# Utility of Edge-wise Funnel Coupling for Asymptotically Solving Distributed Consensus Optimization

Jin Gyu Lee*, Thomas Berger**, Stephan Trenn***, and Hyungbo Shim****

***Abstract*— A new approach to distributed consensus optimization is studied in this paper. The cost function to be minimized is a sum of local cost functions which are not necessarily convex as long as their sum is convex. This benefit is obtained from a recent observation that, with a large gain in the diffusive coupling, heterogeneous multi-agent systems behave like a single dynamical system whose vector field is simply the average of all agents' vector fields. However, design of the large coupling gain requires global information such as network structure and individual agent dynamics. In this paper, we employ a nonlinear time-varying coupling of diffusive type, which we call 'edge-wise funnel coupling.' This idea is borrowed from adaptive control, which enables decentralized design of distributed optimizers without knowledge of global information. Remarkably, without a common internal model, each agent achieves asymptotic consensus to the optimal solution of the global cost. We illustrate this result by a network that asymptotically finds the least-squares solution of a linear equation in a distributed manner.**

## I. Introduction

Recent developments in the fields such as formation control, smart grid, and resilient state estimation have raised the question of how to design a network so that agents collectively find an optimizer [1]–[5], and consensus optimization is a vast research area which studies a subclass of the aforementioned problem. Let the cost function be given by

$$f : \mathbb{R}^n \to \mathbb{R}, \quad f(x) = \sum_{i=1}^{N} f_i(x), \tag{1}$$

which is the sum of $N$ heterogeneous cost functions. The question of how to construct a dynamic system for each node $i \in \mathcal{N} := \{1, \ldots, N\}$ that finds the minimizer $x^* \in \mathbb{R}^n$ of $f(\cdot)$, with each node $i$ having access to its individual cost function $f_i(\cdot)$ only, has been tackled in recent years [6]–[12]. However, most of them, e.g., [6]–[11], assume that the individual cost function $f_i(\cdot)$ is convex. The reason is the need for stability, e.g., passivity, for each node, to achieve consensus.

This work was partially supported by the German Research Foundation (Deutsche Forschungsgemeinschaft) via the grant BE 6263/1-1, by the National Research Foundation of Korea grant funded by the Korea government(MSIP) (No. 2015R1A2A2A01003878 and No. 2019R1A6A3A12032482), and by the NWO vidi grant 639.032.733.

*jgl46@cam.ac.uk, Control Group, Department of Engineering, University of Cambridge, United Kingdom

**thomas.berger@math.upb.de, Institut für Mathematik, Universität Paderborn, Warburger Straße 100, 33098 Paderborn, Germany

***s.trenn@rug.nl, Bernoulli Institute, University of Groningen, Netherlands

****hshim@snu.ac.kr, ASRI, Department of Electrical and Computer Engineering, Seoul National University, Korea

In this paper, we present a network that finds the minimizer $x^*$ of $f(\cdot)$ asymptotically, with the assumption that $f(\cdot)$ is strictly convex even if each function $f_i(\cdot)$ is not necessarily convex. This is obtained from the recent observation that, with a large coupling gain in the diffusive coupling, heterogeneous multi-agent systems behave like a single dynamical system whose vector field is simply the average of all agents' vector fields [13], [14]. By this observation, it is possible to trade stability among agents, and hence, to relax the assumptions on the individual cost functions. However, this approach has some limitations, for instance

1) it only guarantees practical consensus (i.e. for any $\varepsilon > 0$ a coupling strength can be chosen so that the agents' states eventually get $\varepsilon$-close), and
2) the design of the coupling gain that is used for each agent requires global information such as the network structure and the individual agent dynamics.

To resolve these issues, we modify the linear diffusive term of the designed network into a nonlinear time-varying coupling, which we call 'edge-wise funnel coupling.' This idea is motivated by the funnel control methodology (a particular adaptive control method) which was developed in [15], see also the survey [16].

Let us emphasize that we obtain *asymptotic* consensus to the unique minimizer $x^*$ of $f(\cdot)$ by the proposed funnel coupling. However, this seems to violate the common presumption, in the funnel control community, that asymptotic tracking of an arbitrary reference signal with prescribed performance is not possible. A recent discovery in [17] presented a trick how to overcome this restriction, and finally achieved asymptotic tracking via funnel control. In the present paper, we exploit the technique from [17] to achieve asymptotic consensus (without additional dynamics like the PI consensus algorithms or embedding a common internal model).

The paper is organized as follows. In Section II, we give a precise problem formulation and introduce, for a given network graph, agent dynamics that are designed with a constant coupling gain. In order to resolve the above mentioned limitations, the dynamics are modified using the funnel coupling in Section III. In Section IV we illustrate the utility of this design by an example of a distributed least-squares solver. Finally, Section V concludes the paper.

*Notation:* The Laplacian matrix $\mathcal{L} = [l_{ij}] \in \mathbb{R}^{N\times N}$ of a graph is defined as $\mathcal{L} := \mathcal{D} - \mathcal{A}$, where $\mathcal{A} = [\alpha_{ij}]$ is the adjacency matrix of the graph and $\mathcal{D}$ is the diagonal matrix with its $i$-th diagonal entry being $\sum_{j=1}^{N} \alpha_{ij}$. By construction, the Laplacian matrix contains at least one zero eigenvalue with corresponding eigenvector $1_N := [1, \ldots, 1]^\top \in \mathbb{R}^N$, and all other eigenvalues have non-negative real parts. For undirected graphs, the zero eigenvalue is simple if, and only if, the corresponding graph is connected. For vectors or matrices $a$ and $b$ we set $\mathrm{col}(a, b) := [a^\top, b^\top]^\top$. The operation defined by the symbol $\otimes$ is the Kronecker product. The maximum norm of a vector $x$ is defined by $\|x\|_\infty := \max_i |x_i|$, and the Euclidean norm is denoted by $\|x\| := \sqrt{x^\top x}$. The induced maximum norm of a matrix $A$ (the maximum absolute row sum) is $\|A\|_\infty$. The gradient of a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as $\partial f := \mathrm{col}(\partial f/\partial x^1, \ldots, \partial f/\partial x^n)$. The identity matrix of size $m \times m$ is denoted by $I_m$.

## II. Problem Setting and Preliminaries

Consider a network of $N$ agents, whose structure is defined by a graph.

**Assumption 1:** The graph is undirected and connected. //

In the network, each agent $i \in \mathcal{N} = \{1, \ldots, N\}$ has access to its own cost function $f_i : \mathbb{R}^n \to \mathbb{R}$ but not to the other $f_j$, $j \neq i$. Here, $f_i(\cdot)$ satisfies the following.

**Assumption 2:** For each $i \in \mathcal{N}$, $f_i(\cdot)$ is continuously differentiable, and its gradient $\partial f_i(\cdot)$ is globally Lipschitz continuous with Lipschitz constant $L_i > 0$, i.e., $\|\partial f_i(x) - \partial f_i(x')\| \leq L_i \|x - x'\|$ for all $x, x' \in \mathbb{R}^n$. //

The objective is to solve, in a distributed way,

$$\mathrm{minimize}_x \quad f(x) = \sum_{i=1}^{N} f_i(x)$$

under the following assumption.

**Assumption 3:** The sum of the $N$ cost functions,

$$f(x) = \sum_{i=1}^{N} f_i(x)$$

is strictly convex, i.e.,

$$f(tx + (1-t)x') < tf(x) + (1-t)f(x'),$$

for any $t \in (0, 1)$ and $x, x' \in \mathbb{R}^n$ such that $x \neq x'$. Moreover, there exists a point $x^* \in \mathbb{R}^n$ such that $f(x^*) \leq f(x)$ for all $x \in \mathbb{R}^n$. //

By Assumption 3 there exists a unique minimizer $x^* \in \mathbb{R}^n$ of $f(\cdot)$, and hence $\partial f(x)$ becomes zero only at $x^*$. Therefore, the gradient descent algorithm given by

$$\dot{\hat{x}} = -\partial f(\hat{x}) = -\sum_{i=1}^{N} \partial f_i(\hat{x}) \quad \in \mathbb{R}^n \tag{2}$$

solves the optimization problem. In particular, the solution $\hat{x}(\cdot)$ asymptotically converges to the unique minimizer $x^*$.

Motivated by this, we may design a distributed algorithm, in which the individual dynamics of each agent $i \in \mathcal{N}$ are given by

$$\dot{x}_i = -\partial f_i(x_i) + k \sum_{j \in \mathcal{N}_i} (x_j - x_i) \quad \in \mathbb{R}^n \tag{3}$$

where $k > 0$ is a design parameter, and $\mathcal{N}_i$ is a subset of $\mathcal{N}$ whose elements are the indices of those agents which are connected to agent $i$ within the network graph (the neighbors), and are hence able to share information with it.

**Remark 1:** Insight into the proposed network (3) comes from the so-called 'blended dynamics' approach [13], [14]. In this approach, the behavior of heterogeneous multi-agent systems

$$\dot{x}_i = g_i(t, x_i) + k \sum_{j \in \mathcal{N}_i} (x_j - x_i), \quad i \in \mathcal{N},$$

with large coupling gain $k$ is approximated by the behavior of the blended dynamics defined by

$$\dot{\hat{x}} = \frac{1}{N} \sum_{i=1}^{N} g_i(t, \hat{x})$$

under the assumption that these dynamics are stable. In our case, the blended dynamics are given by

$$\dot{\hat{x}} = -\frac{1}{N} \sum_{i=1}^{N} \partial f_i(\hat{x}) = -\frac{1}{N} \partial f(\hat{x})$$

which is the (scaled) gradient descent algorithm (2). //

**Proposition 2 ( [14]):** Let Assumptions 1, 2, and 3 hold. Then, for any compact set $K \subseteq \mathbb{R}^{Nn}$, and for any $\eta > 0$, there exists $k^* > 0$ such that, for each $k > k^*$ and $\mathrm{col}(x_1(0), \ldots, x_N(0)) \in K$, the solution to (3) exists for all $t \geq 0$, and satisfies

$$\forall\, i \in \mathcal{N} : \limsup_{t \to \infty} \|x_i(t) - x^*\| \leq \eta. \qquad //$$

Although this result is already quite powerful, a disadvantage is that the optimizer is not found asymptotically but only approximately. Moreover, for computing the threshold $k^*$, global information such as the network topology and all $f_i$'s is needed, and so the method is not completely decentralized. These drawbacks will be resolved in the next section by choosing the gain $k$ adaptively based on the idea of funnel control.

## III. Edge-wise funnel coupling

Building on the idea of the edge-wise funnel coupling law [18], we propose to replace the static diffusive coupling term $k \sum_{j \in \mathcal{N}_i} (x_j - x_i)$ in (3) by the coupling law

$$\sum_{j \in \mathcal{N}_i} K\left(\frac{x_j - x_i}{\psi(t)}\right) \cdot \frac{x_j - x_i}{\psi(t)}$$

where $\psi : \mathbb{R}_{\geq 0} \to \mathbb{R}_{>0}$ is a so-called funnel boundary function (Figure 1), and $K : \mathbb{R}^n \to \mathbb{R}^{n \times n}$ is defined by

$$K(\eta) := \mathrm{diag}\left(\frac{1}{1 - |\eta^1|}, \cdots, \frac{1}{1 - |\eta^n|}\right). \tag{4}$$

By introducing $e_{ij} := x_j - x_i$, the dynamics of agent $i$ become

$$\dot{x}_i = -\partial f_i(x_i) + \sum_{j \in \mathcal{N}_i} \mathrm{col}\left(\frac{e_{ij}^1}{\psi(t) - |e_{ij}^1|}, \ldots, \frac{e_{ij}^n}{\psi(t) - |e_{ij}^n|}\right) \tag{5}$$

where $x_i = \mathrm{col}(x_i^1, \ldots, x_i^n)$ and $e_{ij}^p = x_j^p - x_i^p$.

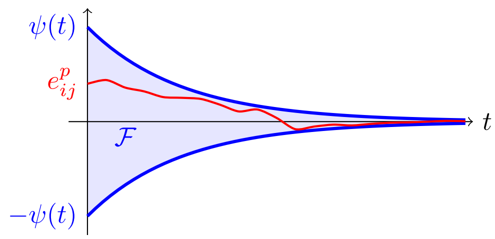


Fig. 1. The funnel: a pre-designed time-varying error bound

The intuition behind the funnel coupling in (5) is as follows. If the $p$-th component of the difference between two agents, $e_{ij}^p(t) = x_j^p(t) - x_i^p(t)$, approaches the funnel boundary $\pm\psi(t)$ so that $\psi(t) - |e_{ij}^p(t)|$ gets close to zero, then the gain associated to $e_{ij}^p(t)$ becomes large. Therefore, if there is only one neighbor, then the state $x_i$ tends to its neighbor $x_j$ since the large coupling term dominates the vector field $-\partial f_i(x_i)$, and the error $e_{ij}^p(t)$ remains inside the funnel. However, with more than one neighbor, the situation is more involved because two neighbors may attract $x_i$ in opposite direction with almost infinite power. Actual analysis shows that all the errors $e_{ij}(t)$ remain inside the funnel, which is however far more complicated. In this paper, we only quote one of the main results in [19].

**Proposition 3:** Let Assumptions 1, 2, and 3 hold. Then, for any bounded continuously differentiable function $\psi : \mathbb{R}_{\geq 0} \to \mathbb{R}_{>0}$ with bounded derivative, and for any initial conditions $x_i(0) \in \mathbb{R}^n$ with $\|x_j(0) - x_i(0)\|_\infty < \psi(0)$ for all $j \in \mathcal{N}_i$, $i \in \mathcal{N}$, the solution to (5) exists for all $t \geq 0$ and satisfies

$$\forall\, t \geq 0 \;\forall\, i \in \mathcal{N} \;\forall\, j \in \mathcal{N}_i : \; \|x_j(t) - x_i(t)\|_\infty < \psi(t).$$

Moreover, if there exists $\overline{M}$ such that $\|x_i(t)\| \leq \overline{M}$ for all $t \geq 0$ and all $i \in \mathcal{N}$, then there exists $\varepsilon > 0$ such that

$$\forall\, t \geq 0 \;\forall\, i \in \mathcal{N} \;\forall\, j \in \mathcal{N}_i : \; \frac{\|x_j(t) - x_i(t)\|_\infty}{\psi(t)} \leq 1 - \varepsilon. \quad (6)$$

//

*Proof:* A brief proof of this proposition is found in the Appendix. ■

According to Proposition 3, if we select $\psi(\cdot)$ such that $\lim_{t\to\infty} \psi(t) = 0$, then we obtain asymptotic consensus, i.e., $\lim_{t\to\infty} \|x_j(t) - x_i(t)\|_\infty = 0$. This in turn implies that, with the new variable $x_{\text{avg}} := (1/N)\sum_{i=1}^N x_i$, each state $x_i(t)$ tends to $x_{\text{avg}}(t)$ as $t \to \infty$. Now, by Assumption 1, we may observe that

$$\sum_{i=1}^N \sum_{j\in\mathcal{N}_i} \frac{e_{ij}^p(t)}{\psi(t) - |e_{ij}^p(t)|} = 0 \quad (7)$$

for $p = 1, \ldots, n$. Therefore, we have that

$$\dot{x}_{\text{avg}} = -\frac{1}{N}\sum_{i=1}^N \partial f_i(x_i) \quad \to \quad -\frac{1}{N}\sum_{i=1}^N \partial f_i(x_{\text{avg}})$$

as $t \to \infty$, and so, intuitively the coupled system (5) will asymptotically find the unique minimizer $x^*$. This intuition is made precise in the following theorem, which is our main result.

**Theorem 4:** Let Assumptions 1, 2, and 3 hold. Then, for any bounded continuously differentiable function $\psi : \mathbb{R}_{\geq 0} \to \mathbb{R}_{>0}$ with bounded derivative which satisfies $\lim_{t\to\infty} \psi(t) = 0$, and for any initial conditions $x_i(0) \in \mathbb{R}^n$ with $\|x_j(0) - x_i(0)\|_\infty < \psi(0)$ for all $j \in \mathcal{N}_i$, $i \in \mathcal{N}$, the solution to (5) exists for all $t \geq 0$ and satisfies

$$\forall\, i \in \mathcal{N} : \; \lim_{t\to\infty} x_i(t) = x^*,$$

i.e., each agent's state converges to the global optimizer. Furthermore, there exists $\varepsilon > 0$ such that (6) holds, i.e., the coupling gain $K$ given by (4) remains bounded. //

*Proof:* Let, according to Proposition 3, $(x_1, \ldots, x_N)$ be the solution of (5) which exists for all $t \geq 0$. Let $L_i > 0$ be a Lipschitz constant of $\partial f_i$ according to Assumption 2, and let $\mathcal{T}$ be an arbitrary spanning tree in the network graph with incidence matrix $T \in \mathbb{R}^{N\times(N-1)}$. Let $\mathsf{t}_i^\top$ be the $i$-th row of $T(T^\top T)^{-1}$ and define

$$\begin{pmatrix} x_{\text{avg}} \\ \tilde{x} \end{pmatrix} := \begin{bmatrix} (1/N)1_N^\top \otimes I_n \\ T^\top \otimes I_n \end{bmatrix} \text{col}(x_1, \ldots, x_N).$$

Since $1_N^\top T = 0$ we find that

$$\begin{bmatrix} (1/N)1_N^\top \\ T^\top \end{bmatrix}^{-1} = \begin{bmatrix} 1_N & T(T^\top T)^{-1} \end{bmatrix},$$

thus it follows that $x_i = x_{\text{avg}} + (\mathsf{t}_i^\top \otimes I_n)\tilde{x}$ for all $i \in \mathcal{N}$, and hence, by (5) and (7),

$$\dot{x}_{\text{avg}} = -\frac{1}{N}\sum_{i=1}^N \partial f_i(x_{\text{avg}} + (\mathsf{t}_i^\top \otimes I_n)\tilde{x}).$$

Note that by Proposition 3, we have that for all $t \geq 0$

$$\|(T^\top \otimes I_n)\text{col}(x_1(t), \ldots, x_N(t))\|_\infty = \|\tilde{x}(t)\|_\infty < \psi(t).$$

Now, let $V(x_{\text{avg}}) := f(x_{\text{avg}}) - f(x^*) = \sum_{i=1}^N (f_i(x_{\text{avg}}) - f_i(x^*))$. Then, due to Assumption 3, there exist class $\mathcal{K}$-functions[1] $\alpha_1(\cdot)$ and $\alpha_2(\cdot)$ such that

$$\forall\, x \in \mathbb{R}^n : \; \alpha_1(\|x - x^*\|) \leq V(x) \leq \alpha_2(\|x - x^*\|).$$

Now, the derivative of $V$ along $x_{\text{avg}}(\cdot)$ satisfies

$$\begin{aligned}
\dot{V} &= \partial f(x_{\text{avg}})^\top \dot{x}_{\text{avg}} \\
&= -\frac{1}{N}\partial f(x_{\text{avg}})^\top \sum_{i=1}^N \partial f_i(x_{\text{avg}} + (\mathsf{t}_i^\top \otimes I_n)\tilde{x}) \\
&= -\frac{1}{N}\|\partial f(x_{\text{avg}})\|^2 \\
&\quad - \frac{1}{N}\partial f(x_{\text{avg}})^\top \sum_{i=1}^N [\partial f_i(x_{\text{avg}} + (\mathsf{t}_i^\top \otimes I_n)\tilde{x}) - \partial f_i(x_{\text{avg}})] \\
&\leq -\frac{1}{N}\|\partial f(x_{\text{avg}})\|^2 + \frac{1}{N}\|\partial f(x_{\text{avg}})\| \sum_{i=1}^N L_i \|(\mathsf{t}_i^\top \otimes I_n)\tilde{x}\| \\
&\leq -\frac{1}{N}\|\partial f(x_{\text{avg}})\| \left(\|\partial f(x_{\text{avg}})\| - L^* \psi(t)\right),
\end{aligned}$$

[1] A continuous function $\alpha : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is called a class $\mathcal{K}$-function, if $\alpha(0) = 0$ and $\alpha$ is strictly monotonically increasing.

where we have used Assumption 2 and

$$\begin{aligned}\|(\mathsf{t}_i^\top \otimes I_n)\tilde{x}\| &\le \sqrt{n}\|(\mathsf{t}_i^\top \otimes I_n)\tilde{x}\|_\infty \\ &\le \sqrt{n}\|\mathsf{t}_i^\top \otimes I_n\|_\infty \|\tilde{x}\|_\infty < \sqrt{n}\|\mathsf{t}_i^\top\|_\infty \psi(t),\end{aligned}$$

whence $L^* := \sqrt{n}\|T(T^\top T)^{-1}\|_\infty \sum_{i=1}^N L_i$.

Seeking a contradiction, assume that $V(x_{\text{avg}}(t)) \not\to 0$ for $t \to \infty$, then there exists $\varepsilon > 0$ and a sequence $(t_i)_{i\in\mathbb{N}}$ with $t_i \nearrow \infty$ such that $V(x_{\text{avg}}(t_i)) > \varepsilon$ for all $i \in \mathbb{N}$. Set

$$t_i^0 := \max\{0, \sup\{\ t \in [0, t_i]\ \mid V(x_{\text{avg}}(t)) \le \varepsilon\ \}\}$$

for $i \in \mathbb{N}$, then $\|x_{\text{avg}}(t) - x^*\| \ge \alpha_2^{-1}\big(V(x_{\text{avg}}(t))\big) > \alpha_2^{-1}(\varepsilon)$ for all $t \in (t_i^0, t_i]$; note that $\alpha_2$ can always be chosen to be unbounded and is hence bijective on $\mathbb{R}_{\ge 0}$. Next, observe that for any $u \in \mathbb{R}^n$ the function $t \mapsto f(x^* + tu)$ is strictly convex, and hence the derivative $t \mapsto \partial f(x^* + tu)u$ is strictly monotonically increasing (and it is zero only for $t = 0$ by Assumption 3), i.e., for any $u \in \mathbb{R}^n$ with $\|u\| = 1$ there exists a class $\mathcal{K}$-function $\alpha_3^u(\cdot)$ such that

$$\forall\, t \in \mathbb{R}:\ \ \|\partial f(x^* + tu)\| \ge |\partial f(x^* + tu)u| \ge \alpha_3^u(|t|).$$

As a consequence, for any $\delta > 0$,

$$\begin{aligned}&\min\{\ \|\partial f(x)\|\ \mid \|x - x^*\| = \delta\ \} \\ &\qquad = \min\{\ \|\partial f(x)\|\ \mid \|x - x^*\| \ge \delta\ \}.\end{aligned}$$

Therefore, it is possible to choose $\eta > 0$ sufficiently small so that $\|\partial f(x_{\text{avg}}(t))\| \ge \eta$ for all $t \in [t_i^0, t_i]$ and all $i \in \mathbb{N}$. Then, choose $i \in \mathbb{N}$ large enough so that $\psi(t) < \eta/(2L^*)$ for all $t \in [t_i^0, t_i]$. Now, we obtain

$$\forall\, t \in [t_i^0, t_i]:\ \ \dot{V}(x_{\text{avg}}(t)) \le -\frac{\eta^2}{2N},$$

which implies

$$\varepsilon < V(x_{\text{avg}}(t_i)) < V(x_{\text{avg}}(t_i^0)) = \varepsilon,$$

a contradiction. Thus we have shown that $\lim_{t\to\infty} V(x_{\text{avg}}(t)) = 0$, which yields $\lim_{t\to\infty} \|x_{\text{avg}}(t) - x^*\| = 0$.

Since $\|\tilde{x}(t)\| < \psi(t)$ and $\psi$ converges to zero we have that $x_i = x_{\text{avg}} + (\mathsf{t}_i^\top \otimes I_n)\tilde{x}$ also converges to $x^*$ and is thus bounded. Then the last statement of the theorem is a consequence of Proposition 3. ■

**Remark 5:** The asymptotic convergence result of Theorem 4 may seem contradictory to a common presumption in the funnel control community that asymptotic tracking of an arbitrary reference signal with prescribed performance is not possible. However, a recent paper [17] has overcome this restriction by a relatively simple trick to rewrite the funnel control law. We inherit the same technique in the present paper. Indeed, the coupling gain $1/\big(\psi(t) - |x_j^p - x_i^p|\big)$ grows unbounded when asymptotic consensus is achieved, because $\psi(t) \to 0$ as $t \to \infty$ and this implies that $\psi(t) - |x_j^p - x_i^p|$ also tends to zero. However, simply rewriting the coupling term as

$$\frac{1}{1 - |x_j^p - x_i^p|/\psi(t)} \cdot \frac{x_j^p - x_i^p}{\psi(t)}$$

we see that by Theorem 4 the fraction $|x_j^p - x_i^p|/\psi(t)$ is bounded away from 1, hence the new gain and the total input are bounded even if $1/\psi(t)$ tends to infinity.

We further emphasize that Theorem 4 may seem to violate another presumption in the synchronization research area that heterogeneous multi-agent systems cannot asymptotically synchronize without a common internal model. This issue is resolved by observing that we use a time-varying coupling law, which is not considered in the framework of the internal model principle for multi-agent systems [20].

Finally, we stress that the difference between asymptotic consensus and practical consensus may not seem very important in practical applications, as long as the residual error in practical consensus is sufficiently small. In view of this, our concern on asymptotic convergence is rather of academic interest. //

## IV. Example: Distributed Least-Squares Solver

As distributed algorithms have been developed in various fields of study so as to divide a large computational problem into small-scale computations, finding the least-squares solution of a given large linear equation in a distributed manner has been tackled in recent years [21]–[24]. Consider the equation

$$Ax = b \in \mathbb{R}^M, \tag{8}$$

where $A \in \mathbb{R}^{M\times n}$ is a matrix with full column rank and $x \in \mathbb{R}^n$. Throughout this section, we suppose that the $M$ equations in (8) are grouped into $N$ equation banks, and the $i$-th equation bank consists of $m_i$ lines of equations so that $\sum_{i=1}^N m_i = M$. We write the $i$-th equation bank as

$$A_i x = b_i \in \mathbb{R}^{m_i}, \quad i = 1, 2, \ldots, N,$$

where $A_i \in \mathbb{R}^{m_i\times n}$ is the $i$-th block row of the matrix $A$, and $b_i \in \mathbb{R}^{m_i}$ is the $i$-th block element of $b$.

Finding the least-squares solution $x^*$ of (8) even when $b \notin \operatorname{im}(A)$ can be cast as a simple optimization problem

$$\text{minimize}_x\ \frac{1}{2}\|Ax - b\|^2 = \sum_{i=1}^N \frac{1}{2}\|A_i x - b_i\|^2.$$

With $f_i(x) = \frac{1}{2}\|A_i x - b_i\|^2$ the problem becomes a consensus optimization. Then, according to the recipe in the previous section, we can find the least-squares solution asymptotically by a network with individual agent dynamics

$$\begin{aligned}\dot{x}_i &= -A_i^\top(A_i x_i - b_i) \\ &+ \sum_{j\in\mathcal{N}_i} \operatorname{col}\left(\frac{x_j^1 - x_i^1}{\psi(t) - |x_j^1 - x_i^1|}, \ldots, \frac{x_j^n - x_i^n}{\psi(t) - |x_j^n - x_i^n|}\right),\end{aligned} \tag{9}$$

where each agent $i$ uses the information of $A_i$ and $b_i$ only.

Now, for a linear equation given by

$$Ax = \begin{bmatrix}1\\1\\2\\2\\1\end{bmatrix} x = \begin{bmatrix}1\\10\\20\\18\\100\end{bmatrix} = b$$

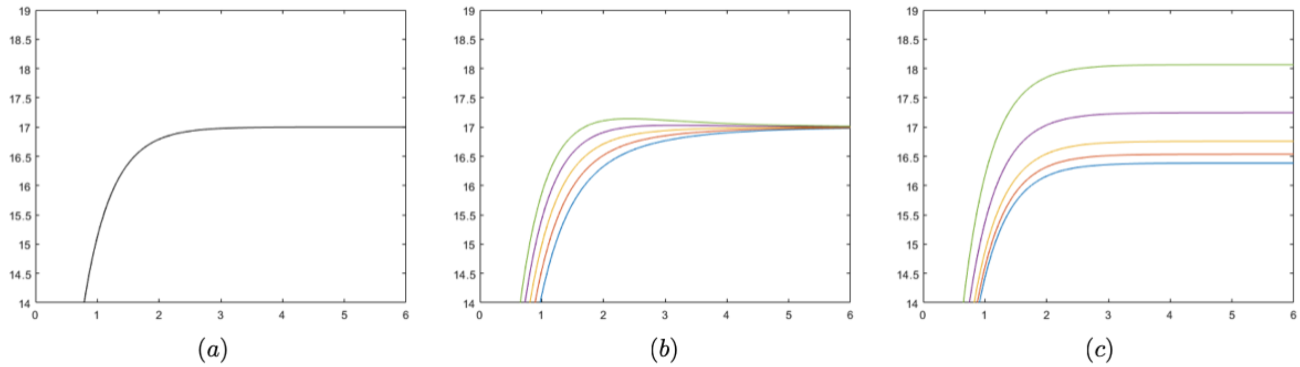


Fig. 2. Solution trajectory of (a) the blended dynamics, (b) the network with edge-wise funnel coupling, and (c) the network with constant coupling gain

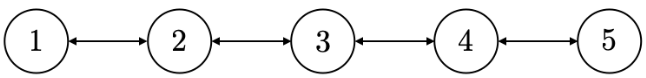


Fig. 3. Underlying graph among five agents

with $N = 5$ and each equation bank consisting of a single equation, the gradient descent algorithm

$$\dot{\hat{x}} = -A^\top (A\hat{x} - b), \quad \hat{x}(0) = 0,$$

results in convergence of its state to the unique minimizer $x^* = 17$. On the other hand, as guaranteed by Theorem 4, the solutions $x_1, \ldots, x_5$ of the system of equations (9) also converges to $x^*$. Indeed, Figure 2.(b) shows a simulation result when the funnel boundary function is chosen as $\psi(t) = \exp(-0.8t)$, the network graph is set to a linear graph as in Figure 3, and the initial conditions are $x_1(0) = 0$, $x_2(0) = 0.1$, $x_3(0) = -0.1$, $x_4(0) = 0.2$, and $x_5(0) = -0.2$.

For comparison, Figure 2(c) shows the trajectory for the agent dynamics

$$\dot{x}_i = -A_i^\top (A_i x_i - b_i) + k \sum_{j \in \mathcal{N}_i} (x_j - x_i), \quad i \in \mathcal{N}, \quad (10)$$

with the constant coupling gain $k = 100$. Figures 2.(b) and 2.(c) clearly show that the network with the constant coupling gain can only achieve practical convergence to the minimizer $x^* = 17$, while asymptotic convergence is obtained by using edge-wise funnel coupling.

We also inspect the derivative of each $x_i(\cdot)$, because the right-hand side of $\dot{x}_i(\cdot)$ can be considered as an input to each agent, and we are interested in the magnitude of their values. Figure 4 shows $\dot{x}_i(\cdot)$ for the edge-wise funnel coupling in (9), while Figure 5 depicts the constant coupling gain in (10). It can be verified that their magnitudes do not differ very much. Note that, for the case of edge-wise funnel coupling, $\dot{x}_i(\cdot)$ is bounded even though the funnel $\psi(\cdot)$ approaches zero. The reason is that, as discussed in Remark 5, the funnel coupling law can be re-written appropriately and the term $|x_j - x_i|/\psi(t)$ is bounded away from 1 by Theorem 4.

Finally, observe that the diffusive coupling term $(x_j - x_i)/(\psi(t) - |x_j - x_i|)$ converges to a specific constant $e_{ij}^*$

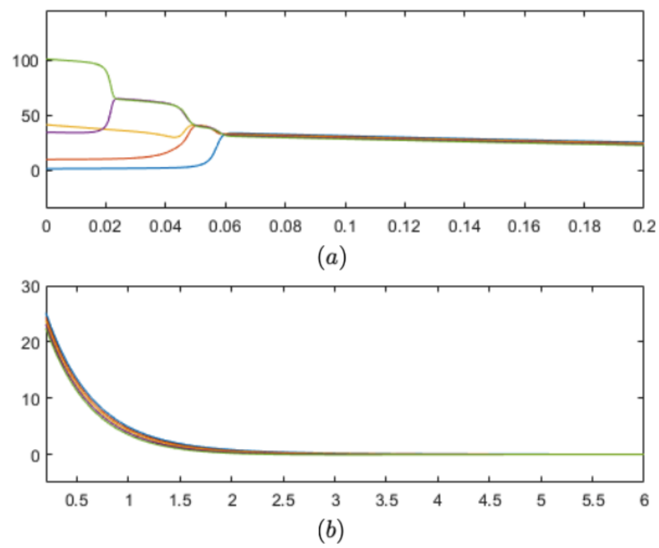


Fig. 4. Plot of $\dot{x}_i(t)$ for the network with edge-wise funnel coupling: (a) $t \in [0, 0.2]$ and (b) $t \in [0.2, 6]$

that cancels the heterogeneity of the individual vector field such that

$$-A_i^\top (A_i x^* - b_i) + \sum_{j \in \mathcal{N}_i} e_{ij}^* = 0, \quad i \in \mathcal{N}.$$

This is clearly indicated in Figure 4, where we observe that $\dot{x}_i(\cdot)$ converges to zero. Hence, we may interpret the edge-wise funnel coupling law as an adaptation scheme. Moreover, in this special case, we may even compute $e_{ij}^*$ as

$$e_{12}^* = 16, \quad e_{23}^* = 23, \quad e_{34}^* = 51, \quad e_{45}^* = 83.$$

This is also shown in Figure 6, which depicts the convergence of the fraction $(x_j(t) - x_i(t))/\psi(t)$. In particular, if we denote $\eta_{ij}^* := \lim_{t\to\infty}(x_j(t) - x_i(t))/\psi(t)$, then we have $\eta_{ij}^*/(1 - |\eta_{ij}^*|) = e_{ij}^*$ for $i \in \mathcal{N}$ and $j \in \mathcal{N}_i$, which gives

$$\eta_{12}^* = \tfrac{16}{17}, \quad \eta_{23}^* = \tfrac{23}{24}, \quad \eta_{34}^* = \tfrac{51}{52}, \quad \eta_{45}^* = \tfrac{83}{84}.$$

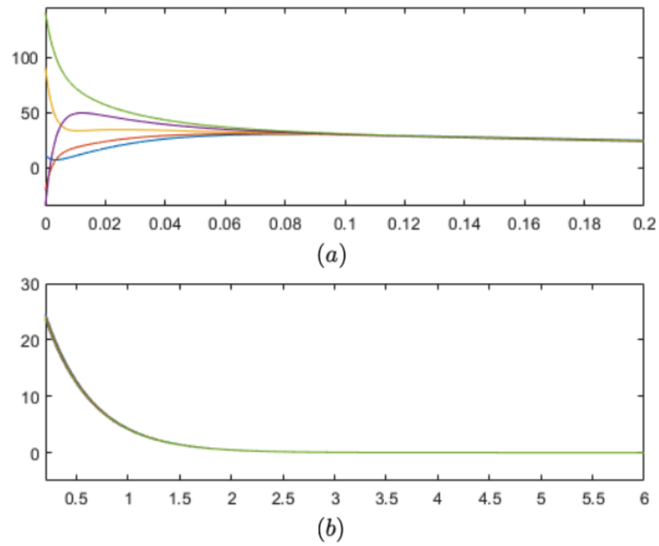


Fig. 5. Plot of $\dot{x}_i(t)$ for the network with constant coupling gain: (a) $t \in [0, 0.2]$ and (b) $t \in [0.2, 6]$

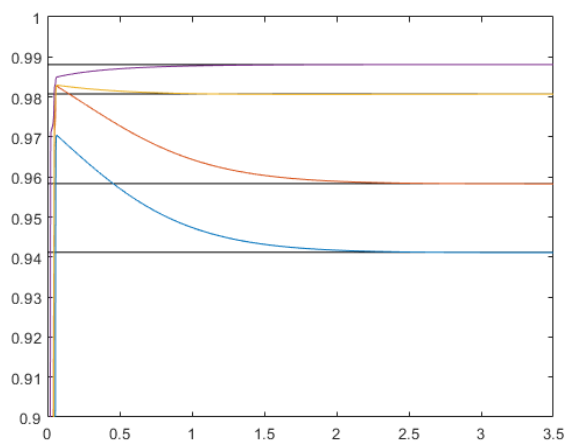


Fig. 6. Trajectory of the fraction $(x_j - x_i)/\psi(t)$

## V. Conclusion

Based on the design philosophy of the blended dynamics induced by a large coupling gain, agent dynamics are designed to solve a distributed consensus optimization by a constant coupling gain for any given undirected connected network graph. Then, to overcome the limitation of the constant gain design, the dynamics are modified by introducing the edge-wise funnel coupling, whose intuition is inherited from adaptive control. As a consequence, we obtain a network that achieves asymptotic convergence to the unique minimizer, which does not require any global information. The utility of the proposed network is illustrated by a distributed least-squares optimization. A detailed comparison of the performance compared to other decentralized optimization algorithms is ongoing research.

## References


[1] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, 2009.

[2] I. Lobel, A. Ozdaglar, and D. Feijer, "Distributed multi-agent optimization with state-dependent communication," *Mathematical Programming*, vol. 129, no. 2, pp. 255–284, 2011.

[3] J. G. Lee, J. Kim, and H. Shim, "Fully distributed resilient state estimation based on distributed median solver," under review for *IEEE Transactions on Automatic Control*, 2019.

[4] F. Wirth, S. Stuedli, J. Y. Yu, M. Corless, and R. Shorten, "Nonhomogeneous place-dependent Markov chains, unsynchronised AIMD, and network utility maximization," *arXiv preprint arXiv: 1404.5064*, 2014.

[5] M. Corless, C. King, R. Shorten, and F. Wirth, *AIMD dynamics and distributed resource allocation*. SIAM, 2016.

[6] G. Qu and N. Li, "Harnessing smoothness to accelerate distributed optimization," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 3, pp. 1245–1260, 2018.

[7] J. Wang and N. Elia, "Control approach to distributed optimization," in *Proceedings of the 48th Annual Allerton Conference on Communication, Control, and Computing*, 2010, pp. 557–561.

[8] T. Hatanaka, N. Chopra, T. Ishizaki, and N. Li, "Passivity-based distributed optimization with communication delays using PI consensus algorithm," *IEEE Transactions on Automatic Control*, vol. 63, no. 12, pp. 4421–4428, 2018.

[9] B. Gharesifard 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.

[10] J. Wang and N. Elia, "A control perspective for centralized and distributed convex optimization," in *Proceedings of the 50th IEEE Conference on Decision and Control*, 2011, pp. 3800–3805.

[11] G. Chen and Z. Li, "A fixed-time convergent algorithm for distributed convex optimization in multi-agent systems," *Automatica*, vol. 95, pp. 539–543, 2018.

[12] Z. Li, Z. Ding, J. Sun, and Z. Li, "Distributed adaptive convex optimization on directed graphs via continuous-time algorithms," *IEEE Transactions on Automatic Control*, vol. 63, no. 5, pp. 1434–1441, 2018.

[13] J. Kim, J. Yang, H. Shim, J.-S. Kim, and J. H. Seo, "Robustness of synchronization of heterogeneous agents by strong coupling and a large number of agents," *IEEE Transactions on Automatic Control*, vol. 61, no. 10, pp. 3096–3102, 2016.

[14] J. G. Lee and H. Shim, "A tool for analysis and synthesis of heterogeneous multi-agent systems under rank-deficient coupling," under review for *Automatica*, available at `arXiv:1804.00638`, 2019.

[15] A. Ilchmann, E. P. Ryan, and C. J. Sangwin, "Tracking with prescribed transient behaviour," *ESAIM: Control, Optimisation and Calculus of Variations*, vol. 7, pp. 471–493, 2002.

[16] A. Ilchmann and E. P. Ryan, "High-gain control without identification: a survey," *GAMM Mitt.*, vol. 31, no. 1, pp. 115–125, 2008.

[17] J. G. Lee and S. Trenn, "Asymptotic tracking via funnel control," accepted for *IEEE Conference on Decision and Control*, 2019.

[18] S. Trenn, "Edge-wise funnel synchronization," in *Proceedings in Applied Mathematics and Mechanics*, vol. 17, pp. 821–822, 2017.

[19] J. G. Lee, T. Berger, S. Trenn, and H. Shim, "Synchronization with prescribed transient behavior: Heterogeneous multi-agent systems under edge-wise funnel coupling," 2019, in preparation.

[20] P. Wieland, J. Wu, and F. Allgöwer, "On synchronous steady states and internal models of diffusively coupled systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 10, pp. 2591–2602, 2013.

[21] X. Wang, J. Zhou, S. Mou, and M. J. Corless, "A distributed linear equation solver for least square solutions," in *Proceedings of the 56th IEEE Conference on Decision and Control*, 2017, pp. 5955–5960.

[22] G. Shi and B. D. O. Anderson, "Distributed network flows solving linear algebraic equations," in *Proceedings of the American Control Conference*, 2016, pp. 2864–2869.

[23] G. Shi, B. D. O. Anderson, and U. Helmke, "Network flows that solve linear equations," *IEEE Transactions on Automatic Control*, vol. 62, no. 6, pp. 2659–2674, 2017.

[24] Y. Liu, Y. Lou, B. D. O. Anderson, and G. Shi, "Network flows as least squares solvers for linear equations," in *Proceedings of 56th IEEE Conference on Decision and Control*, 2017, pp. 1046–1051.

## Appendix: Proof of Proposition 3

We first present a lemma whose proof is omitted due to the page limit.

**Lemma 6:** Consider a digraph defined by $(\mathcal{N}, \mathcal{E})$. If there exists $x \in \mathbb{R}^N$ such that $x_j - x_i > 0$ for all $(j,i) \in \mathcal{E}$, then there is a set of constants $\{\xi_{ij} > 0 : (j,i) \in \mathcal{E}\}$ with which it holds that for all $x \in \mathbb{R}^N$

$$\sum_{(j,i)\in\mathcal{E}} \xi_{ij}(x_j - x_i) = \sum_{(j,i)\in\bar{\mathcal{E}}} \xi_{ij} x_j - \sum_{(j,i)\in\hat{\mathcal{E}}} \xi_{ij} x_i,$$

where $\bar{\mathcal{E}} := \{(j,i) \in \mathcal{E} : \mathcal{N}_j = \emptyset\}$ is the edge set related to all sources and $\hat{\mathcal{E}} := \{(j,i) \in \mathcal{E} : \forall l \in \mathcal{N}, (i,l) \notin \mathcal{E}\}$ is the edge set to all sinks. //

Now, let $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$ be the set of all edges in the graph, i.e., $(j,i) \in \mathcal{E}$ if, and only if, $j \in \mathcal{N}_i$, and let

$$\Omega_\psi := \{(t,x) \in \mathbb{R}_{\geq 0} \times \mathbb{R}^{Nn} : \\ \|x_j - x_i\|_\infty < \psi(t),\ \forall\, i \in \mathcal{N},\ j \in \mathcal{N}_i\}.$$

Since the right-hand side of (5) is locally Lipschitz on $\Omega_\psi \subseteq \mathbb{R}^{Nn+1}$, the standard theory of ODEs yields existence and uniqueness of a maximally extended solution $x : [0,w) \to \mathbb{R}^{Nn}$ such that $(t, x(t)) \in \Omega_\psi$ for all $t \in [0,w)$. If $w = \infty$, then the proof is complete.

Seeking a contradiction, assume that $w < \infty$. Define the edge sets

$$\mathcal{E}_+^p := \left\{(j,i) \in \mathcal{E} : \exists\{t_k\} \to w, \text{ s.t. } \lim_{k\to\infty} \tfrac{e_{ij}^p(t_k)}{\psi(t_k)} = 1\right\}$$
$$\mathcal{E}_-^p := \left\{(j,i) \in \mathcal{E} : \exists\{t_k\} \to w, \text{ s.t. } \lim_{k\to\infty} \tfrac{e_{ij}^p(t_k)}{\psi(t_k)} = -1\right\},$$

where $e_{ij}^p = x_j^p - x_i^p$, $p = 1, 2, \ldots, n$, and $\{t_k\}$ is a strictly increasing time sequence. Then, among these $2n$ edge sets, at least one of them is non-empty.

Under the assumption that $\mathcal{E}_+^1$ is non-empty, for example, we will construct a sequence of strictly increasing edge sets $\mathcal{E}_1 \subsetneq \mathcal{E}_2 \subsetneq \mathcal{E}_3 \subsetneq \ldots$ which are all contained in $\mathcal{E}_+^1$; this then leads to a contradiction to the finiteness of $\mathcal{E}_+^1$. Hence $\mathcal{E}_+^1$ must be empty, and analogous arguments yield that all other $\mathcal{E}_+^p$ and $\mathcal{E}_-^p$ must also be empty and therefore $w < \infty$ is impossible.

So, first, take any element of $\mathcal{E}_+^1$, say $(j_1, i_1)$. Then, by the definition of $\mathcal{E}_+^1$ there exists a strictly increasing time sequence $\{t_k^1\}$ such that $\lim_{k\to\infty} t_k^1 = w$ and satisfies $\lim_{k\to\infty} e_{i_1 j_1}^1(t_k^1)/\psi(t_k^1) = 1$. Then

$$\mathcal{E}_1 := \left\{(j,i) \in \mathcal{E}_+^1 : \lim_{k\to\infty} e_{ij}^1(t_k^1)/\psi(t_k^1) = 1\right\},$$

is non-empty (because it contains $(j_1, i_1)$) and is a subset of $\mathcal{E}_+^1$. Inductively, assume now that for $q \geq 1$ a non-empty index set is given by

$$\mathcal{E}_q := \left\{(j,i) \in \mathcal{E}_+^1 : \lim_{k\to\infty} e_{ij}^1(t_k^q)/\psi(t_k^q) = 1\right\},$$

where $\{t_k^q\}_{k\in\mathbb{N}}$ is a strictly increasing sequence converging to $w$. We now construct a strictly increasing sequence $\{t_k^{q+1}\}$ converging to $w$ as $k \to \infty$ such that the corresponding set $\mathcal{E}_{q+1}$ contains $\mathcal{E}_q$ and there is $(j_{q+1}, i_{q+1}) \in \mathcal{E}_+^1$ which is in $\mathcal{E}_{q+1}$ but not in $\mathcal{E}_q$. Therefore we will first construct a sequence $\{s_r^{q+1}\}_{r\in\mathbb{N}}$ such that

$$\forall (j,i) \in \mathcal{E}_q : \lim_{r\to\infty} \frac{e_{ij}^1(s_r^{q+1})}{\psi(s_r^{q+1})} = 1, \tag{11}$$

and such that for each $r \in \mathbb{N}$ there is an edge $(j_r, i_r) \in \mathcal{E}_+^1 \setminus \mathcal{E}_q$ with

$$\frac{e_{i_r j_r}^1(s_r^{q+1})}{\psi(s_r^{q+1})} > 1 - \delta_r; \tag{12}$$

where $\{\delta_r\}_{r\in\mathbb{N}}$ is some strictly decreasing sequence converging to zero with $\delta_0 > 0$ such that for all $(j,i) \in \mathcal{E} \setminus \mathcal{E}_+^1$ and $t \in [0,w)$ we have $e_{ij}^1(t)/\psi(t) \leq 1 - \delta_0$. Since $\mathcal{E}_+^1 \setminus \mathcal{E}_q$ is finite, we find a subsequence $t_k^{q+1} := s_{r_k}^{q+1}$ and an edge $(j_{q+1}, i_{q+1}) \in \mathcal{E}_+^1 \setminus \mathcal{E}_q$ such that

$$\lim_{k\to\infty} \frac{e_{i_{q+1} j_{q+1}}^1(t_k^{q+1})}{\psi(t_k^{q+1})} = 1;$$

in other words, $(j_{q+1}, i_{q+1}) \in \mathcal{E}_{q+1}$, where $\mathcal{E}_{q+1}$ is defined analogously as $\mathcal{E}_q$ via the sequence $\{t_k^{q+1}\}$. Since (11) also holds for any subsequence, it follows that $\mathcal{E}_q \subseteq \mathcal{E}_{q+1}$. Therefore, it remains to construct the sequence $s_r^{q+1}$ such that (11) and (12) hold.

Towards this goal let

$$W_q(t) := \sum_{(j,i)\in\mathcal{E}_q} \xi_{ij} \frac{e_{ij}^1(t)}{\psi(t)}, \quad t \in [0,w),$$

where $\xi_{ij}$ is given from Lemma 6 by considering the graph $(\mathcal{N}, \mathcal{E}_q)$.[2] Then, by the definition of $\mathcal{E}_q$,

$$\lim_{k\to\infty} W_q(t_k^q) = \sum_{(j,i)\in\mathcal{E}_q} \xi_{ij} =: \bar{\xi}^q$$

and $W_q(t) < \bar{\xi}^q$ for all $t \in [0,w)$. For a suitably chosen strictly decreasing sequence $\{\varepsilon_r^q\}_{r\in\mathbb{N}}$ with $\varepsilon_r^q \to 0$ as $r \to \infty$,[3] we first choose a subsequence $\{t_{k_r}^q\}_{r\in\mathbb{N}}$ of $\{t_k^q\}_{k\in\mathbb{N}}$ such that

$$W_q(t_{k_r}^q) \geq \bar{\xi}^q - \underline{\xi}^q \varepsilon_r^q/2, \quad \forall r \in \mathbb{N},$$

where $\underline{\xi}^q := \min_{(j,i)\in\mathcal{E}_q} \xi_{ij} > 0$. Based on this sequence, we now define a sequence $\{s_r^{q+1}\}_{r\in\mathbb{N}}$ as

$$s_r^{q+1} := \sup\{s \in [0, t_{k_r}^q] : W_q(s) = \bar{\xi}^q - \underline{\xi}^q \varepsilon_r^q\}.$$

By the choice of $s_r^{q+1}$, we have that $W_q(s_r^{q+1}) \to \bar{\xi}^q$ and therefore (11) holds. Furthermore, the choice of $s_r^{q+1}$ also implies $\dot{W}_q(s_r^{q+1}) \geq 0$. Assume now that (12) does *not* hold; we will show in the following that we then arrive at the contradiction that $0 \leq \psi(s_r^{q+1})\dot{W}_q(s_r^{q+1}) < 0$.

[2] The assumption of Lemma 6 is fulfilled because for sufficiently large $k \in \mathbb{N}$ we have $e_{ij}^1(t_k^q) > 0$ for all $(j,i) \in \mathcal{E}_q$.

[3] Without loss of generality we assume $W_q(0) < \bar{\xi}^q - \underline{\xi}^q \varepsilon_0^q$.

The derivative of $W_q$ can be bounded as follows

$$\begin{aligned}\dot{W}_q(t) &= -\frac{\dot{\psi}(t)}{\psi(t)}\sum_{(j,i)\in\mathcal{E}_q}\xi_{ij}\frac{e^1_{ij}(t)}{\psi(t)} \\ &\quad+\frac{1}{\psi(t)}\sum_{(j,i)\in\mathcal{E}_q}\xi_{ij}\left(\frac{\partial f_i}{\partial x^1}(x_i)-\frac{\partial f_j}{\partial x^1}(x_j)\right) \\ &\quad+\frac{1}{\psi(t)}\sum_{(j,i)\in\mathcal{E}_q}\xi_{ij}\sum_{l\in\mathcal{N}_j}\frac{e^1_{jl}(t)}{\psi(t)-|e^1_{jl}(t)|} \\ &\quad-\frac{1}{\psi(t)}\sum_{(j,i)\in\mathcal{E}_q}\xi_{ij}\sum_{l\in\mathcal{N}_i}\frac{e^1_{il}(t)}{\psi(t)-|e^1_{il}(t)|} \\ &\leq \frac{M_0}{\psi(t)}+\frac{1}{\psi(t)}\sum_{(j,i)\in\mathcal{E}_q}\xi_{ij}\sum_{l\in\mathcal{N}_j}\frac{e^1_{jl}(t)}{\psi(t)-|e^1_{jl}(t)|} \\ &\quad-\frac{1}{\psi(t)}\sum_{(j,i)\in\mathcal{E}_q}\xi_{ij}\sum_{l\in\mathcal{N}_i}\frac{e^1_{il}(t)}{\psi(t)-|e^1_{il}(t)|}\end{aligned}$$

with

$$\begin{aligned}M_0 &:= \theta_\psi\sum_{(j,i)\in\mathcal{E}_q}\xi_{ij} \\ &\quad+\sum_{(j,i)\in\mathcal{E}_q}\xi_{ij}\left(\sup_x\|\partial f_i(x)\|_\infty+\sup_x\|\partial f_j(x)\|_\infty\right),\end{aligned}$$

where $|\dot{\psi}(t)|\leq\theta_\psi$ for all $t\in[0,\infty)$ and $\sup_x$ is taken over the compact set $\{x:\|x\|\leq\overline{M}\}$, with $\overline{M}>0$ being the upper bound of the solution trajectory, i.e., $\|x_i(t)\|\leq\overline{M}$ for all $t\in[0,w)$.[4] Invoking now Lemma 6 we get

$$\begin{aligned}\psi(s_r^{q+1})\dot{W}_q(s_r^{q+1}) &\leq M_0+\sum_{(j,i)\in\bar{\mathcal{E}}_q}\xi_{ij}\sum_{l\in\mathcal{N}_j}\frac{e^1_{jl}(s_r^{q+1})}{\psi(s_r^{q+1})-|e^1_{jl}(s_r^{q+1})|} \\ &\quad-\sum_{(j,i)\in\hat{\mathcal{E}}_q}\xi_{ij}\sum_{l\in\mathcal{N}_i}\frac{e^1_{il}(s_r^{q+1})}{\psi(s_r^{q+1})-|e^1_{il}(s_r^{q+1})|}\end{aligned}$$

where $\bar{\mathcal{E}}_q:=\{(j,i)\in\mathcal{E}_q:\forall l\in\mathcal{N}_j,(l,j)\notin\mathcal{E}_q\}$ and $\hat{\mathcal{E}}_q:=\{(j,i)\in\mathcal{E}_q:\forall l\in\mathcal{N}_i,(i,l)\notin\mathcal{E}_q\}$. Noting that for any $(j,i)\in\bar{\mathcal{E}}_q$ and $l\in\mathcal{N}_j$ we have $(l,j)\notin\mathcal{E}_q$ and also that for any $(j,i)\in\hat{\mathcal{E}}_q$ and $l\in\mathcal{N}_i$ we have $(i,l)\notin\mathcal{E}_q$ we can simplify the above inequality as

$$\begin{aligned}\psi(s_r^{q+1})\dot{W}_q(s_r^{q+1}) &\leq M_0+\sum_{(l,j)\in\mathcal{E}\setminus\mathcal{E}_q}\zeta_{jl}\frac{e^1_{jl}(s_r^{q+1})}{\psi(s_r^{q+1})-|e^1_{jl}(s_r^{q+1})|} \\ &\quad-\sum_{(l,i)\in\mathcal{E}_q}\zeta_{il}\frac{e^1_{il}(s_r^{q+1})}{\psi(s_r^{q+1})-|e^1_{il}(s_r^{q+1})|},\end{aligned} \tag{13}$$

where $\zeta_{ij}$ is a non-negative constant given from the above inequality for all $(j,i)\in\mathcal{E}$ which satisfies $\sum_{(l,j)\in\mathcal{E}\setminus\mathcal{E}_q}\zeta_{jl}=:\overline{\zeta}_b>0$ and $\sum_{(l,i)\in\mathcal{E}_q}\zeta_{il}=:\overline{\zeta}_i>0$.[5]

[4]The proof of the existence of $\overline{M}$ is also omitted due to the page limit, however, it can be derived rather easily by the fact that the derivative of $\max_i\max_p x_i^p$ is upper bounded by $-\max_i\max_p\partial f_i^p(x_i)$.

[5]The explicit value of $\zeta_{ij}$ is given as

$$\zeta_{ij}=\sum_{l\in\mathcal{N}_1(i,j)}\xi_{li}+\sum_{l\in\mathcal{N}_2(i,j)}\xi_{jl}+\sum_{l\in\mathcal{N}_3(i,j)}\xi_{il},$$

where $\mathcal{N}_1(i,j):=\{l\in\mathcal{N}:(i,l)\in\bar{\mathcal{E}}_q\}$, $\mathcal{N}_2(i,j):=\{l\in\mathcal{N}:(l,j)\in\hat{\mathcal{E}}_q\}$ if $(i,j)\notin\mathcal{E}_q$ and $\emptyset$ otherwise, and $\mathcal{N}_3(i,j):=\{l\in\mathcal{N}:(l,i)\in\hat{\mathcal{E}}_q\}$ if $(j,i)\in\mathcal{E}_q$ and $\emptyset$ otherwise.

To arrive at the sought contradiction $\psi(s_r^{q+1})\dot{W}_q(s_r^{q+1})<0$ we will exploit that due to (11) the terms $e^1_{il}/(\psi-|e^1_{il}|)$ will be very large for each $(l,i)\in\mathcal{E}_q$ while $e^1_{jl}/(\psi-|e^1_{jl}|)$ for $(l,j)\notin\mathcal{E}_q$ will not be very large. We will now choose the sequence $\{\varepsilon_r^q\}$ (used above to define the sequence $\{s_r^{q+1}\}$) in a suitable way to make this intuition precise; indeed, let each $\varepsilon_r^q>0$ be so small that the following is satisfied.

- $1-\varepsilon_r^q\geq\frac{(M_0+1)/(\underline{\xi}^q\delta_r)}{1+|(M_0+1)/(\underline{\xi}^q\delta_r)|}$.
- $1-\varepsilon_r^q\geq\frac{b_r/((1-\delta_r)\Xi)}{1+|b_r/((1-\delta_r)\Xi)|}$, where $b_r=\frac{1-\delta_r}{1-|1-\delta_r|}$ and $\Xi:=\overline{\zeta}_i/\overline{\zeta}_b>0$.

By rewriting the definition of $W_q$ we get for each pair $(j,i)\in\mathcal{E}_q$

$$\begin{aligned}\frac{e^1_{ij}(s_r^{q+1})}{\psi(s_r^{q+1})} &> \frac{1}{\xi_{ij}}\left(W_q(s_r^{q+1})-(\bar{\xi}^q-\xi_{ij})\right) \\ &= \frac{1}{\xi_{ij}}\left(\bar{\xi}^q-\underline{\xi}^q\varepsilon_r^q-(\bar{\xi}^q-\xi_{ij})\right)\geq 1-\varepsilon_r^q,\end{aligned} \tag{14}$$

and hence, by the choice of $\varepsilon_r^q$ and $\delta_0$ we have for each $(l,j)\in\mathcal{E}\setminus\mathcal{E}_+^1$ and $(o,i)\in\mathcal{E}_q$ that

$$\begin{aligned}&\frac{e^1_{jl}(s_r^{q+1})}{\psi(s_r^{q+1})-|e^1_{jl}(s_r^{q+1})|}\leq\frac{1-\delta_0}{1-|1-\delta_0|}\leq\frac{1-\delta_r}{1-|1-\delta_r|}=b_r \\ &\leq(1-\delta_r)\Xi\frac{1-\varepsilon_r^q}{1-|1-\varepsilon_r^q|}\leq(1-\delta_r)\Xi\frac{e^1_{io}(s_r^{q+1})}{\psi(s_r^{q+1})-|e^1_{io}(s_r^{q+1})|},\end{aligned}$$

and due to assuming that (12) does *n*ot hold, we can conclude the same outer inequality also for all $(l,j)\in\mathcal{E}_+^1\setminus\mathcal{E}_q$. This allows us to bound, in (13), each $e^1_{jl}/(\psi-|e^1_{jl}|)$ by $(1-\delta_r)\Xi e^1_{il}/(\psi-|e^1_{il}|)$, i.e.

$$\begin{aligned}\psi(s_r^{q+1})\dot{W}_q(s_r^{q+1}) &\leq M_0-\sum_{(l,i)\in\mathcal{E}_q}\zeta_{il}\delta_r\frac{e^1_{il}(s_r^{q+1})}{\psi(s_r^{q+1})-|e^1_{il}(s_r^{q+1})|} \\ &\leq M_0-\underline{\xi}^q\delta_r\min_{(l,i)\in\mathcal{E}_q}\frac{e^1_{il}(s_r^{q+1})}{\psi(s_r^{q+1})-|e^1_{il}(s_r^{q+1})|} \\ &\overset{(14)}{<} M_0-\underline{\xi}^q\delta_r\min_{(l,i)\in\mathcal{E}_q}\frac{1-\varepsilon_r^q}{1-|1-\varepsilon_r^q|}\leq -1,\end{aligned}$$

which is the sought contradiction and the proof is complete.

The additional assumption ensures that $M_0$ can be obtained in the time interval $[0,\infty)$. Therefore, we can make a similar argument as in the former proof, with only a slight modification, to prove the existence of $\varepsilon$.

Since we already know that the solution exists for $t\in[0,\infty)$, if all of the sets

$$\begin{aligned}\mathcal{E}_+^p &:= \left\{(j,i)\in\mathcal{E}:\exists\{t_k\}\to\infty,\ \text{s.t.}\ \lim_{k\to\infty}\frac{e^p_{ij}(t_k)}{\psi(t_k)}=1\right\} \\ \mathcal{E}_-^p &:= \left\{(j,i)\in\mathcal{E}:\exists\{t_k\}\to\infty,\ \text{s.t.}\ \lim_{k\to\infty}\frac{e^p_{ij}(t_k)}{\psi(t_k)}=-1\right\},\end{aligned}$$

is empty, then we are done. However, if some sets, e.g., $\mathcal{E}_+^1$ is non-empty, then with new $M_0$, we get a contradiction as above. Now, this concludes the proof.