

## Proper lumping for positive bilinear control systems

Questa è la versione preprint della seguente opera:

*Original*

Proper lumping for positive bilinear control systems / Jimenez-Pastor, A., Toller, D., Tribastone, M., Tschaikowski, M., Vandin, A.. - In: IEEE TRANSACTIONS ON AUTOMATIC CONTROL. - ISSN 0018-9286. - 69:11(2024), pp. 8120-8127. [10.1109/TAC.2024.3410153]

*Availability:*

This version is available at: 20.500.11771/41686

*Publisher:*

IEEE

*Published*

DOI:10.1109/TAC.2024.3410153

*Terms of use:*

This publication is made accessible in accordance with the terms for deposit in the institutional repository, as defined by the IMT School for Advanced Studies Lucca's Open Access Policy. ([https://library.imtlucca.it/sites/default/files/regolamento-policy-open-access-imtlib\\_0.pdf](https://library.imtlucca.it/sites/default/files/regolamento-policy-open-access-imtlib_0.pdf)).

Si prega di consultare le pagine informative dell'editore relative alle politiche di autoarchiviazione.

(Article begins on next page)

# Proper Lumping for Positive Bilinear Control Systems

A. Jiménez-Pastor<sup>1</sup>, D. Toller<sup>1</sup>, M. Tribastone<sup>2</sup>, M. Tschaikowski<sup>1</sup> and A. Vandin<sup>3,4</sup>

<sup>1</sup>Aalborg University, Denmark, <sup>2</sup>IMT Lucca, Italy, <sup>3</sup>Sant'Anna Pisa, Italy, <sup>4</sup>DTU Compute, Denmark

**Abstract**—Positive systems naturally arise in situations where the model tracks physical quantities. Although the linear case is well understood, analysis and controller design for nonlinear positive systems remain challenging. Model reduction methods can help tame this problem. Here we propose a notion of model reduction for a class of positive bilinear systems with (bounded) matrix and exogenous controls. Our reduction, called proper positive lumping, aggregates the original system such that states of the corresponding reduced model represent non-negative linear combinations of original state variables. We prove a characterization result showing that the reductions by proper positive lumping are exactly those preserving the optimality of a suitable class of value functions. Moreover, we provide an efficient polynomial-time algorithm for the computation of the minimal lumping. We numerically evaluate our approach by applying it to a number of benchmark case studies.

## I. INTRODUCTION

Positive systems find applications in many domains where models describe the evolution of variables that live in the non-negative orthant [1], [2], e.g., to track physical quantities [3]–[5]. In this paper we are concerned with controlled positive systems in the form  $\partial_t x(t) = A(t)x(t) + Bu(t)$  with an output map  $z(t) = Ox(t)$ , where  $A(\cdot)$  and  $u(\cdot)$  are controls subject to compact domains. This forms a subset of well-known bilinear control systems [6] with exogenous controls. Although positive linear systems have received considerable attention in the literature (e.g., [7] and references therein), fewer results are available for bilinear control systems. Despite some advances [8]–[10], problems like stability and controller design still remain challenging.

Model reduction methods can simplify those problems by projecting the original model onto a lower-dimensional one with theoretical guarantees of a relationship between the two. Model reduction has a long-standing tradition across many branches of science (e.g., [11]–[13]). Here we are concerned with a specific class of methods known as *proper lumping* (e.g., [14]): the projection is performed by identifying a positive orthogonal projection of the original state variables, and the reduced model tracks the state evolution of the projection. This ensures physically intelligible reductions [15]–[17] that preserve positiveness [13].

The vast majority of lumping algorithms share two main features: i) they do not take controls into account; ii) they consider reductions induced by partitions, where each partition block represents the *sum* of the original variables in that block [15], [18], [19]. In this paper, we develop a lumping method for positive linear systems with matrix and vector controls where each block represents a *non-negative linear combination* of variables; for these two reasons, this method is called *proper positive lumping*.

We define (constrained) proper lumping for a positive control system as a reduction that preserves a given output matrix  $O$  (known as the constraint) and that is a proper lumping for the extremal matrices  $\underline{A}$  and  $\overline{A}$  describing the matrix domain  $A(\cdot) \in [\underline{A}; \overline{A}]$ . We show that the reduced model preserves optimality with respect to the original one under a suitable class of costs. In particular, we provide a result that allows the reconstruction of the original optimal controls from the reduced ones, thus circumventing the original optimal control problem altogether. Importantly, we also give a characterization result: we prove that a reduction is optimality-preserving for the control system if and only if it is a proper lumping of its extremal matrices  $\underline{A}$  and  $\overline{A}$ .

Hence, applying our proper lumping technique, we can compute the minimal control system necessary to obtain the output map  $z(t)$  and perform optimal computations on the reduced system to, then, reconstruct the optimal controls on the original system.

Another important contribution of this paper is of algorithmic nature. We show that the minimal proper lumping, i.e., the one projecting onto the smallest space, can be computed in polynomial time with respect to the dimensionality of the model. The algorithm computes a common minimal proper lumping of the extremal matrices  $\underline{A}$  and  $\overline{A}$  by building on *CLUE* [20], a recent method for lumping ordinary differential equations with polynomial right-hand sides (without controls).

Using CLUE, we extensively test proper lumping on over 1000 models of dynamics on networks taken from a public repository. We show that the notion can be useful in practice as it can reduce at least half of the examined models, with at least a 65% reduction ratio for over 30% of the cases.

*Related work.* Our approach generalizes a recent work that considers uncertain continuous-time Markov chains [21], which are a special case of positive linear systems with interval uncertainties. In addition, the lumping algorithm in [21] only considers blocks that represent sums of variables, unlike our more general case of non-negative linear combinations.

In the field of model reduction, the notions most close to us are those of constrained lumping [20], [22] and consistent abstraction, also known as bisimulation [23]–[26]. Our notion complements those approaches. Indeed, while lumping has been considered for uncontrolled dynamical systems only, the computational aspects of bisimulation have been investigated in full only in the case of additive controls [23]. The nonlinear counterparts [24], [25], instead, provide a mathematical characterization of the bisimulation quotient (via tangential spaces and manifold intersections) but leave the computational aspects unaddressed. Less closely related are [12], [27]–[29]

or [30], [31] as these either do not allow for matrix controls or study network properties.

## II. PRELIMINARIES

*Notation:* We denote by  $A$  a real  $n \times n$  matrix; instead,  $L$  usually denotes a  $k \times n$  matrix with full row rank, whereas  $L^+ = L^T(LL^T)^{-1}$  denotes its generalized right inverse, satisfying  $LL^+ = I_k$  (the  $k \times k$  identity matrix). The row space of  $L$  is denoted by  $\text{rs}(L)$ , while the support of a vector  $v$  is  $\text{supp}(v) = \{i \mid v_i \neq 0\}$ . A matrix function  $A = A(\cdot)$  is called admissible if it is piecewise continuous with right limits and it satisfies  $A(\cdot) \in [\underline{A}; \bar{A}]$  with respect to matrix bounds  $\underline{A} \leq \bar{A}$ , where  $\leq$  is meant component-wise. The technical assumption on right limits is necessary to establish later our characterization results. We often denote by  $x^{(A,u)}$  the solution of  $\partial_t x(t) = A(t) \cdot x(t) + B \cdot u(t)$ .

We begin by reviewing constrained lumping [20], [22].

**Definition 1** (Constrained Lumping). *Fix a linear system with state matrix  $A \in \mathbb{R}^{n \times n}$  and output matrix  $O \in \mathbb{R}^{l \times n}$ :*

$$\partial_t x(t) = A \cdot x(t), \quad z(t) = O \cdot x(t). \quad (1)$$

*A matrix  $L \in \mathbb{R}^{k \times n}$  with rank  $k \leq n$  is called a constrained lumping of (1) if  $\text{rs}(O) \subseteq \text{rs}(L)$  and  $\text{rs}(LA) \subseteq \text{rs}(L)$ .*

In the above setting, we also say that  $L$  is a (constrained) lumping of  $A$  with respect to  $O$ .

**Example 1.** *Consider the linear system*

$$\partial_t x(t) = A \cdot x(t), \quad z(t) = O \cdot x(t),$$

where  $A = \begin{pmatrix} 0 & 2 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$ , and  $O = (0 \ 0 \ 1)$ . Then the matrix  $L = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$  is a constrained lumping.

**Definition 2** (Aggregation and Reduced System). *Fix a matrix  $L \in \mathbb{R}^{k \times n}$  with rank  $k \leq n$ .*

- *The function  $\mathbb{R}^n \rightarrow \mathbb{R}^k$ , mapping  $x \mapsto Lx$ , induced by  $L$  is called the aggregation (map), and  $y = Lx$  are called the aggregated variables.*
- *When  $\text{rs}(O) \subseteq \text{rs}(L)$ , setting  $R = LAL^+$ , we call*

$$\partial_t y(t) = R \cdot y(t), \quad z(t) = OL^+ \cdot y(t) \quad (2)$$

*the reduced system of (1).*

The condition of lumping was recently proved [20] to be equivalent to the following property, motivating our interest in them and the definition of reduced system.

**Proposition 1.** *If  $L \in \mathbb{R}^{k \times n}$  is of rank  $k$  with  $\text{rs}(O) \subseteq \text{rs}(L)$ , then  $L$  is a lumping of (1) if and only if, for any solution  $x$  of (1),  $y = Lx$  is a solution of (2).*

**Remark 1.** *The choice of the output matrix for the reduced system in (2) is motivated by the fact that  $\text{rs}(O) \subseteq \text{rs}(L)$  can be shown to be equivalent to  $O = OL^+L$ , and then*

$$z = Ox = OL^+Lx = OL^+y,$$

*so the original system and the reduced one have the same output values  $z$ .*

**Example 2.** *Continuing Example 1, one has*

$$L^+ = \begin{pmatrix} \frac{1}{2} & 0 \\ \frac{2}{3} & 0 \\ 0 & 1 \end{pmatrix}, \quad LAL^+ = \begin{pmatrix} 2 & 2 \\ 0 & 1 \end{pmatrix}, \quad \text{and } OL^+ = (0 \ 1).$$

*The aggregated variables are  $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 + 2x_2 \\ x_3 \end{pmatrix}$  and the reduced system is*

$$\begin{pmatrix} \partial_t y_1 \\ \partial_t y_2 \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \quad z = (0 \ 1) \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}. \quad (3)$$

Given an output  $O$ , a constrained lumping  $L \in \mathbb{R}^{k \times n}$  with minimal row space  $\text{rs}(L)$  exists and can be computed in polynomial time [20]. We will make use of this result in Section VI.

The matrix  $L$  in Example 1 has non-negative entries, and its rows have pairwise disjoint supports. Matrices with such properties have been studied in the literature [14]; they are fundamental to our work, so we introduce them in the following definition.

**Definition 3** (Proper Matrix and Proper Lumping). *We say a matrix  $L \in \mathbb{R}_{\geq 0}^{k \times n}$  is proper if it has rank  $k \leq n$ , and its rows have pairwise disjoint supports.*

*If  $L$  is proper, and it is a constrained lumping of (1), we call  $L$  a constrained proper lumping of (1) (or simply proper lumping when this creates no ambiguity).*

For a matrix  $L \in \mathbb{R}^{k \times n}$  with rank  $k \leq n$ , the aggregation map  $\mathbb{R}^n \rightarrow \mathbb{R}^k$  is surjective, as  $L$  has full row rank by definition, and transforms the  $n$ -dimensional problem (1) to the  $k$ -dimensional one (2). If  $L$  has any zero columns, then  $L$  essentially deletes the corresponding state variables, as they do not appear in the aggregation  $Lx$ ; obviously  $L$  induces an aggregation of the remaining variables. Each of the rows of  $L$  represents a state variable for the reduced system, obtained as the linear combination of the original state variables using that row of coefficients. The additional algebraic property of  $L$  to be proper translates to the fact that such linear combinations are taken over disjoint subsets of variables, so each original variable influences at most one variable of the reduced system.

## III. LUMPING POSITIVE CONTROL SYSTEMS

We begin by introducing the notion of Positive Controlled Systems (PCS). Following the literature [10], the system (1) is called positive if it satisfies the following: for every non-negative initial condition  $x[0] \in \mathbb{R}_{\geq 0}^n$ , the solution  $x(\cdot)$  of (1) remains non-negative, i.e.  $x(t) \in \mathbb{R}_{\geq 0}^n$  for every  $t \geq 0$ . It is well known that this holds if and only if  $A_{i,j} \geq 0$  for every  $1 \leq i \neq j \leq n$ , and a matrix with this property is called Metzler matrix. Adhering to this setting, we thus consider in the following Metzler matrices, non-negative input matrices  $B$ , and non-negative control inputs  $u$ .

**Definition 4** (Positive Controlled System, PCS). A PCS, denoted by  $([\underline{A}; \bar{A}], O, B, U)$ , is given by

$$\partial_t x(t) = A(t) \cdot x(t) + B \cdot u(t), \quad z(t) = O \cdot x(t) \quad (4)$$

where  $O \in \mathbb{R}^{l \times n}$  is the output matrix,  $B \in \mathbb{R}_{\geq 0}^{n \times m}$ , and

- $\underline{A}, \bar{A} \in \mathbb{R}^{n \times n}$  are Metzler matrices with  $\underline{A} \leq \bar{A}$ ;
- $U \subseteq \mathbb{R}_{\geq 0}^m$  is compact and contains the origin;
- $A(\cdot) \in [\underline{A}; \bar{A}]$  and  $u(\cdot) \in U$  are admissible.

We denote by  $x[0] \in \mathbb{R}_{\geq 0}^n$  the initial condition of the PCS.

As  $\underline{A}$  is Metzler, all admissible matrices  $A(\cdot) \geq \underline{A}$  are Metzler too, and so given  $x[0] \in \mathbb{R}_{\geq 0}^n$  all solutions to (4) remain in  $\mathbb{R}_{\geq 0}^n$ . We remark that PCSs constitute a family of bilinear control systems [6] with exogenous linear controls, i.e., both the matrix  $A(t)$  and  $u(t)$  are control variables. Being interested in the optimal control of a PCS, we next introduce the value function associated to it [32].

**Definition 5** (Value Function). For time  $\tau \in \mathbb{R}_{\geq 0}$  and vector  $x[0] \in \mathbb{R}_{\geq 0}^n$  describing the initial condition of (4), we define

- for any  $A(\cdot) \in [\underline{A}; \bar{A}]$  and  $u(\cdot) \in U$ , the cost as

$$J_{(A,u)}(x[0], \tau) = \Phi(x^{(A,u)}(\tau), u(\tau)) + \int_0^\tau \Psi(t, x^{(A,u)}(t), u(t)) dt,$$

where  $x^{(A,u)}$  is the solution of (4) with given  $A$  and  $u$ , the continuous function  $\Psi : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$  is called running cost, and the continuous function  $\Phi : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$  is called final cost.

- the optimal values as

$$V^{\inf}(x[0], \tau) = \inf \{ J_{(A,u)}(x[0], \tau) \mid A \in [\underline{A}; \bar{A}], u \in U \},$$

$$V^{\sup}(x[0], \tau) = \sup \{ J_{(A,u)}(x[0], \tau) \mid A \in [\underline{A}; \bar{A}], u \in U \}.$$

The compactness of  $[\underline{A}; \bar{A}]$  and  $U$  ensure the well-definedness of the value functions. While we consider costs  $\Psi, \Phi$  that do not depend directly on  $A$ , we argue that this is a mild restriction because the control set  $[\underline{A}; \bar{A}]$  is bounded. We note in particular that we can encode reachability, bounded affine controls  $(A, u)$  and costs quadratic in state  $x$  and control  $u$ , as studied in the linear quadratic regulator [32].

We next introduce the main concept of the present work, by lifting the classical notion of constrained lumping to a PCS. Due to the non-negativity of the solutions of a PCS, we consider aggregation maps given by non-negative matrices.

**Definition 6** (Constrained Proper Positive Lumping). Fix a PCS  $([\underline{A}; \bar{A}], O, B, U)$ .

- We call  $L$  a constrained proper positive lumping of the PCS if  $L \in \mathbb{R}_{\geq 0}^{k \times n}$  is proper and is a constrained lumping of  $\underline{A}, \bar{A}$  with respect to  $O$ .
- The reduced PCS is  $([L\underline{A}L^+; L\bar{A}L^+], OL^+, LB, U)$ , with initial condition  $y[0] = Lx[0] \in \mathbb{R}_{\geq 0}^k$ .

Often, we simply call proper lumping a matrix  $L$  as above.

**Example 3.** Consider the PCS that arises by perturbing  $A$  from Example 1 by a factor of 10%, yielding  $\underline{A} = 0.9A$

and  $\bar{A} = 1.1A$ . Let  $O = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}$ , and consider the PCS  $([\underline{A}; \bar{A}], O, 0, \{0\})$ . Since  $L$  from Example 1 is a constrained proper lumping for both  $\underline{A}$  and  $\bar{A}$ , it is constrained proper lumping for the above PCS. The reduced PCS is given by

$$L\underline{A}L^+ = 0.9 \begin{pmatrix} 2 & 2 \\ 0 & 1 \end{pmatrix} \quad L\bar{A}L^+ = 1.1 \begin{pmatrix} 2 & 2 \\ 0 & 1 \end{pmatrix}$$

$$OL^+ = \begin{pmatrix} 0 & 1 \end{pmatrix} \quad LB = 0.$$

Next we introduce the natural candidates for cost functions and optimal values of the reduced PCS. To this end, note that given  $L \in \mathbb{R}^{k \times n}$  one has  $L^+ \in \mathbb{R}^{n \times k}$ , so its associated linear map is  $L^+ : \mathbb{R}^k \rightarrow \mathbb{R}^n$ .

**Definition 7** (Reduced optimal values and cost preservation). Fix a PCS and a matrix  $L \in \mathbb{R}^{k \times n}$  of rank  $k \leq n$ .

- Final and running costs  $\Phi, \Psi$  are called  $L$ -invariant if  $\Phi(x, u) = \Phi(x', u)$  and  $\Psi(t, x, u) = \Psi(t, x', u)$  whenever  $x, x' \in \mathbb{R}^n$  satisfy  $Lx = Lx'$ . In this case, we consider the reduced final and running costs  $\hat{\Phi} : \mathbb{R}^k \times \mathbb{R}^m \rightarrow \mathbb{R}$  and  $\hat{\Psi} : \mathbb{R} \times \mathbb{R}^k \times \mathbb{R}^m \rightarrow \mathbb{R}$  defined by

$$\hat{\Phi}(y, u) = \Phi(L^+y, u) \quad \text{and} \quad \hat{\Psi}(t, y, u) = \Psi(t, L^+y, u).$$

Using them, the values for the reduced PCS are as in Definition 5, and we denote them by  $\hat{V}^{\inf}$  and  $\hat{V}^{\sup}$ .

- We say that  $L$  preserves the costs if for any  $x[0] \in \mathbb{R}_{\geq 0}^n$ , it holds that  $y[0] = Lx[0]$  satisfies

$$V^{\inf}(x[0], \tau) = \hat{V}^{\inf}(y[0], \tau),$$

$$V^{\sup}(x[0], \tau) = \hat{V}^{\sup}(y[0], \tau)$$

for every  $\tau \in \mathbb{R}_{\geq 0}$  and for every final cost  $\Phi$  and running cost  $\Psi$  that are  $L$ -invariant.

We remark that the above definitions are well-posed as the matrix  $L$  has full row rank, and both  $\hat{\Phi}$  and  $\hat{\Psi}$  are continuous.

## IV. MAIN RESULTS

We show first that a constrained proper lumping of a PCS preserves the family of its solutions. As a consequence, it also preserves its costs.

**Theorem 1** (Cost Preservation). Assume that  $L \in \mathbb{R}_{\geq 0}^{k \times n}$  is a constrained proper lumping of the PCS  $([\underline{A}; \bar{A}], O, B, U)$ . Then for any initial condition  $x[0] \in \mathbb{R}_{\geq 0}^n$ , and any  $u(\cdot) \in U$ :

- for any  $A(\cdot) \in [\underline{A}; \bar{A}]$ , there is  $R(\cdot) \in [L\underline{A}L^+; L\bar{A}L^+]$  such that  $Lx^{(A,u)} = y^{(R,u)}$ ;
- for any  $R(\cdot) \in [L\underline{A}L^+; L\bar{A}L^+]$ , there is  $A(\cdot) \in [\underline{A}; \bar{A}]$  such that  $Lx^{(A,u)} = y^{(R,u)}$ .

In particular,  $L$  preserves the costs of the PCS.

In the next result we give the converse to Theorem 1, thus providing a complete characterization of a constrained proper lumping in terms of preservation of costs.

**Theorem 2** (Cost Characterization). Let  $L \in \mathbb{R}_{\geq 0}^{k \times n}$  be a proper matrix. If  $\text{rs}(O) \subseteq \text{rs}(L)$  and  $L$  preserves the costs of a PCS  $([\underline{A}; \bar{A}], O, B, U)$ , then  $L$  is a constrained proper lumping of the PCS.

### A. Control reconstruction and proofs of Theorem 1 and 2

In this section we give some insight about the results previously presented.

**Definition 8.** Given a proper matrix  $L \in \mathbb{R}_{\geq 0}^{k \times n}$ , for every integer  $1 \leq r \leq k$ , let us introduce the disjoint subsets of indexes (called blocks)  $H_r = \text{supp}(r\text{-th row of } L)$ , and  $H_0 = \{i \mid \text{the } i\text{-th column of } L \text{ is zero}\}$ . Finally, for every  $1 \leq i \leq n$ , let us denote by  $[i]$  the unique set of the above such that  $i \in H_{[i]}$ .

In the above terminology, note that  $H_0$  is the complement of the disjoint union  $\bigcup_{r=1}^k H_r$ . A proper lumping deletes the state variables indexed by  $H_0$  and aggregates the remaining variables indexed by  $\bigcup_{r=1}^k H_r$ . Then for every  $x \in \mathbb{R}^n$  the aggregation  $Lx$  has the form

$$Lx = \left( \sum_{i \in H_1} \lambda_i x_i, \dots, \sum_{i \in H_k} \lambda_i x_i \right)^T$$

for positive weights  $\lambda_i$ .

Proposition 2 below formalizes this discussion, by ensuring that  $L$  can be written as a product  $CD$ , where  $C$  encodes the blocks, while  $D$  captures the weight associated by  $L$  to each variable. Moreover, we identify the matrices of this form that are lumpings.

**Proposition 2.** If  $L \in \mathbb{R}_{\geq 0}^{k \times n}$  is a proper matrix, then it can be written as  $L = CD$ , where

- $C \in \{0, 1\}^{k \times n}$  is given by  $C_{r,i} = 1$  if  $L_{r,i} > 0$  and  $C_{r,i} = 0$  otherwise, for any  $1 \leq r \leq k$  and  $1 \leq i \leq n$ ;
- $D = \text{diag}(\lambda_i)_{1 \leq i \leq n} \in \mathbb{R}_{> 0}^{n \times n}$  is the diagonal matrix whose  $(i, i)$ -th entry  $\lambda_i$  is one of the positive entries of  $L$  if  $i \notin H_0$ , while  $\lambda_i = 1$  otherwise.

Given matrices  $A$  and  $O$ ,  $L$  is a constrained proper lumping of  $A$  (wrt  $O$ ) if and only if  $C$  is a constrained proper lumping of  $DAD^{-1}$  (wrt  $OD^{-1}$ ).

A proper matrix  $L$  and the matrix  $C$  introduced in Proposition 2 define the same blocks, introduced as in Definition 8. Indeed, for every  $r \leq k$ , the  $r$ -th row of  $L$  and the  $r$ -th row of  $C$  have the same support. So there is no ambiguity if we write  $[i]$  for the block containing an index  $i$ .

The next results can be proven by generalizing the proof for uncertain continuous-time Markov chains (UCTMCs) from [21] and by invoking Proposition 2. According to [21], a UCTMC is a continuous-time Markov chain whose transition rates from state  $j$  into state  $i$  are subject to the constraint  $A_{i,j}(\cdot) \in [\underline{A}_{i,j}; \overline{A}_{i,j}]$ .

Moreover, while [21] confined itself to lumpings  $C \in \{0, 1\}^{k \times n}$  that encode blocks, Proposition 2 allows us to consider proper lumpings  $L$  by expressing these as suitably stretched versions of aforementioned matrices  $C$ , i.e.,  $L = CD$ .

**Proposition 3** (Cost Preservation – special case). Fix a PCS  $([\underline{A}; \overline{A}], O, B, U)$ , and a proper matrix  $C \in \{0, 1\}^{k \times n}$ . Then the following conditions are equivalent:

- 1)  $C$  is a proper lumping of both  $\underline{A}$  and  $\overline{A}$ ;
- 2)  $C$  preserves the costs of  $([\underline{A}; \overline{A}], O, B, U)$ ;

3) for any initial condition  $x[0] \in \mathbb{R}_{\geq 0}$  and  $u(\cdot) \in U$ :

- i) Any  $A(\cdot) \in [\underline{A}; \overline{A}]$  has an  $R(\cdot) \in [C\underline{A}C^+; C\overline{A}C^+]$  such that  $Cx^{(A,u)} = y^{(R,u)}$ ,
- ii) Any  $R(\cdot) \in [C\underline{A}C^+; C\overline{A}C^+]$  has an  $A(\cdot) = (A_{i,j}(\cdot))_{i,j} \in [\underline{A}; \overline{A}]$ , defined as follows, such that  $Cx^{(A,u)} = y^{(R,u)}$  and  $C$  is a lumping for  $A$ . Let  $\underline{R} = C\underline{A}C^+$ , and for  $i, j$  with  $[i], [j] \neq 0$ , we define

$$A_{i,j}(t) = \underline{A}_{i,j} + \frac{(\overline{A}_{i,j} - \underline{A}_{i,j})(R_{[i],[j]}(t) - \underline{R}_{[i],[j]})}{\sum_{k \in [i]} (\overline{A}_{k,j} - \underline{A}_{k,j})}$$

provided the division is well-defined; in all other cases, we set  $A_{i,j}(t) = \underline{A}_{i,j}$ .

The proofs of Theorems 1 and 2 consist in reducing the general case to the case considered in Proposition 3. Additionally, here follows separately a more detailed statement about how to recover optimal controls for the original system (from ones of the reduced system), providing an explicit formula for them.

**Proposition 4** (Control Reconstruction). Assume that  $L$  is a proper positive lumping of the PCS  $([\underline{A}; \overline{A}], O, B, U)$ . Write  $L = CD$  according to Proposition 2, so that  $C$  is a lumping of both  $\underline{M} = D\underline{A}D^{-1}$  and  $\overline{M} = D\overline{A}D^{-1}$ , and let  $\underline{R} = L\underline{A}L^+$ .

Then, for any  $R(\cdot) \in [L\underline{A}L^+; L\overline{A}L^+]$ , we can define  $A(\cdot) = (A_{i,j}(\cdot))_{i,j} \in [\underline{A}; \overline{A}]$  as follows. If  $[i], [j] \neq 0$ , set

$$A_{i,j}(t) = \lambda_i^{-1} \lambda_j \left( \underline{M}_{i,j} + \frac{(\overline{M}_{i,j} - \underline{M}_{i,j})(R_{[i],[j]}(t) - \underline{R}_{[i],[j]})}{\sum_{k \in [i]} (\overline{M}_{k,j} - \underline{M}_{k,j})} \right)$$

provided the division is well-defined; in all other cases, we set  $A_{i,j}(t) = \lambda_i^{-1} \lambda_j \underline{M}_{i,j}$ .

Then  $L$  is a lumping for  $A(\cdot)$  and  $R(\cdot) = LA(\cdot)L^+$ , so  $Lx^{(A,u)} = y^{(R,u)}$  for any  $x[0] \in \mathbb{R}_{\geq 0}^n$  and any  $u(\cdot) \in U$ .

## V. COMPUTATION OF CONSTRAINED PROPER LUMPING

This section studies the computation of PCS lumpings. We recall that finding a PCS lumping amounts to finding a proper lumping of the linear dynamical systems  $\partial_t x = \underline{A}x(t)$  and  $\partial_t x = \overline{A}x(t)$ , constrained with respect to an output matrix  $O$ . This can be achieved by making use of the efficient algorithm CLUE from [20] which computes the minimal constrained lumping  $L$  for a system of polynomial differential equations

$$\partial_t x(t) = p(x(t)), \quad z(t) = O \cdot x(t). \quad (5)$$

Indeed, as shown in [20],  $L$  is a constrained lumping of (5) if and only if  $L$  is a constrained lumping of a certain family of linear dynamical systems that arise from the total derivative of  $p$ . Feeding CLUE with  $\underline{A}$  and  $\overline{A}$  instead those linear systems yields then the following.

**Proposition 5** ([20]). For a PCS  $([\underline{A}; \overline{A}], O, B, U)$ :

- Applying CLUE to  $\underline{A}, \overline{A}$  gives a constrained lumping  $L \in \mathbb{R}^{k \times n}$  of both  $\underline{A}$  and  $\overline{A}$  (wrt  $O$ ), and with row space dimension  $k$ ;
- such  $k$  is minimal satisfying all the above properties;
- the matrix  $L$  is in reduced echelon-form;

- the computation of  $L$  requires at most  $\mathcal{O}(kn(e+k))$  steps, where  $e$  is the total number of non-zero entries in  $\underline{A}$ ,  $\bar{A}$  and  $O$ .

**Example 4.** For the PCS in Example 3, CLUE produces the lumping  $L = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ . All entries of  $L$  are non-negative and the supports of its rows are disjoint, hence  $L$  is a constrained proper positive lumping of the PCS.

In general,  $L$  from Proposition 5 may have negative entries, hence it may not be a constrained proper lumping. The following result certifies that  $L$  is the only candidate to check for being a constrained proper lumping with minimum row dimension.

**Theorem 3.** Fix a PCS  $([\underline{A}; \bar{A}], O, B, U)$ , and let  $L \in \mathbb{R}^{k \times n}$  be the result of CLUE as in Proposition 5 for  $\underline{A}$  and  $\bar{A}$  with output map  $O$ . If  $L$  is not proper, then there is no constrained proper lumping with row space dimension  $k$ .

The above result ensures that, in order to compute a constrained proper lumping, we can apply the algorithm from [20] to get a matrix  $L$ . If  $L$  is a constrained proper lumping for  $[\underline{A}; \bar{A}]$ , we terminate; otherwise, we are guaranteed that there is no proper lumping with the same row space of  $L$ .

**Remark 2.** In all models in Section VI, if the matrix  $L$  from Proposition 5 has rows with disjoint supports, then it also has non-negative entries, and so it is proper.

## VI. EXPERIMENTAL EVALUATION

We implement PCS proper lumping based on CLUE [20]. The implementation can be found in the *cpl* branch (see <https://github.com/Antonio-JP/CLUE/tree/cpl>). All examples are available in the *tests/examples/cpl* folder in the aforementioned repository. This repository also include the scripts necessary to run all the experiments described in this article. Reported experiments refer to a common laptop (Lenovo with 2.8GHz i7 processor) with 23 GB RAM and not exploiting parallelization.

**Set up.** The models have been taken from the repository of (positive) weighted networks in <https://networks.skewed.de/>. We have considered networks in the repository up to 100000 vertices. For each of them we have considered a set of PCS in the form  $([0.9A; 1.1A], O_i, 0, \{0\})$  where  $A$  is the weighted adjacency matrix of the network and  $O_i = e_i^T$  for all  $i$  such that the  $i$ -th row of  $A$  is not zero (i.e., the variables that are not constant in the system). Similarly to [21], this setting creates interval uncertainties of  $\pm 10\%$  around the original values of  $A$ .

**Results.** We summarize the results across all matrices  $L$  computed by Proposition 5 on the considered PCS as follows:

- *Proper lumping:* the obtained lumping is a constrained proper lumping with respect to the given  $O_i$  and it reduces the size of the original system.
- *General lumping:* the obtained lumping is not proper.
- *No reduction:* the obtained lumping does not provide any reduction, i.e., the lumping is the identity matrix.

- *Timeout examples:* in all experiments, we imposed an arbitrary timeout of 10 minutes. With this, more than 96% of the models could be processed.

In Table I we summarize the obtained results, by splitting them into the different cases described above. From this table we can see that our method usually finds a proper lumping (60.8% of the cases). It is interesting to remark that in the remaining cases, for the majority of them (35.5%) we obtain a non-proper lumping. These reductions are not covered by the theory of the current paper. Only the 3.7% of cases were unsuccessful, either not finding any reduction of the models, or hitting the timeout of 10 minutes.

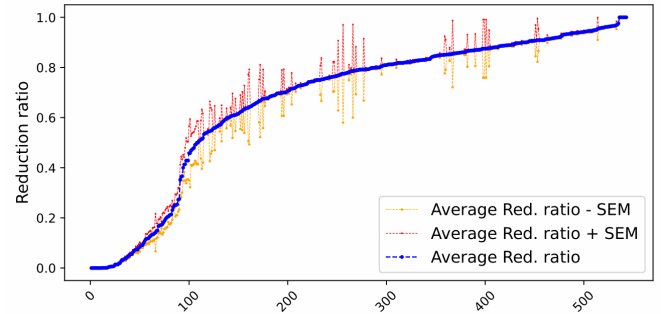


Fig. 1. Average reduction ratio for models with some proper lumpings

In Figure 1 we show, for each model, the average reduction ratio (i.e., number of reduced species over number of original ones in the system) obtained through all possible output matrices  $O$ . In this figure we have only considered the models where there is at least one case with a proper lumping, and removing all cases with general lumpings (i.e., with a reduction we can not interpret). We also considered the cases with no reduction. The standard error of the mean (SEM) is also included for each model.

From this figure we can see that for more than half of the networks (around 600 models) have at least one case with a proper lumping. Specifically, around 30% of models have a reduction ratio of 65% or less, thus providing a significant evidence of the usefulness of the proposed method.

## VII. CONCLUSION

In the paper we introduced constrained proper lumping, a model reduction technique for a class of bilinear control systems described by positive linear control systems (PCS) whose matrices take the role of control inputs. Our main result extends [21] and ensures that a so-called proper aggregation map  $x \mapsto Lx$  is optimality-preserving if and only if the matrix constraints of the PCS  $A(t) \in [\underline{A}; \bar{A}]$  are such that  $L$  is a bisimulation (aka lumping) [22], [23] of the *uncontrolled* dynamical systems  $\partial_t x = \underline{A}x$  and  $\partial_t x = \bar{A}x$ . Building upon [20], this allowed for the efficient computation of minimal proper lumpings with respect to given output maps [23], [24]. The practical applicability of the framework was demonstrated on a number of benchmark case studies. Future work will consider nonlinear reductions.

| No. of PCS (no. of models) | Proper        | General       | No reduction | Timeout     |
|----------------------------|---------------|---------------|--------------|-------------|
| 110808 (1352)              | 67371 (60.8%) | 39318 (35.5%) | 2495 (2.2%)  | 1624 (1.5%) |

TABLE I  
SUMMARY OF RESULTS FROM EXPERIMENTS

## REFERENCES

- [1] L. Farina and S. Rinaldi, *Positive linear systems: theory and applications*. John Wiley & Sons, 2000, vol. 50.
- [2] H. L. Smith, *Monotone Dynamical Systems: An Introduction to the Theory of Competitive and Cooperative Systems*. American Mathematical Society, Mar. 2008, no. 41.
- [3] R. Shorten, F. Wirth, and D. Leith, “A positive systems model of TCP-like congestion control: asymptotic results,” *IEEE/ACM transactions on networking*, vol. 14, no. 3, pp. 616–629, 2006.
- [4] W. M. Haddad, V. Chellaboina, and Q. Hui, *Nonnegative and compartmental dynamical systems*. Princeton University Press, 2010.
- [5] T. Kaczorek, “Positive electrical circuits and their reachability,” *Archives of Electrical Engineering*, vol. 60, no. 3, pp. 283–301, 2011.
- [6] G. Pappas and S. Simic, “Consistent abstractions of affine control systems,” *IEEE TAC*, vol. 47, no. 5, 2002.
- [7] A. Rantzer and M. E. Valcher, “A tutorial on positive systems and large scale control,” in *2018 IEEE Conference on Decision and Control (CDC)*, 2018, pp. 3686–3697.
- [8] F. Amato, M. Ariola, and P. Dorato, “Finite-time control of linear systems subject to parametric uncertainties and disturbances,” *Automatica*, vol. 37, no. 9, pp. 1459–1463, Sept. 2001.
- [9] M. A. Rami and F. Tadeo, “Controller synthesis for positive linear systems with bounded controls,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 54, no. 2, pp. 151–155, 2007.
- [10] E. Weiss and M. Margaliot, “A generalization of linear positive systems with applications to nonlinear systems: Invariant sets and the Poincaré-Bendixson property,” *Automatica*, vol. 123, p. 109358, 2021.
- [11] M. Aoki, “Control of large-scale dynamic systems by aggregation,” *IEEE Trans. Automat. Contr.*, vol. 13, no. 3, pp. 246–253, 1968.
- [12] A. C. Antoulas, *Approximation of large-scale dynamical systems*. SIAM, 2005.
- [13] T. J. Snowden, P. H. van der Graaf, and M. J. Tindall, “Methods of model reduction for large-scale biological systems: A survey of current methods and trends,” *Bull. Math. Biol.*, vol. 79, no. 7, 2017.
- [14] M. S. Okino and M. L. Mavrouniotis, “Simplification of mathematical models of chemical reaction systems,” *Chemical Reviews*, vol. 2, no. 98, pp. 391–408, 1998.
- [15] L. Cardelli, M. Tribastone, M. Tschaikowski, and A. Vandin, “Maximal aggregation of polynomial dynamical systems,” *Proceedings of the National Academy of Sciences*, vol. 114, no. 38, 2017.
- [16] —, “Guaranteed Error Bounds on Approximate Model Abstractions Through Reachability Analysis,” in *QEST*, 2018, pp. 104–121.
- [17] —, “Symbolic computation of differential equivalences,” *Theor. Comput. Sci.*, vol. 777, pp. 132–154, 2019.
- [18] S. Derisavi, H. Hermans, and W. H. Sanders, “Optimal state-space lumping in Markov chains,” *Information Processing Letters*, vol. 87, no. 6, pp. 309–315, 2003.
- [19] A. Dokoumetzidis and L. Aarons, “Proper lumping in systems biology models,” *IET Systems Biology*, vol. 3, no. 1, pp. 40–51, 2009.
- [20] A. Ovchinnikov, I. Pérez Verona, G. Pogudin, and M. Tribastone, “CLUE: exact maximal reduction of kinetic models by constrained lumping of differential equations,” *Bioinformatics*, vol. 37, no. 19, pp. 3385–3385, 08 2021.
- [21] L. Cardelli, R. Grosu, K. G. Larsen, M. Tribastone, M. Tschaikowski, and A. Vandin, “Algorithmic minimization of uncertain continuous-time Markov chains,” *IEEE Trans. Automat. Contr.*, pp. 1–16, 2023.
- [22] G. Li, H. Rabitz, and J. Tóth, “A general analysis of exact nonlinear lumping in chemical kinetics,” *Chemical Engineering Science*, vol. 49, no. 3, pp. 343–361, 1994.
- [23] G. J. Pappas, G. Lafferriere, and S. Sastry, “Hierarchically consistent control systems,” *IEEE Trans. Automat. Contr.*, vol. 45, no. 6, pp. 1144–1160, 2000.
- [24] A. van der Schaft, “Equivalence of dynamical systems by bisimulation,” *IEEE Trans. Automat. Contr.*, vol. 49, pp. 2160–2172, 2004.
- [25] G. J. Pappas and S. Simic, “Consistent abstractions of affine control systems,” *IEEE Trans. Automat. Contr.*, vol. 47, no. 5, 2002.
- [26] P. Tabuada and G. J. Pappas, “Abstractions of Hamiltonian control systems,” *Automatica*, vol. 39, no. 12, pp. 2025–2033, 2003.
- [27] V. Mehrmann and T. Stykel, “Balanced truncation model reduction for large-scale systems in descriptor form,” in *Lecture Notes in Computational Science and Engineering*. Springer-Verlag, 2005, pp. 83–115. [Online]. Available: [https://doi.org/10.1007/3-540-27909-1\\_3](https://doi.org/10.1007/3-540-27909-1_3)
- [28] T. Reis and E. Virnik, “Positivity preserving model reduction,” in *Positive Systems*. Springer Berlin Heidelberg, 2009, pp. 131–139. [Online]. Available: [https://doi.org/10.1007/978-3-642-02894-6\\_13](https://doi.org/10.1007/978-3-642-02894-6_13)
- [29] P. Li, J. Lam, Z. Wang, and P. Date, “Positivity-preserving  $H_\infty$  model reduction for positive systems,” *Automatica*, vol. 47, no. 7, pp. 1504–1511, 2011.
- [30] T. Ishizaki, K. Kashima, A. Girard, J. Imura, L. Chen, and K. Aihara, “Clustered model reduction of positive directed networks,” *Automatica*, vol. 59, pp. 238–247, Sept. 2015.
- [31] H. Sandberg and R. M. Murray, “Model reduction of interconnected linear systems using structured gramians,” *IFAC Proceedings Volumes*, vol. 41, no. 2, pp. 8725–8730, 2008.
- [32] D. Liberzon, *Calculus of Variations and Optimal Control Theory: A Concise Introduction*, Princeton, NJ, USA, 2011.

## APPENDIX

*Proof of Proposition 2.* In the definition of  $D$ , for an index  $i \in H_0$  we arbitrarily defined  $\lambda_i = 1$ . This choice ensures that  $D$  is invertible but any other choice of  $\lambda_i \neq 0$  would not affect the aggregation, as such indexes are the ones that do not belong to the support of any rows of  $L$  (nor of  $C$ ), and so they correspond to variables that do not appear in the aggregation  $Lx$ . The only not self-evident part of the statement is that  $CD$  is a lumping of  $A$  wrt  $O$  if and only if  $C$  is a lumping of  $DAD^{-1}$  wrt  $OD^{-1}$ . It follows from the definition of lumping and from the fact that  $(CD)^+ = D^{-1}C^+$ .  $\square$

*Proof sketch of Proposition 3.* The direction from 1) to 3) follows by repeating the argumentation of [21, Proposition 4]. Specifically, in the special case where  $u(\cdot) = 0$ , the argumentation carries over verbatim by noting that transient state probabilities  $\pi_i$  can be replaced with non-negative state variables  $x_i$ , while lower and upper bounds can be given by Metzler rather than transition rate matrices. The general case where  $u(\cdot) \in U$ , instead, is then reduced to foregoing special case as follows. First, we pick an arbitrary  $A(\cdot) \in [\underline{A}; \overline{A}]$  and  $u(\cdot) \in U$ . Using the solution  $x^{(A,u)}(\cdot)$  and  $A(\cdot)$ , we construct  $R(\cdot) \in [\underline{R}; \overline{R}]$  as in the case of  $u(\cdot) = 0$ . Then, the proof of the special case ensures that  $Cx^{(A,u)} = y^{(R,u)}$ . Conversely, for any  $R(\cdot) \in [\underline{R}; \overline{R}]$  and  $u(\cdot) \in U$ , we construct  $A(\cdot)$  from  $R(\cdot)$  and  $y^{(R,u)}(\cdot)$  as in the special case. Then, the proof ensures that  $Cx^{(A,u)} = y^{(R,u)}$ . The direction 3) to 2) follows readily because  $\Psi$  and  $\Phi$  are  $C$ -invariant by assumption. This leaves us with the direction from 2) to 1). To this end, let us fix some  $1 \leq r \leq k$  and consider the running cost  $\Psi = 0$  and final cost  $\Phi(x, u) = \sum_{j \in H_r} x_j + \|u\|_2$ . Obviously,  $\Phi$  and  $\Psi$  are  $C$ -invariant. Hence, 2) ensures that  $V^{\text{inf}}(x[0], \tau) =$

$V^{\inf}(x[0]', \tau)$  and  $V^{\sup}(x[0], \tau) = V^{\sup}(x[0]', \tau)$  where  $x[0] = e_i$  and  $x[0]' = e_{i'}$  for some  $i, i' \in H_{r'}$  with  $r \neq r'$  and  $i \neq i'$ , while  $e_j$  denotes the  $j$ -th unit vector of  $\mathbb{R}^n$ . Similarly to [21, Proposition 5], we next observe that  $x[0] = e_i$  yields

$$\sum_{j \in H_r} x_j(t) = \left[ \sum_{j \in H_r} A_{j,i}(0) + \sum_{j \in H_r} \sum_{\nu=1}^m B_{j,\nu} u_\nu(0) \right] t + o(t)$$

because  $A(t)$  and  $u(t)$  are piecewise continuous with right limits. Since  $u(\cdot) \geq 0$  and  $B \in \mathbb{R}_{\geq 0}^{n \times m}$ , the infimum is attained for  $u(0) = 0$  and  $A(0) = \underline{A}$ , that is

$$V^{\inf}(x[0], \tau) = \tau \sum_{j \in H_r} \underline{A}_{j,i} + o(\tau).$$

Since a similar reasoning applies to  $V^{\inf}(x[0]', \tau)$ , letting  $\tau \rightarrow 0$  yields  $\sum_{j \in H_r} \underline{A}_{j,i} = \sum_{j \in H_r} \underline{A}_{j,i'}$ . Since the choice of  $r \neq r'$ ,  $H_r$  and  $i, i' \in H_{r'}$  were arbitrary, this implies that  $C$  is a forward equivalence [15], hence a lumping of  $\underline{A}$ . The proof that  $C$  is a lumping of  $\bar{A}$  proceeds in a similar fashion by considering the final cost  $\Phi(x, u) = \sum_{j \in H_r} x_j - \eta \|u\|_2$ . Here,  $\eta \geq 0$  can be chosen sufficiently large to ensure that  $\sum_{j \in H_r} \sum_{\nu=1}^m B_{j,\nu} u_\nu(0) \leq \eta \|u(0)\|_2$  because  $U$  is bounded. With this, it follows that  $V^{\sup}(x[0], \tau) = \tau \sum_{j \in H_r} \bar{A}_{j,i} + o(\tau)$  and  $V^{\sup}(x[0]', \tau) = \tau \sum_{j \in H_r} \bar{A}_{j,i'} + o(\tau)$ , allowing us to conclude that  $\sum_{j \in H_r} \bar{A}_{j,i} = \sum_{j \in H_r} \bar{A}_{j,i'}$ . Then again [15] yields that  $C$  is a lumping of  $\bar{A}$ .  $\square$

To increase readability, in the following we often drop the dependence on time  $t$  from the functions.

**Lemma 1.** *Consider the system of ODEs  $\partial_t x = Ax + Bu$  in  $n$  variables, and let  $D$  be an invertible  $n \times n$  matrix. A function  $x(t)$  satisfies  $\partial_t x = Ax + Bu$  if and only if  $z(t) := Dx(t)$  satisfies the system  $\partial_t z = DAD^{-1}z + DBu$ .*

*Proof.* The function  $z(t) = Dx(t)$  satisfies  $\partial_t z = DAD^{-1}z + DBu$  if and only if  $\partial_t(Dx) = DAD^{-1}Dx + DBu$ , if and only if  $D\partial_t x = DAx + DBu$ , if and only if  $\partial_t x = Ax + Bu$ .  $\square$

*Proof of Theorem 1.* We first prove the special case when  $U = \{0\}$ , and to increase readability we shall write  $x^{(A)}$  and  $J_A$  for  $x^{(A,0)}$  and  $J_{(A,0)}$ , respectively.

By assumption,  $L$  is a lumping of  $\underline{A}$  and  $\bar{A}$ . Using Proposition 2, write  $L = CD$ , where  $D$  is a diagonal matrix with positive entries, and  $C \in \{0, 1\}^{k \times n}$  is a proper lumping of both  $D\underline{A}D^{-1}$  and  $D\bar{A}D^{-1}$ .

Fix an initial condition  $x[0] \in \mathbb{R}_{\geq 0}^n$ , and let  $\mathcal{S} = \{Lx^{(A)} \mid A \in [\underline{A}; \bar{A}]\}$ . Then Lemma 1 yields  $Dx^{(A)} = x^{(DAD^{-1})}$  for every  $A \in [\underline{A}; \bar{A}]$ , and so

$$\begin{aligned} \mathcal{S} &= \{CDx^{(A)} \mid A \in [\underline{A}; \bar{A}]\} = \{Cx^{(DAD^{-1})} \mid A \in [\underline{A}; \bar{A}]\} \\ &= \{Cx^{(DAD^{-1})} \mid DAD^{-1} \in D[\underline{A}; \bar{A}]D^{-1}\}. \end{aligned}$$

As  $D$  has positive entries, one can easily verify that  $D\underline{A}D^{-1} \leq D\bar{A}D^{-1}$ , and that the family of matrices  $D[\underline{A}; \bar{A}]D^{-1}$  coincides with  $[D\underline{A}D^{-1}; D\bar{A}D^{-1}]$ . Then

$$\begin{aligned} \mathcal{S} &= \{Cx^{(DAD^{-1})} \mid DAD^{-1} \in [D\underline{A}D^{-1}; D\bar{A}D^{-1}]\} \\ &= \{Cx^{(M)} \mid M \in [D\underline{A}D^{-1}; D\bar{A}D^{-1}]\}. \end{aligned}$$

As  $C$  is a proper lumping of the bounds in the above set, we can apply Proposition 3 to obtain that

$$\begin{aligned} \mathcal{S} &= \{y^{(R)} \mid R \in [C(D\underline{A}D^{-1})C^+; C(D\bar{A}D^{-1})C^+]\} \\ &= \{y^{(R)} \mid R \in [L\underline{A}L^+; L\bar{A}L^+]\}, \end{aligned}$$

where in the last step we used the fact that  $L^+ = (CD)^+ = D^{-1}C^+$ . This proved the first part of the statement.

Now from it we deduce that  $L$  preserves the costs of the PCS, so fix an initial condition  $x[0] \in \mathbb{R}_{\geq 0}^n$ , a time  $\tau \in \mathbb{R}$ , and  $\Phi, \Psi$  final and running costs that are  $L$ -invariant. As here we are assuming  $U = \{0\}$ , we drop the dependence of  $\Phi, \Psi$  on  $u \in U$ . For every  $A \in [\underline{A}; \bar{A}]$ , let  $R \in [L\underline{A}L^+; L\bar{A}L^+]$  be the matrix provided by the first part of the proof, and satisfying  $Lx^{(A)} = y^{(R)}$ . Then  $\Phi(x^{(A)}(\tau)) = \hat{\Phi}(Lx^{(A)}(\tau)) = \hat{\Phi}(y^{(R)}(\tau))$ , and  $\Psi(t, x^{(A)}(t)) = \hat{\Psi}(t, Lx^{(A)}(t)) = \hat{\Psi}(t, y^{(R)}(t))$  for every  $t \geq 0$ , giving  $J_A(x[0], \tau) = J_R(Lx[0], \tau)$ . The same argument as above shows that for every  $R \in [L\underline{A}L^+; L\bar{A}L^+]$  there is  $A \in [\underline{A}; \bar{A}]$  such that  $J_A(x[0], \tau) = J_R(Lx[0], \tau)$ , so the final coincidence of the optimal values is then obvious.

We now deal with the general case, so we drop the assumption of having  $U = \{0\}$ . Pick arbitrary  $A(\cdot) \in [\underline{A}; \bar{A}]$  and  $u(\cdot) \in U$ . Using the solution  $x^{(A,u)}(\cdot)$  and  $A(\cdot)$ , we construct  $R(\cdot) \in [\underline{R}; \bar{R}]$  as in the case of  $U = \{0\}$ . Then, the proof of the special case ensures that  $Lx^{(A,u)} = y^{(R,u)}$ . Conversely, for any  $R(\cdot) \in [\underline{R}; \bar{R}]$  and  $u(\cdot) \in U$ , we construct  $A(\cdot)$  from  $R(\cdot)$  and  $y^{(R,u)}(\cdot)$ , as in the shown special case. Then, the proof ensures that  $Lx^{(A,u)} = y^{(R,u)}$ . From this, the coincidence of the optimal values follows immediately.  $\square$

*Proof of Proposition 4.* As  $L^+ = (CD)^+ = D^{-1}C^+$ , we have  $R(\cdot) \in [L\underline{A}L^+; L\bar{A}L^+] = [C\underline{M}C^+; C\bar{M}C^+]$ . As  $C$  is a lumping of both  $\underline{M}$  and  $\bar{M}$ , we can apply Proposition 3, and let  $\tilde{A}(\cdot) \in [\underline{M}; \bar{M}]$  be the matrix defined in item 3ii), thus satisfying  $Cx^{(\tilde{A},u)} = y^{(R,u)}$ . It is easy to check that for every  $i, j$  we have  $A_{i,j}(\cdot) = \lambda_i^{-1} \lambda_j \tilde{A}_{i,j}(\cdot) = (D^{-1}\tilde{A}(\cdot)D)_{i,j}$ , i.e.  $A(\cdot)$  in our statement satisfies  $A(\cdot) = D^{-1}\tilde{A}(\cdot)D$ . By construction,  $\tilde{A}(\cdot) \in [\underline{M}; \bar{M}] = [D\underline{A}D^{-1}; D\bar{A}D^{-1}] = D[\underline{A}; \bar{A}]D^{-1}$ , so  $A(\cdot) = D^{-1}\tilde{A}(\cdot)D \in [\underline{A}; \bar{A}]$ . The final part is easy to check, and Lemma 1 yields  $Lx^{(A,u)} = CDx^{(A,u)} = Cx^{(DAD^{-1},u)} = Cx^{(\tilde{A},u)} = y^{(R,u)}$ .  $\square$

*Proof of Theorem 2.* We have to prove that  $L$  is a proper lumping of both  $\underline{A}$  and  $\bar{A}$ . Define  $C$  and  $D$  as in Proposition 2, so that  $L = CD$  and  $\text{rs}(OD^{-1}) \subseteq \text{rs}(C)$ ; now it is sufficient to prove that  $C$  is a proper lumping of both  $D\underline{A}D^{-1}$  and  $D\bar{A}D^{-1}$ . In view of Proposition 3, we check that  $C$  preserves the costs of the PCS  $([D\underline{A}D^{-1}; D\bar{A}D^{-1}], OD^{-1}, DB, U)$ . This means checking that the latter PCS and its reduction by  $C$ , that is  $([L\underline{A}L^+; L\bar{A}L^+], OL^+, LB, U)$ , have the same optimal values (to write the above reduced PCS we used that  $L^+ = D^{-1}C^+$ ). To this end, fix a time  $\tau \in \mathbb{R}$ , an arbitrary  $z[0] \in \mathbb{R}_{\geq 0}^n$ , and arbitrary final and running costs  $\phi, \psi$  that are  $C$ -invariant. First, consider the new final cost function  $\Phi: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$  defined by  $\Phi(x, u) = \phi(Dx, u)$  for  $x \in \mathbb{R}^n$  and  $u \in \mathbb{R}^m$ . It is easy to check that  $\Phi$  is  $CD$ -invariant. Similarly, we denote by  $\Psi$  the new running cost function

$\mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$  defined by  $\Psi(t, x, u) = \psi(t, Dx, u)$ , and it holds that also  $\Psi$  is  $L$ -invariant; moreover, it holds that

$$\hat{\phi} = \hat{\Phi} \text{ and } \hat{\psi} = \hat{\Psi}. \quad (6)$$

Fix  $M \in [D\underline{A}D^{-1}; D\overline{A}D^{-1}]$  and  $u \in U$ . Let  $A \in [\underline{A}; \overline{A}]$  be such that  $M = DAD^{-1}$ . Using the initial condition  $x[0] = D^{-1}z[0] \in \mathbb{R}_{\geq 0}^n$ , time  $\tau$ , and the costs  $\Phi, \Psi$ , apply the assumption on  $L$  to preserve the costs. This way, we get  $R \in [L\underline{A}L^+; L\overline{A}L^+]$  and  $v \in U$  such that

$$J_{(A,u)}(x[0], \tau) = J_{(R,v)}(Lx[0], \tau). \quad (7)$$

Our goal now is to prove that

$$J_{(M,u)}(z[0], \tau) = J_{(R,v)}(Cz[0], \tau); \quad (8)$$

to deduce (8) from (7), we show that both their left-hand sides and right-hand sides coincide. We begin by studying their left-hand sides. Let us denote by  $z$  the solution of  $\partial_t z = Mz + DBu$ , with initial condition  $z[0]$  (and along which the functional  $J_{(M,u)}(z[0], \tau)$  is computed, by definition). By Lemma 1, we have  $z = Dx^{(A,u)}$ , and then  $\Phi(x^{(A,u)}(\tau), u(\tau)) = \phi(Dx^{(A,u)}(\tau), u(\tau)) = \phi(z(\tau), u(\tau))$ . Similarly,  $\Psi(t, x^{(A,u)}(t), u(t)) = \psi(t, z(t), u(t))$  for every  $t \geq 0$ , giving  $J_{(A,u)}(x[0], \tau) = J_{(M,u)}(z[0], \tau)$ . Now we compare the functionals in the right-hand sides of (7) and (8). First, both of these functionals are computed along the same trajectory  $y^{(R,v)}$ , solution of the problem  $\partial_t y = Ry + LBu$ , and with the same initial condition  $Lx[0] = CDx[0] = Cz[0]$ . The evaluation of  $J_{(R,v)}(Lx[0], \tau)$  involves the reduced costs  $\hat{\Phi}$  and  $\hat{\Psi}$ , while  $J_{(R,v)}(Cz[0], \tau)$  involves the reduced costs  $\hat{\phi}$  and  $\hat{\psi}$ , so we conclude by applying (6). With a similar argument, one can see that also for every  $z[0] \in \mathbb{R}_{\geq 0}^n$ , and  $R \in [L\underline{A}L^+; L\overline{A}L^+]$ , and  $v \in U$ , there are  $M \in [D\underline{A}D^{-1}; D\overline{A}D^{-1}]$  and  $u \in U$  such that (8) holds for all costs that are  $C$ -invariant, so to finally establish that  $C$  preserves the costs of the PCS  $([D\underline{A}D^{-1}; D\overline{A}D^{-1}], OD^{-1}, DB, U)$ .  $\square$

### A. Proof of Theorem 3

To prove Theorem 3 we need some technical results first. For  $L, W \in \mathbb{R}^{k \times n}$ , the property  $\text{rs}(L) = \text{rs}(W)$  means that the rows of  $L$  and those of  $W$  span the same vector space, i.e., there is an invertible matrix  $S \in \mathbb{R}^{k \times k}$  such that  $L = SW$ .

Let us denote by  $e_r$  the  $r$ -th unit column vector in  $\mathbb{R}^k$ , having all 0 entries but the  $r$ -th one, which is 1.

**Lemma 2** (Disjoint supports). *Let  $L, W \in \mathbb{R}^{k \times n}$  with rank  $k$  be such that  $L = SW$ , for an invertible matrix  $S$ . If the rows of  $W$  have pairwise disjoint supports, then the columns of  $L$  are multiples of the columns of  $S$ .*

*Proof.* Since  $L = SW$ , we have that the  $i$ -th column  $L_i$  of  $L$  can be written as  $L_i = SW_i$ . Since the rows of  $W$  have disjoint supports, there is at most one  $r$  in  $\{1, \dots, k\}$  such that  $W_{r,i} \neq 0$ . If there is no such  $r$ , then  $0 = W_i = L_i$  and we are done. Otherwise,  $W_i = W_{r,i}e_r$ , and so  $L_i = SW_{r,i}e_r = W_{r,i}S_r$ .  $\square$

This result is enough to characterize when, given a lumping  $L \in \mathbb{R}^{k \times n}$ , there is another lumping  $W \in \mathbb{R}^{k \times n}$  with  $\text{rs}(W) = \text{rs}(L)$  whose rows have disjoint supports.

**Definition 9** (Equivalence over columns). *Let  $L \in \mathbb{R}^{k \times n}$  have rank  $k \leq n$ . Let  $Z = \{1, 2, \dots, n\} \setminus H_0 = \{1 \leq i \leq n \mid L_i \neq 0\}$  (i.e., the set of indices of the columns that are not identically zero) and write  $i \sim j$  for any  $i, j \in Z$  whenever the columns  $L_i, L_j$  are multiples of each other.*

**Proposition 6.** *Let  $L, Z$  be as in Definition 9. Then  $|Z/\sim| = k$  if and only if there is  $W \in \mathbb{R}^{k \times n}$  with  $\text{rs}(W) = \text{rs}(L)$  whose rows have pairwise disjoint supports.*

*Proof.* Assume such  $W$  exists. Since  $\text{rs}(W) = \text{rs}(L)$ , there is an invertible  $S$  with  $L = SW$ . By Lemma 2, the columns of  $L$  are multiples of the columns of  $S$ . Hence,  $|Z/\sim| \leq k$ . On the other hand, since  $W$  has rank  $k$ , all its rows have at least one non-zero element. Thus, for every column of  $S$  there is a column of  $L$  that is a multiple of its, meaning that  $|Z/\sim| \geq k$ . Now, assume that  $|Z/\sim| = k$  and let  $\mathcal{H} = \{H_1, \dots, H_k\}$  be the partition of  $Z$  obtained by  $\sim$ . For each  $r = 1, \dots, k$ , we fix one representative  $i(r) \in H_r$ . For  $i, j \in H \in \mathcal{H}$ , we denote by  $q(i, j)$  the ratio between the  $i$ -th and  $j$ -th column of  $L$  (note that neither  $L_i$  nor  $L_j$  is the zero column). Set  $W$  as follows:  $W_{r,j} = q_{i(r),j}$  if  $j \in H_r$ , and zero otherwise. By construction, the rows of  $W$  have disjoint supports (since the support of the  $r$ -th row of  $W$  is  $H_r$ ). Moreover, if we define  $S = (L_{i(1)} \dots L_{i(k)})^T$  we have  $L = SW$ , so  $\text{rs}(L) = \text{rs}(W)$ .  $\square$

*Proof of Theorem 3.* Assume that there is a constrained proper lumping  $W \in \mathbb{R}_{\geq 0}^{k \times n}$  of the PCS such that  $\text{rs}(W)$  has dimension  $k$ . By construction,  $L$  from Proposition 5 has the property that  $\text{rs}(L)$  has also dimension  $k$ , and it is minimal satisfying those conditions, so we have  $\text{rs}(L) = \text{rs}(W)$ . To prove that  $L$  is a constrained proper lumping, we first show that its rows have disjoint supports. Consider  $Z$  as in Definition 9 for the lumping  $L$ . By Proposition 6 we get  $|Z/\sim| = k$ . By Proposition 5,  $L$  is in reduced echelon-form and so the  $k$  vectors in  $\{e_r \mid 1 \leq r \leq k\}$  are columns of  $L$ . We can take them as representatives of the equivalence classes of  $\sim$ . Then for every non-zero column  $L_i$  of  $L$ , there is  $r \in \{1, \dots, k\}$  with  $L_i \sim e_r$ , and in particular  $L_i$  only has a non-zero entry in the  $r$ -th row. Thus, the rows of  $L$  have pairwise disjoint supports. Next, we show that  $L \in \mathbb{R}_{\geq 0}^{k \times n}$ . To this end, assume by contradiction that  $L$  has a negative entry, say  $L_{r,i} < 0$ . From  $\text{rs}(L) = \text{rs}(W)$ , there is an invertible matrix  $S$  with  $W = SL$ . Since the rows of  $L$  have pairwise disjoint supports, we can apply Lemma 2 in this setting. As the column  $L_i$  is not zero, the proof of Lemma 2 shows that  $W_i = L_{r,i}S_j$ . On the other hand,  $L$  is in reduced echelon-form so it has a column, say  $L_j$ , such that  $L_j = e_r$ . Then  $W = SL$  gives  $W_j = SL_j = Se_r = S_r$ . So we have deduced that  $W$  has as columns the vectors  $S_r$  and  $L_{r,i}S_j$ . Since  $W$  is a proper lumping we obtain that  $S_r = 0$ , but this contradicts the hypothesis of  $S$  being invertible.  $\square$