A Lie bracket approximation approach to distributed optimization over directed graphs *

Simon Michalowsky[1], Bahman Gharesifard[2], and Christian Ebenbauer[1]

[1]Institute for Systems Theory and Automatic Control, University of Stuttgart, Germany
{michalowsky,ce}@ist.uni-stuttgart.de
[2]Department of Mathematics and Statistics, Queen's University, Canada
bahman@mast.queensu.ca

Abstract. We consider a group of computation units trying to cooperatively solve a distributed optimization problem with shared linear equality and inequality constraints. Assuming that the computation units are communicating over a network whose topology is described by a time-invariant directed graph, by combining saddle-point dynamics with Lie bracket approximation techniques we derive a methodology that allows to design distributed continuous-time optimization algorithms that solve this problem under minimal assumptions on the graph topology as well as on the structure of the constraints. We discuss several extensions as well as special cases in which the proposed procedure becomes particularly simple.

1. Introduction

Driven by new applications and advancing communication technologies, the idea of solving optimization problems in a distributed fashion using a group of agents interchanging information over a communication network has gained a lot of interest during the last decades. Application examples include, among others, optimal power dispatch problems in Smart Grids [16], distributed machine learning [5] or formation control problems [7]. Besides several results on distributed computation [10], controllability and stabilization [3, 8, 17], there also exists a vast body of literature on distributed optimization algorithms, both in discrete- [27, 5] as well as continuous-time [15, 33, 13, 28, 18, 31], where in the present work we will focus on the latter one. While in most of the works a consensus-based approach is used where all agents aim to agree on a common solution of the overall optimization problem, in the last years other solutions have been proposed as well [28]. However, it is usually assumed that the underlying communication network is of undirected nature and it has turned out that establishing distributed optimization algorithms in the presence of directed communication structures is much more difficult. While there exist some approaches aiming to address this problem [18, 31], these are limited to unconstrained optimization problems using a consensus-based approach.

The contribution of this work is to provide a novel approach to the design of continuous-time distributed optimization algorithms applicable to a very general class of constrained optimization problems under mild assumptions on the possibly directed underlying communication network. The main idea of our approach is to employ classical saddle-point dynamics with proven convergence guarantees in a centralized setting and derive distributed approximations thereof. To this end, we follow a two step procedure where we first propose suitable Lie bracket representations of saddle-point dynamics and then use ideas from geometric control theory to design distributed approximations thereof. This idea has already been employed in previous works using a consensus-based approach [14] and for more general optimization problems with linear equality constraints in a gradient-free setting [25]. However, the focus in both works was on the first step of rewriting the saddle-point dynamics and the second step of designing distributed approximations was rarely treated. In the present paper we further contribute to both steps: on the one hand, we extend the class of optimization problems the approach is applicable to, and, on the other, we present an algorithm for the design of suitable approximations. While we limit ourselves to convex optimization problems with linear equality and inequality constraints, we emphasize that the same techniques may be used for a much larger class of optimization problems.

* This article is a sligthly extended version of [26] with an extra illustration Figure 2 and an additional result Lemma 4.

2. Preliminaries

2.1. Notation

We denote by $\mathbb{R}^n$ the set of n-dimensional real vectors, by $\mathbb{R}^n_+$ those with positive entries and by $\mathbb{R}^n_{++}$ those with strictly positive entries. We further write $\mathcal{C}^p$, $p \in \mathbb{N}$, for the set of p-times continuously differentiable real-valued functions. The gradient of a function $f : \mathbb{R}^n \to \mathbb{R}$, $f \in \mathcal{C}^1$, with respect to its argument $x \in \mathbb{R}^n$, will be denoted by $\nabla_x f : \mathbb{R}^n \to \mathbb{R}^n$; we often omit the subscript, if it is clear from the context. We denote the (i, j)th entry of a matrix $A \in \mathbb{R}^{n\times m}$ by a_{ij}, and sometimes denote A by $A = [a_{ij}]$. The rank of A is denoted by $\mathrm{rank}(A)$. We use e_i to denote the vector with the ith entry equal to 1 and all other entries equal to 0, and also use the short-hand notation $\mathbf{1}_n = [1, \dots, 1]^T$. For a vector $\lambda \in \mathbb{R}^n$ we let $\mathrm{diag}(\lambda) \in \mathbb{R}^{n\times n}$ denote the diagonal matrix whose diagonal entries are the entries of λ. Given two continuously differentiable vector fields $f_1 : \mathbb{R}^n \to \mathbb{R}^n$ and $f_2 : \mathbb{R}^n \to \mathbb{R}^n$, the Lie bracket of f_1 and f_2 evaluated at x is defined to be

$$[f_1, f_2](x) := \frac{\partial f_2}{\partial x}(x) f_1(x) - \frac{\partial f_1}{\partial x}(x) f_2(x). \tag{1}$$

For a set of vector fields $\Phi = \{f_1, f_2, \dots, f_p\}$, $f_i : \mathbb{R}^n \to \mathbb{R}^n$, $f_i \in \mathcal{C}^1$, we denote by $\mathcal{LB}(\Phi)$ the set of Lie brackets generated by Φ. For an (iterated) Lie bracket $B = [B_1, B_2]$, $B_1, B_2 \in \mathcal{LB}(\Phi)$, we then let $\mathrm{left}(B) = B_1$, $\mathrm{right}(B) = B_2$ denote the left and right factor of B, respectively. We further define the degree of a Lie bracket $B \in \mathcal{LB}(\Phi)$ as $\delta(B) = \tilde{\delta}_\Phi(B)$ and the degree of the kth vector field as $\delta_k(B) = \tilde{\delta}_{\{f_k\}}(B)$, where

$$\tilde{\delta}_{\mathcal{S}}(B) = \begin{cases} 1 & \text{if } B \in \mathcal{S} \\ \tilde{\delta}_{\mathcal{S}}(\mathrm{left}(B)) + \tilde{\delta}_{\mathcal{S}}(\mathrm{right}(B)) & \text{otherwise,} \end{cases}$$

with $\mathcal{S} \subseteq \Phi$. We denote the sign function by $\mathrm{sgn} : \mathbb{R} \to \{-1, 0, 1\}$, where $\mathrm{sgn}(-a) = -1$, $\mathrm{sgn}(a) = 1$ for any $a > 0$ and $\mathrm{sgn}(0) = 0$. For a vector $x = [x_1, \dots, x_n]^T \in \mathbb{R}^n$ and a finite set $S \subset \{1, \dots, n\}$, we denote by x_S the set of all x_i with $i \in S$. We also denote the complement of a set $S \subset \mathbb{R}^n$ by S^c.

2.2. Basics on graph theory

We recall some basic notions on graph theory, and refer the reader to [4] or other standard references for more information. A directed graph (or simply digraph) is an ordered pair $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{v_1, v_2, \dots, v_n\}$ is the set of nodes and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges, i.e. $(v_i, v_j) \in \mathcal{E}$ if there is an edge from node v_i to v_j. In our setup the edges encode to which other agents some agent has access to, i.e. $(v_i, v_j) \in \mathcal{E}$ means that node v_i receives information from node v_j. We say that node v_j is an out-neighbor of node v_i if there is an edge from node v_i to node v_j. The adjacency matrix $\mathbf{A} = [\mathbf{a}_{ij}] \in \mathbb{R}^{n\times n}$ associated to $\mathcal{G}$ is defined as

$$\mathbf{a}_{ij} = \begin{cases} 1 & \text{if } i \neq j \text{ and } (v_i, v_j) \in \mathcal{E}, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

We also define the out-degree matrix $\mathrm{D} = [\mathrm{d}_{ij}]$ associated to $\mathcal{G}$ as

$$\mathrm{d}_{ij} = \begin{cases} \sum_{k=1}^n \mathbf{a}_{ik} & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Finally, we call $G = \mathrm{D} - \mathbf{A} = [g_{ij}] \in \mathbb{R}^{n\times n}$ the Laplacian of $\mathcal{G}$. A digraph is said to be undirected if $(v_i, v_j) \in \mathcal{E}$ implies that $(v_j, v_i) \in \mathcal{E}$, or, equivalently, if $G = G^\top$. Further, a digraph $\mathcal{G}$ is called weight-balanced if $\mathbf{1}_n^T G = 0$. A directed path in $\mathcal{G}$ is a sequence of nodes connected by edges and we write $p_{i_1 i_r} = \langle v_{i_1} | v_{i_2} | \dots | v_{i_r} \rangle$ for a path from node v_{i_1} to node v_{i_r}. We further denote by $\mathrm{head}(p_{i_1 i_r}) = i_1$ and $\mathrm{tail}(p_{i_1 i_r}) = i_r$ the head and the tail of a path $p_{i_1 i_r}$, respectively. We also let $\ell(p_{i_1 i_r}) = r - 1$ denote the length of the path. A digraph $\mathcal{G}$ is said to be strongly connected (or simply connected in case of undirected graphs) if there is a directed path between any two nodes. For a path p_{ij} from node v_i to node v_j we denote by $\mathrm{subpath}_{i\bullet}(p_{ij})$ and $\mathrm{subpath}_{\bullet j}(p_{ij})$ the set of all subpaths of p_{ij} (not including p_{ij} itself) which, respectively, start at v_i or end at v_j. Given a subpath $q \in \mathrm{subpath}_{i\bullet}(p_{ij})$, we denote by q^c the path in $\mathrm{subpath}_{\bullet j}(p_{ij})$ whose composition with q gives p_{ij}.

3. Problem setup

Consider an optimization problem of the form

$$\begin{aligned} \min_x \quad & F(x) = \sum_{i=1}^n F_i(x_i) \\ \text{s.t} \quad & a_i x - b_i = 0, \qquad i \in \mathcal{I}_{\text{eq}} \subseteq \{1, 2, \dots, n\}, \\ & c_i x - d_i \leq 0, \qquad i \in \mathcal{I}_{\text{ineq}} \subseteq \{1, 2, \dots, n\}, \end{aligned} \tag{4}$$

where $x = [x_1, \dots, x_n]^\top \in \mathbb{R}^n$, $a_i \in \mathbb{R}^{1\times n}$, $c_i \in \mathbb{R}^{1\times n}$, and the $F_i : \mathbb{R} \to \mathbb{R}$, $F_i \in \mathcal{C}^2$, are assumed to be strictly convex functions. We assume further that the feasible set of (4) is non-empty and that there exists a unique solution $x^* \in \mathbb{R}^n$ to (4).

The problem can be interpreted as having n computation units or agents available, each one trying to optimize its own objective function F_i while, if $i \in \mathcal{I}_{\text{ineq}}$ or $i \in \mathcal{I}_{\text{eq}}$, respecting the ith global constraints among all agents. It is reasonable to assume that the constraints are associated to the agents in such a way that the constraint corresponding to agent i involves its own state. This is ensured by the following assumption on the set of constraints:

Assumption 1. For $a_i = [a_{i1}, \dots, a_{in}] \neq 0$, $c_i = [c_{i1}, \dots, c_{in}] \neq 0$, $i = 1, \dots, n$, we have that $a_{ii} \neq 0$ and $c_{ii} \neq 0$ for all $i = 1, \dots, n$. •

It should be noted that, for the ease of presentation, we limit ourselves to the case that each agent has at most one equality and one inequality constraint but the following results apply with some modifications to the case where each agent has several constraints, i.e., $a_i \in \mathbb{R}^{M_i \times n}$, $c_i \in \mathbb{R}^{m_i \times n}$ for some $m_i, M_i \in \mathbb{N}$. Our intention is to focus on presenting our results in a more understandable fashion and avoid complicated notations introduced when considering more general problem setups.

Going along that direction of a simpler notation, we augment the problem (4) by non-restrictive constraints such that exactly one equality and one inequality constraint is associated to each agent, i.e., we consider the augmented problem

$$\begin{aligned} \min_x \quad & F(x) = \sum_{i=1}^n F_i(x_i) \\ \text{s.t} \quad & a_i x - b_i = 0, \qquad i = 1, 2, \dots, n, \\ & c_i x - d_i \leq 0, \qquad i = 1, 2, \dots, n, \end{aligned} \tag{5}$$

where $a_i = 0, b_i = 0$ for $i \notin \mathcal{I}_{\text{eq}}$ and $c_i = 0, d_i > 0$ for $i \notin \mathcal{I}_{\text{ineq}}$, such that the feasible set as well as the solution of (4) and (5) are the same.

In the following, we wish to design continuous-time algorithms that "converge" to an arbitrarily small neighborhood of the solution of (5) and that can be implemented in a *distributed* fashion, i.e., each agent only uses information of its own state and objective function F_i as well as those of its out-neighbors, where out-neighboring agents are defined by a communication graph.

More precisely, we assume that the communication topology is given by some directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{v_1, v_2, \dots, v_n\}$ is a finite set of nodes and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges between the nodes. In our setup, the nodes play the role of the n agents and the edges define the allowed communication links between the agents, i.e., if there exists an edge from agent i to agent j, then agent i has access to the state of agent j. Using the graph Laplacian $G = [g_{ij}]$ associated to $\mathcal{G}$, we then have the following definition of a distributed algorithm:

Definition 1. We say that a continuous-time algorithm with agent dynamics of the form

$$\dot{z}_j = f_j(t, z), \tag{6}$$

$j = 1, 2, \dots, N$, $z = [z_1, z_2, \dots, z_N]^\top \in \mathbb{R}^N$, $f_j : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$, is distributed w.r.t. the graph $\mathcal{G}$ if it can equivalently be written as

$$\dot{z}_j = \tilde{f}_j(t, z_{\mathcal{N}(i)}), \tag{7}$$

where $\mathcal{N}(i) := \{j = 1, 2, \dots, N : g_{ij} \neq 0\}$ is the set of indices of all out-neighboring agents. •

In words, f_i may only depend on z_i and all states z_j whose corresponding agent j have a communication link to agent j, i.e., the algorithm obeys the communication topology defined by the directed graph $\mathcal{G}$.

Our approach relies on the use of saddle-point dynamics, i.e. algorithms that utilize the saddle point property of the Lagrangian. The Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n_+ \to \mathbb{R}$ associated to (5) is given by

$$\begin{aligned} L(x, \nu, \lambda) &= \sum_{i=1}^n F_i(x_i) + {\nu_i}^\top (a_i x - b_i) + {\lambda_i}^\top (c_i x - d_i) \\ &= F(x) + \nu^\top (Ax - b) + \lambda^\top (Cx - d), \end{aligned} \tag{8}$$

where we have used the stacked matrices

$$\begin{aligned} C &= \begin{bmatrix} {c_1}^\top & \dots & {c_n}^\top \end{bmatrix}^\top, & d &= \begin{bmatrix} d_1 & \dots & d_n \end{bmatrix}^\top, \\ A &= \begin{bmatrix} {a_1}^\top & \dots & {a_n}^\top \end{bmatrix}^\top, & b &= \begin{bmatrix} b_1 & \dots & b_n \end{bmatrix}^\top, \\ \lambda &= \begin{bmatrix} \lambda_1 & \dots & \lambda_n \end{bmatrix}^\top, & \nu &= \begin{bmatrix} \nu_1 & \dots & \nu_n \end{bmatrix}^\top, \end{aligned} \tag{9}$$

with $\nu \in \mathbb{R}^n, \lambda \in \mathbb{R}^n$ being the associated Lagrange multipliers. Here, a point $(x^\star, \nu^\star, \lambda^\star) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n_+$ is said to be a (global) saddle point of L if for all $x \in \mathbb{R}^n$, $\nu \in \mathbb{R}^n$, $\lambda \in \mathbb{R}^n_+$ we have

$$L(x^\star, \nu, \lambda) \leq L(x^\star, \nu^\star, \lambda^\star) \leq L(x, \nu^\star, \lambda^\star). \tag{10}$$

It is well-known that if the Lagrangian has some saddle point $(x^\star, \nu^\star, \lambda^\star)$, then $x^\star$ is a solution of (5). In the present setup, since (5) is a convex problem and the feasible set is non-empty, the existence of a saddle point is ensured (cf. e.g. [20]) such that finding a saddle point of L is equivalent to finding a solution to (5). We further require the following regularity assumption to hold:

Assumption 2. The constraints in (4) fulfill the Mangasarian-Fromovitz constraint qualifications at the optimal solution $x^\star$, i.e., the vectors a_i, $i \in \mathcal{I}_{\text{eq}}$, are linearly independent and there exists $q \in \mathbb{R}^n$ such that $c_i q < 0$ for all $i \in \mathcal{I}_{\text{ineq}}$ for which $c_i x^\star - d_i = 0$ and $a_i q = 0$ for all $i \in \mathcal{I}_{\text{eq}}$. •

This assumption ensures that the set of saddle points of the Lagrangian associated to (4) is non-empty and compact, see [32, Theorem 1]. Note that, due to the augmentation of the optimization problem, the set of saddle points of the Lagrangian L associated to (5) is in general not compact and we will care for that by modifying the saddle-point dynamics. To be more precise, in the following Lemma we propose a modified saddle-point dynamics, which is an extension of the one proposed in [13], and show asymptotic stability of a compact subset of the set of saddle points.

Lemma 1. Consider the following modified saddle-point dynamics

$$\dot{x} = -\nabla_x L(x, \nu, \lambda) \qquad = -\nabla F(x) - A^\top \nu - C^\top \lambda \quad (11a)$$
$$\dot{\nu} = \nabla_\nu L(x, \nu, \lambda) + w(\nu) \quad = Ax - b + w(\nu) \quad (11b)$$
$$\dot{\lambda} = \mathrm{diag}(\lambda) \nabla_\lambda L(x, \nu, \lambda) \quad = \mathrm{diag}(\lambda)(Cx - d), \quad (11c)$$

where $F : \mathbb{R}^n \to \mathbb{R}$, $F \in \mathcal{C}^2$, is strictly convex and where $w : \mathbb{R}^n \to \mathbb{R}^n$ is defined as

$$w(\nu) = -\sum_{\substack{i=1 \\ i \notin \mathcal{I}_{\mathrm{eq}}}}^{n} \nu_i e_i \quad (12)$$

with $e_i \in \mathbb{R}^n$ being the ith unit vector. Let

$$\begin{aligned} \mathcal{M} := \big\{ &(x, \nu, \lambda) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n_+ : \\ &x = x^*, \nu_i = 0 \text{ for } i \notin \mathcal{I}_{\mathrm{eq}}, \lambda_i = 0 \text{ for } i \notin \mathcal{I}_{\mathrm{ineq}}, \\ &L(x^*, \nu, \lambda) \le L(x^*, \nu^*, \lambda^*) \le L(x, \nu^*, \lambda^*) \big\} \end{aligned} \quad (13)$$

and suppose that Assumption 2 holds. Then the set $\mathcal{M}$ is asymptotically stable for (11) with region of attraction

$$\mathcal{R}(\mathcal{M}) \subseteq \big\{ (x, \nu, \lambda) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n : \lambda \in \mathbb{R}^n_{++} \big\}. \quad (14)$$

•

Remark 1. Since a point in $\mathcal{M}$ might as well lie on the boundary of $\mathcal{R}(\mathcal{M})$, one needs to modify the corresponding notions of stability accordingly, by restricting the neighborhoods to the set of admissible initial conditions (cf. [11]); from now on, we assume that this is understood, without stating it. •

Remark 2. The term w in (11b) is usually not included in saddle-point dynamics. Here, it is used to render the dynamics of the additional dual variables introduced due to the augmentation asymptotically stable. It should be noted that the augmentation might lead to a significantly larger state vector of (11) compared to the saddle-point dynamics corresponding to the original optimization problem (4). However, it should be kept in mind that, besides possible performance benefits (cf. Remark 9), the main reason for the augmentation is a significantly simpler notation and it is not crucial for the following methodology to apply (cf. Remark 4). •

A proof of Lemma 1 is given in Appendix A.1. While (11) converges to a solution of (4), it is in general not distributed in the aforementioned sense. Note that if the underlying graph is undirected and the constraints are only imposed between neighboring agents, then (11) is indeed distributed. In the following, we wish to derive dynamics that "*approximate*" those of (11) arbitrarily close, in a sense that will be made precise shortly, and are additionally distributed, even when the underlying graph is directed. To be more precise, we consider agent dynamics of the form

$$\dot{x}_i^\sigma = u_{x,i}^\sigma(t, x_{\mathcal{N}(i)}^\sigma, \nu_{\mathcal{N}(i)}^\sigma, \lambda_{\mathcal{N}(i)}^\sigma) \quad (15a)$$
$$\dot{\nu}_i^\sigma = u_{\nu,i}^\sigma(t, x_{\mathcal{N}(i)}^\sigma, \nu_{\mathcal{N}(i)}^\sigma, \lambda_{\mathcal{N}(i)}^\sigma) \quad (15b)$$
$$\dot{\lambda}_i^\sigma = u_{\lambda,i}^\sigma(t, x_{\mathcal{N}(i)}^\sigma, \nu_{\mathcal{N}(i)}^\sigma, \lambda_{\mathcal{N}(i)}^\sigma), \quad (15c)$$

where $i = 1, 2, \ldots, n$, $\sigma \in \mathbb{R}^+$ is a parameter and

$$\mathcal{N}(i) := \{ j = 1, 2, \ldots, n : g_{ij} \neq 0 \} \quad (16)$$

is the set of indices of all out-neighboring agents of the ith agent. Note that (15) is obviously distributed according to Definition 1. Our objective is then to design functions $u_{x,i}^\sigma$, $u_{\nu,i}^\sigma$, $u_{\lambda,i}^\sigma$, $i = 1, 2, \ldots, n$, parametrized by $\sigma \in \mathbb{R}^+$, such that the trajectories $\big(x^\sigma(t), \nu^\sigma(t), \lambda^\sigma(t)\big)$ of (15) uniformly converge to the trajectories $\big(x(t), \nu(t), \lambda(t)\big)$ of (11) with increasing σ. To this end, the main idea of the proposed methodology is to rewrite the right-hand side of (11) in terms of Lie brackets of *admissible* vector fields, i.e., vector fields that can be computed locally by the nodes, and then employ ideas from geometric control theory to derive suitable approximations.

4. Main results

Consider the saddle-point dynamics (11). As a first step, we separate the right-hand side into admissible and non-admissible vector fields, where admissible refers to the part of the dynamics that can be computed locally by the nodes. For the ease of presentation we assume in the following that the constraints of agent i are only imposed to its out-neighboring agents, i.e., we impose the following assumption on the constraints:

Assumption 3. For $a_i = [a_{i1}, \ldots, a_{in}]$, $c_i = [c_{i1}, \ldots, c_{in}]$, $i = 1, 2, \ldots, n$, we have for each $j = 1, \ldots, n$, that $a_{ij} \neq 0$ or $c_{ij} \neq 0$ only if $g_{ij} \neq 0$. •

In other words, we thereby assume that the constraints match the communication topology induced by the graph[1]. Under this assumption, the right-hand side of (11b), (11c) is admissible, while parts of the right-hand side of (11a) are not. Note that the gradient of F is admissible, since F is a separable function; the remaining terms, however, are not necessarily admissible, since the underlying communication graph is directed. Now, for $A_i = [a_{ij}]$, $C_i = [c_{ij}]$, we define the admissible part of

[1] It should be noted that the following results can be extended to problems where this assumption does not hold, cf. Remark 6 as well as in the example in Section 5.2.

$A^\top, C^\top$ as

$$\tilde{A}_{\mathrm{adm}} = \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{sgn}(|g_{ij}|) a_{ji} e_i e_j^\top, \tag{17}$$

$$\tilde{C}_{\mathrm{adm}} = \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{sgn}(|g_{ij}|) c_{ji} e_i e_j^\top, \tag{18}$$

where $\mathrm{sgn} : \mathbb{R} \to \{-1, 0, 1\}$ is the sign function and e_i is the ith unit vector. Observe that $\tilde{A}_{\mathrm{adm}}, \tilde{C}_{\mathrm{adm}}$ correspond to the admissible part of $A^\top$ and $C^\top$, respectively. We then let

$$\tilde{A}_{\mathrm{rest}} = A^\top - \tilde{A}_{\mathrm{adm}}, \quad \tilde{C}_{\mathrm{rest}} = C^\top - \tilde{C}_{\mathrm{adm}}, \tag{19}$$

and define the complete state of (11) as

$$z := [x^\top, \nu^\top, \lambda^\top]^\top \in \mathbb{R}^{3n}. \tag{20}$$

Hence, we can write the saddle-point dynamics (11) as

$$\dot{z} = f_{\mathrm{adm}}(z) + \begin{bmatrix} -\tilde{A}_{\mathrm{rest}}\nu - \tilde{C}_{\mathrm{rest}}\lambda \\ 0 \\ 0 \end{bmatrix}, \tag{21}$$

where $f_{\mathrm{adm}} : \mathbb{R}^{3n} \to \mathbb{R}^{3n}$ is defined as

$$f_{\mathrm{adm}}(z) = \begin{bmatrix} -\nabla F(x) - \tilde{A}_{\mathrm{adm}}\nu - \tilde{C}_{\mathrm{adm}}\lambda \\ Ax - b + w(\nu) \\ \mathrm{diag}(\lambda)(Cx - d) \end{bmatrix}. \tag{22}$$

There, f_{adm} is admissible whereas the second part in (21) is not. The essential idea to derive suitable distributed approximations is to rewrite the non-admissible part in terms of Lie brackets of admissible vector fields and we will elaborate on that in the following.

4.1. Rewriting the non-admissible vector fields

We first define the index set

$$\mathcal{I}(i) := \{i, n+i, 2n+i\}, \tag{23}$$

$i = 1, 2, \ldots, n$, associating the components of z to the ith agent, i.e., $z_{\mathcal{I}(i)}$ is the state of agent i. We then define a set of vector fields $h_{i,j} : \mathbb{R}^{3n} \to \mathbb{R}^{3n}$, $i, j = 1, 2, \ldots, 3n$, as

$$h_{i,j}(z) = z_i e_j, \tag{24}$$

where $e_j \in \mathbb{R}^{3n}$ is the jth unit vector. Observe that $h_{i,j}$ is an admissible vector field if and only if there exist ℓ, k such that $i \in \mathcal{I}(\ell)$, $j \in \mathcal{I}(k)$ and $g_{k\ell} \neq 0$. Before we present a general construction rule, let us first illustrate the main idea by means of a simple example.

Example 1. Consider the graph shown in Figure 1 with $n = 5$ nodes. Let $h_{i,j}$ be defined as in (24) and observe that $h_{n+3,n+2}, h_{n+2,1}$ are admissible. Consider the Lie bracket

$$\begin{aligned} &[h_{n+3,n+2}, h_{n+2,1}](z) \\ &= e_1 e_{n+2}^\top z_{n+3} e_{n+2} - e_{n+2} e_{n+3}^\top z_{n+2} e_1 \\ &= z_{n+3} e_1, \end{aligned} \tag{25}$$

which, according to (24), is equal to $h_{n+3,1}(z)$, i.e., a non-admissible vector field. Given the graphical representation in Figure 1, this can be interpreted as a "fictitious" edge from agent 1 to agent 3, generated by the Lie bracket of two admissible vector fields. This observation is of key importance in the rest of the paper. More generally, we can observe that

$$[h_{i,j}, h_{j,k}](z) = h_{i,k}(z) \tag{26}$$

for any $i, j, k = 1, 2, \ldots, 3n$. •

Next, we generalize this idea. Let $p_{ij} = \langle v_{i_1} | \ldots | v_{i_r} \rangle$ be a path in $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ from node v_i to node v_j, i.e. $i = i_1$, $j = i_r$, $v_{i_1}, \ldots, v_{i_r} \in \mathcal{V}$, $r \geq 2$, and let $\ell(p_{ij}) = r - 1$ denote its length. We now, recursively, define a mapping R_{k_1,k_2}, $k_1, k_2 = 1, 2, \ldots, 3n$, from a given path p_{ij} in $\mathcal{G}$ to the set of vector fields on $\mathbb{R}^{3n}$:

- for $\ell(p_{ij}) = 1$, we define
$$R_{k_1,k_2}(p_{ij}) = h_{k_1,k_2}. \tag{27}$$
- for $\ell(p_{ij}) \geq 2$, we show, c.f. Lemma 2, that for all $q_1, q_2 \in \mathrm{subpath}_{i\bullet}(p_{ij})$ and $k_1, k_2, s_1, s_2 = 1, 2, \ldots, 3n$, we have that
$$[R_{k_1,s_1}(q_1^c), R_{s_1,k_2}(q_1)] = [R_{k_1,s_2}(q_2^c), R_{s_2,k_2}(q_2)]; \tag{28}$$
hence we define
$$R_{k_1,k_2}(p_{ij}) = [R_{k_1,s}(q^c), R_{s,k_2}(q)], \tag{29}$$
where q is any subpath in $\mathrm{subpath}_{i\bullet}(p_{ij})$ and $s \in \mathcal{I}(\mathrm{tail}(q))$.

Remark 3. Observe that R_{k_1,k_2} is independent of the path p_{ij} according to the definition (27). However, the path comes into play when it gets to choosing k_1, k_2 such that the resulting Lie bracket is a Lie bracket of admissible vector fields, cf. Lemma 2.

Using these definitions, we next state a result that extends the ideas from Example 1.

Lemma 2. Consider a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ of n nodes. Let p_{ij} be a path between v_i and v_j, $v_i, v_j \in \mathcal{V}$, and let R_{k_1,k_2} be defined as in (27), (29). Then, if $k_1 \neq k_2$, we have for all $z \in \mathbb{R}^{3n}$

$$R_{k_1,k_2}(p_{ij})(z) = z_{k_1} e_{k_2} = h_{k_1,k_2}(z), \tag{30}$$

and, if $k_1 \in \mathcal{I}(\mathrm{tail}(p_{ij}))$, $k_2 \in \mathcal{I}(\mathrm{head}(p_{ij}))$, then $R_{k_1,k_2}(p_{ij})$ is a Lie bracket of admissible vector fields. •

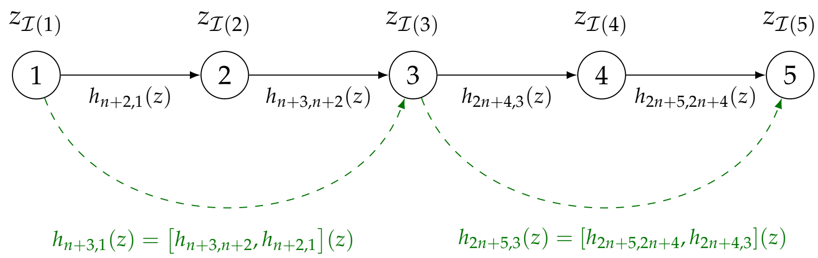

Figure 1. A communication structure with $n = 5$ nodes is depicted. The arrows indicate to which agent state some agent has access to, e.g., agent 1 has access to the state of agent 2 given by $z_{\mathcal{I}(2)} = [x_2, \nu_2, \lambda_2]^\top$ but not the other way round. The dotted green arrow shows a fictitious edge with associated vector fields created by Lie brackets of admissible vector fields.

Proof. We prove the result by induction. For paths of the form $p_{i_1 i_2} = \langle v_{i_1} | v_{i_2} \rangle$, i.e., $\ell(p_{i_1 i_2}) = 1$, by (24) and (27) equation (30) follows immediately. Further we observe that the vector field (27) is admissible if $k_1 \in \mathcal{I}(j), k_2 \in \mathcal{I}(i)$ and $g_{ij} \neq 0$, which is true since p_{ij} is a path in $\mathcal{G}$. Suppose now that the result holds for all paths p with $\ell(p) \leq \bar{\ell}$, $\bar{\ell} \geq 2$. Let $p_{i_1 i_\ell} = \langle v_{i_1} | v_{i_2} | \dots | v_{i_k} \rangle$ be any path with $\ell(p_{i_1 i_k}) = \bar{\ell} + 1$. Let further $q_r \in \mathrm{subpath}_{i_1 \bullet}(p_{i_1 i_k})$ be a subpath of $p_{i_1 i_k}$ that ends at v_r, $r = i_2, i_3, \dots, i_{k-1}$. Then, since $\ell(q_r) \leq \bar{\ell}$, $\ell(q_r^c) \leq \bar{\ell}$, we have by (29) and the induction hypothesis

$$
\begin{aligned}
R_{k_1,k_2}(p_{i_1 i_k})(z) &= \left[R_{k_1,s}(q_r^c), R_{s,k_2}(q_r)\right](z) \\
&= \left[h_{k_1,s}, h_{s,k_2}\right](z) \\
&= e_{k_2} e_s^\top z_{k_1} e_s - e_s e_{k_1}^\top z_s e_{k_2} \\
&= z_{k_1} e_{k_2}, \qquad (31)
\end{aligned}
$$

where $s \in \mathcal{I}(\mathrm{tail}(q_r))$ and where we used that $k_1 \neq k_2$. This proves (30). Further, if $k_1 \in \mathcal{I}(\mathrm{tail}(p_{i_1 i_k}))$, i.e., $k_1 \in \mathcal{I}(\mathrm{tail}(q_r^c))$, then, by the induction hypothesis and with $s \in \mathcal{I}(\mathrm{tail}(q_r)) = \mathcal{I}(\mathrm{head}(q_r^c))$, $R_{k_1,s}(q_r^c)$ is a Lie bracket of admissible vector fields. Similarly, if $k_2 \in \mathcal{I}(\mathrm{head}(p_{i_1 i_k}))$, by the induction hypothesis and with $s \in \mathcal{I}(\mathrm{tail}(q_r))$, also $R_{s,k_2}(q_r)$ is a Lie bracket of admissible vector fields. Thus, $R_{k_1,k_2}(p_{i_1 i_k})$ is a Lie bracket of admissible vector fields as well which concludes the proof. □

Remark 4. The same result holds true if we drop the assumption that each agent has exactly one equality and one inequality constraint, since this only leads to a reformulation of the index sets $\mathcal{I}(i)$, $i = 1, 2, \dots, n$. Interestingly, additional constraints also introduce additional degrees of freedom in rewriting the non-admissible vector fields since the index set $\mathcal{I}(\mathrm{tail}(q))$ grows. •

Remark 5. It is worth pointing out that vector fields of the form (24) are not the only ones that can be used to rewrite non-admissible vector fields in terms of Lie brackets of admissible vector fields. In fact, there exists a whole class of vector fields which can be employed for this purpose. Similar as in [19], some of them might have beneficial properties in terms of the approximation. •

While Lemma 2 holds for any directed path in $\mathcal{G}$, in the following, we will use the shortest path since this leads to iterated Lie brackets of smallest degree. We do not discuss how to compute the paths here since this is a problem on its own but refer the reader to standard algorithms, see, e.g., [9]. Further, the choice of subpath and the state index s in the recursion (29) is arbitrary as well. In Lemma 3 in Section 4.2, we provide a particular choice that turns out to be beneficial in the construction of the approximating input sequences. The next result is an immediate consequence of Lemma 2.

Proposition 1. *Suppose that Assumption 3 holds. Suppose $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is strongly connected and, for all $i, j = 1, \dots, n$ let p_{ij}, denote a path from node v_i to node v_j, where $v_i, v_j \in \mathcal{V}$. Then, with $z = [x^\top, \nu^\top, \lambda^\top]^\top$, the dynamics (21) can equivalently be written as*

$$
\begin{aligned}
\dot{z} = f_{\mathrm{adm}}(z) &- \sum_{i=1}^{n} \sum_{j=1}^{n} \tilde{a}_{\mathrm{rest},ij}\, R_{n+j,i}(p_{ij})(z) \\
&- \sum_{i=1}^{n} \sum_{j=1}^{n} \tilde{c}_{\mathrm{rest},ij} R_{2n+j,i}(p_{ij})(z)
\end{aligned} \qquad (32)
$$

and the right-hand side is a linear combination of Lie brackets of admissible vector fields. •

Remark 6. If Assumption 3 does not hold the terms $[0, f_{\mathrm{adm},2}(z), 0]^\top$, $[0, 0, f_{\mathrm{adm},3}(z)]^\top$ may no longer be admissible. While $[0, f_{\mathrm{adm},2}(z), 0]^\top$ can be rewritten using Lemma 2, for $[0, 0, f_{\mathrm{adm},3}(z)]^\top$ different construction techniques are required, since $f_{\mathrm{adm},3}$ is bilinear as a function of x and λ. However, it should be noted that it is as well possible to rewrite these terms in a similar manner. •

Remark 7. In general, having a strongly connected graph is sufficient but not necessary. In fact, it is sufficient that there exists a path from node v_i to node v_j for all i, j such that $\tilde{a}_{\mathrm{rest},ij} \neq 0$ or $\tilde{c}_{\mathrm{rest},ij} \neq 0$. •

Now that we have rewritten the non-admissible vector fields in terms of iterated Lie brackets of admissible vector fields, there is still the issue of how to generate suitable functions $u^{\sigma}_{x,i}, u^{\sigma}_{v,i}, u^{\sigma}_{\lambda,i}$ to be addressed. We will touch upon this in the next section and thereby provide a result on how (15) and (32) are related in terms of their stability properties under a suitable choice of the input functions.

4.2. Construction of distributed control laws

Our main objective in this section is to elaborate on how to construct suitable input functions $u^{\sigma}_{x,i}, u^{\sigma}_{v,i}, u^{\sigma}_{\lambda,i}$ such that the trajectories of (15) uniformly converge to those of (32) as we increase σ. The following procedure is based on the results presented in [21], [30], [22]. In [22], the relation between the trajectories of a system of the form

$$\dot{z}^{\sigma} = f_0(z^{\sigma}) + \sum_{k=1}^{M} \phi_k(z^{\sigma}) U^{\sigma}_k(t), \quad z^{\sigma}(0) = z_0, \tag{33}$$

where $f_0, \phi_k : \mathrm{R}^N \to \mathrm{R}^N$, $U^{\sigma}_k : \mathrm{R} \to \mathrm{R}$, $z_0 \in \mathrm{R}^N$ and the trajectories of an associated *extended system*

$$\dot{z} = f_0(z) + \sum_{B \in \mathcal{B}} v_B B(z), \quad z(0) = z_0, \tag{34}$$

is studied, where $\mathcal{B}$ is a set of Lie brackets of the vector fields ϕ_k, $k = 1, \ldots, M$, and $v_B \in \mathrm{R}$ is the corresponding coefficient. In our setup, (15) will play the role of (33) with ϕ_k being the admissible vector fields and (32) plays the role of (34) with $\mathcal{B}$ being the set of Lie brackets of admissible vector fields required to rewrite the non-admissible vector fields. It is shown in [22] that, under a suitable choice of the input functions U^{σ}_k, the solutions of (33) uniformly converge to those of (34) on compact time intervals for increasing σ, i.e., for each $\varepsilon > 0$ and for each $T \geq 0$, there exists $\sigma^* > 0$ such that for all $\sigma > \sigma^*$ and $t \in [0, T]$ we have that

$$\|z(t) - z^{\sigma}(t)\| \leq \varepsilon. \tag{35}$$

An algorithm for constructing suitable input functions U^{σ}_k that fulfill these assumptions is presented in [21] as well as in a brief version in [30]; we will follow this idea in here, however, given that in [21] the input functions are not given in explicit form, here we exploit the special structure of the admissible vector fields in order to simplify this procedure and arrive at explicit formulas for a large class of scenarios, applicable to our work.

4.2.1. *Writing the Lie brackets in terms of a P. Hall basis*

The algorithm presented in [22] requires the brackets B in (34) to be brackets in a so-called P. Hall basis; we need to "project" the brackets in (32) to such a basis, in the sense that will be made precise shortly. We first recall the definition of a P. Hall basis; we let $\delta(B)$ to denote the degree of a bracket B.

Definition 2. [P. Hall basis of a Lie algebra] Let $\Phi = \{\phi_1, \phi_2, \ldots, \phi_m\}$ be a set of smooth vector fields. A P. Hall basis $\mathcal{PH}(\Phi) = (\mathrm{P}, \prec)$ of the Lie algebra generated by Φ is a set P of brackets equipped with a total ordering $\prec$ that fulfills the following properties:

[PH1] Every ϕ_k is in P.

[PH2] $\phi_k \prec \phi_j$ if and only if $k < j$.

[PH3] If $B_1, B_2 \in \mathrm{P}$ and $\delta(B_1) < \delta(B_2)$, then $B_1 \prec B_2$.

[PH4] Each $B = [B_1, B_2] \in \mathrm{P}$ if and only if

- [PH4.a] $B_1, B_2 \in \mathrm{P}$ and $B_1 \prec B_2$
- [PH4.b] either $\delta(B_2) = 1$ or $B_2 = [B_3, B_4]$ for some B_3, B_4 such that $B_3 \preceq B_1$. •

Note that [PH2] is usually not included in the definition of a P. Hall basis, but it is common to include it for the approximation problem at hand. Further, the construction rule [PH4] ensures that no brackets are included in the basis that are related to other brackets in the basis by the Jacobi identity or skew-symmetry; thus the brackets are in this sense independent. However, this does not mean that, when evaluating the brackets, the resulting vector fields are independent, which we will exploit later. It is as well worth mentioning that the ordering fulfilling the properties [PH1] - [PH4] is in general not unique, i.e., for a given set of vector fields Φ, there may exist several P. Hall bases.

Let us now return to our setup. Let Φ be given by the set of admissible vector fields defined as

$$\Phi := \{h_{i,j} : \exists k_1, k_2 \in \{1, 2, \ldots, n\} \text{ such that } i \in \mathcal{I}(k_1), \\ j \in \mathcal{I}(k_2), g_{k_2 k_1} \neq 0\}, \tag{36}$$

where $h_{i,j}$ is defined in (24). Every bracket in the set of Lie brackets of admissible vector fields $\mathcal{B}$ can then be written as a linear combination of elements of a corresponding P. Hall basis by successively resorting the brackets, making use of skew-symmetry and the Jacobi identity. Such a projection algorithm is for example given in [29] and in the following we let

$$\mathrm{proj}_{\mathcal{PH}} : \mathcal{B} \to \mathrm{P} \tag{37}$$

denote the projection operator for a given P. Hall basis $\mathcal{PH} = (\mathrm{P}, \prec)$. However, for brackets of higher degree, this might be tedious and results in a large number of brackets; we hence propose an alternative approach. Instead of resorting the complete brackets appearing in (32), we suggest to reduce the resorting steps by a proper choice of the subpaths in the construction procedure presented in Lemma 2. The main idea is to choose the subpath q in (29) in such a way that, in each recursion step, the degree of the left factor of the bracket is strictly smaller than the degree of the right factor and such that the degree of the left factor of the right factor is smaller than that of the left factor of the original bracket such that [PH4.a] and [PH4.b] are automatically fulfilled. Since the degree directly corresponds to the length of the subpath this can be achieved by choosing the subpath appropriately, see also Figure 2. We make this idea more precise in the following Lemma.

Lemma 3. Consider a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ of n nodes. Let the set of admissible vector fields be defined according to (36). Let some P. Hall basis $\mathcal{PH}(\Phi) = (\mathrm{P}, \prec)$ be given and let $\mathrm{proj}_{\mathcal{PH}} : \mathcal{B} \to \mathrm{P}$ denote a projection operator onto the P. Hall basis. Let $p_{i_1 i_r}$ be a path from node $v_{i_1} \in \mathcal{V}$ to node $v_{i_r} \in \mathcal{V}$ and define

$$\tilde{R}_{k_1,k_2}(p_{i_1 i_r}) = \begin{cases} R_{k_1,k_2}(p_{i_1 i_r}) & \text{if } \ell(p_{i_1 i_r}) = 1, \\ \mathrm{proj}_{\mathcal{PH}}([R_{k_1,s}(q^c), R_{s,k_2}(q)]) & \text{if } \ell(p_{i_1 i_r}) = 2,3,4,6, \\ [\tilde{R}_{k_1,s}(q^c), \tilde{R}_{s,k_2}(q)] & \text{otherwise,} \end{cases} \tag{38}$$

where

$$s = \begin{cases} n + i_{\theta(p_{i_1 i_r})} & \text{if } 1 \leq k_1 \leq 2n \\ n + i_{\theta(p_{i_1 i_r})} & \text{if } 2n+1 \leq k_1 \leq 3n \end{cases} \tag{39}$$

$$q = p_{i_1 i_{\theta(p_{i_1 i_r})}} \in \mathrm{subpath}_{i_1 \bullet}(p_{i_1 i_r}) \tag{40}$$

$$\theta(p_{i_1 i_r}) = \begin{cases} \frac{1}{2}\ell(p_{i_1 i_r}) + 1 & \text{if } \ell(p_{i_1 i_r}) = 2,4, \\ \lfloor \frac{1}{2}\ell(p_{i_1 i_r}) \rfloor + 2 & \text{otherwise,} \end{cases} \tag{41}$$

with $\lfloor a \rfloor$ being the largest integer value smaller than $a \in \mathrm{R}^+$. Then $\tilde{R}_{k_1,k_2}(p_{i_1 i_r})(z) = R_{k_1,k_2}(p_{i_1 i_r})(z)$ for all $z \in \mathrm{R}^{3n}$ and $\tilde{R}_{k_1,k_2}(p_{i_1 i_r}) \in \mathrm{P}$ for all $k_1 \in \mathcal{I}(\mathrm{tail}(p_{i_1 i_r})), k_2 \in \mathcal{I}(\mathrm{head}(p_{i_1 i_r}))$. •

Proof. A proof is given in Appendix A.2. □

Remark 8. It should be noted that the projection can be computed easily in the given case. To this end, first notice that – by the choice of subpaths – for $\ell(p_{i_1 i_r}) = 2, 3$, the brackets admit the following structure

$$R_{k_1,k_2}(p_{i_1 i_r}) = \begin{cases} [\phi_{a_1}, \phi_{a_2}] & \text{if } \ell(p_{i_1 i_r}) = 2 \\ [\phi_{a_1}, [\phi_{a_2}, \phi_{a_3}]] & \text{if } \ell(p_{i_1 i_r}) = 3 \end{cases} \tag{42}$$

for some $a_{1/2/3} \in \mathrm{N}^+$ depending on $k_1, k_2, p_{i_1 i_r}$, where $\phi_{a_i} \in \Phi$, $i = 1, 2, 3$. For such brackets, the projection on the P. Hall basis $\mathcal{PH}$ is easily computed making use of skew-symmetry and the Jacobi-identity and we obtain

$$\mathrm{proj}_{\mathcal{PH}}([\phi_{a_1}, \phi_{a_2}]) = \begin{cases} [\phi_{a_1}, \phi_{a_2}] & \text{if } a_1 < a_2, \\ -[\phi_{a_2}, \phi_{a_1}] & \text{if } a_1 > a_2, \end{cases} \tag{43}$$

and

$$\mathrm{proj}_{\mathcal{PH}}([\phi_{a_1}, [\phi_{a_2}, \phi_{a_3}]]) = \begin{cases} [\phi_{a_2}, [\phi_{a_1}, \phi_{a_3}]] - [\phi_{a_3}, [\phi_{a_1}, \phi_{a_2}]] & \text{if } a_1 = \min\limits_{i=1,2,3} a_i, \\ [\phi_{a_1}, [\phi_{a_2}, \phi_{a_3}]] & \text{if } a_2 = \min\limits_{i=1,2,3} a_i, \\ -[\phi_{a_1}, [\phi_{a_3}, \phi_{a_2}]] & \text{if } a_3 = \min\limits_{i=1,2,3} a_i. \end{cases} \tag{44}$$

In the same manner, for $\ell(p_{i_1 i_r}) = 4, 6$, we have

$$R_{k_1,k_2}(p_{i_1 i_r}) = \begin{cases} [B_{a_1}, B_{a_2}] & \text{if } \ell(p_{i_1 i_r}) = 4, \\ [B_{a_1}, [B_{a_2}, B_{a_3}]] & \text{if } \ell(p_{i_1 i_r}) = 6, \end{cases} \tag{45}$$

where the B_{a_i} are Lie brackets of the ϕ_i with $\delta(B_{a_i}) = 2$, $i = 1, 2, 3$. The projection is then done by first projection the inner brackets B_{a_i} on the P. Hall basis using (43) and then resorting $R_{k_1,k_2}(p_{i_1 i_r})$ as in (43), (44). •

Remark 9. It is worth pointing out, as become clear in the proof, that the aforementioned result is independent of the choice of s as given in (39); in fact, any $s \in \mathcal{I}(i_{\theta(r)}) = \{i_{\theta(r)}, n + i_{\theta(r)}, 2n + i_{\theta(r)}\}$ can be taken. Although the particular choice made does not make any difference in rewriting the non-admissible vector fields, it becomes relevant in designing suitable approximating inputs. In particular, the choice of s controls in which components of the complete state the perturbing inputs are injected. The specific choice (39) is motivated by the idea of injecting the most perturbation in the dual variables. Observe that the degrees of freedom for s increase with the number of constraints of each agent. In particular, it might as well happen that there is no degree of freedom if we do not augment the optimization problem (4). •

Using Lemma 3 we can then write (32) as

$$\begin{aligned} \dot{z} = f_{\mathrm{adm}}(z) &- \sum_{i=1}^{n}\sum_{j=1}^{n} c_{\mathrm{rest},ij} \tilde{R}_{n+j,i}(p_{ij})(z) \\ &- \sum_{i=1}^{n}\sum_{j=1}^{n} a_{\mathrm{rest},ij} \tilde{R}_{2n+j,i}(p_{ij})(z). \end{aligned} \tag{46}$$

and we can identify the set of brackets $\mathcal{B}$ in (34) with

$$\begin{aligned} \mathcal{B} = &\{\tilde{R}_{n+j,i}(p_{ij}) : a_{\mathrm{rest},ij} \neq 0, i, j = 1, \ldots, n\} \\ &\cup \{\tilde{R}_{2n+j,i}(p_{ij}) : c_{\mathrm{rest},ij} \neq 0, i, j = 1, \ldots, n\}, \end{aligned} \tag{47}$$

where now $\mathcal{B} \subset \mathrm{P}$ for some P. Hall basis $\mathcal{PH} = (\mathrm{P}, \prec)$, and for the coefficients we have

$$v_{\tilde{R}_{n+i,j}(p_{ij})} = -a_{\mathrm{rest},ij}\mathrm{sign}(\tilde{R}_{n+i,j}(p_{ij})(1)) \tag{48a}$$

$$v_{\tilde{R}_{2n+i,j}(p_{ij})} = -c_{\mathrm{rest},ij}\mathrm{sign}(\tilde{R}_{n+i,j}(p_{ij})(1)). \tag{48b}$$

We are now ready to apply the algorithm presented in [21] to construct suitable approximating inputs and we will discuss that in the following section.

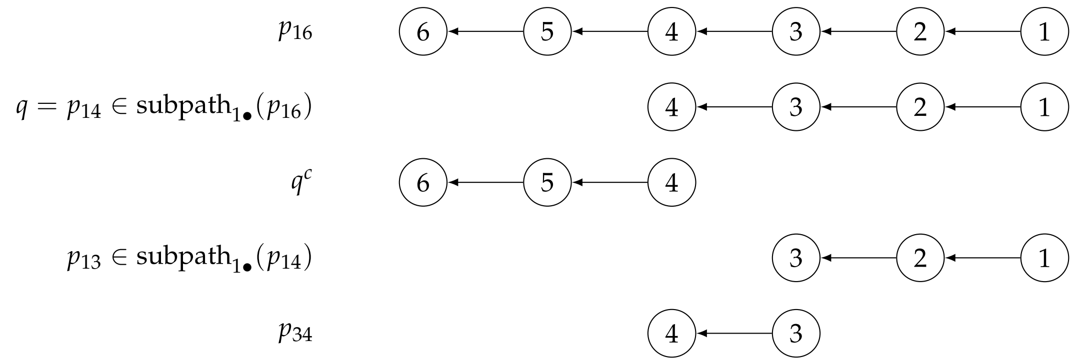

Figure 2. An illustration of the idea of choosing the subpaths. The complement of the subpath q^c is strictly shorter than the subpath q such that in the recursion (29) the left factor of the bracket has strictly smaller degree than the right factor, hence [PH4.a] in the Definition of a P. Hall basis holds. Also, the subpath p_{34} of the subpath q is strictly shorter than q^c such that in the recursion (29) the left factor of the right factor of the bracket has strictly smaller degree than the left factor of the bracket, thus making sure that [PH4.b] holds as well.

4.2.2. Approximating input sequences

We now consider the collection of all agent dynamics (15) given by

$$\dot{z}^\sigma = u^\sigma(t, z^\sigma) = \begin{bmatrix} u_x^\sigma(t, x^\sigma, \nu^\sigma, \lambda^\sigma) \\ u_\nu^\sigma(t, x^\sigma, \nu^\sigma, \lambda^\sigma) \\ u_\lambda^\sigma(t, x^\sigma, \nu^\sigma, \lambda^\sigma) \end{bmatrix}, \tag{49}$$

where $z^\sigma = [x^{\sigma\top}, \nu^{\sigma\top}, \lambda^{\sigma\top}]^\top$, $x^\sigma \in \mathbb{R}^n$, and $\nu^\sigma \in \mathbb{R}^n$, $\lambda^\sigma \in \mathbb{R}^n$ are the stacked vectors of all $x_i^\sigma, \nu_i^\sigma, \lambda_i^\sigma$, $i = 1, 2, \ldots, n$, respectively, and $u_x^\sigma, u_\nu^\sigma, u_\lambda^\sigma : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ are the stacked vectors of all $u_{x,i}^\sigma, u_{\nu,i}^\sigma, u_{\lambda,i}^\sigma$, $i = 1, 2, \ldots, n$, respectively. Following the algorithm presented in [21], we let the input take the form

$$u^\sigma(t, z^\sigma) = f_{\mathrm{adm}}(z^\sigma) + \sum_{k=1}^{M} \phi_k(z^\sigma) U_k^\sigma(t), \tag{50}$$

where $\Phi = \{\phi_1, \phi_2, \ldots, \phi_M\}$ is the set of admissible vector fields defined in (36) and where $\phi_k \in \mathrm{P}$ for some P. Hall basis $\mathcal{PH}(\Phi) = (\mathrm{P}, \prec)$. Further, $U_k^\sigma : \mathbb{R} \to \mathbb{R}$, $k = 1, \ldots, M$, are so-called *approximating input sequences* with sequence parameter $\sigma \in \mathbb{N}^+$ which in the following we aim to construct in such a way that the solutions of (49) uniformly converge to those of (46) with increasing σ. The algorithm in [21] relies on a "superposition principle", i.e., we group all brackets in $\mathcal{B}$ defined by (47) into equivalence classes, which we later denote by E, and associate to each class an input $U_{k,E}^\sigma$ and then add them up as

$$U_k^\sigma(t) = \sum_{E \in \mathcal{E}} U_{k,E}^\sigma(t), \tag{51}$$

where $\mathcal{E}$ is the set of all equivalence classes in $\mathcal{B}$. Roughly speaking, two brackets are said to be equivalent if each vector field appears the same number of times in the bracket but possibly in a different order. A precise definition of the equivalence relation is given in Definition 3. For each bracket $E \in \mathcal{E}$ and $k = 1, \ldots, M$ we then define the corresponding input $U_{k,E}^\sigma(t)$ as follows:

- If $\delta_k(E) = 0$: $U_{k,E}^\sigma(t) = 0$.
- If $\delta(E) = 2, \delta_k(E) = 1$:

$$U_{k,E}^\sigma(t) = 2\sqrt{\sigma}\,\mathrm{Re}\big(\eta_{E,k}(\omega_E) e^{\mathrm{i}\sigma\omega_E t}\big). \tag{52}$$

- If $\delta(E) = N$, $\delta_k(E) = 1$:

$$U_{k,E}^\sigma(t) = 2\sigma^{\frac{N-1}{N}} \sum_{\rho=1}^{|E|} \mathrm{Re}\big(\eta_E(\omega_{E,\rho,k}) e^{\mathrm{i}\sigma\omega t}\big). \tag{53}$$

Here, it is $\delta(E) = \delta(B)$, $\delta_k(E) = \delta_k(B)$ for any $B \in \mathcal{E}$. Further, $\omega_E, \omega_{E,\rho,k} \in \mathbb{R}$ are frequencies we will specify later, $\eta_{E,k}, \eta_E : \mathbb{R} \to \mathbb{C}$ are coefficients to be chosen in dependence of the frequencies, and $\mathrm{i} \in \mathbb{C}$ is the imaginary unit. However, the superposition principle does not hold as desired but there are two major issues one has to take care of:

1. The input sequences $U_{k,E}^\sigma$ may not interfere with each other in a way which ensures that the superposition principle holds; this can be dealt with by a proper choice of the frequencies.
2. Each input sequence $U_{k,E}^\sigma$ not only excites the desired brackets $E \cap \mathcal{B}$, but also all other equivalent brackets in E; we can overcome this by a proper choice of the coefficients $\eta_{\omega,k}, \eta_\omega$. The idea behind this is to also excite the undesired equivalent brackets on purpose, which itself excite the desired brackets, in such a way that the undesired equivalent brackets all cancel out.

While the problem at hand does not allow for simplifications in the choice of the frequencies, the calculation of proper coefficients $\eta_{\omega,k}, \eta_\omega$ can be simplified drastically by exploiting some structural properties of the set of brackets $\mathcal{B}$. More precisely, there are two properties that turn out to be beneficial: First, in each bracket $B \in \mathcal{B}$ each vector field ϕ_k appears only once, i.e., $\delta_k(B) \in \{0,1\}$, for any $B \in \mathcal{B}$, $k = 1, \ldots, M$, and second, for any bracket $B \in \mathcal{B}$, all equivalent brackets either evaluate to the same vector field as B or vanish, see Lemma 4. We present and discuss the simplified algorithm in Appendix A.4.

4.3. Distributed algorithm

We next state our main result which relates the solutions of (11) with those of (49) in closed-loop with the distributed control input (50)-(53). We use the notion *practically uniformly asymptotically stability from [11, 12], without explicitly defining it here.*

Theorem 1. Consider the distributed optimization problem (4) and suppose that the communication topology is given by a strongly connected digraph with n nodes. Assume that F is strictly convex and suppose further that Assumption 1 - 3 hold. Consider the agent dynamics (49) with the control law (50)-(53). Then, for each $\varepsilon > 0$ and for each initial condition $z^\sigma(0) = z(0) = z_0 \in \mathcal{R}(\mathcal{M})$, with $\mathcal{R}(\mathcal{M})$ given in (14), there exists $\sigma^* > 0$ such that for all $\sigma > \sigma^*$ the following holds:

1. For all $t \geq 0$, we have

$$\left\|z^\sigma(t) - z(t)\right\| \leq \varepsilon, \tag{54}$$

   where $z^\sigma(t)$ is the solution of (49) with the control law (50) - (53) and $z(t) = \big(x(t), \nu(t), \lambda(t)\big)$ is the solution of (11), with initial condition $z^\sigma(0) = z(0) = z_0$.

2. The set $\mathcal{M}$ defined by (13) is practically uniformly asymptotically stable. •

We postpone the proof of this result to Appendix A.5 and focus on its useful implications in the next section.

5. Special cases and examples

In this section we discuss special cases in which the inputs can be given in explicit form and present several simulation examples illustrating the previous results.

5.1. Explicit representation of approximating inputs for low order brackets

While the algorithm given in Appendix A.4 can in general be complicated to implement, this procedure becomes particularly simple to implement in scenarios where the set of brackets $\mathcal{B}$ defined in (47) only contains brackets of degree less or equal than three. As stated in our next result, in this case the set of equivalent brackets only contains the bracket itself but no other bracket, thus the second issue (2) in Section 4.2.2 does not come into play.

Proposition 2. *Consider (46) and assume that all paths p_{ij} fulfill $\ell(p_{ij}) \leq 3$. Let $\mathcal{PH}(\Phi) = (\mathrm{P}, \prec)$ be any P. Hall basis of Φ defined by (36) that fulfills $h_{k_1,k_2} \prec h_{k_3,k_4}$ for all $k_4 > k_2$. Then, for any path p_{ij} with $\ell(p_{ij}) \leq 3$, we have that*

$$\begin{aligned} E\big(\tilde{R}_{r+j,i}(p_{ij})\big) &= \{B \in \mathrm{P} : B \sim \tilde{R}_{r+j,i}(p_{ij}), B(z) \not\equiv 0\} \\ &= \tilde{R}_{r+j,i}(p_{ij}) \end{aligned} \tag{55}$$

for $r \in \{n, 2n\}$, where the equivalence relation $\sim$ is defined by Definition 3.

Remark 10. It should be noted that the ordering of the P. Hall basis is important for this result to hold. Further, if Assumption 3 does not hold, different brackets are introduced in (46) which still are of degree three under the assumption that all paths p_{ij} fulfill $\ell(p_{ij}) \leq 3$ but have a different structure. Hence, the assumption on the ordering usually is not sufficient anymore.

A proof of this result can be found in Appendix A.3. The condition that all paths p_{ij} in (46) are of length less or equal than three holds, for example, if the longest cordless cycle in $\mathcal{G}$ is of length 4. Using the result of Proposition 2 and following the algorithm presented in Appendix A.4, we obtain

- if $E = B = [\phi_{k_1}, \phi_{k_2}]$:

$$\begin{aligned} &U^\sigma_{k,E}(t) \\ &= \begin{cases} -\sqrt{2\sigma}\frac{1}{\beta_E}\sqrt{|v_B\omega_E|}\cos(\sigma\omega_B t) & \text{if } k = k_1 \\ \operatorname{sgn}(v_B\omega_B)\sqrt{2\sigma}\beta_E\sqrt{|v_B\omega_E|}\sin(\sigma\omega_B t) & \text{if } k = k_2 \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \tag{56}$$

- if $E = B = \big[\phi_{k_1}, [\phi_{k_2}, \phi_{k_3}]\big]$:

$$\begin{aligned} &U^\sigma_{k,E}(t) \\ &= \begin{cases} -\sigma^{\frac{2}{3}} 2\beta_E(\omega_{E,k_1}\omega_{E,k_2})^{\frac{1}{3}}\cos(\sigma\omega_{E,k}t) & \text{if } k = k_1, k_3 \\ -\sigma^{\frac{2}{3}} 2\frac{1}{\beta_E^2}(\omega_{E,k_1}\omega_{E,k_2})^{\frac{1}{3}}\cos(\sigma\omega_{E,k_2}t) & \text{if } k = k_2 \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \tag{57}$$

where $\beta_E \neq 0$ is a design parameter. The frequencies $\omega_E, \omega_{E,k} \in \mathrm{R} \setminus \{0\}$ need to be chosen such that they fulfill the following properties:

- All frequencies ω_E, $E \in \mathcal{E}$, $\delta(E) = 2$, are distinct.
- For each $E = B = \big[\phi_{k_1}, [\phi_{k_2}, \phi_{k_3}]\big]$, the set of frequencies $\{\omega_{E,k_1}, \omega_{E,k_2}, \omega_{E,k_3}\}$ is minimally canceling, see Definition 4.

- The collection of sets

$$\{\{\omega_E\}_{E\in\mathcal{E},\delta(E)=2},\{\omega_{E,k_1},\omega_{E,k_2},\omega_{E,k_3}\}_{E\in\mathcal{E},\delta(E)=3}\}$$

  is an independent collection, see Definition 5.

It should be noted that similar explicit formulas can as well be obtained for brackets of higher degree but they become more complicated. The main reason is that, while for brackets of degree strictly less than four all equivalent brackets evaluate to zero (cf. Table 2), this is no longer the case for brackets of higher degree such that now the second issue discussed in Section 4.2.2 needs to be taken care of.

5.2. Simulation examples

Next, we present some simulated examples to illustrate our results: We consider an optimization problem of the form (4) with $n = 5$ agents, where, for $i = 1, 2, \ldots, 5$, $F_i(x_i) = (x_i - i)^2$, and the constraints are given by

$$x_1 - x_2 \leq -10, \quad x_2 = x_3 + 1, \tag{58a}$$
$$x_4 + x_3 \leq -3, \quad x_5 - x_2 = 7, \tag{58b}$$

such that after augmentation we have for the matrices that define the constraints in (5)

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 7 \end{bmatrix}, \tag{59}$$

$$C = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad d = \begin{bmatrix} -10 \\ K \\ K \\ -3 \\ K \end{bmatrix}, \tag{60}$$

where $K = 3$ but can as well be chosen arbitrary as long as $K > 0$. We consider two different communication graphs as depicted in Figure 3, where graph (b) is the same as graph (a) except that the communication link from agent 5 to agent 2 got broken, thus an additional fictitious edge is required. While the constraints match the communication topology of graph (a), i.e., Assumption 3 holds, this is not the case for graph (b) due to the last constraint in (58). We first consider the case that graph (a) represents the communication topology. In this case, the graph Laplacian is given by

$$G = \begin{bmatrix} 2 & -1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & -1 & 2 \end{bmatrix} \tag{61}$$

and hence

$$\tilde{A}_{\text{adm}} = \begin{bmatrix} 0&0&0&0&0 \\ 0&1&0&0&0 \\ 0&0&0&0&0 \\ 0&0&0&0&0 \\ 0&0&0&0&1 \end{bmatrix}, \tilde{C}_{\text{adm}} = \begin{bmatrix} 1&0&0&0&0 \\ 0&0&0&0&0 \\ 0&0&0&0&0 \\ 0&0&0&1&0 \\ 0&0&0&0&0 \end{bmatrix}. \tag{62}$$

The saddle-point dynamics (21) are then given by

$$\begin{aligned} \dot{z} &= f_{\text{adm}}(z) - (-e_3 z_7 - e_2 z_{11} - e_2 z_{10} + e_3 z_{14}) \\ &= f_{\text{adm}}(z) + h_{n+2,3}(z) + h_{2n+1,2}(z) \\ &\qquad + h_{n+5,2}(z) - h_{2n+4,3}(z), \end{aligned} \tag{63}$$

where the admissible part $f_{\text{adm}} : \mathrm{R}^{15} \to \mathrm{R}^{15}$ is defined by (22) and the remaining four vector fields are non-admissible. Following Lemma 2 and choosing the subpaths as suggested in Lemma 3 we then rewrite the non-admissible vector fields as given in Table 1.

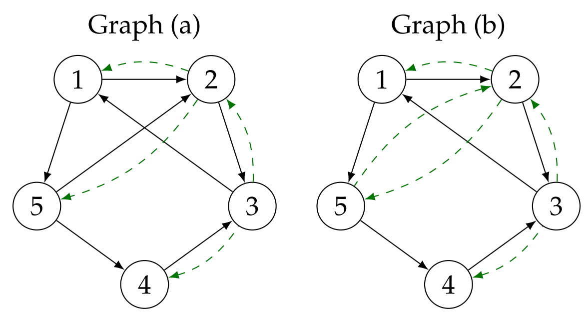

Figure 3. Two communication graphs (a) and (b) for the simulation example from Section 5.2. The dashed green arrows indicate the required fictitious edges, respectively.

vector field	corresponding path	Lie bracket representation
$h_{n+2,3}$	$\langle v_3 \| v_1 \| v_2 \rangle$	$[h_{n+2,n+1}, h_{n+1,3}]$
$h_{2n+1,2}$	$\langle v_1 \| v_3 \| v_1 \rangle$	$[h_{2n+1,2n+3}, h_{2n+3,3}]$
$h_{n+5,2}$	$\langle v_2 \| v_3 \| v_1 \| v_5 \rangle$	$[h_{n+5,n+1}, [h_{n+1,n+3}, h_{n+3,2}]]$
$h_{2n+4,3}$	$\langle v_3 \| v_1 \| v_5 \| v_4 \rangle$	$[h_{2n+4,2n+5}, [h_{2n+5,2n+1}, h_{2n+1,3}]]$

Table 1. The results of applying Lemma 2 to rewrite the non-admissible vector fields in the example from Section 5.2 in terms of Lie brackets of admissible vector fields ($n = 5$).

As a next step, we need to project the Lie brackets on any P. Hall basis $\mathcal{PH}(\Phi) = (\mathrm{P}, \prec)$ with

$$\Phi = \{h_{n+2,n+1}, h_{n+1,3}, h_{2n+1,2n+3}, h_{2n+3,3}, h_{n+5,n+1}, h_{n+1,n+3}, h_{n+3,2}, h_{2n+4,2n+5}, h_{2n+5,2n+1}, h_{2n+1,3}\}.$$

In general, we can choose any P. Hall basis and then make use of Remark 8 for the projection. However, in this case it is also easily possible to properly choose the ordering of the P. Hall basis in such a way that the brackets in Table 1 are already in P. More precisely, we only have to make sure that $h_{n+2,n+1} \prec h_{n+1,3}$, $h_{2n+1,2n+3} \prec h_{2n+3,3}$, $h_{n+1,n+3} \prec h_{n+3,2}$, $h_{n+1,n+3} \prec h_{n+5,n+1}$, $h_{2n+5,2n+1} \prec h_{2n+1,3}$, $h_{2n+5,2n+1} \prec h_{2n+4,2n+5}$. Note that this is in general not possible, since the conditions might be conflicting and – to keep this example more general – we do not adapt the ordering in that way in our implementation.

We are now ready to apply the algorithm presented in Appendix A.4. We do not discuss the resulting input sequences in detail here and also do not provide the complete simulation results due to space limitations, but instead do this for the case that the communication graph is given by graph (b). We refer the interested reader to employ the provided Matlab implementation [23]. We next discuss the implications of having the communication graph given by graph (b) in Figure 3 instead of graph (a). Since the link from node 2 to node 3 is missing in the graph, Assumption 3 does no longer hold. In particular, the vector field $h_{n+2,n+5}(z) = z_{n+2}e_{n+5}$, which is included in the admissible vector field f_{adm} in case the communication is given by graph (a), now is non-admissible. Despite Assumption 3 not being fulfilled, we can still use Lemma 2 to rewrite $h_{2,n+5}$, since the result is completely independent of this assumption. Indeed, the corresponding path is given by $p_{52} = \langle v_5|v_4|v_3|v_1|v_2\rangle$ and we obtain

$$\begin{aligned} &h_{2,n+5}(z) \\ &= \big[[h_{2,n+1}, h_{n+1,n+3}], [h_{n+3,n+4}, h_{n+4,n+5}]\big](z). \end{aligned} \tag{64}$$

We can then follow the same procedure as discussed before to project on any P. Hall basis, where Φ now additionally includes the vector fields $h_{2,n+1}$, $h_{n+1,n+3}$, $h_{n+3,n+4}$, and $h_{n+4,n+5}$, and then apply the algorithm presented in Appendix A.4. The corresponding simulation results are depicted in Figure 4.

6. Conclusion and outlook

We presented a new approach to distributed optimization problems where the communication topology is given by a directed graph. Our approach is based on a two-step procedure where in a first step first we derived suitable Lie bracket representations of saddle-point dynamics and then used Lie bracket approximations techniques from geometric control theory to obtain distributed control laws. While we limited ourselves to the class of convex problems with separable cost function and linear equality and inequality constraints that match the communication topology, we emphasize that the methodology is applicable to a much larger class of optimization problems, including, for example, non-linear constraints, constraints not compatible with the graph structure or non-separable cost functions. Additionally, similar techniques can be applied to distributed control problems. We also presented a simplified algorithm for the design of approximating inputs that exploits the problem structure. Summarizing, the presented approach provides a systematic way to address distributed optimization problems under mild assumptions on the communication graph as well as the problem structure. However, the design of suitable approximating inputs with improved transient and asymptotic behavior is complex and still an important issue to be addressed. While filters can be used as a simple remedy to this problem, there are also two other ways we plan to approach this problem: (1) altering the choice of admissible vector fields and (2) modifying the design of the approximating inputs including an optimal choice of parameters.

References

[1] The On-Line Encyclopedia of Integer Sequences, A000048. `https://oeis.org/A000048`.

[2] The On-Line Encyclopedia of Integer Sequences, A006788. `https://oeis.org/A006788`.

[3] M.-A. Belabbas. Sparse stable systems. *Systems & Control Letters*, 62(10):981 – 987, 2013.

[4] Norman Biggs. *Algebraic graph theory*. Cambridge university press, 1993.

[5] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.

[6] Lev Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR computational mathematics and mathematical physics*, 7(3):200–217, 1967.

[7] F. Bullo, J. Cortés, and S. Martínez. *Distributed Control of Robotic Networks*. Applied Mathematics Series. Princeton University Press, 2009.

[8] X. Chen, M. A. Belabbas, and T. Başar. Controllability of formations over directed graphs. In *2015 54th IEEE Conference on Decision and Control (CDC)*, pages 4764–4769, 2015.

[9] Thomas H Cormen. *Introduction to algorithms*. MIT press, 2009.

[10] Z. Costello and M. Egerstedt. The degree of nonholonomy in distributed computations. In *53rd IEEE Conference on Decision and Control*, pages 6092–6098, 2014.

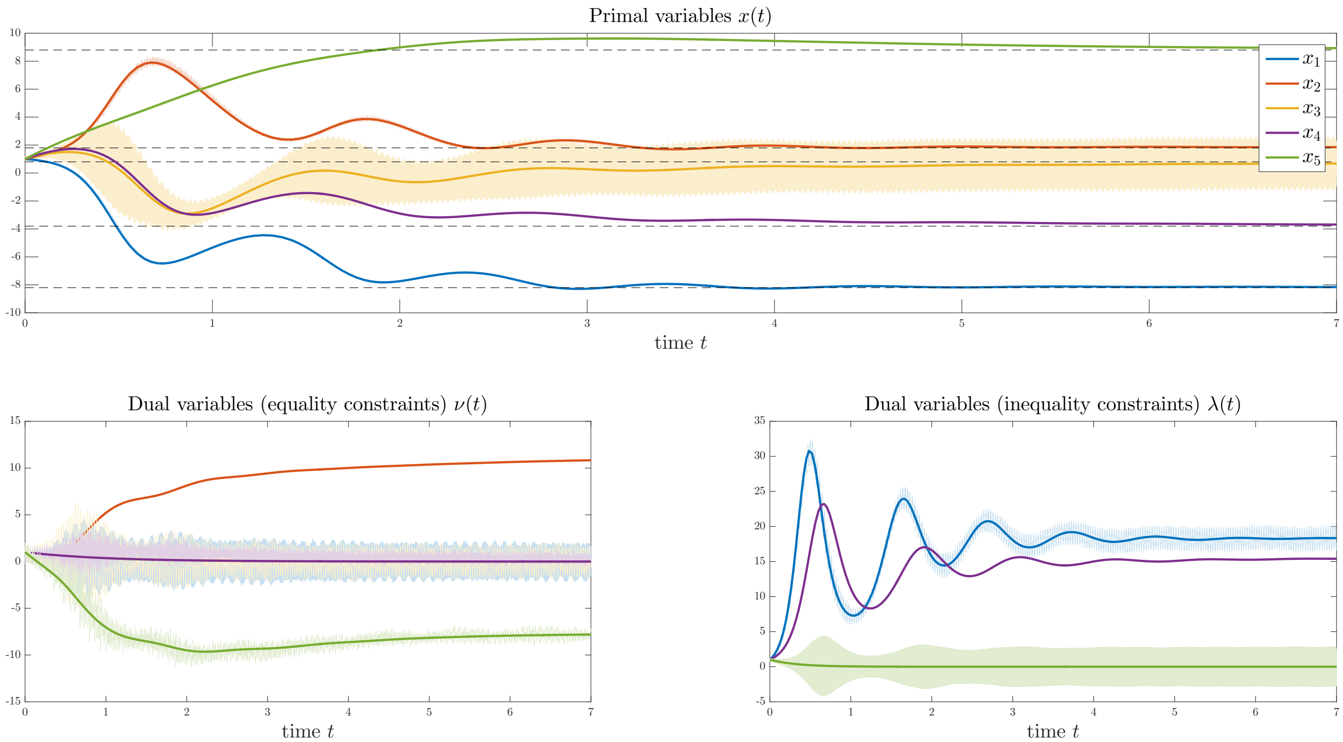

Figure 4. Simulation results for the example of Section 5.2 with communication graph (b) given in Figure 3. The thick lines depict the trajectories of the (non-distributed) saddle-point dynamics with initial condition $z(0) = \mathbf{1} \in \mathbb{R}^{15}$, whereas the thinner oscillating lines depict the solution of the distributed approximation with the same initial condition $z^\sigma(0) = z(0)$. Where no oscillating lines are visible, they are covered by the corresponding component of the solution $z(\cdot)$. The dashed black lines indicate the optimal solution of the optimization problem given by $x^\star = [-8.2,\ 1.8,\ 0.8,\ -3.8,\ 8.8]^\top$. The frequencies required in the approximating inputs were chosen according to some heuristics with absolute value in the range between 2.6309 and 62.6381 and making sure that the minimally canceling property is fulfilled. Further, the sequence parameter was set to $\sigma = 1000$.

[11] Hans-Bernd Dürr. *Constrained Extremum Seeking: A Lie bracket and singular perturbation approach*. Phd thesis, University of Stuttgart, 2015.

[12] Hans-Bernd Dürr, Milos S. Stankovic, Christian Ebenbauer, and Karl Henrik Johansson. Lie bracket approximation of extremum seeking systems. *Automatica*, 49(6):1538 – 1552, 2013.

[13] Hans-Bernd Dürr and Christian Ebenbauer. On a class of smooth optimization algorithms with applications in control. *IFAC Proceedings Volumes*, 45(17):291 – 298, 2012. 4th IFAC Conference on Nonlinear Model Predictive Control.

[14] Christian Ebenbauer, Simon Michalowsky, Victoria Grushkovskaya, and Bahman Gharesifard. Distributed optimization over directed graphs with the help of Lie brackets. In *Proc. 20th IFAC World Congress*, pages 15908–15913, 2017.

[15] Diego Feijer and Fernando Paganini. Stability of primal-dual gradient dynamics and applications to network optimization. *Automatica*, 46(12):1974–1981, 2010.

[16] Martin Geidl and Göran Andersson. Optimal power dispatch and conversion in systems with multiple energy carriers. In *Proc. 15th Power Systems Computation Conference (PSCC)*. Citeseer, 2005.

[17] Bahman Gharesifard. Stabilization of bilinear sparse matrix control systems using periodic inputs. *Automatica*, 77(Supplement C):239 – 245, 2017.

[18] Bahman Gharesifard and Jorge Cortés. Distributed continuous-time convex optimization on weight-balanced digraphs. *IEEE Transactions on Automatic Control*, 59(3):781–786, 2014.

[19] Victoria Grushkovskaya, Alexander Zuyev, and Christian Ebenbauer. On a class of generating vector fields for the extremum seeking problem: Lie bracket approximation and stability properties. *arXiv preprint arXiv:1703.02348*, 2017.

[20] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Convex analysis and minimization algorithms I: Fundamentals*, volume 305. Springer science & business media, 2013.

[21] Wensheng Liu. An approximation algorithm for nonholonomic systems. *SIAM Journal on Control and Optimization*, 35(4):1328–1365, 1997.

[22] Wensheng Liu. Averaging theorems for highly oscillatory differential equations and iterated Lie brackets. *SIAM journal on control and optimization*, 35(6):1989–2020, 1997.

[23] Simon Michalowsky. A MATLAB toolbox for the Lie bracket approximation approach to distributed optimization. http://www.ist.uni-stuttgart.de/public/SM17-toolboxDistOpt.zip, 2017.

[24] Simon Michalowsky, Bahman Gharesifard, and Christian Ebenbauer. Approximating sequences for the excitation of iterated Lie brackets – A reformulation of the Liu/Sussmann algorithm. http://www.ist.uni-stuttgart.de/public/SM17-approxSeq.pdf, 2017.

[25] Simon Michalowsky, Bahman Gharesifard, and Christian Ebenbauer. Distributed extremum seeking over directed graphs. In *56th IEEE Conference on Decision and Control, to appear*, 2017.

[26] Simon Michalowsky, Bahman Gharesifard, and Christian Ebenbauer. A Lie bracket approximation approach to distributed optimization over directed graphs. *Submitted*, 2017.

[27] Angelia Nedić and Alex Olshevsky. Distributed optimization over time-varying directed graphs. *IEEE Transactions on Automatic Control*, 60(3):601–615, 2015.

[28] S. K. Niederländer and J. Cortés. Distributed coordination for separable convex optimization with coupling constraints. In *54th IEEE Conference on Decision and Control (CDC)*, pages 694–699, Dec 2015.

[29] Christophe Reutenauer. Free Lie algebras. *Handbook of algebra*, 3:887–903, 2003.

[30] Héctor J Sussmann and Wensheng Liu. Limits of highly oscillatory controls and the approximation of general paths by admissible trajectories. In *30th IEEE Conference on Decision and Control*, pages 437–442. IEEE, 1991.

[31] Behrouz Touri and Bahman Gharesifard. Saddle-point dynamics for distributed convex optimization on general directed graphs. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 862–866, 2016.

[32] Gerd Wachsmuth. On LICQ and the uniqueness of Lagrange multipliers. *Operations Research Letters*, 41(1):78–80, 2013.

[33] Jing Wang and Nicola Elia. A control perspective for centralized and distributed convex optimization. In *2011 IEEE 50th Conference on Decision and Control and European Control Conference (CDC-ECC)*, pages 3800–3805, 2011.

A. Appendix

A.1. Proof of Lemma 1

Proof. The proof follows a similar argument as the one in [11, Theorem 5.1.3]. First, using (11c), we have

$$\lambda_i(t) = \exp\big(\int_0^t \big(a_i x(\tau) - b_i\big)\mathrm{d}\tau\big)\lambda_i(0), \tag{65}$$

for all $i = 1, 2, \ldots, n$; hence, $\lambda_i(0) > 0$ implies that $\lambda_i(t) > 0$, for all $t \geq 0$, and consequently, the set $\mathcal{R}(\mathcal{M})$ is positively invariant. Let $(x^\star, \nu^\star, \lambda^\star)$ be an arbitrary point in $\mathcal{M}$. Consider the candidate Lyapunov function $V : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n_+ \to \mathbb{R}_{+,0}$ defined as

$$\begin{aligned} V(x,\nu,\lambda) &= \tfrac{1}{2}\|x - x^\star\|^2 + \tfrac{1}{2}\|\nu - \nu^\star\|^2 \\ &\quad + \sum_{i=1}^n (\lambda_i - \lambda_i^\star) - \sum_{i:\lambda_i^\star \neq 0}^n \lambda_i^\star \ln\big(\tfrac{\lambda_i}{\lambda_i^\star}\big). \end{aligned} \tag{66}$$

We first observe that V is positive definite with respect to $(x^\star, \nu^\star, \lambda^\star)$ on $\mathcal{R}(\mathcal{M})$, and that all the level sets are compact. To see this, note that according to [6, p. 207, eq. (1.5)], the function $D : \mathbb{R}^n_{++} \times \mathbb{R}^n_{++} \to \mathbb{R}$ defined as

$$D(\lambda^\star, \lambda) = \sum_{i=1}^n \big(\lambda_i - \lambda_i^\star + \lambda_i^\star(\ln(\lambda_i^\star) - \ln(\lambda_i))\big) \tag{67}$$

$$= \sum_{i=1}^n (\lambda_i - \lambda_i^\star) - \sum_{i:\lambda_i^\star \neq 0}^n \lambda_i^\star \ln\big(\tfrac{\lambda_i}{\lambda_i^\star}\big) \tag{68}$$

is positive for all $(\lambda^\star, \lambda) \in \mathbb{R}^n_{++} \times \mathbb{R}^n_{++}$ and zero if and only if $\lambda = \lambda^*$ [6, Condition I.] and its level sets are compact [6, Condition V.]. Thus, with $V(x,\nu,\lambda)$ additionally being quadratic in x and ν, positive definiteness and compactness of all level sets follows and hence V is uniformly unbounded on $\mathcal{R}(\mathcal{M})$. The derivative of V along the trajectories of (11) is then given by

$$
\begin{aligned}
&\dot{V}(x,\nu,\lambda) &(69)\\
&= -(x-x^\star)^\top\left(\nabla F(x) + A^\top \nu + C^\top \lambda\right)\\
&+ (\nu-\nu^\star)^\top(Ax-b)\\
&- \sum_{\substack{i=1\\ i\notin\mathcal{I}_{\mathrm{eq}}}}^{n} \nu_i^2 + \sum_{i=1}^{n}\lambda_i(c_ix-d_i) - \sum_{i:\lambda_i^\star\neq 0}^{n}\lambda_i(c_ix-d_i)\\
&= -(x-x^\star)^\top\nabla F(x) - \nu^\top\left(Ax-b-(Ax^\star-b)\right)\\
&- \lambda^\top\left(Ax-b-(Ax^\star-b)\right) + (\nu-\nu^\star)^\top(Ax-b)\\
&- \sum_{\substack{i=1\\ i\notin\mathcal{I}_{\mathrm{eq}}}}^{n} \nu_i^2 + \sum_{i=1}^{n}(\lambda_i-\lambda_i^\star)(c_ix-d_i) + F(x) - F(x),
\end{aligned}
$$

Using strict convexity of F, we now have that $-(x-x^\star)^\top\nabla F(x) < F(x^\star) - F(x)$, for all $x \neq x^\star$ and hence we obtain for all $x \neq x^\star$

$$
\begin{aligned}
&\dot{V}(x,\nu,\lambda) &(70)\\
&< F(x^\star) - F(x) - \nu^\top\left(Ax-b-(Ax^\star-b)\right)\\
&- \lambda^\top\left(Cx-d-(Cx^\star-d)\right) + (\nu-\nu^\star)^\top(Ax-b)\\
&- \sum_{\substack{i=1\\ i\notin\mathcal{I}_{\mathrm{eq}}}}^{n} \nu_i^2 + \sum_{i=1}^{n}(\lambda_i-\lambda_i{}^\star)(c_ix-d_i) + F(x) - F(x)\\
&= L(x^\star,\nu,\lambda) - L(x,\nu,\lambda) + L(x,\nu,\lambda) - L(x,\nu^\star,\lambda^\star)\\
&- \sum_{\substack{i=1\\ i\notin\mathcal{I}_{\mathrm{eq}}}}^{n} \nu_i^2\\
&= L(x^\star,\nu,\lambda) - L(x,\nu^\star,\lambda^\star) - \sum_{\substack{i=1\\ i\notin\mathcal{I}_{\mathrm{eq}}}}^{n} \nu_i^2. &(71)
\end{aligned}
$$

Due to the saddle point property (10) the derivative of V along the flow is strictly negative, for all (x,ν,λ) except for $(x,\nu,\lambda)\in\mathcal{M}$; thus, $(x^\star,\nu^\star,\lambda^\star)$ is stable according to [11, Theorem 2.2.2]. This procedure can be repeated for any point $(x^\star,\nu^\star,\lambda^\star)\in\mathcal{M}$, hence $\mathcal{M}$ is stable. Let L_{orig} denote the Lagrangian associated to the original problem (4) and let $\mathcal{S}_{\mathrm{orig}}$ denote the corresponding set of saddle points. Observe that $L(x,\nu,\lambda) = L_{\mathrm{orig}}(x,\nu_{\mathcal{I}_{\mathrm{eq}}},\lambda_{\mathcal{I}_{\mathrm{ineq}}}) - \sum_{i=1,i\notin\mathcal{I}_{\mathrm{ineq}}}^{n}\lambda_i d_i$ such that $\lambda_i^\star = 0$ for all $i = 1,2,\ldots,n$, $i\notin\mathcal{I}_{\mathrm{ineq}}$, for any saddle point $(x^\star,\nu^\star,\lambda^\star)$ of L, since $b_i > 0$ for $i = 1,2,\ldots,n$, $i\notin\mathcal{I}_{\mathrm{ineq}}$. Thus, the set of saddle points of L is given by

$$
\begin{aligned}
\mathcal{S} = \{\ &(x,\nu,\lambda)\in\mathrm{R}^n\times\mathrm{R}^n\times\mathrm{R}^n_+ : &(72)\\
&(x,\nu_{\mathcal{I}_{\mathrm{eq}}},\lambda_{\mathcal{I}_{\mathrm{ineq}}})\in\mathcal{S}_{\mathrm{orig}}, \lambda_i = 0 \text{ for } i\notin\mathcal{I}_{\mathrm{ineq}}\}
\end{aligned}
$$

and hence, $\mathcal{M} = \{(x,\nu,\lambda)\in\mathrm{R}^n\times\mathrm{R}^n\times\mathrm{R}^n : (x,\nu_{\mathcal{I}_{\mathrm{eq}}},\lambda_{\mathcal{I}_{\mathrm{ineq}}})\in\mathcal{S}$ and $\nu_i = 0$ for $i\notin\mathcal{I}_{\mathrm{eq}}, \lambda_i = 0$ for $i\notin\mathcal{I}_{\mathrm{ineq}}\}$. Since $\mathcal{S}_{\mathrm{orig}}$ is compact due to Assumption 2, the set $\mathcal{M}$ is compact as well. The same argument as the one in the proof of [11, Theorem 5.1.3] then yields that the set of saddle points is asymptotically stable with respect to the set of initial conditions $\mathcal{R}(\mathcal{M})$. □

A.2. Proof of Lemma 3

Proof. We first observe first that (38) is the same as (27), (29) with a special choice of the subpath as well as an additional projection with the property $\mathrm{proj}_{\mathcal{PH}}(B)(z) = B(z)$ for all $z\in\mathrm{R}^{3n}$. Hence, it immediately follows that $\tilde{R}_{k_1,k_2}(p_{i_1i_r})(z) = R_{k_1,k_2}(p_{i_1i_r})(z)$. In the same manner, we also have that

$$\delta\big(\tilde{R}_{k_1,k_2}(p_{i_1i_r})\big) = \delta\big(R_{k_1,k_2}(p_{i_1i_r})\big) = \ell(p_{i_1i_r}). \qquad (73)$$

We show the second part by induction. First observe that for paths $p_{i_1i_r}$ with $\ell(p_{i_1i_r}) = 1$ it is clear that $\tilde{R}_{k_1,k_2}(p_{i_1i_r})\in\mathrm{P}$ since $R_{k_1,k_2}(p_{i_1i_r})$ is an admissible vector field by Lemma 2 and all admissible vector fields are in P. Further, for paths $p_{i_1i_r}$ with $\ell(p_{i_1i_r})\in\{2,3,4,6\}$ it is also follows from the definition of the projection operator that $\tilde{R}_{k_1,k_2}(p_{i_1i_r})\in\mathrm{P}$. Suppose now that the result holds true for all paths p with $\ell(p) = \bar{\ell}$, where $\bar{\ell}\geq 2$, and consider a path $p_{i_1i_r}$ with $\ell(p_{i_1i_r}) = \bar{\ell}+1$. Observe that all subbrackets of $\tilde{R}_{k_1,k_2}(p_{i_1i_r})$ are in P by the induction hypothesis and hence, by [PH3], [PH4.a], [PH4.b], we have $\tilde{R}_{k_1,k_2}(p_{i_1i_r})\in\mathrm{P}$ if

$$\delta\big(\mathrm{left}(\tilde{R}_{k_1,k_2}(p_{i_1i_r}))\big) < \delta\big(\mathrm{right}(\tilde{R}_{k_1,k_2}(p_{i_1i_r}))\big) \qquad (74)$$

$$\delta\big(\mathrm{left}(\mathrm{right}(\tilde{R}_{k_1,k_2}(p_{i_1i_r})))\big) < \delta\big(\mathrm{left}(\tilde{R}_{k_1,k_2}(p_{i_1i_r}))\big); \qquad (75)$$

we will show next that these conditions are fulfilled for for the above choice of subpaths. By (38) and (40) we have that

$$
\begin{aligned}
\delta\big(\mathrm{right}(\tilde{R}_{k_1,k_2}(p_{i_1i_r}))\big) &= \delta\big(\tilde{R}_{s,k_2}(q)\big) = \ell(q)\\
\delta\big(\mathrm{left}(\tilde{R}_{k_1,k_2}(p_{i_1i_r}))\big) &= \delta\big(\tilde{R}_{k_1,s}(q^c)\big) = \ell(p_{i_1i_r}) - \ell(q).
\end{aligned}
$$

Since $\lfloor\frac{a}{b}\rfloor \geq \frac{a-b+1}{b}$, for all $a\in\mathrm{Z}, b\in\mathrm{N}$, we have that

$$\ell(q) = \theta(p_{i_1i_r}) - 1 \geq \frac{\ell(p_{i_1i_r})+1}{2}, \qquad (76)$$

for $\ell(p_{i_1i_r})\geq 5$, and hence we obtain

$$
\begin{aligned}
&\delta\big(\mathrm{right}(\tilde{R}_{k_1,k_2}(p_{i_1i_r}))\big) - \delta\big(\mathrm{left}(\tilde{R}_{k_1,k_2}(p_{i_1i_r}))\big)\\
&\qquad \geq \ell(p_{i_1i_r}) + 1 - \ell(p_{i_1i_r}) > 0. \qquad (77)
\end{aligned}
$$

Thus, (74) holds. For (75), we first note that

$$\delta\big(\mathrm{left}(\mathrm{right}(\tilde{R}_{k_1,k_2}(p_{i_1i_r})))\big) = \delta\big(\mathrm{left}(\tilde{R}_{s,k_2}(q))\big) \qquad (78)$$

and, since $\text{left}(\tilde{R}_{s,k_2}(q)) \in \mathrm{P}$ by the induction hypothesis, it is $\delta\big(\text{left}(\tilde{R}_{s,k_2}(q))\big) \leq \delta\big(\text{right}(\tilde{R}_{s,k_2}(q))\big) = \ell(q) - \delta\big(\text{left}(\tilde{R}_{s,k_2}(q))\big)$ according to [PH4.a]. Hence, we obtain

$$\delta\big(\text{left}(\text{right}(\tilde{R}_{k_1,k_2}(p_{i_1 i_r})))\big) \leq \tfrac{\ell(q)}{2}. \tag{79}$$

As a result, (75) is fulfilled when

$$\tfrac{\ell(q)}{2} \leq \ell(p_{i_1 i_r}) - \ell(q). \tag{80}$$

We now compute

$$\tfrac{3}{2}\ell(q) = \tfrac{3}{2}\lfloor \tfrac{\ell(p_{i_1 i_r})}{2} \rfloor + \tfrac{3}{2} \leq \tfrac{3}{4}\ell(p_{i_1 i_r}) + \tfrac{3}{2} \leq \ell(p_{i_1 i_r}),$$

for $\ell(p_{i_1 i_r}) \geq 6$; for $\ell(p_{i_1 i_r}) = 5$, we have that $\frac{3}{2}\ell(q) = \frac{9}{2} < \ell(p_{i_1 i_r})$, thus (80) holds for all considered $p_{i_1 i_r}$ which proves that (75) holds; this concludes the proof. □

A.3. Proof of Proposition 2

Proof. It is clear that (55) holds for $\ell(p_{ij}) = 2$, since $\tilde{R}_{r+j,i}(p_{ij})$ is a bracket of degree two, i.e., a bracket of the form $[\phi_{k_1}, \phi_{k_2}]$, $k_1 \neq k_2$, such that

$$E\big(\text{proj}_{\mathcal{PH}}([\phi_{k_1}, \phi_{k_2}])\big) = \begin{cases} [\phi_{k_1}, \phi_{k_2}] \text{ if } k_1 < k_2 \\ [\phi_{k_2}, \phi_{k_1}] \text{ if } k_2 < k_1. \end{cases} \tag{81}$$

Consider now a path $p_{i_1 i_4} = \langle v_{i_1} | v_{i_2} | v_{i_3} | v_{i_4} \rangle$, $i_1 \neq i_2 \neq i_3 \neq i_4$, i.e., $\ell(p_{i_1 i_4}) = 3$. Then

$$\begin{aligned} &\tilde{R}_{r+i_4,i_1}(p_{i_1 i_4}) \\ &= \text{proj}_{\mathcal{PH}}\big(R_{r+i_4,i_1}(p_{i_1 i_4})\big) \\ &= \text{proj}_{\mathcal{PH}}\big([h_{r+i_4,r+i_3}, [h_{r+i_3,r+i_2}, h_{r+i_2,i_1}]]\big) \\ &= -\big[h_{r+i_4,r+i_3}, [h_{r+i_2,i_1}, h_{r+i_3,r+i_2}]\big], \end{aligned} \tag{82}$$

where we have used the assumption on the ordering of the P. Hall basis. The only equivalent bracket in P is then given by $B = \big[h_{r+i_3,r+i_2}, [h_{r+i_2,i_1}, h_{r+i_4,r+i_3}]\big]$, but we have that $B(z) \equiv 0$, since

$$\begin{aligned} &[h_{r+i_2,i_1}, h_{r+i_4,r+i_3}](z) \\ &= e_{r+i_3} e_{r+i_4}^\top e_{i_1} z_{r+i_2} - e_{i_1} e_{r+i_2}^\top e_{r+i_3} z_{+i_2} = 0. \end{aligned} \tag{83}$$

Thus, the claim follows. □

A.4. A simplified algorithm for the construction of approximating sequences

Our objective in this section is to provide a modified version of the construction procedure from [21] using the structural properties of the problem at hand, which leads to considerable simplifications. Given the scopes of this paper and the complicated nature of the subject, we do not discuss this algorithm in detail; we refer the reader to [24], as well as the original work [21]. We first provide a formal definition of the already mentioned equivalence relation on the set of Lie brackets:

Definition 3. [Equivalent brackets] Let $\mathcal{PH} = (\mathrm{P}, \prec)$ be a P. Hall basis of $\Phi = \{\phi_1, \ldots, \phi_M\}$ and let $\delta_k(B)$ denote the degree of the vector field ϕ_k in the bracket $B \in \mathcal{PH}$. We say that two brackets $B_1, B_2 \in \mathcal{P}$ are equivalent, denoted by $B_1 \sim B_2$, if $\delta_k(B_1) = \delta_k(B_2)$ for all $k = 1, \ldots, M$. •

For the construction of the sets of frequencies, we also need the following two definitions:

Definition 4. [Minimally canceling] A set $\Omega = \{\omega_1, \ldots, \omega_m\}$ is called minimally canceling if for each collection of integers $\{y_i\}_{i=1}^m$, such that $\sum_{k=1}^m |y_k| \leq m$ we have $\sum_{k=1}^m y_k \omega_k = 0$ if and only if all y_k are equal. •

Definition 5. [Independent collection] A finite collection of sets $\{\Omega_\lambda\}_{\lambda=1}^N$, where $\Omega_\lambda = \{\omega_{\lambda,1}, \omega_{\lambda,2} \ldots, \omega_{\lambda,M_\lambda}\}$, is called independent if the followings hold:

1. the sets Ω_λ are pairwise disjoint, and
2. for each collection of integers $\{y_{i,k}\}_{i=1}^N$, $k = 1, \ldots, M_i$, such that
$$\sum_{i=1}^N \sum_{k=1}^{M_i} y_{i,k}\omega_{i,k} = 0 \quad \text{and} \quad \sum_{i=1}^N \sum_{k=1}^{M_i} |y_{i,k}| \leq \sum_{i=1}^N M_i$$
we have
$$\sum_{k=1}^{M_i} y_{i,k}\omega_{i,k} = 0, \tag{84}$$
for each $i = 1, 2, \ldots, N$. •

Consider now an extended system of the form

$$\dot{z} = f_0(z) + \sum_{\substack{B \in \mathcal{B} \\ \delta(B) \geq 2}} v_B B(z), \tag{85}$$

where $f_0 : \mathrm{R}^N \to \mathrm{R}^N$, $\mathcal{B} \subset \mathrm{P}$ for some P. Hall basis $\mathcal{PH}(\Phi) = (\mathrm{P}, \prec)$, $\Phi = \{\phi_1, \phi_2, \ldots, \phi_M\}$, $\phi_k : \mathrm{R}^N \to \mathrm{R}^N$, f_0, ϕ_k sufficiently smooth, and $v_B \in \mathrm{R} \setminus \{0\}$. Suppose that for any $B \in \mathcal{B}$, we have that $\delta_k(B) \in \{0, 1\}$, $k = 1, 2, \ldots, M$. Consider the system

$$\dot{X}^\sigma = f_0(X^\sigma) + \sum_{k=1}^M \phi_k(X^\sigma) U_k^\sigma(t). \tag{86}$$

The following algorithm allows to compute suitable input functions U_k^σ such that the solutions of (86) uniformly converge to those of (85) with increasing σ. It should as well be mentioned that we also provide an exemplary implementation of the algorithm in Matlab which is available at [23].

$\delta(B)$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$\lvert E_{\text{full}}(B)\rvert$	1	2	3!	4!	5!	6!	7!	8!	9!	10!	11!	12!	13!	14!	15!	16!
$\lvert E(B)\rvert$	1	1	2	3	5	9	16	28	51	93	170	315	585	1089	2048	3855

Table 2. A comparison of $|E_{\text{full}}(B)|$ and $|E(B)|$ for a specific choice of the P. Hall basis that fulfills the assumptions as in Proposition 2. The numbers were obtained by symbolically computing the resulting vector fields using a computer algebra system. Interestingly, the sequence of $|E(B)|$ has two matching sequences [1] and [2] except for the value for $\delta(B) = 15$ which should be 1091 or 1092, thus we conjecture that these sequences are a good upper bound for $|E(B)|$.

Algorithm

Step 1: For all $B \in \mathcal{B}$, determine the associated (reduced) equivalence class

$$\begin{aligned} E(B) &= \{\tilde{B} \in \mathcal{P} : \tilde{B} \sim B, \tilde{B}(z) \not\equiv 0\} \\ &= \{\tilde{B}_{E,1}, \tilde{B}_{E,2}, \dots, \tilde{B}_{E,|E(B)|}\}, \end{aligned}$$

and let $\mathcal{E} = \{E(B), B \in \mathcal{B}\}$. For each $E \in \mathcal{E}$, set

$$\tilde{v}_B = \begin{cases} v_B & \text{if } B \in \mathcal{B} \\ 0 & \text{otherwise.} \end{cases}$$

Step 2: For all $E \in \mathcal{E}$, $\delta(E) = 2$, choose $|\mathcal{E}_2|$ distinct frequencies $\omega_E \in \mathrm{R}$, and for all $E \in \mathcal{E}$, $\delta(E) \geq 3$ choose $|E|$ sets

$$\begin{aligned} \Omega^+_{E,\rho,k} &= \begin{cases} \omega_{E,\rho,k} & \text{if } \delta_k(E) = 1 \\ \emptyset & \text{if } \delta_k(E) = 0 \end{cases} \\ \Omega^-_{E,\rho,k} &= -\Omega^+_{B,\rho,k}, \end{aligned}$$

$k = 1, \dots, M$, $\rho = 1, \dots, |E(B)|$, such that

1. For each $E \in \mathcal{E}$, $\delta(E) \geq 3$, and each $\rho = 1, \dots, |E|$, the set $\Omega^+_{E,\rho} = \bigcup_{k=1}^{M} \Omega^+_{E,\rho,k}$ is minimally canceling.
2. The collection of sets
$$\Big\{ \{\omega_E, -\omega_E\}_{E \in \mathcal{E}_2}, \{\Omega^+_{E,\rho} \cup \Omega^-_{E,\rho}\}_{\substack{E \in \mathcal{E}, \delta(E) \geq 3, \\ \rho = 1,\dots,|E(B)|}} \Big\}$$
is independent.

Step 3: For all $E \in \mathcal{E}$ with $\delta(E) \geq 3$, compute

$$\Xi_E = \begin{bmatrix} \xi^+_{\tilde{B}_{E,1},1} & \xi^+_{\tilde{B}_{E,1},2} & \cdots & \xi^+_{\tilde{B}_{E,1},|E|} \\ \xi^+_{\tilde{B}_{E,2},1} & \xi^+_{\tilde{B}_{E,2},2} & \cdots & \xi^+_{\tilde{B}_{E,2},|E|} \\ \vdots & \vdots & \ddots & \vdots \\ \xi^+_{\tilde{B}_{E,|E|},1} & \xi^+_{\tilde{B}_{E,|E|},2} & \cdots & \xi^+_{\tilde{B}_{E,|E|},|E|} \end{bmatrix},$$

where

$$\xi^+_{B,\rho} = \hat{g}_B(\omega_{E,\rho,\theta_B(1)}, \omega_{E,\rho,\theta_B(2)}, \dots, \omega_{E,\rho,\theta_B(\delta(B))}),$$

with $\theta_B(i) = k$ if the ith vector field in B is ϕ_k and where $\hat{g}_B : \mathrm{R}^{\delta(B)} \to \mathrm{R}$ is defined as follows:

- If $\delta(B) = 1$, then $\hat{g}_B(\tilde{\omega}_1) = 1$.
- If $B = [B_1, B_2]$, then

$$\begin{aligned} \hat{g}_B(\tilde{\omega}_1, \tilde{\omega}_2, \dots, \tilde{\omega}_{\delta(B)}) &= \frac{\hat{g}_{B_2}(\tilde{\omega}_1, \tilde{\omega}_2, \dots, \tilde{\omega}_{\delta(B_1)})}{\sum_{i=1}^{\delta(B_1)} \tilde{\omega}_i} \\ &\times \hat{g}_{B_2}(\tilde{\omega}_{\delta(B_1)+1}, \tilde{\omega}_{\delta(B_1)+2}, \dots, \tilde{\omega}_{\delta(B_1)+\delta(B_2)}). \end{aligned}$$

Step 4: For all $E \in \mathcal{E}$ with $\delta(E) = 2$, i.e., $E(B) = \{B\} = [\phi_{k_1}, \phi_{k_2}]$, set

$$\begin{aligned} \eta_{E,k_1}(\omega_E) &= \mathrm{i}\tfrac{1}{\beta_E}\operatorname{sign}(\tilde{v}_B\omega_E)\sqrt{\tfrac{1}{2}|\tilde{v}_B\omega_E|} \\ \eta_{E,k_2}(\omega_E) &= \beta_E\sqrt{\tfrac{1}{2}|\tilde{v}_B\omega_E|}, \end{aligned}$$

where $\beta_E \neq 0$. For all $E \in \mathcal{E}$ with $\delta(E) \geq 3$ let

$$\begin{bmatrix} \gamma_{E,1} \\ \gamma_{E,2} \\ \vdots \\ \gamma_{E,|E|} \end{bmatrix} = \Xi_E^{-1} \begin{bmatrix} \tilde{v}_{\tilde{B}_{E,1}} \\ \tilde{v}_{\tilde{B}_{E,2}} \\ \vdots \\ \tilde{v}_{\tilde{B}_{E,|E|}} \end{bmatrix}$$

and compute $\eta_E(\omega)$ as follows:

- If $\delta(E)$ is odd, take
$$\eta_E(\omega) = \beta_{E,\omega}\left(\tfrac{1}{2}\gamma_{E,\rho}\mathrm{i}^{\delta(E)-1}\right)^{\frac{1}{\delta(E)}}$$
for all $\omega \in \Omega^+_{E,\rho}$, and
- if $\delta(E)$ is even, take
$$\eta_E(\tilde{\omega}) = \mathrm{i}\beta_{E,\omega}\operatorname{sign}(\gamma_{E,\rho}(t)\mathrm{i}^{\delta(B)-2})\left|\tfrac{1}{2}\gamma_{E,\rho}(t)\mathrm{i}^{\delta(B)-2}\right|^{\frac{1}{\delta(B)}}$$
for some $\tilde{\omega} \in \Omega^+_{E,\rho}$ and
$$\eta_E(\omega) = \beta_{E,\omega}\left|\tfrac{1}{2}\gamma_{E,\rho}(t)\mathrm{i}^{\delta(B)-2}\right|^{\frac{1}{\delta(B)}}$$
for all $\omega \in \Omega^+_{E,\rho} \setminus \{\tilde{\omega}\}$.

In both cases $\beta_{E,\omega} \in \mathrm{R}$ can be chosen freely such that it fulfills $\prod_{\omega \in \Omega^+_{E,\rho}} \beta_{E,\omega} = 1$.

Step 5: Compute the input according to $U^\sigma_k(t) = \sum_{E \in \mathcal{E}} U^\sigma_{k,E}(t)$ with $U^\sigma_{k,E} : \mathrm{R} \to \mathrm{R}$ being defined as follows:

- If $\delta_k(E) = 0$: $U^\sigma_{k,E}(t) = 0$.
- If $\delta(E) = 2, \delta_k(E) = 1$:
$$U^\sigma_{k,E}(t) = 2\sqrt{\sigma}\mathrm{Re}\big(\eta_{E,k}(\omega_E)e^{\mathrm{i}\sigma\omega_E t}\big).$$
- If $\delta(E) = N$, $\delta_k(E) = 1$:
$$U^\sigma_{k,E}(t) = 2\sigma^{\frac{N-1}{N}} \sum_{\rho=1}^{|E|} \mathrm{Re}\big(\eta_E(\omega_{E,\rho,k})e^{\mathrm{i}\sigma\omega t}\big).$$

Note that this algorithm is a reformulation of the one presented in [21] (see [24] for a derivation) exploiting two structural properties of the problem at hand: (1) each $B \in \mathcal{B}$ fulfills $\delta_k(B) \in \{0, 1\}$ for all $k = 1, 2, \ldots, M$ and (2) a large number of the equivalent brackets evaluate to zero (see Table 2). Note that (1) simplifies the calculation of $\xi^+_{B,\rho}$ in step 3 and (2) reduces the cardinality of each $E(B)$ in step 1, where usually the full equivalence class $E_{\mathrm{full}}(B) = \{\tilde{B} \in \mathcal{B} : \tilde{B} \sim B\}$ is used, thus leading to a reduction of the dimension of Ξ_E in step 3 and hence also simplifying step 4. In fact, we can derive the following result on the equivalent brackets:

Lemma 4. Consider graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ of n nodes. Let $p_{i_1 i_r} = (v_{i_1}, v_{i_2}, \ldots, v_{i_r})$ be the shortest path between v_{i_1} and v_{i_r}, $v_{i_k} \in \mathcal{V}$ for $k = 1, 2, \ldots, r$, $r \geq 3$. Let $\Phi = \{\phi_{a_1}, \phi_{a_2}, \ldots, \phi_{a_{r-1}}\}$ be a set of vector fields with

$$\phi_{a_j} \in \{h_{k_1,k_2} : k_1 \in \mathcal{I}(i_{j+1}), k_2 \in \mathcal{I}(i_j)\}, \tag{87}$$

for $j = 1, 2, \ldots, r-1$. Denote some given P. Hall basis of Φ by $\mathcal{PH}(\Phi) = (\mathrm{P}, \prec)$. Let $B \in \mathrm{P}$ and suppose that $\delta_{a_j}(B) \in \{0, 1\}$ for $j = 1, 2, \ldots, r-1$. Define $\mathcal{J}(B) = \{j = 1, 2, \ldots, r-1 : \delta_{a_j}(B) = 1\}$ and further denote

$$j_{\min}(B) = \min_{j \in \mathcal{J}(B)} \{j\}, \quad j_{\max}(B) = \max_{j \in \mathcal{J}(B)} \{j\}. \tag{88}$$

Then, if $\mathcal{J}(B)$ is a connected set, i.e., $\mathcal{J}(B) = \{j_{\min}(B), j_{\min}(B) + 1, \ldots, j_{\min}(B) + \delta(B) - 1\}$ and $j_{\max}(B) = j_{\min}(B) + \delta(B) - 1$, for any $k_1 \in \mathcal{I}(i_{j_{\max}(B)+1})$, $k_2 \in \mathcal{I}(i_{j_{\min}(B)})$ and for all $z \in \mathrm{R}^{3n}$, we have that

$$B(z) = \pm h_{k_1,k_2}(z) \quad \text{or} \quad B(z) = 0. \tag{89}$$

If $\mathcal{J}(B)$ is not a connected set, we have $B(z) = 0$ for all $z \in \mathrm{R}^{3n}$. •

Proof. We prove this result by induction. Suppose first that $\delta(B) = 1$. Then $\mathcal{J}(B) = \{j_{\min}\} = \{j_{\max}\}$, which means it has only one element. Hence, the claim is obviously true. Since the case of $\mathcal{J}(B)$ not being a connected set does not appear for $\delta(B) = 1$, we also look at $\delta(B) = 2$. Let $\mathcal{J}(B) = \{j_1, j_2\}$, $j_1 \neq j_2$. Observe that, for all $j_1, j_2 = 1, 2, \ldots, r-1$, $j_1 \neq j_2$, and $j_1 \leq r-2$ (or $j_2 \leq r-1$), we have

$$\begin{aligned} B(z) &= [\phi_{a_{j_1}}, \phi_{a_{j_2}}](z) \\ &= [h_{k_1,k_2}, h_{k_3,k_4}](z) \\ &= [z_{k_1} e_{k_2}, z_{k_3} e_{k_4}] \\ &= e_{k_4} e_{k_3}^\top e_{k_2} z_{k_1} - e_{k_2} e_{k_1}^\top e_{k_4} z_{k_3}, \end{aligned} \tag{90}$$

where $k_1 \in \mathcal{I}(i_{j_1+1}), k_2 \in \mathcal{I}(i_{j_1}), k_3 \in \mathcal{I}(i_{j_2+1})$, and $k_4 \in \mathcal{I}(i_{j_2})$. We then compute

$$[\phi_{a_{j_1}}, \phi_{a_{j_2}}](z) = \begin{cases} z_{k_1} e_{k_4} & \text{if } k_2 = k_3 \\ -z_{k_3} e_{k_2} & \text{if } k_1 = k_4 \\ 0 & \text{otherwise.} \end{cases} \tag{91}$$

Note that $k_2 = k_3$ only if $i_{j_1} = i_{j_2+1}$, i.e., $j_1 = j_2 + 1 = j_{\max}$, $j_{\min} = j_2$, and $k_1 = k_4$ only if $i_{j_2} = i_{j_1+1}$, i.e., $j_2 = j_1 + 1 = j_{\max}$, $j_1 = j_{\min}$; hence $B(z)$ is non-zero only if $\mathcal{J}(B) = \{j_1, j_2\}$ is connected, which proves that the claim is true for $\delta(B) = 2$. Note also that the case $k_1 = k_4, k_2 = k_3$ cannot occur since $j_1 \neq j_2$. The second claim (89) follows immediately from these considerations. To proceed with our induction argument, suppose now that the claim is true for all $B \in \mathrm{P}$ that fulfill the assumptions with $\delta(B) \leq \delta^*$, $\delta^* \leq r-1$. Consider now some $B \in \mathrm{P}$ with $\delta(B) = \delta^* + 1 > 2$. Every B can be written as $B = [B_1, B_2]$, where $\delta(B_1), \delta(B_2) \leq \delta^*$. Let $\mathcal{J}(B) = \{j_1, j_2, \ldots, j_{\delta(B)}\}$ and assume, without loss of generality, that $j_k < j_{k+1}$, for all $k = 1, \ldots, \delta(B) - 1$. By the induction hypothesis, $B_1(z)$ and $B_2(z)$ are non-zero only if $\mathcal{J}(B_1)$ and $\mathcal{J}(B_2)$ are both connected sets. Since $\mathcal{J}(B_2) = \mathcal{J}(B) \setminus \mathcal{J}(B_1)$ this is the case if and only if

$$\begin{aligned} \mathcal{J}(B_1) &= \begin{cases} \{j_1, j_2, \ldots, j_{\delta(B_1)}\} \quad \text{or} \\ \{j_{\delta(B)-\delta(B_1)+1}, j_{\delta(B)-\delta(B_1)+2}, \ldots, j_{\delta(B)}\} \end{cases} \\ &= \begin{cases} \{j_1, j_1 + 1, \ldots, j_1 + \delta(B_1) - 1\}, \quad \text{or} \\ \{j_{\delta(B)-\delta(B_1)+1}, j_{\delta(B)-\delta(B_1)+1} + 1, \ldots, \\ \qquad j_{\delta(B)-\delta(B_1)+1} + \delta(B_1) - 1\}. \end{cases} \end{aligned}$$

We only consider the first case here, since the second case can be treated analogously. Using the first equality above, for $k_1 \in \mathcal{I}(i_{j_1+\delta(B_1)}), k_2 \in \mathcal{I}(i_{j_1})$, and $k_3 \in \mathcal{I}(i_{j_{\delta(B)}+1}), k_4 \in \mathcal{I}(i_{j_{\delta(B_1)+1}})$, we have by the induction hypothesis that

$$B_1(z) = \pm h_{k_1,k_2}(z) \quad \text{or} \quad B_1(z) = 0 \tag{92}$$

$$B_2(z) = \pm h_{k_3,k_4}(z) \quad \text{or} \quad B_2(z) = 0. \tag{93}$$

Obviously, following our previous calculations, $[B_1, B_2]$ is non-zero only if $k_2 = k_3$, meaning that $j_1 = j_{\delta(B)} + 1$, or if $k_1 = k_4$, meaning that $j_1 + \delta(B_1) = j_{\delta(B_1)+1}$. The first case cannot occur, since $\delta(B) > 2$ and $j_{k+1} > j_k$; the second case holds true if and only if $\mathcal{J}(B)$ is connected, thus showing that $B(z)$ is non-zero only if $\mathcal{J}(B)$ is connected. To show that (89) holds, consider the case that $\mathcal{J}(B)$ is connected, i.e., $\mathcal{J}(B) = \{j_1, j_1+1, \ldots, j_1 + \delta(B)\}$, $j_{\min}(B) = j_1$, $j_{\max}(B) = j_1 + \delta(B)$, and $k_1 = k_4$. Then, following the same arguments as before, we have that $B(z) = \pm h_{k_3,k_2}(z)$ for $k_3 \in \mathcal{I}(i_{j_{\delta(B)+1}}) = \mathcal{I}(i_{j_{\max}(B)+1})$, $k_2 \in \mathcal{I}(i_{j_1}) = \mathcal{I}(i_{j_{\min}(B)})$, which concludes the proof. □

Remark 11. The condition that $\mathcal{J}(B)$ must be a connected set can be interpreted as follows: Each admissible vector field ϕ_{a_j} can be associated to an edge in the communication graph $\mathcal{G}$. The condition then means that the vector fields in the bracket must be ordered along a path. •

The algorithm presented beforehand still includes several degrees of freedom, namely the specific choice of frequencies in step 2 as well as the scalings $\beta_E, \beta_{E,\omega}$ in step 4. While the conditions on the frequencies are not hard to satisfy and in fact, are not restrictive, it turns out that their choice is crucial in practical implementations. There is still no constructive way of choosing "good" frequencies that we are aware of in the literature. The situation is similar as it comes to the choice of scalings, but here a heuristic way of how to choose them is to distribute the energy of the approximating inputs among different admissible input vector fields ϕ_k. In this spirit, we suggest decreasing the amplitudes of the approximating inputs entering in the primal variables, which will lead to an increase of the amplitudes of the inputs entering in the dual variables. Our simulations results indicate that this procedure usually leads to a better transient and asymptotic behavior of the primal variables, which we are typically most interested in.

A.5. Proof of Theorem 1

The proof of Theorem 1 relies on the next general stability result. The proof follows in the same lines as the proof of in [12, Theorem 1], and is omitted here.

Lemma 5. Consider

$$\dot{z} = f(t, z), \qquad z(t_0) = z_0, \tag{94}$$

and a one-parameter family of dynamics

$$\dot{z}^\sigma = f^\sigma(t, z^\sigma), \qquad z^\sigma(t_0) = z_0, \tag{95}$$

where $f, f^\sigma : \mathrm{R} \times \mathrm{R}^n \to \mathrm{R}^n$, $f, f^\sigma \in \mathcal{C}^1$ and $\sigma \in \mathrm{N}^+$ is a parameter. Suppose that

1. a compact set $\mathcal{S}$ is locally uniformly asymptotically stable for (94) with region of attraction $\mathcal{R}(\mathcal{S}) \subseteq R^n$;
2. the region of attraction $\mathcal{R}(\mathcal{S})$ is positively invariant for (95);
3. for every $\varepsilon > 0$, for every $T > t_0$ and for every $z_0 \in \mathcal{R}(\mathcal{S})$ there exists $\sigma^* > 0$ such that, for all $\sigma > \sigma^*$ and for all $t_0 \in \mathrm{R}$, there exist unique solutions $z(t), z^\sigma(t)$ of (94) and (95) that fulfill for all $t \in [t_0, T]$

$$\|z(t) - z^\sigma(t)\| \leq \varepsilon. \tag{96}$$

Then the set $\mathcal{S}$ is locally practically uniformly asymptotically stable for (95) and $z^\sigma(t)$ uniformly converges to $z(t)$ on $[t_0, \infty)$ for increasing σ. •

We are now ready to prove Theorem 1 making use of Lemma 5.

Proof of Theorem 1. Since the control law (50) is obtained from the construction procedure presented in [21], it follows directly from [21, Theorem 8.1] that for each $\varepsilon > 0$, for each $T > 0$ and for each initial condition $z^\sigma(0) = z_0 \in \mathcal{R}(\mathcal{M})$, there exists $\sigma^* > 0$ such that for all $\sigma \geq \sigma^*$ and for all $t \in [0, T]$ the inequality (54) holds. Moreover, note that the set $\mathcal{M}$ defined by (13) is compact by Assumption 2 (see also the proof of Lemma 1) and asymptotically stable for (11) with region of attraction $\mathcal{R}(\mathcal{M}) = \{(x, \nu, \lambda) \in \mathrm{R}^n \times \mathrm{R}^n \times \mathrm{R}^n : \lambda \in \mathrm{R}^n_{++}\}$, according to Lemma 1. Also, by the same argumentation as the one in the proof of Lemma 1, the set $\mathcal{R}(\mathcal{M})$ is positively invariant for (49) together with the control law (50) - (53). Hence, all assumptions from Lemma 5 are fulfilled and the result follows. □