let $\mathbf{v}^{(k)}$ be a generalized eigenvector of $A \in \mathbb{R}^{n\times n}$ of order k corresponding to the eigenvalue λ and set $\mathbf{v}^{(k-i)} := (A - \lambda I_n)^i \mathbf{v}^{(k)}$, for $i = 1, 2, \dots, k-1$. Then, at every time instant $t \in \mathbb{R}$, we have*

$$e^{At}\mathbf{v}^{(k)} = e^{\lambda t}\mathbf{v}^{(k)} + t \cdot e^{\lambda t}\mathbf{v}^{(k-1)} + \cdots + \frac{t^{k-1}}{(k-1)!} \cdot e^{\lambda t}\mathbf{v}^{(1)}.$$

Lemma A.2 *Let A be a Metzler matrix, then*

i) $e^{At} \geq 0, \ \forall\, t \geq 0$, *and if A is irreducible then* $e^{At} \gg 0$ *for every* $t > 0$*;*

ii) $\mathrm{ZP}(e^{At}) = \mathrm{ZP}(e^{A})$ *for every* $t > 0$*;*

iii) if a column of e^A is an ith monomial vector, then it must be the ith column (the nonzero entry must be on the main diagonal of e^A).

iv) if $\mathbf{v}$ and $\mathbf{w}$ are two nonnegative vectors with the same nonzero pattern (i.e., $\overline{\mathrm{ZP}}(\mathbf{v}) = \overline{\mathrm{ZP}}(\mathbf{w})$), then $\overline{\mathrm{ZP}}(e^{At}\mathbf{v}) = \overline{\mathrm{ZP}}(e^{At}\mathbf{w})$ for every $t \geq 0$. So, in particular, if $\mathcal{S} := \overline{\mathrm{ZP}}(\mathbf{v})$, then $\overline{\mathrm{ZP}}(e^{At}\mathbf{v}) = \overline{\mathrm{ZP}}(e^{At}\mathbf{e}_{\mathcal{S}})$ for every $t \geq 0$.

Proof: *i*) has been proved in [4].

To prove *ii*), assume $A = A_+ - \alpha I_n$, with $A_+ \geq 0$ and $\alpha \geq 0$. Clearly, the nonzero pattern of

$$e^{At} = e^{-\alpha t} e^{A_+ t} = e^{-\alpha t}\left[I_n + A_+ t + A_+^2 \frac{t^2}{2!} + \dots\right] \tag{A.1}$$

remains the same for every $t > 0$, so, in particular, it coincides with $\overline{\mathrm{ZP}}(e^{At})$ for $t = 1$. Eq. (A.1) clearly shows that at least the ith entry of each ith column of e^A must always be nonzero, hence *iii*) follows; *iv*) is obvious. ■

The previous lemma allows us to address the case when A is reducible.

Proposition A.3 *Let A be an $n \times n$ Metzler matrix in Frobenius normal form (1.2). Then, at every time instant $t > 0$*

$$e^{At} =: \mathcal{A}(t) = \begin{bmatrix} \mathcal{A}_{11}(t) & \mathcal{A}_{12}(t) & \dots & \mathcal{A}_{1\ell}(t) \\ & \mathcal{A}_{22}(t) & \dots & \mathcal{A}_{2\ell}(t) \\ & & \ddots & \vdots \\ & & & \mathcal{A}_{\ell\ell}(t) \end{bmatrix}, \tag{A.2}$$

where $\mathcal{A}_{ii}(t)$ is strictly positive for every i, while for $i \neq j$ the matrix $\mathcal{A}_{ij}(t)$ is either strictly positive or zero. Specifically,

$$\mathcal{A}_{ij}(t) = \begin{cases} \gg 0, & \text{if } i \in \mathcal{D}(\mathcal{C}_j) \ (\Leftrightarrow j \in \mathcal{A}(\mathcal{C}_i)); \\ 0, & \text{otherwise.} \end{cases}$$

Proof: The block-triangular structure of $\mathcal{A}(t)$ (and hence the fact that $\mathcal{A}_{ij}(t) = 0$ for $i > j$) is obvious, so we are remained to showing the nonzero pattern properties of the blocks $\mathcal{A}_{ij}(t)$ for $i \le j$. Condition $i \in \mathcal{D}(\mathcal{C}_j)$ holds if and only if for every vertex r in $\mathcal{C}_i$ and every vertex s in $\mathcal{C}_j$ there is a path of length say $k = k(r, s)$ from s to r, namely $[A^k]_{rs} > 0$ for some $k \in \mathbb{Z}_+$, or, equivalently, $[e^A]_{rs} > 0$. This amounts to saying that $i \in \mathcal{D}(\mathcal{C}_j)$ if and only if $\mathcal{A}_{ij}(1) \gg 0$ and hence, by Lemma A.2 point *ii*), if and only if $\mathcal{A}_{ij}(t) \gg 0$ for every $t > 0$. ∎

Remark A.4 *The previous result is consistent with the fact that the graph of the positive matrix* e^A *(and hence of* e^{At}, $\forall\, t > 0$*) is just the reflexive and transitive closure of* $\mathcal{G}(A)$*. Indeed, from equation (A.1), evaluated at* $t = 1$*, it immediately follows that* $\mathcal{G}(e^A)$ *is obtained from* $\mathcal{G}(A)$ *by adding loops and every arc* (j, i) *such that the vertex* j *has access to* i *in* $\mathcal{G}(A)$*.*

The "on/off" situation of the blocks of $\mathcal{A}(t)$ (by this meaning that they are either strictly positive or zero) entails immediate consequences on the zero pattern of the free state evolution of system (2.1), starting from any nonnegative initial condition.

Corollary A.5 *Let* A *be an* $n \times n$ *Metzler matrix in Frobenius normal form (1.2), and let* $\mathbf{v}$ *be a positive vector in* $\mathbb{R}^n_+$*. Set*[1]

$$J := \{j \in \langle \ell \rangle : \mathcal{C}_j \cap \overline{\mathrm{ZP}}(\mathbf{v}) \neq \emptyset\} \qquad \text{and} \qquad I := \cup_{j \in J} \mathcal{D}(\mathcal{C}_j).$$

Then, for every $t > 0$, $\overline{\mathrm{ZP}}(e^{At}\mathbf{v}) = \cup_{i \in I} \mathcal{C}_i$*.*

Proof: As nonzero blocks in $\mathcal{A}(t)$ are strictly positive, we have that for any $i \in \langle \ell \rangle$

$$\mathrm{block}_i[e^{At}\mathbf{v}] = \sum_{j=i}^{\ell} \mathcal{A}_{ij}(t) \cdot \mathrm{block}_j[\mathbf{v}] \neq 0$$

if and only if there exists $j \in \{i, i+1, \dots, \ell\}$ such that both $\mathcal{A}_{ij}(t) \neq 0$ (hence $\mathcal{A}_{ij}(t) \gg 0$) and $\mathrm{block}_j[\mathbf{v}] \neq 0$ (and if so, $\mathrm{block}_i[e^{At}\mathbf{v}] \gg 0$). This amounts

[1] The set J is also known [22, 47] as the *support* of $\mathbf{v}$. Notice that, in the general case, $\mathrm{supp}(\mathbf{v}) \neq \overline{\mathrm{ZP}}(\mathbf{v})$ and they coincide for each vector $\mathbf{v}$ if and only if each class consists of a single vertex, namely $\mathcal{R}(A) = \mathcal{G}(A)$.

to saying that there exists $j \in \langle \ell \rangle$ such that $i \in \mathcal{D}(\mathcal{C}_j)$ and $\mathcal{C}_j \cap \overline{\mathrm{ZP}}(\mathbf{v}) \neq \emptyset$. So, we have shown that $\mathrm{block}_i[e^{At}\mathbf{v}] \neq 0$ if and only if $i \in I$, and when so, $\mathrm{block}_i[e^{At}\mathbf{v}] \gg 0$. ∎

Remark A.6 *From the previous corollary, it immediately follows that the nonzero pattern of* $e^{At}\mathbf{v}$ *is the same at every time instant* $t > 0$*, and it always includes* $\overline{\mathrm{ZP}}(\mathbf{v})$*. Even more,* $\overline{\mathrm{ZP}}(e^{At}\mathbf{v})$ *is always the union of the indices corresponding to the classes* $\mathcal{C}_i, i \in I$*, which, in turn, includes* $\cup_{j\in J}\mathcal{C}_j$*. So, as a general result, we have for* $t > 0$

$$\overline{\mathrm{ZP}}(e^{At}\mathbf{v}) = \cup_{i\in I}\mathcal{C}_i \supseteq \cup_{j\in J}\mathcal{C}_j \supseteq \overline{\mathrm{ZP}}(\mathbf{v}).$$

Lemma A.7 *Let* A *be an* $n \times n$ *Metzler matrix in Frobenius normal form (1.2). If* $\mathbf{v} \in \mathbb{R}^n_+$ *and* $\mathcal{S} \subseteq \langle n \rangle$*, then*

$$\overline{\mathrm{ZP}}(e^{A\bar{t}}\mathbf{v}) = \mathcal{S},\ \exists \bar{t} > 0 \qquad \Rightarrow \qquad \begin{cases} \overline{\mathrm{ZP}}(e^{At}\mathbf{e}_{\mathcal{S}}) = \mathcal{S},\ \forall t > 0, \\ \\ \overline{\mathrm{ZP}}(\mathbf{v}) \subseteq \mathcal{S}. \end{cases} \tag{A.3}$$

Proof: By Lemma A.2, part *ii)*, if $\overline{\mathrm{ZP}}(e^{A\bar{t}}\mathbf{v}) = \mathcal{S},\ \exists \bar{t} > 0$, then $\overline{\mathrm{ZP}}(e^{At}\mathbf{v}) = \mathcal{S},\ \forall t > 0$. So, from Corollary A.5, it immediately follows that $\mathcal{S} = \overline{\mathrm{ZP}}(e^{At}\mathbf{v}) \supseteq \overline{\mathrm{ZP}}(\mathbf{v})$. To prove that $\overline{\mathrm{ZP}}(e^{At}\mathbf{e}_{\mathcal{S}}) = \mathcal{S}$ for every $t > 0$, set, as in Corollary A.5,

$$J := \{j \in \langle \ell \rangle : \mathcal{C}_j \cap \overline{\mathrm{ZP}}(\mathbf{v}) \neq \emptyset\} \qquad \text{and} \qquad I := \cup_{j\in J}\mathcal{D}(\mathcal{C}_j).$$

We know that $\overline{\mathrm{ZP}}(e^{At}\mathbf{v}) = \mathcal{S} = \cup_{i\in I}\mathcal{C}_i$. On the other hand, we can analogously define

$$J' := \{j \in \langle \ell \rangle : \mathcal{C}_j \cap \overline{\mathrm{ZP}}(\mathbf{e}_{\mathcal{S}}) \neq \emptyset\} \qquad \text{and} \qquad I' := \cup_{j\in J'}\mathcal{D}(\mathcal{C}_j),$$

so that $\overline{\mathrm{ZP}}(e^{At}\mathbf{e}_{\mathcal{S}}) = \cup_{i\in I'}\mathcal{C}_i$. We want to prove that $I' = I$. This follows immediately from the fact that

$$J' = \{j \in \langle \ell \rangle : \mathcal{C}_j \cap \left(\cup_{i\in I}\mathcal{C}_i\right) \neq \emptyset\} = I,$$

and hence $I' = \cup_{j\in I}\mathcal{D}(\mathcal{C}_j) = I$. ∎

Lemma A.8 *Every $n \times n$ Metzler matrix A in Frobenius normal form (1.2) admits an echelon basis.*

Proof: Since the diagonal block $A_{11} \in \mathbb{R}^{n_1 \times n_1}$ admits n_1 linearly independent (generalized) eigenvectors, $\mathbf{v}_1^{[i]}, i = 1, \ldots, n_1$, we first select the n_1 (generalized) eigenvectors of A

$$\mathbf{v}_i := \begin{bmatrix} \mathbf{v}_1^{[i]} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \qquad i = 1, \ldots, n_1.$$

Consider, now, the square matrix

$$\begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \in \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)}.$$

This matrix has, in turn, $n_1 + n_2$ linearly independent (generalized) eigenvectors, of size $n_1 + n_2$. If we choose the vectors $\begin{bmatrix} \mathbf{v}_1^{[i]} \\ 0 \end{bmatrix}$, $i = 1, \ldots, n_1$, as the first n_1 of them, then other n_2 generalized eigenvectors can be found and they necessarily have at least one nonzero entry corresponding to the indices of $\mathcal{C}_2$ (otherwise they could not be linearly independent from the first ones). So, they take the following form:

$$\mathbf{v}_i := \begin{bmatrix} \mathbf{v}_1^{[i]} \\ \mathbf{v}_2^{[i]} \end{bmatrix}, \qquad \mathbf{v}_2^{[i]} \neq 0, \ i = n_1 + 1, \ldots, n_1 + n_2.$$

It is then straightforward to complete these last n_2 vectors, by means of $n - (n_1 + n_2)$ zeros, to (generalized) eigenvectors of A. So far, we have obtained $n_1 + n_2$ linearly independent (generalized) eigenvectors of A. By proceeding in this way we get the desired family of generalized eigenvectors. ■

The preceding Lemma can also be obtained as a corollary of the Extension Lemma given in [22], which can be restated, according to the previous setting and notation, as follows:

Let $\tilde{A}$ be an $n \times n$ singular Metzler matrix in Frobenius normal form (1.2), let i be an index in $\langle \ell \rangle$ such that $\tilde{A}_{ii}$ is singular, and let j be in $\mathcal{C}_i$. For each vector, say $\mathbf{v}_i^{[j]}$, in the generalized nullspace (i.e. the generalized eigenspace *corresponding to the zero eigenvalue) of* $\tilde{A}_{ii}$ *there exists a vector* $\mathbf{v}_j$ *in the generalized nullspace of* $\tilde{A}$ *which is a weak* i*-combinatorial extension of* $\mathbf{v}_i^{[j]}$*, by this meaning that*

- $\mathrm{block}_i[\mathbf{v}_j] = \mathbf{v}_i^{[j]}$;
- $\overline{\mathrm{ZP}}(\mathbf{v}_j) \subseteq \cup_{k\in\mathcal{D}(\mathcal{C}_i)}\mathcal{C}_k \subseteq \cup_{k\le i}\mathcal{C}_k$.

So, once this lemma is applied to each singular Metzler matrix $A - \lambda I_n$, $\lambda \in \sigma(A)$, the result follows.

Remark A.9 *i) The result could also be obtained as an extension of the weakly preferred basis theorem (Theorem 4.9 in [22]) stating that if* $\tilde{A}$ *is a singular (upper) block-triangular matrix, then a basis for the generalized eigenspace of* $\tilde{A}$ *corresponding to its zero eigenvalue can be found, whose vectors satisfy the zero pattern constraints described within the previous proof (together with additional conditions, which are of no interest for the present analysis). Clearly, by applying the Theorem to each matrix* $A - \lambda I_n$*, as* λ *varies within* $\sigma(A)$*, we immediately get the desired echelon basis.*

ii) It can be shown that not every family of n *linearly independent generalized eigenvectors of* A *can be reduced, by means of a simple permutation, to an echelon basis for* A*. In particular, not every Jordan basis for* A*, namely a family of* n *linearly independent generalized eigenvectors ordered by chains, i.e.,* $\mathcal{B} = \{\mathbf{v}_h^{(k)}\}_{\substack{h=1,2,\dots,q\\ k=1,2,\dots,n_h}}$*, with* $\mathbf{v}_h^{(k)}$ *a generalized eigenvector of order* k *corresponding to the eigenvalue* λ_h*, and* $\mathbf{v}_h^{(k-1)} = (A - \lambda_h I_n)\mathbf{v}_h^{(k)}$*, is equivalent, up to a permutation, to an echelon basis. This is the case, for instance, of the diagonal matrix*

$$A = \begin{bmatrix} 1 & 0 & 0\\ 0 & 2 & 0\\ 0 & 0 & 1\end{bmatrix} \text{ which admits as a Jordan basis, for instance,}$$

$$\mathcal{B} = \left\{ \begin{bmatrix}1\\0\\1\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}2\\0\\1\end{bmatrix} \right\},$$

which cannot be reduced, by simple permutation, to an echelon basis.

However, it is not difficult to show (one may resort, for instance, to Lemma 3.6 in [22]) that an echelon basis for A *which is also a Jordan basis always exists.*

An echelon basis $\mathcal{B}_e$ can be endowed with nice additional properties.

Proposition A.10 *Let A be an $n \times n$ Metzler matrix in Frobenius normal form (1.2), and let*

$$\mathcal{B}_e := \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{n_1}, \mathbf{v}_{n_1+1}, \dots, \mathbf{v}_n\} = \cup_{i=1}^{\ell} \cup_{j\in\mathcal{C}_i} \mathbf{v}_j$$

be any echelon basis for A (satisfying the ordering and nonzero pattern assumptions of Lemma A.8). Assume that the first vector of each class $\mathcal{C}_i, i \in \langle \ell \rangle$, namely $\mathbf{v}_{n_1+\dots+n_{i-1}+1}$, is a (possibly generalized) eigenvector of A corresponding to $\lambda_{\max}(A_{ii})$, and denote it in the sequel as

$$\mathbf{v}_{n_1+\dots+n_{i-1}+1} =: \mathbf{v}_{\mathcal{C}_i} = \begin{bmatrix} w_1^{[i]} \\ w_2^{[i]} \\ \vdots \\ w_i^{[i]} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \qquad \textit{where} \qquad w_i^{[i]} \neq 0. \tag{A.4}$$

Then, (possibly modulo a change of sign of all the entries of $\mathbf{v}_{\mathcal{C}_i}$) we have that

i) $w_i^{[i]}$ is a strictly positive eigenvector of A_{ii} corresponding to $\lambda_{\max}(A_{ii})$, and

ii) for $i \geq 2$, if it is possible to define the index

$$\begin{aligned} k_i := \min\{j < i : \ & \textit{for every } r \in \{j, j+1, \dots, i-1\} \\ & \textit{either } (a)\ r \in \mathcal{D}(\mathcal{C}_i) \textit{ and } \lambda_{\max}(A_{rr}) < \lambda_{\max}(A_{ii}) \\ & \textit{or } (b)\ r \notin \mathcal{D}(\mathcal{C}_i) \textit{ and } \lambda_{\max}(A_{ii}) \notin \sigma(A_{rr})\}, \end{aligned}$$

then in (A.4) all blocks $w_r^{[i]}$, for $r = k_i, k_i+1, \dots, i-1$, satisfy the following condition:

$$w_r^{[i]} = \begin{cases} \gg 0, & \text{if } r \text{ falls in case (a);} \\ 0, & \text{if } r \text{ falls in case (b).} \end{cases} \tag{A.5}$$

Proof: Let m be the order of $\mathbf{v}_{\mathcal{C}_i}$ as generalized eigenvector of A corresponding to $\lambda_{\max}(A_{ii})$.

i) Of course, $(A - \lambda_{\max}(A_{ii})I_n)^m \mathbf{v}_{\mathcal{C}_i} = 0$ implies, in particular, $(A_{ii} - \lambda_{\max}(A_{ii})I)^m w_i^{[i]} = 0$. This means that $w_i^{[i]} \neq 0$ is a generalized eigenvector (of some order k, with $1 \leq k \leq m$) of A_{ii} corresponding to $\lambda_{\max}(A_{ii})$. Since $\lambda_{\max}(A_{ii})$ is a simple eigenvalue of A_{ii}, it follows that $w_i^{[i]}$ is an eigenvector. On the other hand, since A_{ii} admits a strictly positive eigenvector corresponding to $\lambda_{\max}(A_{ii})$ (which is uniquely determined up to positive multiplicative coefficients), $w_i^{[i]}$ can always be assumed strictly positive.

ii) Suppose that $i \geq 2$ and the index k_i may be defined, and consider the submatrix $A_{\{k_i,i\}}$ of A (see (1.1)). By definition of k_i, $\lambda_{\max}(A_{ii})$ is a simple eigenvalue of $A_{\{k_i,i\}}$. So, since $\mathbf{v}_{\mathcal{C}_i}$ is a generalized eigenvector of order m of A corresponding to $\lambda_{\max}(A_{ii})$, then

$$\begin{bmatrix} w_{k_i}^{[i]} \\ w_{k_i+1}^{[i]} \\ \vdots \\ w_i^{[i]} \end{bmatrix}$$

is an eigenvector of $A_{\{k_i,i\}}$ corresponding to the same eigenvalue. Even more, $\mathcal{C}_i$ represents a distinguished class of the directed graph associated with $A_{\{k_i,i\}}$. So, we may apply[2] Theorem 3.7 in [47] (see, also, [19, 29]), and deduce that since $w_i^{[i]} \gg 0$, then all blocks $w_r^{[i]}$, for $r = k_i, k_i+1, \dots, i-1$, (which are uniquely determined by $w_i^{[i]}$ because $\lambda_{\max}(A_{ii}) \notin \sigma(A_{\{k_i,i-1\}})$) satisfy (A.5). This completes the proof. ∎

Remark A.11 *i) Theorem 3.1 in [47] for M-matrices can be obtained as a corollary of the previous proposition in the special case, when we consider only classes $\mathcal{C}_i$ for which $k_i = 1$.*

ii) One may wonder under what conditions there exists an echelon basis of A whose vectors (either eigenvectors or generalized eigenvectors) are all positive. It is clear that since the last nonzero block of a vector $\mathbf{v}_i$ in $\mathcal{B}_e$,

[2] As a matter of fact, the result was obtained for nonnegative matrices and hence it applies, in its original formulation, to any matrix $A_+ \geq 0$ such that $A = A_+ - \alpha I_n, \exists\, \alpha \geq 0$, and, correspondingly, to the real eigenvalue $\alpha + \lambda_{\max}(A_{ii}) \in \sigma(A_+)$. However, its adjustment to the case of a Metzler matrix is rather straightforward.

corresponding to some class $\mathcal{C}_j$, is a (possibly generalized) eigenvector of the irreducible matrix A_{jj}, the only way to ensure that $\text{block}_j[\mathbf{v}_i] \geq 0$ *for every choice of $\mathbf{v}_i$ is to impose that A_{jj} has size $n_j = 1$. So, a necessary condition for an echelon basis to have positive vectors is that all communicating classes consist of a single vertex (i.e., A, in Frobenius normal form, is upper triangular); equivalently, $\ell = n$ and hence $\mathbf{v}_i = \mathbf{v}_{\mathcal{C}_i}$ $\forall$ i.*

Assume, now, that A is upper triangular. If $k_i = 1$ for every $i \in \{2, \dots, \ell\} = \{2, \dots, n\}$(i.e. each class $\mathcal{C}_i$ is a distinguished class for the graph associated with the submatrix $A_{\{1,i\}}$, defined in (1.1)), then all nonzero entries (= blocks of unitary size) of $\mathbf{v}_i$ are necessarily positive. So, we have shown that if $\ell = n$ and $k_i = 1$ for every $i \geq 2$, then A admits a positive echelon basis. Even more, the basis thus obtained consists of positive eigenvectors and hence is a Jordan basis.

If all the eigenvalues of A are distinct, the previous one is also a necessary condition: indeed, a positive echelon basis exists only if $\ell = n$ and $k_i = 1$ for every $i \geq 2$. In the general case, however, this is not true and we can derive a weaker sufficient condition by resorting to the preferred basis theorem for Z-matrices *as it has been derived in [23]. Indeed, by restating that result according to our notation and for Metzler matrices, we get:*

FACT: (Corollary 5.16 in [23]) Let A be an $n \times n$ Metzler matrix in Frobenius normal form (1.2), and let $\lambda \in \sigma(A)$ be a real eigenvalue of A. Let $E_\lambda(A)$ be the generalized eigenspace of A corresponding to λ and set $S := \{i \in \langle \ell \rangle : \lambda \in \sigma(A_{ii})\}$. Then the following facts are equivalent:

i) $E_\lambda(A)$ has a nonnegative basis;

ii) $E_\lambda(A)$ has an S-preferred basis[3];

iii) for every $j \in \cup_{i \in S} \mathcal{D}(\mathcal{C}_i)$, $\lambda_{\max}(A_{jj}) \leq \lambda$.

(Notice that, as an immediate consequence of point iii) in the previous corollary, an S-preferred basis may exist only corresponding to those eigenvalues $\lambda \in \sigma(A)$ which are dominant for the diagonal blocks in which they appear, which means that $S \equiv \{i \in \langle \ell \rangle : \lambda = \lambda_{\max}(A_{ii})\}$). So, to ensure the existence of an echelon basis for A consisting of positive vectors, it is sufficient that (the Frobenius normal form of) A is an upper triangular matrix,

[3]For the formal definition of an S-preferred basis we refer to Definitions 4.1 and 4.7 in [23]. For the present discussion, and according to the previous notation, it is sufficient to recall that if $\{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^r\}$ is an S-preferred basis of $E_\lambda(A)$ then, in particular, for every $i \in \langle r \rangle$, $\text{block}_j[\mathbf{x}^i]$ is strictly positive if $j \in \mathcal{D}(\mathcal{C}_i)$ and zero otherwise.

and that for every $i \in \langle n \rangle$, the vertex i has access only to vertices $j \leq i$ with $a_{jj} \leq a_{ii}$. It is worthwhile noticing, however, that the echelon basis obtained under these assumptions may not be a Jordan basis. Consider, for instance, the two simple triangular matrices:

$$A_1 = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \qquad \textit{and} \qquad A_2 = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

For neither of them the indices k_2 and k_3 can be defined, but it is easily seen that both A_1 and A_2 satisfy the previous sufficient condition and hence admit a nonnegative echelon basis: the canonical basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$. However, such a basis is a Jordan basis just for A_1, while any echelon basis of A_2 which satisfies the constraints of Proposition A.10 and is a Jordan basis takes the following form:

$$\mathcal{B}_e = \left\{ \begin{bmatrix} a \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} * \\ b \\ 0 \end{bmatrix}, \begin{bmatrix} * \\ -c \\ c \end{bmatrix} \} \right\},$$

where a, b and c are positive real numbers, while $*$ denotes an arbitrary real number, and hence it cannot be nonnegative.

iii) It is worthwhile to conclude the section by providing a general comment about the relationship between the echelon basis here introduced and the preferred basis treated in [23], for instance. First of all, an echelon basis of a Metzler matrix A is a basis of the whole vector space $\mathbb{R}^n$ consisting of (generalized) eigenvectors. It yields information on the zero/nonzero patterns of all the eigenvectors in the basis. It also yields information on the positive entries of eigenvectorscorresponding to the dominant eigenvalues $\lambda_{\max}(A_{ii})$ of the diagonal blocks in a Frobenius normalform. Moreover, it exists for any Metzler matrix.

On the other hand, the existence of a preferred basis, as considered in [23], is always ensured only for the generalized eigenspace corresponding to $\lambda_{\max}(A)$. It may exist, under suitable conditions, also for the generalized eigenspaces corresponding to the other eigenvalues $\lambda_{\max}(A_{ii})$, but in the general case we cannot ensure the existence of a basis of $\mathbb{R}^n$ which is obtained as the union of preferred bases.

Also, a preferred basis consists of positive vectors, but it is not necessarily a Jordan basis (this problem has been investigated in [42] for M-matrices and dominant eigenvalues). An echelon basis which is a Jordan basis always exists and Proposition A.10 points out which structural and positivity properties can always be ensured.

Proposition A.12 *Let $A \in \mathbb{R}^{n\times n}$ be an irreducible Metzler matrix and let $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a basis of $\mathbb{R}^n$ consisting of eigenvectors and generalized eigenvectors of A, with $\mathbf{v}_1$ being the strictly positive eigenvector of A corresponding to $\lambda_{\max}(A)$. Then every $\mathbf{v} \in \mathbb{R}^n_+, \mathbf{v} > 0$, can be expressed as $\mathbf{v} = c_1\,\mathbf{v}_1 + \ldots + c_n\,\mathbf{v}_n$, for suitable complex coefficients c_i, with $c_1 > 0$.*

Proof: Suppose that $\mathbf{v} = c_1\,\mathbf{v}_1 + \ldots + c_n\,\mathbf{v}_n$ is a positive vector and let $\mathbf{w} \gg 0$ be a left eigenvector of A corresponding to $\lambda_{\max}(A)$. Since all vectors $\mathbf{v}_i, i \geq 2$, correspond to eigenvalues distinct from $\lambda_{\max}(A)$, then $\mathbf{w}^T\mathbf{v}_i = 0$ for every $i \geq 2$. Consequently, conditions $\mathbf{w}^T\mathbf{v}_1 > 0$ and $0 < \mathbf{w}^T\mathbf{v} = c_1 \cdot \mathbf{w}^T\mathbf{v}_1$, together, ensure that $c_1 > 0$. ■

Corollary A.13 *Let A be an $n \times n$ Metzler matrix in Frobenius normal form (1.2) and let $\mathcal{B}_e = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ be an echelon basis for A satisfying the assumptions of Proposition A.10. If $\tilde{\mathbf{z}} = \sum_{i=1}^n c_i\mathbf{v}_i$, $c_i \in \mathbb{C}$, is a positive vector with $\overline{\mathrm{ZP}}(\tilde{\mathbf{z}}) \subseteq \cup_{i=1}^k \mathcal{C}_i$ and $\overline{\mathrm{ZP}}(\tilde{\mathbf{z}}) \cap \mathcal{C}_k \neq \emptyset$, then the (possibly generalized) eigenvector of A corresponding to $\lambda_{\max}(A_{kk})$ and to the class $\mathcal{C}_k$, i.e. $\mathbf{v}_{\mathcal{C}_k} := \mathbf{v}_{n_1+\dots+n_{k-1}+1}$, is weighted with a positive coefficient in the expression of $\tilde{\mathbf{z}}$.*

Proof: Assume that $\tilde{\mathbf{z}}$ is block partitioned as follows

$$\tilde{\mathbf{z}}^T = \begin{bmatrix} \mathbf{z}_1^T \; \mathbf{z}_2^T \; \dots \; \mathbf{z}_k^T \; 0 \dots \; 0 \end{bmatrix}.$$

It is clearly seen that $c_i = 0$ for all $i > n_1 + n_2 + \dots + n_k$, since all (linearly independent) vectors $\mathbf{v}_i$ for $i > n_1 + n_2 + \dots + n_k$ have nonzero components corresponding to the classes $\mathcal{C}_j$, $j > k$. Moreover, it is worth to observe that, due to the structure of the echelon basis,

$$\mathbf{z}_k = \mathrm{block}_k[\tilde{\mathbf{z}}] = \sum_{i \in \mathcal{C}_k} c_i \cdot \mathrm{block}_k[\mathbf{v}_i].$$

Recall now that all subvectors $\mathrm{block}_k[\mathbf{v}_i], i \in \mathcal{C}_k$, are the (possibly generalized) eigenvectors of A_{kk}. As a consequence, since $\mathbf{z}_k > 0$, from Proposition A.12 we can conclude that $c_{n_1+\dots+n_{k-1}+1} > 0$ and hence the vector $\mathbf{v}_{\mathcal{C}_k}$ is weighted by a positive coefficient. ■

Before proceeding, we need to recall here a known result (see [43] and, for instance, [47], Corollary 7.5, and references therein) which, within our framework and according to our notation, easily leads to the following:

Lemma A.14 *Given a Metzler matrix A in Frobenius normal form (1.2), the size of the largest Jordan block relative to the dominant eigenvalue of A, $\lambda_{\max}(A)$, is equal to the maximum number of classes $\mathcal{C}_k$ with $\lambda_{\max}(A_{kk}) = \lambda_{\max}(A)$ that lie in a single chain in $\mathcal{R}(A)$.*

Proposition A.15 *Let $A \in \mathbb{R}^{n\times n}$ be a Metzler matrix in Frobenius normal form (1.2), and let i and j be indices in $\langle \ell \rangle$ such that $\mathcal{A}(\mathcal{C}_i) \cap \mathcal{D}(\mathcal{C}_j) \neq \emptyset$. Then*

i) the only modes $\dfrac{t^m}{m!}e^{\lambda t}$ appearing in $\mathcal{A}_{ij}(t)$ are those corresponding to eigenvalues $\lambda \in \sigma(A_{kk})$, with $k \in \mathcal{A}(\mathcal{C}_i) \cap \mathcal{D}(\mathcal{C}_j)$.

Moreover, set

$$\lambda^*_{i,j} := max\{\lambda_{\max}(A_{kk}) : \ k \in \mathcal{A}(\mathcal{C}_i) \cap \mathcal{D}(\mathcal{C}_j)\},$$

and let $\bar{m}_{i,j} + 1$ be the maximum number of classes $\mathcal{C}_k$ with $\lambda_{\max}(A_{kk}) = \lambda^*_{i,j}$ that lie in a single chain from $\mathcal{C}_j$ to $\mathcal{C}_i$ in $\mathcal{R}(A)$. Then,

ii) for each $h \in \mathcal{C}_i$ and $k \in \mathcal{C}_j$ we have

$$[e^{At}]_{h,k} \sim \frac{t^{\bar{m}_{i,j}}}{\bar{m}_{i,j}!}e^{\lambda^*_{i,j}t},$$

namely, $\dfrac{t^{\bar{m}_{i,j}}}{\bar{m}_{i,j}!}e^{\lambda^*_{i,j}t}$ is the dominant mode in the expression of the (h,k)th entry of e^{At}.

Proof: $i)$ Partition the set $\langle \ell \rangle$ into the following three disjoint sets

$$\begin{aligned} R &:= A(\mathcal{C}_i) \cap \mathcal{D}(\mathcal{C}_j) = \{k_1, k_2, \dots, k_r\},\\ R_1 &:= \mathcal{D}(\mathcal{C}_j) \setminus R,\\ R_3 &:= \langle \ell \rangle \setminus (R \cup R_1), \end{aligned}$$

with $i = k_1 < \ldots < k_r = j$. If $r_i := |R_i|$, $i = 1, 3$, then $r_1 + r + r_3 = \ell$. Consider now a permutation matrix P such that in

$$\hat{A} := P^T A P = \begin{bmatrix} \hat{A}_{11} & \hat{A}_{12} & \hat{A}_{13} \\ 0 & \hat{A}_{22} & \hat{A}_{23} \\ 0 & 0 & \hat{A}_{33} \end{bmatrix} \tag{A.6}$$

- $\hat{A}_{11}$ is block-triangular, and its diagonal blocks are the r_1 matrices A_{ii} with $i \in R_1$,
- $\hat{A}_{33}$ is block-triangular, and its diagonal blocks are the r_3 matrices A_{ii} with $i \in R_3$, and
- $$\hat{A}_{22} = \begin{bmatrix} A_{ii} & A_{ik_2} & \dots & \dots & A_{ij} \\ 0 & A_{k_2k_2} & \dots & \dots & A_{k_2j} \\ 0 & 0 & \ddots & \dots & \vdots \\ 0 & 0 & 0 & A_{k_{r-1}k_{r-1}} & A_{k_{r-1}j} \\ 0 & 0 & 0 & 0 & A_{jj} \end{bmatrix}.$$

Correspondingly we get

$$e^{\hat{A}t} = \begin{bmatrix} e^{\hat{A}_{11}t} & * & * \\ 0 & e^{\hat{A}_{22}t} & * \\ 0 & 0 & e^{\hat{A}_{33}t} \end{bmatrix}, \tag{A.7}$$

and since $\mathcal{A}_{ij}(t) = \text{block}_{(i,j)}[e^{At}] = \text{block}_{(r_1+1,\,r_1+r)}[e^{\hat{A}t}] = \text{block}_{(1,\,r)}[e^{\hat{A}_{22}t}]$ it is easy to conclude that the only modes $\frac{t^m}{m!}e^{\lambda t}$ appearing in $\mathcal{A}_{ij}(t)$ are those corresponding to eigenvalues $\lambda \in \sigma(A_{kk})$, with $k \in \mathcal{A}(\mathcal{C}_i) \cap \mathcal{D}(\mathcal{C}_j) = \{k_1, k_2, \dots, k_r\}$.

Since the expression of $\mathcal{A}_{ij}(t)$ is completely determined by the time evolution of the matrix $e^{\hat{A}_{22}t}$, in the sequel of the proof we will uniquely focus on this latter, and simplify our notation by assuming $A = \hat{A}_{22}$, $(i,j) = (1,\ell)$ and $\mathcal{A}(\mathcal{C}_1) \cap \mathcal{D}(\mathcal{C}_\ell) = \langle \ell \rangle$. Consequently, $\lambda^*_{i,j}$ will be replaced by $\lambda_{\max}(A)$ and $\bar{m}_{i,j}$ by $\bar{m}$, the maximum number of classes $\mathcal{C}_k$ with $\lambda_{\max}(A_{kk}) = \lambda_{\max}(A)$ that lie in a single chain in $\mathcal{R}(A)$ minus 1.

ii) By Lemma A.14, none of the elementary modes $\frac{t^m}{m!}e^{\lambda t}$ appearing in the expression of the entries of e^{At} can dominate $\frac{t^{\bar{m}}}{\bar{m}!}e^{\lambda_{\max}(A)t}$. So, in particular, for every $h \in \mathcal{C}_1$ and $k \in \mathcal{C}_\ell$:

$$\lim_{t\to\infty} \frac{[e^{At}]_{h,k}}{\frac{t^{\bar{m}}}{\bar{m}!}e^{\lambda_{\max}(A)t}} < \infty. \tag{A.8}$$

Let $\mathcal{B}_e = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be an echelon basis for A which is a Jordan basis and satisfies the additional conditions of Proposition A.10. Let $\mathbf{v}_l \in \mathcal{B}_e$ be the generalized eigenvector of order $\bar{m}+1$ corresponding to $\lambda_{max}(A)$ of smallest index (i.e. l is minimum among the indices of all vectors in $\mathcal{B}_e$ which are generalized eigenvectors of order $\bar{m}+1$ corresponding to $\lambda_{max}(A)$). Clearly, $\mathbf{v}_l = \mathbf{v}_{\mathcal{C}_g}$ for some class $\mathcal{C}_g$. Moreover $\mathbf{v} := (A - \lambda_{\max}(A)I_n)^{\bar{m}}\mathbf{v}_{\mathcal{C}_g}$ is still in $\mathcal{B}_e$ and, precisely, it is an eigenvector of A corresponding to $\lambda_{\max}(A)$ and to some class $\mathcal{C}_b$ (so that $\mathbf{v} = \mathbf{v}_{\mathcal{C}_b}$). Notice that, by Lemma A.1,

$$\begin{aligned} e^{At}\mathbf{v}_{\mathcal{C}_g} &= e^{\lambda_{\max}(A)t}\mathbf{v}_{\mathcal{C}_g} + t\cdot e^{\lambda_{\max}(A)t}(A-\lambda_{\max}(A)I_n)\mathbf{v}_{\mathcal{C}_g} + \dots \\ &\quad + \frac{t^{\bar{m}}}{\bar{m}!}\cdot e^{\lambda_{\max}(A)t}(A-\lambda_{\max}(A)I_n)^{\bar{m}}\mathbf{v}_{\mathcal{C}_g} \\ &= e^{\lambda_{\max}(A)t}\mathbf{v}_{\mathcal{C}_g} + \dots + \frac{t^{\bar{m}}}{\bar{m}!}\cdot e^{\lambda_{\max}(A)t}\mathbf{v}_{\mathcal{C}_b}. \end{aligned}$$

On the other hand, since the class $\mathcal{C}_1$ is accessible from every other class, and hence, in particular, from $\mathcal{C}_b$, and $\lambda_{\max}(A_{bb}) = \lambda_{\max}(A) > \lambda_{\max}(A_{hh})$ for every $h < b$ such that $h \in \mathcal{D}(\mathcal{C}_b)$ (if not, we would have more than $\bar{m}_{i,j}+1$ classes in the chain corresponding to $\lambda_{\max}(A)$, a contradiction), by Proposition A.10, $\mathrm{block}_1[\mathbf{v}_{\mathcal{C}_b}] \gg 0$.

Let k be an arbitrary index in $\mathcal{C}_\ell$, and set $e^A e_k =: \left[w_1^T \ \dots \ w_g^T \ \dots \ w_\ell^T\right]^T$. Since every class is accessible from $\mathcal{C}_\ell$, $e^A e_k \gg 0$. By Corollary A.13 , the vector

$$\tilde{\mathbf{z}}^T := \left[w_1^T \ \dots \ w_g^T \ 0 \ \dots \ 0\right]^T$$

has a positive projection on the generalized eigenvector $\mathbf{v}_{\mathcal{C}_g}$. Consequently, for every $h \in \mathcal{C}_1$ and every $k \in \mathcal{C}_\ell$, and sufficiently large t, we get

$$\begin{aligned} [e^{At}]_{h,k} &= \mathbf{e}_h^T e^{At}\mathbf{e}_k = \mathbf{e}_h^T e^{A(t-1)}\left[e^A\mathbf{e}_k\right] \geq \mathbf{e}_h^T e^{A(t-1)}\tilde{\mathbf{z}} \\ &\sim \mathbf{e}_h^T\left[\frac{t^{\bar{m}}}{\bar{m}!}e^{\lambda_{\max}(A)t}\mathbf{v}_{\mathcal{C}_b}\right] = \frac{t^{\bar{m}}}{\bar{m}!}e^{\lambda_{\max}(A)t}\left[\mathrm{block}_1[\mathbf{v}_{\mathcal{C}_b}]\right]_h. \end{aligned} \tag{A.9}$$

So, putting together (A.8) and (A.9), we get the result. ■

As an immediate corollary of Proposition A.15, we obtain the following result, which extends to the continuous-time case a result that in the discrete time case was always true for (nonnegative) matrices with primitive diagonal blocks, but could not be true in the general case (unless introducing a smoothing factor, which compensates for the periodic patterns due to nontrivial cyclicity indices of the diagonal blocks). The result was nicely described in [47] and we will paraphrase here Schneider's comment: the pseudo-exponential growth of the (i,j)th block of e^{At} is determined by the hardest path from $\mathcal{C}_j$ to $\mathcal{C}_i$ in $\mathcal{R}(A)$: a chain that not only reaches the highest peaks (of dominant eigenvalue $\lambda^*_{i,j}$) but also the maximum number of peaks of that height ($\bar{m}_{i,j}+1$).

Theorem A.16 *Let $A \in \mathbb{R}^{n\times n}$ be a Metzler matrix in Frobenius normal form (1.2). For any pair of indices i and j in $\langle \ell \rangle$, we have:*

- *if $\mathcal{A}(\mathcal{C}_i) \cap \mathcal{D}(\mathcal{C}_j) = \emptyset$, then $\mathcal{A}_{ij}(t) = 0$;*
- *if $\mathcal{A}(\mathcal{C}_i) \cap \mathcal{D}(\mathcal{C}_j) \neq \emptyset$, then $\mathcal{A}_{ij}(t) \sim e^{\lambda^*_{i,j} t}\dfrac{t^{\bar{m}_{i,j}}}{\bar{m}_{i,j}!}$, where $\lambda^*_{i,j}$ and $\bar{m}_{i,j}$ are defined as in Proposition A.15.*

An interesting decomposition of the exponential matrix e^{At}, which highlights the dominant mode of each column and the expression of the associated vector coefficient, can be obtained as an immediate corollary of Theorem A.16.

Proposition A.17 *Let A be an $n \times n$ Metzler matrix in Frobenius normal form (1.2). Then there exist (not necessarily distinct) positive eigenvectors of unitary modulus of A, $\tilde{\mathbf{v}}_j \in \mathbb{R}^n_+$, and real modes $m_j(t) = \dfrac{t^{\bar{m}_j}}{\bar{m}_j!} e^{\lambda^*_j t}$, with $\lambda^*_j \in \mathbb{R}$ and $\bar{m}_j \in \mathbb{Z}_+$, $j \in \langle \ell \rangle$, and strictly positive row vectors $\mathbf{c}_i \in \mathbb{R}^{1\times n_i}_+$ such that*

$$\mathcal{A}(t) = e^{At} = \begin{bmatrix} \tilde{\mathbf{v}}_1 & \tilde{\mathbf{v}}_2 & \dots & \tilde{\mathbf{v}}_\ell \end{bmatrix} \begin{bmatrix} m_1(t) & & & \\ & m_2(t) & & \\ & & \ddots & \\ & & & m_\ell(t) \end{bmatrix} \begin{bmatrix} \mathbf{c}_1 & & & \\ & \mathbf{c}_2 & & \\ & & \ddots & \\ & & & \mathbf{c}_\ell \end{bmatrix} + \mathcal{A}_{lc}(t), \tag{A.10}$$

and for every $i \in \langle n \rangle$ if we let $\mathcal{C}_j$ be the class of vertex i, then

$$\lim_{t\to+\infty} \frac{\mathcal{A}_{lc}(t)\mathbf{e}_i}{m_j(t)} = 0.$$

Moreover,

$$\lambda^*_j = \max\{\lambda_{\max}(A_{kk}) : k \in \mathcal{D}(\mathcal{C}_j)\},$$

and $\bar{m}_j + 1$ is the maximum number of classes $\mathcal{C}_k$ with $\lambda_{\max}(A_{kk}) = \lambda^*_j$ that lie in a single chain from $\mathcal{C}_j$ in $\mathcal{R}(A)$. Also, $\tilde{\mathbf{v}}_j$ is a positive eigenvector of A corresponding to λ^*_j.

Remark A.18 *i) Observe that, by referring to Definition 8.1,* $\tilde{\mathbf{v}}_j = \mathbf{v}_j^\infty$*. As a consequence,* $\mathrm{Cone}_\infty(e^{At}) = \mathrm{Cone}(\begin{bmatrix}\tilde{\mathbf{v}}_1 & \tilde{\mathbf{v}}_2 & \dots & \tilde{\mathbf{v}}_\ell\end{bmatrix})$*.*

ii) A weaker formulation of Proposition A.17, which does not require the matrix A to be in Frobenius normal form (1.2), is the following

Proposition A.19 *Let* A *be an* $n \times n$ *Metzler matrix. Then there exist nonnegative eigenvectors of* A*,* $\mathbf{v}_j \in \mathbb{R}^n_+, \mathbf{v}_j \neq 0$*, and real modes* $m_j(t) = \frac{t^{k_j}}{k_j!}e^{\lambda_j t}$*, with* $\lambda_j \in \sigma(A) \cap \mathbb{R}$ *and* $k_j \in \mathbb{Z}_+$*,* $j = 1, 2, \dots, n$*, such that*

$$e^{At} = \begin{bmatrix}\mathbf{v}_1 & \dots & \mathbf{v}_n\end{bmatrix} \begin{bmatrix} m_1(t) & & \\ & \ddots & \\ & & m_n(t) \end{bmatrix} + \mathcal{A}_{lc}(t), \tag{A.11}$$

$$\text{with} \lim_{t\to+\infty} \frac{\mathcal{A}_{lc}(t)\mathbf{e}_j}{m_j(t)} = 0, \qquad \forall\, j \in \{1, \dots, n\}.$$

Even more, we can always assume $\mathbf{v}_j = \mathbf{v}_j^\infty$ *for every index* $j \in \{1, \dots, n\}$*.*

Lemma A.20 *Given an* n*-dimensional Metzler matrix* A *in Frobenius normal form (1.2), let* $V := \{\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \dots, \tilde{\mathbf{v}}_\ell\}$ *be the set of all the asymptotic directions of the columns of* e^{At}*, corresponding to the various classes* $\mathcal{C}_1, \dots, \mathcal{C}_\ell$*, as they are defined in Proposition A.17. Then either all vectors in* V *are linearly independent or, if they are not, one at least of such vectors can be expressed as a positive combination of the others.*

Proof: We know from Proposition A.17 that each $\tilde{\mathbf{v}}_j$ is a positive eigenvector of A, corresponding to some eigenvalue λ_j^* and to some class $\mathcal{C}_j$. Observe that eigenvectors corresponding to distinct eigenvalues are linearly independent. As a consequence, V is linearly dependent only if there exists some eigenvalue λ such that the set of (not necessarily distinct) eigenvectors in V which correspond to λ, say $V_\lambda := \{\tilde{\mathbf{v}}_{i_1}, \dots, \tilde{\mathbf{v}}_{i_s}\}$, $s \leq \ell$, is linearly dependent.

Express the set V_λ as the union of two subsets $V_\lambda := V_{\mathrm{dist}} \cup V_{\mathrm{rem}}$, with V_{dist} containing those eigenvectors of V_λ which correspond to some distinguished class $\mathcal{C}_i$, and the second one including those eigenvectors of V_λ which do not correspond to a distinguished class.

Since the asymptotic direction of a column corresponding to a distinguished class $\mathcal{C}_j$ always exhibits a strictly positive jth block, and from any distinguished class $\mathcal{C}_j$ one cannot access any other distinguished class corresponding to the same eigenvalue (thus having the corresponding entries all equal to zero (see [45])), it necessarily follows that the vectors in V_{dist} are all linearly independent. So, if V_λ is a set of linearly dependent vectors, then it must be $V_{\text{rem}} \neq \emptyset$.

Choose $\tilde{\mathrm{v}}_j \in V_{\text{rem}}$, and let $\mathcal{C}_j$ be the communicating class corresponding to $\tilde{\mathrm{v}}_j$. Since $\tilde{\mathrm{v}}_j$ is an eigenvector corresponding to λ, $\mathcal{C}_j$ must access at least one distinguished class whose dominant eigenvalue is λ. If $\bar{m}_j + 1$ is the maximum number of distinguished class with dominant eigenvalue λ which can be encountered along a chain of classes starting from $\mathcal{C}_j$ and there exist k such chains, then by Proposition A.17, $\tilde{\mathrm{v}}_j$ is necessarily a linear combination of the k linearly independent eigenvectors (belonging to V_{dist}) which correspond to λ and to those k distinguished classes.

Now, since each one of these eigenvectors, as previously observed, has one strictly positive block which is zero in all the other eigenvectors, in order for $\tilde{\mathrm{v}}_j$ to be positive such a linear combination must have only positive coefficients. By applying this reasoning to all vectors in V_{rem}, we can claim that the cone generated by the vectors in V_λ is equal to the cone generated by the vectors in V_{dist} alone. ■

Lemma A.21 *Resorting to the same hypothesis and notation of Proposition A.17, the minimum number r of distinct eigenvectors $\{\hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2, \dots, \hat{\mathbf{v}}_r\}$ one can pick up in the set $\{\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \dots, \tilde{\mathbf{v}}_\ell\}$ so that* $\text{Cone}\left(\{\hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2, \dots, \hat{\mathbf{v}}_r\}\right) = \text{Cone}\left(\{\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \dots, \tilde{\mathbf{v}}_\ell\}\right)$ *coincides with the number of distinguished classes in $\mathcal{R}(A)$.*

Proof: Note that, given any index $i \in \{1, \dots, n\}$ such that $i \in \mathcal{C}_j$, the vector $\tilde{\mathbf{v}}_j$ represents the normalized asymptotic value of the ith column of e^{At}, namely

$$\tilde{\mathbf{v}}_j = \lim_{t \to +\infty} \frac{e^{At}\mathbf{e}_i}{\|e^{At}\mathbf{e}_i\|}. \tag{A.12}$$

Hence, by referring to the decomposition of the monomial vector $\mathbf{e}_i$ as the linear combination of (possibly generalized) eigenvectors of the Echelon basis $\mathcal{B}_e$, and to Lemma A.1, we can immediately conclude that each eigenvector $\tilde{\mathbf{v}}_j$ must be the linear combination of nonnegative eigenvectors of A, relative to possibly different distinguished classes, all sharing the same spectral radius

λ_j^*. It can be proved that whenever $\mathbf{v}_j$ turns out to be the linear combination of more than one nonnegative eigenvectors, all the non-zero coefficient of such combination are positive (see Lemma A.20). So we can remove all the linearly dependent vectors from $\{\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \ldots, \tilde{\mathbf{v}}_\ell\}$ thus showing that the cone generated by the set obtained by picking only one eigenvector for every distinguished class in $\mathcal{R}(A)$ is equivalent to $\mathrm{Cone}\left(\{\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \ldots, \tilde{\mathbf{v}}_\ell\}\right)$. ∎

Remark A.22 *By the previous comment we also deduce that a necessary and sufficient condition for* $\mathrm{Cone}_\infty(e^{At})$ *to be solid (namely to have n linearly independent generators) is that the influence graph associated with the matrix A has all classes which consist of a single vertex and each of them is distinguished.*

Lemma A.23 Let $\mathbf{v} \in \mathbb{R}^n_+$ be strictly positive and set

$$\mathbf{v}_{\min} := \min_{i=1,2,\ldots,n} [\mathbf{v}]_i > 0.$$

Let $A \in \mathbb{R}^{n\times n}_+$ be a nonsingular square matrix, with strictly positive diagonal entries, i.e. $[A]_{ii} \gg 0 \; \forall i \in \{1, \ldots, n\}$, and off-diagonal entries satisfying

$$[A]_{ij} \le \frac{v_{\min}}{\sqrt{n}\|A^{-1}\| \, \|v\|}, \qquad \forall\, i \ne j, \tag{A.13}$$

where $\|\cdot\|$ is the euclidean norm. Then $A^{-1}\mathbf{v} \ge 0$ or, equivalently, the equation $A\mathbf{x} = \mathbf{v}$ in the indeterminate x has a (uniquely determined) positive solution.

Proof: Since A is nonsingular, the equation $A\mathbf{x} = \mathbf{v}$ necessarily has a unique solution $\mathbf{x} = A^{-1}\mathbf{v}$. It only remains to show that $\mathbf{x} \in \mathbb{R}^n_+$. Note, first, that $\|\mathbf{x}\| = \|A^{-1}\mathbf{v}\| \le \|A^{-1}\| \, \|\mathbf{v}\|$. Now, $\forall\, j \in \{1, \ldots, n\}$, $[\mathbf{v}]_j = \sum_{k=1}^n [A]_{jk} \; [\mathbf{x}]_k$, and hence

$$[\mathbf{x}]_j = \frac{[\mathbf{v}]_j - ([A]_{j1}[\mathbf{x}]_1 + \ldots + [A]_{j\; j-1}[\mathbf{x}]_{j-1} + [A]_{j\; j+1}[\mathbf{x}]_{j+1} + \ldots + [A]_{jn}[\mathbf{x}]_n)}{[A]_{jj}}.$$

Consequently, $[\mathbf{x}]_j \ge 0 \Leftrightarrow ([\mathbf{v}]_j - ([A]_{j1}[\mathbf{x}]_1 + \ldots + [A]_{j\; j-1}[\mathbf{x}]_{j-1} + [A]_{j\; j+1}[\mathbf{x}]_{j+1} + \ldots + [A]_{jn}[\mathbf{x}]_n)) \ge 0$. Upon setting $A_{max} := \max\{[A]_{hk} \; : \; h \ne k\}$, and by resorting to the inequality $\sum_{i=1}^n |[\mathbf{x}]_i| \le \sqrt{n}\; \|\mathbf{x}\|$, after some computations one gets $[\mathbf{v}]_j - ([A]_{j1}[\mathbf{x}]_1 + \ldots + [A]_{j,j-1}[\mathbf{x}]_{j-1} + [A]_{j,j+1}[\mathbf{x}]_{j+1} + \ldots + [A]_{jn}[\mathbf{x}]_n) \ge v_{\min} - (A_{max} \sum_{i=1}^n |[\mathbf{x}]_i|) \ge v_{\min} - (A_{max}\sqrt{n}\|\mathbf{x}\|) \ge v_{\min} - (A_{max}\sqrt{n}\|A^{-1}\| \, \|\mathbf{v}\|) \ge 0$, and hence the claim is proved. ∎

Lemma A.24 *Given an $n \times n$ Metzler matrix A, for every $\varepsilon > 0$ there exists $t > 0$ such that $(1-\varepsilon)I_n \leq e^{At} \leq I_n + \varepsilon \mathbf{1}_n \mathbf{1}_n^T$, namely*

$$\begin{cases} (1-\varepsilon) \leq [e^{At}]_{ii} \leq (1+\varepsilon), & \\ 0 \leq [e^{At}]_{ij} \leq \varepsilon, & \text{for } i \neq j. \end{cases}$$

Proof: Being Metzler, A can be expressed as $A = -\alpha I + A_+$, for some $\alpha \geq 0$ and $A_+ \geq 0$. Thus, $[e^{At}]_{ij} = e^{-\alpha t}[e^{A_+ t}]_{ij} = e^{-\alpha t}\Big[I_n + A_+ t + A_+^2 \frac{t^2}{2!} + \dots\Big]_{ij}$. Finally, the continuity of the function e^{At}, together with the fact that $e^{A0} = I_n$ and $e^{A_+ t} - I_n \geq 0$ for $t > 0$, ensure that (A.24) holds for some $t > 0$. ∎

Lemma A.25 *Referring to the setting and notation of Theorem 9.1,* (9.20) *is always solvable.*

Proof: Let $w := \frac{1}{2}g_2^q - (M_{12}^q)^T x_{\text{pl}}(t_0) - M_{23}^q x_{\text{copy}}(t_0)$ and assume by contradiction that there is no solution $x_{\text{cn}}^*(t_0)$ to the equation $M_{22}^q x_{\text{cn}}^*(t_0) = w$, which means that $w \notin \operatorname{Im} M_{22}^q = (\operatorname{Ker} M_{22}^q)^\perp$, where $(\operatorname{Ker} M_{22}^q)^\perp$ denotes the orthogonal complement of $\operatorname{Ker} M_{22}^q$. In this case, there must exist a vector $z \in \operatorname{Ker} M_{22}^q$ such that $w^T z \neq 0$. If we then set $x_{\text{cn}}(t_0) = \alpha z$ in (9.19), we obtain $J = -2\alpha w^T z + *$. However, this would lead to a negative value for J for α sufficiently large, which leads to a contradiction since $J \geq 0$. ∎

Lemma A.26 *Referring to the setting and notation of Theorem 10.3,* (10.21) *is always solvable.*

Proof: The proof goes exactly along the lines of the proof of Lemma A.25. ∎

Bibliography

[1] & M. Saldago A. Feuer, G. Goodwin. Potential benefits of hybrid control for linear time invariant plants. In *Proc. of the American Control Conference*, pages 2790–2794, 1997.

[2] Egerstedt. M. & Wardi Y. Azuma, S. Output-based optimal timing control of switched systems. In *Hybrid Systems: Computation and Control*, pages 64–78. Springer-Verlag, 2006.

[3] Borrelli F. Morari M. Bemporad, A. Piecewise linear optimal controllers for hybrid systems. In *Proceedings of the 2000 American Control Conference*, pages 1190–1194, 2000.

[4] A. Berman and R.J. Plemmons. *Nonnegative matrices in the mathematical sciences.* Academic Press, New York (NY), 1979.

[5] V. D. Blondel and J. N. Tsitsiklis. When is a pair of matrices mortal? *Information Processing Letters*, 63, no. 5:283286, 1997.

[6] O. Bournez and M. S. Branicky. On matrix mortality in low dimensions. In V. D. Blondel et al., editor, *Open Problems in Mathematical Systems Theory*, pages 67–70. Springer and Verlag, 1999.

[7] & Barrat C. H. Boyd, S. P. *Linear controller design: Limits of performance.* New Jersey: Prentice Hall, 1991.

[8] Vandenberghe L. Boyd S. *Convex Optimization.* Cambridge University Press, 2004.

[9] R. Bru, S. Romero, and E. Sanchéz. Canonical forms for positive discrete-time linear control systems. *Linear Algebra & its Appl.*, 310:49–71, 2000.

[10] R.A. Brualdi and H.J. Ryser. *Combinatorial matrix theory.* Cambridge Univ.Press, Cambridge (GB), 1991.

[11] E. Carson and C. Cobelli. *Modelling Methodology for Physiology and Medicine*. Academic Press, San Diego, 2001.

[12] J. Clegg. A nonlinear integrator for servomechanisms. *Transactions A.I.E.E.*, 77:41–42, 1958.

[13] P.G. Coxson and H. Shapiro. Positive reachability and controllability of positive systems. *Lin. Alg. Appl.*, 94:35–53, 1987.

[14] Branicky M.S. Petterson S. & Lennartson B. De Carlo, R.A. Perspectives and results on the stabilizability of hybrid systems. *Proc. of the IEEE*, 88 (7):1069 –1082, 2000.

[15] Paganini F. Dullerud, G. *A course in robust control theory*. Spriger, 1999.

[16] M. Egerstedt and M. Babaali. On observability and reachability in a class of discrete-time switched linear systems. In *Proc. of the 2005 American Control Conf.*, pages 1179–1180, Portland, Orgeon, 2005.

[17] L. Farina and S. Rinaldi. *Positive linear systems: theory and applications*. Wiley-Interscience, Series on Pure and Applied Mathematics, New York, 2000.

[18] S. Friedland and H. Schneider. The growth of powers of a non-negative matrix. *SIAM J. Algebraic Discrete Methods*, 1:185–200, 1980.

[19] G.F. Frobenius.

[20] S.S. Ge, Z. Sun, and T.H. Lee. Reachability and controllability of switched linear discrete-time systems. *IEEE Trans. Aut. Contr.*, 46, no. 9:1437–1441, 2001.

[21] D. Hershkowitz, U.G. Rothblum, and H. Schneider. Characterizations and classifications of M-matrices using generalized nullspaces. *Linear Algebra and Appl.*, 109:59–69, 1988.

[22] D. Hershkowitz, U.G. Rothblum, and H. Schneider. The combinatorial structure of the generalized nullspace of a block triangular matrix. *Linear Algebra and Appl.*, 116:9–26, 1989.

[23] D. Hershkowitz and H. Schneider. On the generalized nullspace of M-matrices and Z-matrices. *Linear Algebra Appl.*, 106:5–23, 1988.

[24] & Morse A. S. Hespanha, J. P. Towards the high performance control of uncertain process via supervision. In *Proc. of the 30th annual conf. on information sciences and systems*, pages 405–410, 1996.

[25] J.P. Hespanha and A.S. Morse. Switching between stabilizing controllers. *Automatica*, 38:1905–1917, 2002.

[26] R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge Univ. Press, Cambridge (GB), 1985.

[27] I. Horowitz and P. Rosenbaum. Non-linear design for cost of feedback reduction in systems with large parameter uncertainty. *International Journal of Control*, 21:977–1001, 1975.

[28] J.A. Jacquez. *Compartmental analysis in biology and medicine*. Elsevier, Amsterdam (NL), 1972.

[29] H.D. Victory Jr. On nonnegative solutions to matrix equations. *SIAM J. Algebraic Discrete Methods*, 6:406–412, 1985.

[30] L.T. Conner Jr. and D.P. Stanford. State deadbeat response and observability in multi-modal systems. *SIAM J. Contr. Optimiz.*, 22(4):630–644, 1984.

[31] L.T. Conner Jr. and D.P. Stanford. The structure of the controllable set for multi-modal systems. *Linear Algebra & its Appl.*, 95:171–180, 1987.

[32] K. Krishnan and I. Horowitz. Synthesis of a non-linear feedback system with significant plant-ignorance for prescribed system tolerances. *International Journal of Control*, 19:689–706, 1974.

[33] D. Liberzon. *Switching in Systems and Control*. Birkhauser, Boston, 2003.

[34] D. Liberzon and A.S. Morse. Basic problems in stability and design of switched systems. *IEEE Contr. Syst. Magazine*, 19:59–70, 1999.

[35] & Antsaklis P. J. Lin, H. Stability and stabilizability of switched linear systems: a short survey of recent results. In *Proc. of the 2005 IEEE Int. Symp. on intelligent Control*, pages 24–29, Cyprus, 2005.

[36] & G. Goodwin M. Seron, J. Braslavsky. *Fundamental Limitations in Filtering and Control*. London: Springer-Verlag, 1997.

[37] H. Minc. *Nonnegative Matrices*. J.Wiley & Sons, New York, 1988.

[38] & Morse A.S. Pait, F.M. A cycling switching strategy for parameter-adaptive control. *IEEE Transaction on Automatic Control*, 39:1172–1183, 1994.

[39] B. Piccoli. Hybrid systems and optimal control. In *Proc. of the 37th IEEE Conf. on Decision and Control*, pages 13–18, 1998.

[40] C. Hollot Q. Chen and Y. Chait. Bibo stability of a class of reset control system. In *Proc. 2000 Conf. Information Sciences and Systems*, pages TP8–39, 2000.

[41] C. Hollot Q. Chen and Y. Chait. Stability and asymptotic performance analysis of a class of reset control systems. In *Proc. 39th Conf. Decis. Contr.*, pages 251–256, 2000.

[42] D. Richman and H. Schneider. On the singular graph and the Weyr characteristic of an M-matrix. *Aequationes Math.*, 17:208–234, 1978.

[43] U.G. Rothblum. Algebraic eigenspaces of non-negative matrices. *Linear Algebra Appl.*, 12:281–292, 1975.

[44] A. Salomaa and M. Soittola. *Automata theoretic aspects of formal power series*. Springer-Verlag, 1978.

[45] P. Santesso and M.E. Valcher. On the zero pattern properties and asymptotic behavior of continuous-time positive system trajectories. *Linear Algebra and its Applications*, 425:283–302, 2007.

[46] H. Schneider. The elementary divisors associated with 0 of a singular M-matrix. *Proc. Edinburgh Math. Sot. Ser. 2*, 10:108–122, 1956.

[47] H. Schneider. The influence of the marked reduced graph of a nonnegative matrix on the jordan form and on related properties: a survey. *Linear Algebra and Appl.*, 84:161–189, 1986.

[48] N.K. Son and D. Hinrichsen. Robust stability of positive continuous time systems. *Numer. Funct. Anal. and Optimiz.*, 17 (5 & 6):649–659, 1996.

[49] D.P. Stanford and L.T. Conner Jr. Controllability and stabilizability in multi-pair systems. *SIAM J. Contr. Optimiz.*, 18(5):488–497, 1980.

[50] G. A. Stewart, G. E. & Dumont. Finite horizon based switching between stabilizing controllers. In *Proc. American Control Conference*, pages 1550–1556, 2006.

[51] Z. Sun and D. Zheng. On reachability and stabilization of switched linear systems. *IEEE Trans. Aut. Contr.*, 46, no. 2:291–295, 2001.

[52] Ge S.S. Sun Z. *Switched Linear Systems: control and design.* Springer, 2005.

[53] M.E. Valcher. Controllability and reachability criteria for discrete time positive systems. *Int. J. of Control*, 65:511–536, 1996.

[54] Y. Wang, G. Xie, and L. Wang. Reachability of switched discrete-time systems under constrained switching. In *Proc. of the 42nd IEEE Conf. on Decision and Control*, pages 5765–5770, Maui, Hawaii, 2003.

[55] M.A. Wicks, P. Peleties, and R.A. De Carlo. Switched controller synthesis for the quadratic stabilization of a pair of unstable linear systems. *European J. of Control*, 4, no. 2:140–147, 1998.

[56] G. Xie and L. Wang. Reachability realization and stabilizability of switched linear discrete-time systems. *J. Math. Anal. Appl.*, 280:209–220, 2003.

[57] P.J. Xu, X. & Antsaklis. Optimal control of switched systems via nonlinear optimization based on direct differentiation of value functions. *Int. Journal of Control*, 75:1406–1426, 2002.

[58] X. Xu and P.J. Antsaklis. Optimal control of switched systems based on parametrization of the switching instants. *IEEE Transactions on Automatic Control*, 49, no.1:2–16, 2004.