

On the Lie bracket approximation approach to distributed optimization: Extensions and limitations

Simon Michalowsky, Bahman Ghahesifard and Christian Ebenbauer

Abstract—We consider the problem of solving a smooth convex optimization problem with equality and inequality constraints in a distributed fashion. Assuming that we have a group of agents available capable of communicating over a communication network described by a time-invariant directed graph, we derive distributed continuous-time agent dynamics that ensure convergence to a neighborhood of the optimal solution of the optimization problem. Following the ideas introduced in our previous work, we combine saddle-point dynamics with Lie bracket approximation techniques. While the methodology was previously limited to linear constraints and objective functions given by a sum of strictly convex separable functions, we extend these result here and show that it applies to a very general class of optimization problems under mild assumptions on the communication topology.

I. INTRODUCTION

Over the last decades, distributed optimization has been an active area of research with high practical relevance, see, e.g., [1]–[3] for applications. In these type of problems, the goal is to cooperatively solve an optimization problem using a group of agents communicating over a network. While discrete-time algorithms for distributed optimization constitute the majority in the existing literature, we focus on continuous-time algorithms which have regained interest in the last decades, [4]–[9]. These algorithms require strong assumptions either on the structure of the optimization problem or on the communication network. Recently, a novel approach to continuous-time distributed optimization based on Lie bracket approximations has been proposed in [10], [11] that has the potential to relax these assumptions.

In the present work we want to extend these results and show that the approach can be applied to a large class of optimization problems under mild assumptions on the communication network. While in [10], [11], [12] only optimization problems with linear constraints and objective functions in the form of a sum of separable functions were considered, in the present work we enhance the methodology to general convex optimization problems. The main idea of the approach is to use Lie bracket approximation techniques to find distributed approximations of non-distributed saddle-point dynamics. While in the previous works only certain Lie bracket approximations were used, we further show here that a whole class is applicable.

S. Michalowsky and C. Ebenbauer are with the Institute for Systems Theory and Automatic Control, University of Stuttgart, 70550 Stuttgart, Germany (e-mail: {simon.michalowsky, ce}@ist.uni-stuttgart.de). B. Ghahesifard is with the Department of Mathematics and Statistics, Queen's University, Kingston, ON, Canada (e-mail: bahman.ghahesifard@queensu.ca).

II. PRELIMINARIES

A. Notation

We denote by \mathbb{R}^n the set of n -dimensional real vectors and 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 $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f \in \mathcal{C}^1$, with respect to its argument $x \in \mathbb{R}^n$, will be denoted by $\nabla_x f : \mathbb{R}^n \rightarrow \mathbb{R}^n$; 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 $A = [a_{ij}]$. We use e_i to denote the vector with the i th entry equal to 1 and all other entries equal to 0. We do not specify the dimension of e_i but expect it to be clear from the context. For a vector $\lambda \in \mathbb{R}^n$ we let $\text{diag}(\lambda) \in \mathbb{R}^{n \times n}$ denote the diagonal matrix whose diagonal entries are the entries of λ . Given two continuously differentiable vector fields $f_1 : \mathbb{R}^n \rightarrow \mathbb{R}^n$ and $f_2 : \mathbb{R}^n \rightarrow \mathbb{R}^n$, the Lie bracket of f_1 and f_2 evaluated at x is defined to be

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

With a slight abuse of notation we sometimes also write $[f_1(x), f_2(x)]$. For a vector $x = [x_1, x_2, \dots, x_n]^\top \in \mathbb{R}^n$ and a finite set $S \subseteq \{1, 2, \dots, n\}$, we denote by $x_S \in \mathbb{R}^{|S|}$ the ordered stacked vector of all x_i with $i \in S$. For example, if $n = 5$ and $S = \{2, 4\}$, then $x_S = [x_2, x_4]^\top$.

B. Basics on graph theory

We recall some basic notions on graph theory, and refer the reader to [13] 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 $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{n \times n}$ associated to \mathcal{G} is defined as

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

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

$$d_{ij} = \begin{cases} \sum_{k=1}^n a_{ik} & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

Finally, we call $G = D - A = [g_{ij}] \in \mathbb{R}^{n \times n}$ the Laplacian of \mathcal{G} . 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 say that a path is simple if the sequence contains no node more than once.

III. PROBLEM SETUP

Consider the following convex optimization problem

$$\begin{aligned} \min \quad & F(x) \\ \text{s.t.} \quad & a_i(x) = 0, \quad i \in \mathcal{I}_{\text{eq}} \subseteq \{1, 2, \dots, n\}, \\ & c_i(x) \leq 0, \quad i \in \mathcal{I}_{\text{ineq}} \subseteq \{1, 2, \dots, n\}, \end{aligned} \quad (4)$$

where $F : \mathbb{R}^n \rightarrow \mathbb{R}$, $a_i : \mathbb{R}^n \rightarrow \mathbb{R}^{n_{\text{eq}_i}}$, $c_i : \mathbb{R}^n \rightarrow \mathbb{R}^{n_{\text{ineq}_i}}$, $n_{\text{eq}_i}, n_{\text{ineq}_i} \geq 1$, $F \in \mathcal{C}^2$ is strictly convex, the functions $a_i, i \in \mathcal{I}_{\text{eq}}$, are affine and the $c_i \in \mathcal{C}^2, i \in \mathcal{I}_{\text{ineq}}$, are convex. We assume further that F, a_i, c_i are such that the feasible set of (4) is non-empty and that the problem has a unique solution.

Our goal is to design continuous-time optimization algorithms that converge to an arbitrarily small neighborhood of the unique global optimizer of (4) and that can be implemented in a distributed fashion. More precisely, we assume that we have a group of n agents available, each capable of interchanging information over a communication network described by a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with graph Laplacian $G = [g_{ij}]$, where $\mathcal{V} = \{v_1, v_2, \dots, v_n\}$ is a set of n nodes and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the edge set between the nodes. In the present setup, each node v_i represents an agent and the edges define the existing communication links between the agents, i.e., if there is an edge from node i to node j then agent i has access to the information provided by agent j . We then say that an algorithm is distributed if each agent only uses its own information as well as that provided by its out-neighboring agents.

Let $L : \mathbb{R}^n \times \mathbb{R}^{n_{\text{eq}}} \times \mathbb{R}^{n_{\text{ineq}}} \rightarrow \mathbb{R}$, $n_{\text{eq}} = \sum_{i \in \mathcal{I}_{\text{eq}}} n_{\text{eq}_i}$, $n_{\text{ineq}} = \sum_{i \in \mathcal{I}_{\text{ineq}}} n_{\text{ineq}_i}$, denote the Lagrangian associated to (4), i.e.,

$$L(x, \nu, \lambda) = F(x) + \nu^\top a(x) + \lambda^\top c(x), \quad (5)$$

where $a = [a_i]_{i \in \mathcal{I}_{\text{eq}}}$, $c = [c_i]_{i \in \mathcal{I}_{\text{ineq}}}$ are the stacked vectors of all a_i and c_i , respectively, and $\nu \in \mathbb{R}^{n_{\text{eq}}}$, $\lambda \in \mathbb{R}^{n_{\text{ineq}}}$ are the associated Lagrange multipliers. In the sequel, we assume that the state of the i th agent comprises of x_i as well as the dual variables associated to the constraints a_i, c_i , given they exist. Without loss of generality we then put the following assumption on the indexing of the constraints:

Assumption 1. For any $i \in \mathcal{I}_{\text{eq}}$ and any $j \in \mathcal{I}_{\text{ineq}}$, there exists an $x = [x_1, x_2, \dots, x_n]^\top \in \mathbb{R}^n$ such that $\frac{\partial a_i}{\partial x_i}(x) \neq 0$ and $\frac{\partial c_j}{\partial x_j}(x) \neq 0$.

In a nutshell, this assumption guarantees that the i th constraints a_i, c_i are both functions of x_i .

It is well-known that if the Lagrangian has a saddle point (x^*, ν^*, λ^*) , then x^* is an optimizer of (4). We say that a point (x^*, ν^*, λ^*) is a saddle point if for all $x \in \mathbb{R}^n$, $\nu \in \mathbb{R}^{n_{\text{eq}}}$,

$\lambda \in \mathbb{R}_+^{n_{\text{ineq}}}$, we have

$$L(x^*, \nu, \lambda) \leq L(x^*, \nu^*, \lambda^*) \leq L(x, \nu^*, \lambda^*). \quad (6)$$

The following saddle-point dynamics adapted from [14] is known to converge to a saddle point of the Lagrangian (see [12] for a proof), thus providing a solution to (4)

$$\dot{x} = -\nabla F(x) - \frac{\partial a}{\partial x}(x)^\top \nu - \frac{\partial c}{\partial x}(x)^\top \lambda \quad (7a)$$

$$\dot{\nu} = a(x) \quad (7b)$$

$$\dot{\lambda} = \text{diag}(\lambda)c(x). \quad (7c)$$

However, (7) is in general not distributed in the aforementioned sense, since the right-hand side of (7) is not composed only of *admissible* vector fields, i.e., vector fields that can be computed locally by the nodes. For example, if $g_{12} \neq 0$ and $g_{13} = 0$, then the vector field $[e_1^\top x_2, 0, 0]^\top$ is admissible for (7), while $[e_1^\top x_3, 0, 0]^\top$ is not.

Recently, a novel approach to distributed optimization has been proposed that employs Lie bracket approximation techniques to derive distributed approximations of (7). The idea is to write the right-hand side of (7) by means of Lie brackets of admissible vector fields. If we have achieved to rewrite (7) in this form, i.e., we have

$$\dot{z} = \sum_{B \in \mathcal{B}} v_B B(z), \quad (8)$$

where $z = [x^\top, \nu^\top, \lambda^\top]^\top$, $v_B \in \mathbb{R}$ and \mathcal{B} is a set of Lie brackets of admissible vector fields $\{\phi_1, \phi_2, \dots, \phi_M\}$, $\phi_i : \mathbb{R}^{n+n_{\text{eq}}+n_{\text{ineq}}} \rightarrow \mathbb{R}^{n+n_{\text{eq}}+n_{\text{ineq}}}$, then we can employ Lie bracket approximation techniques from [15], [16] to derive distributed approximations of (8). More precisely, we can find a family of functions $u_i^\sigma : \mathbb{R} \rightarrow \mathbb{R}$ parametrized by $\sigma > 0$ such that the trajectories of

$$\dot{z}^\sigma = \sum_{i=1}^M \phi_i(z) u_i^\sigma(t) \quad (9)$$

uniformly converge to those of (8) as σ increases, given (8) and (9) are initialized equally. A general algorithm to compute suitable functions u_i^σ is presented in [15] and we will present a modified version thereof tailored to the problem at hand in [12]. In the present paper we do not discuss the second step of how to design these functions u_i^σ but focus on rewriting the right-hand side of (7) in terms of Lie brackets of admissible vector fields.

IV. MAIN RESULTS

Consider the saddle-point dynamics (7) and observe that the right-hand side is a sum of vector fields of the form $e_i \psi(x)$ and $e_i z_j \psi(x)$, $i, j = 1, 2, \dots, n + n_{\text{eq}} + n_{\text{ineq}}$, $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$, where

$$z = [x^\top, \nu^\top, \lambda^\top]^\top \in \mathbb{R}^N, \quad N = n + n_{\text{eq}} + n_{\text{ineq}}, \quad (10)$$

is the complete state and $e_i \in \mathbb{R}^N$ is the i th unit vector. These vector fields might either be admissible or not, depending on the communication graph as well as the problem structure. In the following, we wish to discuss how to write vector fields

of this form by means of Lie brackets of admissible vector fields. For the purpose of notation, for each $i = 1, 2, \dots, n$, we define the index set

$$\begin{aligned} \bar{\mathcal{I}}(i) := \{i\} \cup \{j = n + \sum_{k=1}^{i-1} n_{\text{eq}_k} + \ell, \ell = 1, 2, \dots, n_{\text{eq}_i}\} \\ \cup \{j = n + n_{\text{eq}} + \sum_{k=1}^{i-1} n_{\text{ineq}_k} + \ell, \ell = 1, 2, \dots, n_{\text{ineq}_i}\} \end{aligned} \quad (11)$$

associating the components of the complete state z to the i th agent meaning that z_j is part of the state of agent i for all $j \in \bar{\mathcal{I}}(i)$. Hence, the state vector of the i th agent is given by $z_{\bar{\mathcal{I}}(i)}$. Based on this, for all $j = 1, 2, \dots, N$, we define

$$\mathcal{I}(j) = \bar{\mathcal{I}}(i) \quad (12)$$

for some i such that $j \in \bar{\mathcal{I}}(i)$, i.e., $\mathcal{I}(j)$ is the set of all indices which are associated to the same agent as the j th index. Note that $\bar{\mathcal{I}}(j) = \mathcal{I}(j)$, for all $j = 1, 2, \dots, n$.

A. Lie brackets of admissible vector fields

In the following, we first want to discuss which kind of vector fields can be written in terms of Lie brackets of admissible vector fields. For $i, j = 1, 2, \dots, N$, define

$$h_{i,j}(z) = e_j f_j(z_{\mathcal{I}(j)}, z_{\mathcal{I}(i)}), \quad (13)$$

where $f_j : \mathbb{R}^{|\mathcal{I}(j)|} \times \mathbb{R}^{|\mathcal{I}(i)|} \rightarrow \mathbb{R}$, $f_j \in \mathcal{C}^1$. Observe that $h_{i,j}$ is admissible if and only if there exist $\ell, k \in \{1, 2, \dots, n\}$ such that $i \in \mathcal{I}(\ell)$, $j \in \mathcal{I}(k)$ and $g_{k\ell} \neq 0$. In the next result we consider Lie brackets of admissible vector fields of the form (13).

Lemma 1. Consider a graph \mathcal{G} of n nodes and let $p_{i_1 i_r} = \langle v_{i_1} | v_{i_2} | \dots | v_{i_r} \rangle$ denote a simple path in \mathcal{G} from v_{i_1} to v_{i_r} . Then, for any $j_k \in \mathcal{I}(i_k)$, $k = 1, 2, \dots, r$, we have

$$\begin{aligned} \left[h_{j_r, j_{r-1}}, \left[h_{j_{r-1}, j_{r-2}}, \left[\dots, [h_{j_3, j_2}, h_{j_2, j_1}] \dots \right] \right] \right] (z) \quad (14) \\ = e_{j_1} f_{j_{r-1}}(z_{\mathcal{I}(j_{r-1})}, z_{\mathcal{I}(j_r)}) \prod_{k=1}^{r-2} \frac{\partial f_{j_k}}{\partial z_{j_{k+1}}}(z_{\mathcal{I}(j_k)}, z_{\mathcal{I}(j_{k+1})}) \end{aligned}$$

and the left-hand side is a Lie bracket of admissible vector fields. •

A proof is given in [17]. By the above Lemma, each non-admissible vector field that takes the same form as the right-hand side of (14) can be written in terms of a Lie bracket of admissible vector fields. It is worth mentioning that this does not classify the whole set of vector fields that can be written as a Lie bracket of admissible vector fields since we limited ourselves to a single path. We next discuss a special case that is of particular importance for the application at hand.

Proposition 1. Consider a graph \mathcal{G} of n nodes and let $p_{i_1 i_r} = \langle v_{i_1} | v_{i_2} | \dots | v_{i_r} \rangle$ denote a simple path in \mathcal{G} from v_{i_1} to v_{i_r} . Let $j_k \in \mathcal{I}(i_k)$, $k = 1, 2, \dots, r$, be any set of indices and

suppose that

$$f_{j_k}(z_{\mathcal{I}(j_k)}, z_{\mathcal{I}(j_{k+1})}) = f_{j_k}^{(1)}(z_{\mathcal{I}(j_k)}) f_{j_k}^{(2)}(z_{\mathcal{I}(j_{k+1})}) \quad (15)$$

with

$$f_{j_k}^{(1)}(z_{\mathcal{I}(j_k)}) \frac{\partial f_{j_{k-1}}^{(2)}}{\partial z_{j_k}}(z_{\mathcal{I}(j_k)}) = 1 \quad (16)$$

for all $z_{\mathcal{I}(j_k)} \in \mathbb{R}^{|\mathcal{I}(j_k)|}$, $k = 2, 3, \dots, r-1$. Then

$$\begin{aligned} \left[h_{j_r, j_{r-1}}, \left[h_{j_{r-1}, j_{r-2}}, \left[\dots, [h_{j_3, j_2}, h_{j_2, j_1}] \dots \right] \right] \right] (z) \quad (17) \\ = e_{j_1} f_{j_1}^{(1)}(z_{\mathcal{I}(j_1)}) f_{j_{r-1}}^{(2)}(z_{\mathcal{I}(j_r)}) \end{aligned}$$

for all $z \in \mathbb{R}^N$ and the left-hand side is a Lie bracket of admissible vector fields. •

A proof is given in [17]. In view of (14), the constraint (16) ensures that all terms depending on $z_{\mathcal{I}(j_k)}$, $k = 2, 3, \dots, r-1$, cancel out. Equation (17) is of particular interest since the non-admissible vector fields often take the form as its right-hand side. According to (16), there exists a whole class of vector fields $h_{j_k, j_{k-1}}$, or equivalently, functions f_{j_k} , such that (17) holds. A particularly simple choice that has been utilized in the previous works [10], [11] is to take

$$f_{j_k}^{(1)}(z_{\mathcal{I}(j_k)}) = 1, \quad k = 2, 3, \dots, r-1, \quad (18a)$$

$$f_{j_k}^{(2)}(z_{\mathcal{I}(j_{k+1})}) = z_{j_{k+1}}, \quad k = 1, 2, \dots, r-2; \quad (18b)$$

hence leading to $h_{j_{k+1}, j_k}(z_{\mathcal{I}(j_k)}, z_{\mathcal{I}(j_{k+1})}) = e_{j_k} z_{j_{k+1}}$. However, in view of (9) where the ϕ_i are given by admissible vector fields of the form (13), it is often desired that the admissible vector fields have certain properties such as boundedness in order to simplify the calculation of the approximating inputs u_i^σ and improve the transient behavior of (9). However, as we see next, it is not possible to render all vector fields bounded.

Lemma 2. Suppose that all assumptions from Proposition 1 are fulfilled. Then there exists no set of bounded vector fields $h_{j_k, j_{k-1}} \in \mathcal{C}^1$ such that (17) holds. •

Remark 1. The same holds true if we do not assume the structure (15). In fact, the structure is required for all other variables except of $z_{\mathcal{I}(j_1)}$, $z_{\mathcal{I}(j_r)}$ to cancel out in (17). •

The proof is given in [17]. We next discuss how we can make use of the previous results to rewrite more general vector fields. While Lemma 1 enables us to write products of functions of variables of nodes which lie on the same path as Lie brackets of admissible vector fields, we cannot directly use this result to rewrite functions which do not fulfill this property. To make this clearer, consider the following example:

Example 1. Consider the graph shown in Fig. 1 and assume for the sake of simplicity that $\mathcal{I}(j) = j$ for $j = 1, \dots, 5$ and $z \in \mathbb{R}^5$. By (14) we can write non-admissible vector fields of the form $e_1 \varphi_1(z_2) \varphi_2(z_3)$ as well as $e_1 \varphi_3(z_4) \varphi_4(z_5)$ and sums thereof in terms of Lie brackets of admissible vector fields as long as φ_1, φ_3 admit an analytic expression of their

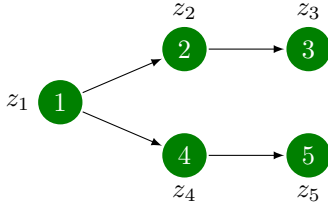


Fig. 1: The graph considered in Example 1 where the i th agent has an associated state z_i .

antiderivatives $\int \varphi_1(z_2)dz_2$, $\int \varphi_3(z_4)dz_4$. In fact, with

$$h_{2,1}(z) = e_1 \int \varphi_1(z_2)dz_2, \quad h_{3,2}(z) = e_2 \varphi_2(z_3), \quad (19a)$$

$$h_{4,3}(z) = e_1 \int \varphi_3(z_4)dz_4, \quad h_{5,4}(z) = e_4 \varphi_4(z_5), \quad (19b)$$

we have for all $z \in \mathbb{R}^5$

$$e_1 \varphi_1(z_2) \varphi_2(z_3) = [h_{3,2}, h_{2,1}](z) \quad (20a)$$

$$e_1 \varphi_3(z_4) \varphi_4(z_5) = [h_{5,4}, h_{4,3}](z). \quad (20b)$$

However, we cannot directly use (14) to rewrite a non-admissible vector field of the form $e_1 \varphi_2(z_3) \varphi_4(z_5)$. •

In the next result, we wish to overcome this limitation and show how to not only write sums of the vector fields in the form of the right-hand side of (14) in terms of Lie brackets of admissible vector fields, but also products thereof.

Proposition 2. Let $\eta_k : \mathbb{R} \rightarrow \mathbb{R}$, $\eta_k \in \mathcal{C}^1$, $k = 1, 2, \dots, m$, $m \geq 2$, and define

$$\psi_1(z) = e_i \eta_1(z_{j_1}) \int \eta_0(z_i) dz_i, \quad (21a)$$

$$\psi_k(z) = e_i z_i \eta_k(z_{j_k}), \quad k = 2, \dots, m-1, \quad (21b)$$

$$\psi_m(z) = e_i \eta_m(z_{j_m}), \quad (21c)$$

$i, j_k \in \mathbb{N}$, $\psi_k : \mathbb{R}^N \rightarrow \mathbb{R}^N$. Then, for any $j_k \neq i$, $k = 1, \dots, m$, we have for all $z \in \mathbb{R}^N$

$$\begin{aligned} & \left[\left[\dots \left[[\psi_m, \psi_{m-1}], \psi_{m-2} \right], \dots \right], \psi_1 \right](z) \\ &= e_i \eta_0(z_i) \prod_{k=1}^m \eta_k(z_{j_k}). \end{aligned} \quad (22)$$

A proof is given in [17]. The vector fields ψ_k defined by (21) are in general non-admissible; thus, the left-hand side of (22) is not a Lie bracket of admissible vector fields. However, observing that the vector fields ψ_k take the same form as the right-hand side of (17), we can make use of Proposition 1 to write the left-hand side of (22) as a Lie bracket of admissible vector fields as long as, for any $k \in \{1, \dots, m\}$, there exists a path from the i th node to the node that is associated to state z_{j_k} . We illustrate that by means of Example 1.

Example 1 (continued). Reconsider Example 1 and suppose we want to rewrite the non-admissible vector field $e_1 \varphi_2(z_3) \varphi_4(z_5)$. Following Proposition 2 we let $\psi_1(z) = e_1 z_1 \varphi_2(z_3)$, $\psi_2(z) = e_1 \varphi_3(z_5)$, and observe that for all $z \in \mathbb{R}^N$ we have

$$[\psi_2, \psi_1](z) = e_1 \varphi_2(z_3) \varphi_4(z_5). \quad (23)$$

Further, following Proposition 1 and choosing

$$h_{2,1}(z) = e_1 z_1, \quad h_{3,2}(z) = e_2 \varphi_2(z_3), \quad (24a)$$

$$h_{4,3}(z) = e_1, \quad h_{5,4}(z) = e_4 \varphi_4(z_5), \quad (24b)$$

we have for all $z \in \mathbb{R}^N$

$$\psi_1(z) = [h_{3,2}, h_{2,1}](z), \quad \psi_2(z) = [h_{5,4}, h_{4,3}](z). \quad (25)$$

Using this in (23) we finally managed to rewrite $e_1 \varphi_2(z_3) \varphi_4(z_5)$ in terms of admissible vector fields. •

Remark 2. Instead of realizing the multiplication by means of Lie brackets another way is to augment the agent state by estimates of the respective state of the other agent. More precisely, in Example 1, we augment the state of agent 1 by ξ_3 and ξ_5 that are estimates of z_3 and z_5 , respectively, and let

$$\begin{bmatrix} \dot{z}_1 \\ \dot{\xi}_3 \\ \dot{\xi}_5 \end{bmatrix} = \begin{bmatrix} w(z_1, z_2, z_4, \xi_3, \xi_5) \\ -\mu \xi_3 \\ -\mu \xi_5 \end{bmatrix} + e_2 z_5 + e_3 z_5, \quad (26)$$

where $w : \mathbb{R}^5 \rightarrow \mathbb{R}$ and $\mu > 0$ is sufficiently large, hence $\xi_3(t) \approx z_3(t)$, $\xi_5(t) \approx z_5(t)$. The resulting non-admissible vector fields in the complete augmented system can then be written in terms of Lie brackets of admissible vector fields using Proposition 1. However, in the application at hand this alters the saddle-point dynamics (7) which necessitates a stability analysis of the augmented system.

Hence, under suitable assumptions on the communication graph, this allows us to write vector fields whose components are sums of products of arbitrary functions in terms of Lie brackets of admissible vector fields. This observation gives rise to the next Lemma.

Lemma 3. Consider a strongly connected graph \mathcal{G} of n nodes and let $\varphi : \mathbb{R}^N \rightarrow \mathbb{R}$ be an analytic function. Then any vector field $\psi(z) = e_i \varphi(z)$, $e_i \in \mathbb{R}^N$, $i = 1, 2, \dots, N$, can be written as a possibly infinite sum of Lie brackets of admissible vector fields. •

A proof is given in [17]. While the result might be more of a theoretical nature for the application at hand, it nevertheless shows that the proposed approach in principle applies to a large class of problems.

B. Distributed optimization via Lie brackets

In the sequel we apply the results from the last section to the problem at hand and rewrite the saddle-point dynamics (7) by means of Lie brackets of admissible vector fields. For the sake of a simpler notation we assume in the following that $n_{\text{eq}_i} = 1$ for all $i \in \mathcal{I}_{\text{eq}}$ and $n_{\text{ineq}_i} = 1$ for all $i \in \mathcal{I}_{\text{ineq}}$. We further assume that each agent has an associated equality and inequality constraint, i.e., $\mathcal{I}_{\text{eq}} = \{1, 2, \dots, n\}$, $\mathcal{I}_{\text{ineq}} = \{1, 2, \dots, n\}$. This can always be achieved by augmenting the optimization problem (4) by constraints that do not alter the feasible set. We emphasize that this is not necessary for the methodology to apply as we will illustrate in the example in Section V. We can then write the saddle-point

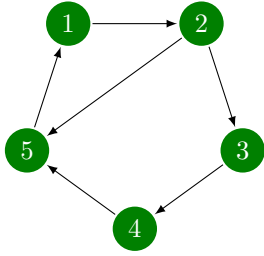


Fig. 2: The communication graph for the example considered in Section V.

Vector field	Corresponding path	Lie bracket representation
$e_5 \nu_2$	$p_{52} = \langle v_5 v_1 v_2 \rangle$	$[e_j z_6, e_5 z_j]$ $j \in \mathcal{I}(1)$
$e_2 x_2 \lambda_1$	$p_{21} = \langle v_2 v_5 v_1 \rangle$	$[e_j z_7, e_2 z_2 z_j]$ $j \in \mathcal{I}(5)$
$e_1 \lambda_3$	$p_{13} = \langle v_1 v_2 v_3 \rangle$	$[e_j \lambda_3, e_1 z_j]$ $j \in \mathcal{I}(2)$
$e_4 x_1 \lambda_4$	$p_{41} = \langle v_4 v_5 v_1 \rangle$	$[e_j x_1, e_4 \lambda_4 z_j]$ $j \in \mathcal{I}(5)$
$e_1 x_4 \lambda_4$	$p_{14} = \langle v_1 v_2 v_3 v_4 \rangle$	$[e_{j_2} \lambda_4 x_4, [e_{j_1} z_{j_2}, e_1 z_{j_1}]]$ $j_1 \in \mathcal{I}(2), j_2 \in \mathcal{I}(3)$
$e_1 x_1 \lambda_4$	$p_{14} = \langle v_1 v_2 v_3 v_4 \rangle$	$[e_{j_2} \lambda_4, [e_{j_1} z_{j_2}, e_1 x_1 z_{j_1}]]$ $j_1 \in \mathcal{I}(2), j_2 \in \mathcal{I}(3)$
$e_8 \lambda_3 x_1$	$p_{31} = \langle v_3 v_4 v_5 v_1 \rangle$	$[e_{j_2} x_1, [e_{j_1} z_{j_2}, e_8 \lambda_2 z_{j_1}]]$ $j_1 \in \mathcal{I}(4), j_2 \in \mathcal{I}(5)$
$e_9 \lambda_4 x_4 x_1$	$p_{41} = \langle v_4 v_5 v_1 \rangle$	$[e_j x_1, e_9 z_j x_4 \lambda_4]$ $j \in \mathcal{I}(5)$
$e_9 \lambda_4 x_1^2$	$p_{41} = \langle v_4 v_5 v_1 \rangle$	$[e_j x_1^2, e_9 \lambda_4 z_j]$ $j \in \mathcal{I}(5)$

TABLE 1: An overview of the non-admissible vector fields in (31) and their Lie bracket representations. Here, the index sets are $\mathcal{I}(1) = \{1, 7\}$, $\mathcal{I}(2) = \{2, 6\}$, $\mathcal{I}(3) = \{3, 8\}$, $\mathcal{I}(4) = \{4, 9\}$, $\mathcal{I}(5) = \{5\}$.

dynamics (7) equivalently as

$$\dot{x} = - \sum_{i=1}^n e_i \left(\frac{\partial F}{\partial x_i}(x) + \sum_{k \in \mathcal{I}_{\text{eq}}} \bar{a}_{ik} \nu_k + \sum_{k \in \mathcal{I}_{\text{ineq}}} \frac{\partial c_k}{\partial x_k}(x) \lambda_k \right) \quad (27a)$$

$$\dot{\nu} = \sum_{k \in \mathcal{I}_{\text{eq}}} e_k \left(\sum_{i=1}^n \bar{a}_{ki} x_i + \bar{b}_k \right) \quad (27b)$$

$$\dot{\lambda} = \sum_{k \in \mathcal{I}_{\text{ineq}}} e_k \lambda_k c_k(x), \quad (27c)$$

where $a_i(x) = \bar{a}_i x + \bar{b}_i$, $\bar{a}_i = [\bar{a}_{i1}, \bar{a}_{i2}, \dots, \bar{a}_{in}] \in \mathbb{R}^{1 \times n}$, $\bar{b}_i \in \mathbb{R}$. Motivated by our previous discussions, we assume in the following that the objective function as well as the inequality constraints are sums of products of separable functions, i.e.,

$$F(x) = \sum_{\ell=1}^{n_F} \prod_{j \in \mathcal{J}_F^{(\ell)}} F_j^{(\ell)}(x_j), \quad (28)$$

$$c_k(x) = \sum_{\ell=1}^{n_{c_k}} \prod_{j \in \mathcal{J}_{c_k}^{(\ell)}} c_{k,j}^{(\ell)}(x_j), \quad (29)$$

where the $F_j^{(\ell)} : \mathbb{R} \rightarrow \mathbb{R}$, $j \in \mathcal{J}_F^{(\ell)} \subseteq \{1, 2, \dots, n\}$, $\ell = 1, 2, \dots, n_F$, $n_F > 0$, are strictly convex functions and the $c_{k,j}^{(\ell)} : \mathbb{R} \rightarrow \mathbb{R}$, $k \in \mathcal{I}_{\text{ineq}}$, $j \in \mathcal{J}_{c_k}^{(\ell)} \subseteq \{1, 2, \dots, n\}$, $\ell = 1, 2, \dots, n_{c_k}$, $n_{c_k} > 0$, are convex. Observe that, if n_F and n_{c_k} are infinite and $F_j^{(\ell)}$, $c_{k,j}^{(\ell)}$ are monomials, this includes all analytic functions F , c_k . If $\mathcal{J}_F^{(\ell)} = \{\ell\}$, $\mathcal{J}_{c_k}^{(\ell)} = \{\ell\}$, we obtain the particularly important special case that both the objective function and the constraints are a sum of separable functions; hence also the case of linear constraints considered in [10], [11] is covered here. Depending on the communication graph as well as the structure of the constraints and the objective function, the vector fields in (27) can either be admissible or not. In particular, the vector fields in (27b), (27c) are admissible if the constraints are compatible with the communication topology defined by the graph \mathcal{G} , i.e., if the following assumption holds:

Assumption 2. For all $i, j = 1, 2, \dots, n$ with $g_{ij} = 0$ we

have $\frac{\partial a_i}{\partial x_j}(x) \equiv 0$ as well as $\frac{\partial c_i}{\partial x_j}(x) \equiv 0$. •

We point out that all non-admissible vector fields in (27) can be written in terms of Lie brackets of admissible vector fields under appropriate assumptions on the communication graph, see also [17, Table 1]. Specifically, if the graph is strongly connected, then all non-admissible vector fields can be rewritten independent of the objective function as well as the constraints, given that they admit the structure (28), (29). In most cases, however, much less restrictive requirements on the communication graph are sufficient. We do not explicitly discuss how to rewrite the non-admissible vector fields using Proposition 1 and Proposition 2 in general but illustrate this by means of an example in Section V.

V. EXAMPLE

In this section we illustrate the previous results by means of an example. Consider the following optimization problem

$$\begin{aligned} \min_{x \in \mathbb{R}^5} \quad & F(x) = \sum_{i=1}^5 F_i(x_i) \\ \text{s.t.} \quad & a_2(x) = 2x_2 - x_5 = 0 \\ & c_1(x) = x_1^2 + x_2^2 - 4 \leq 0 \\ & c_3(x) = x_1 + x_3 - 2 \leq 0 \\ & c_4(x) = x_4^2 - x_4 x_1 + x_1^2 - 100 \leq 0, \end{aligned} \quad (30)$$

where $F_i(x_i) = (x_i - i)^2$, $x = [x_1, x_2, x_3, x_4, x_5]^\top \in \mathbb{R}^5$. We assume that the communication topology is described by the graph in Fig. 2. Observe that the constraints c_3 and c_4 are not compatible with the graph topology, hence Assumption 2 is not fulfilled. The corresponding saddle-point dynamics (7) are then given by

$$\dot{x} = -\nabla F(x) - (2e_2 - e_5)\nu_2 - 2(x_1 e_1 + x_2 e_2)\lambda_1 - (2x_4 - x_1)e_4 \lambda_4 - (2x_1 - x_4)e_1 \lambda_4 - (e_1 + e_3)\lambda_3 \quad (31a)$$

$$\dot{\nu}_2 = 2x_2 - x_5 \quad (31b)$$

$$\begin{aligned} \dot{\lambda} = & \lambda_1 e_1 (x_1^2 + x_2^2 - 4) + \lambda_3 e_2 (x_1 + x_3 - 2) \\ & + \lambda_4 e_3 (x_4^2 - x_4 x_1 + x_1^2 - 100), \end{aligned} \quad (31c)$$

where $x \in \mathbb{R}^5$, $v_2 \in \mathbb{R}$, and $\lambda = [\lambda_1, \lambda_3, \lambda_4]^\top \in \mathbb{R}^3$. We next rewrite all non-admissible vector fields in (31) using Proposition 1. We will thereby follow the choice (18) to make sure that (16) holds. We do not discuss how to rewrite each non-admissible vector field in detail, but limit ourselves to the vector field $e_1 x_4 \lambda_4$ from (31a). Comparing with (17), we have $j_1 = 1$, $j_r = 4$, and $f_{j_1}^{(1)}(z_{\mathcal{I}(j_1)}) = 1$, $f_{j_r-1}^{(2)}(z_{\mathcal{I}(j_r)}) = x_4 \lambda_4$. Following Proposition 1, we require a path from node 1 to node 4 that is here given by $p_{14} = \langle v_1 | v_2 | v_3 | v_4 \rangle$; thus $r = 4$. We then obtain $e_1 x_4 \lambda_4 = [e_{j_2} \lambda_4 x_4, [e_{j_1} z_{j_2}, e_1 z_{j_1}]]$, where $j_1 \in \mathcal{I}(2) = \{2, 6\}$, $j_2 \in \mathcal{I}(3) = \{3, 8\}$, and the right-hand side is a Lie bracket of admissible vector fields. All other non-admissible vector fields in (31) can be treated similarly and we sum up the resulting Lie brackets in Table 1. For the simulation we let each j, j_1, j_2 be the largest index in its respective index set. By that choice, we inject less perturbation in the primal and more in the dual variables which is also visible in the simulation results depicted in Fig. 3. As to be seen, the distributed algorithm approximates the trajectories of the non-distributed saddle-point dynamics (31) and converges to a neighborhood of the optimizer $x^* = [0, 2, 2, 4, 4]^\top$. We also included simulation results with additional low-pass filters in the distributed x -, v - and λ -dynamics. While the effect on the primal variables is small since we already reduced the oscillations by our design choice, the dual variables show significantly less oscillations and better approximate the non-distributed trajectories. A rigorous stability analysis of the augmented distributed dynamics and a performance-oriented design of the filters is up to future work.

VI. CONCLUSIONS AND OUTLOOK

We considered a convex optimization problem and showed how distributed optimization algorithms can be designed for a quite general class of problems with little structural requirements under mild assumptions on the communication network. We therefore extended the Lie bracket approximation approach to distributed optimization proposed in [10], [11] and discussed which kind of vector fields can in principle be written in terms of Lie brackets of admissible vector fields. We did not discuss the construction of approximating inputs but postpone this to [12] where we will present a modified version of the general algorithm from [15] that exploits the structure of the problem at hand.

REFERENCES

- [1] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [2] F. Bullo, J. Cortés, and S. Martínez, *Distributed Control of Robotic Networks*, ser. Applied Mathematics Series. Princeton University Press, 2009.
- [3] J. Zhao and F. Dörfler, "Distributed control and optimization in DC microgrids," *Automatica*, vol. 61, no. Supplement C, pp. 18–26, 2015.
- [4] D. Feijer and F. Paganini, "Stability of primaldual gradient dynamics and applications to network optimization," *Automatica*, vol. 46, no. 12, pp. 1974–1981, 2010.
- [5] J. Wang and N. 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)*, 2011, pp. 3800–3805.

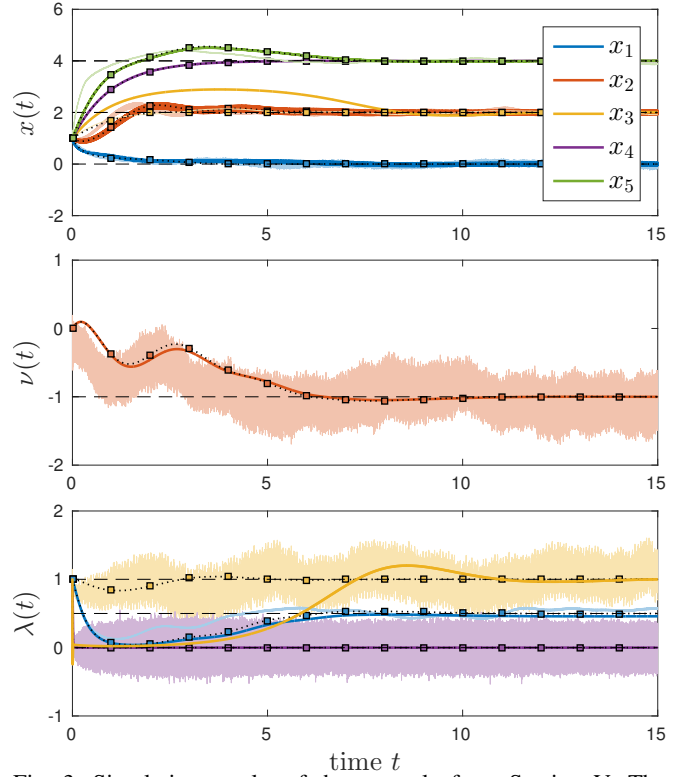


Fig. 3: Simulation results of the example from Section V. The upper plot shows the primal variable $x(t)$, the middle one the dual variable $v_2(t)$ and the lower one the dual variable $\lambda(t)$. In each plot, the dotted black lines marked with squares depict the trajectories of the non-distributed saddle-point dynamics (31) while the thinner oscillating ones depict the trajectories of the distributed approximation. The corresponding thick lines depict the distributed approximation with additional low-pass filters. The dashed black lines indicate the desired equilibrium of (31).

- [6] H.-B. Dürr, C. Zeng, and C. Ebenbauer, "Saddle point seeking for convex optimization problems," *IFAC Proceedings Volumes*, vol. 46, no. 23, pp. 540–545, 2013.
- [7] B. Ghareisifard and J. Cortés, "Distributed continuous-time convex optimization on weight-balanced digraphs," *IEEE Transactions on Automatic Control*, vol. 59, no. 3, pp. 781–786, 2014.
- [8] S. K. Niederländer and J. Cortés, "Distributed coordination for separable convex optimization with coupling constraints," in *2015 54th IEEE Conference on Decision and Control (CDC)*, Dec 2015, pp. 694–699.
- [9] B. Touri and B. Ghareisifard, "Saddle-point dynamics for distributed convex optimization on general directed graphs," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, 2016, pp. 862–866.
- [10] C. Ebenbauer, S. Michalowsky, V. Grushkovskaya, and B. Ghareisifard, "Distributed optimization over directed graphs with the help of Lie brackets," in *Proc. 20th IFAC World Congress*, 2017, pp. 15 908–15 913.
- [11] S. Michalowsky, B. Ghareisifard, and C. Ebenbauer, "Distributed extremum seeking over directed graphs," in *2017 IEEE 56th Conference on Decision and Control (CDC)*, 2017, pp. 2095–2101.
- [12] —, "A Lie bracket approximation approach to distributed optimization over directed graphs," *ArXiv e-prints arXiv:1711.05486 [math.OC]*, 2017, <https://arxiv.org/abs/1711.05486>.
- [13] N. Biggs, *Algebraic graph theory*. Cambridge university press, 1993.
- [14] H.-B. Dürr and C. Ebenbauer, "On a class of smooth optimization algorithms with applications in control," *IFAC Proceedings Volumes*, vol. 45, no. 17, pp. 291–298, 2012, 4th IFAC Conference on Nonlinear Model Predictive Control.
- [15] W. Liu, "An approximation algorithm for nonholonomic systems," *SIAM Journal on Control and Optimization*, vol. 35, no. 4, pp. 1328–1365, 1997.
- [16] H. J. Sussmann and W. Liu, "Limits of highly oscillatory controls and the approximation of general paths by admissible trajectories," in *1991 IEEE 30th Conference on Decision and Control (CDC)*, 1991, pp. 437–442.
- [17] S. Michalowsky, B. Ghareisifard, and C. Ebenbauer, "On the Lie bracket approximation approach to distributed optimization: Extensions and limitations," *ArXiv e-prints arXiv:1802.10519 [math.OC]*, 2018, <https://arxiv.org/abs/1802.10519>.