Decentralized optimization via funnel coupling

Stephan Trenn

Jan C. Willems Center for Systems and Control

University of Groningen, Netherlands

Joint work with **Jin Gyu Lee** (U Cambridge, UK), **Thomas Berger** (U Paderborn, Germany) and **Hyungbo Shim** (SNU, Korea)

DMV Annual Meeting 2020, Chemnitz, Germany (online), 15 September 2020, 12:00 - 12:30
Minisymposium "The many facets of mathematical systems theory"

The problem

Minimization problem

For some differentiable and convex $F : \mathbb{R}^n \to \mathbb{R}$ find x^* which solves

$$\min_{x \in \mathbb{R}^n} F(x)$$

Challenge

$$F(x) = \sum_{i=1}^{N} F_i(x)$$

Find x^* in a decentralized way with N agents who only know their $F_i : \mathbb{R}^n \to \mathbb{R}$, which are also not assumed to be convex.

Some approaches available already, new approach: Utilize ideas from funnel control

Local gradient descent and coupling

Lemma (Global gradient descent)

$$\dot{x}(t) = -\eta \nabla F(x(t)), \quad x(0) = x_0 \in \mathbb{R}^n$$

Then $x(t) \to x^*$ *for all* x_0 *and all* $\eta > 0$*,*

Local gradient descent + coupling

$$\dot{x}_i = -\nabla F_i(x_i) + u_i \qquad x_i(0) = x_{0,i} \in \mathbb{R}^n$$

Find suitable control law

$x_i(t) \xrightarrow{u_i} x^*$ in particular, synchronization / consensus

Contents

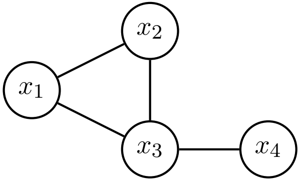
Problem statement

Given

- N agents with individual n-dimensional dynamics:

$$\dot{x}_i = f_i(t, x_i) + u_i$$

- undirected connected coupling-graph $G = (V, E)$
- local feedback $u_i = \gamma_i(x_i, x_{\mathcal{N}_i})$



$$u_1 = \gamma_1(x_1, x_2, x_3)$$
$$u_2 = \gamma_2(x_2, x_1, x_3)$$
$$u_3 = \gamma_3(x_3, x_1, x_2)$$
$$u_4 = \gamma_4(x_4, x_3)$$

Desired

Control design for practical synchronization

$$x_1 \approx x_2 \approx \ldots \approx x_n$$

A „high-gain“ result

Let $\mathcal{N}_i := \{j \in V \mid (j,i) \in E\}$ and $d_i := |\mathcal{N}_i|$ and $\mathcal{L}$ be the Laplacian of G.

Diffusive coupling

$$u_i = -k \sum_{j \in \mathcal{N}_i} (x_i - x_j) \quad \text{or, equivalently,} \quad u = -k\, \mathcal{L}\, x$$

Theorem (Practical synchronization[1])

Assumptions: G connected, all solutions of average dynamics

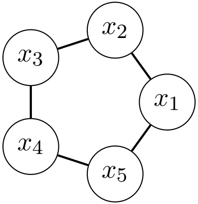
$$\dot{s}(t) = \frac{1}{N} \sum_{i=1}^{N} f_i(t, s(t))$$

remain bounded. Then $\forall \varepsilon > 0\ \exists K_\varepsilon > 0\ \forall k \geq K_\varepsilon$: Diffusive coupling results in

$$\boxed{\limsup_{t\to\infty} \|x_i(t) - x_j(t)\| < \varepsilon \quad \forall i,j \in V}$$

[1] J. Kim, J. Yang, H. Shim, J.-S. Kim, and J.H. Seo. IEEE Transactions on Automatic Control 2016.

Example[2]



Simulations in the following for $N = 5$ agents with dynamics

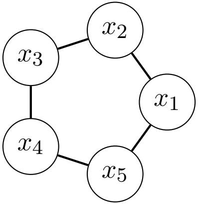
$$f_i(t, x_i) = (-1+\delta_i)x_i + 10\sin t + 10 m_i^1 \sin(0.1t + \theta_i^1) + 10 m_i^2 \sin(10t + \theta_i^2),$$

with randomly chosen parameters $\delta_i, m_i^1, m_i^1 \in \mathbb{R}$ and $\theta_i^1, \theta_i^2 \in [0, 2\pi]$.

Parameters identical in all following simulations, in particular $\delta_2 > 1$, hence agent 2 has unstable dynamics (without coupling).

[2] Shim, H. and Trenn, S. Proc. 54th IEEE Conf. Decision Control (CDC) 2015.

Example[2]

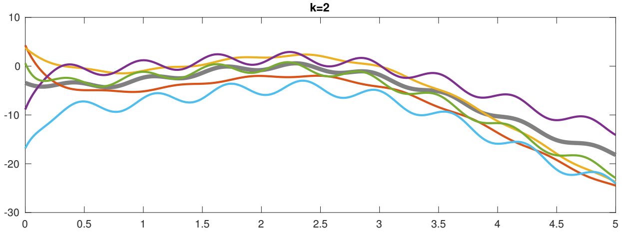


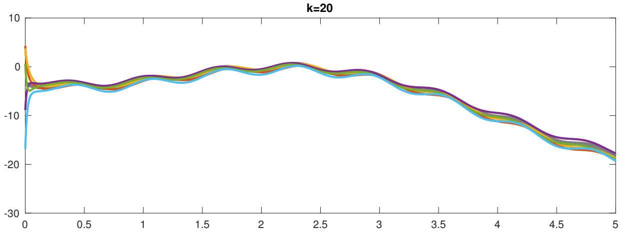
$$u = -k\,\mathcal{L}\,x$$

gray curve:

$$\dot{s}(t) = \frac{1}{N}\sum_{i=1}^{N} f_i(t, s(t))$$

$$s(0) = \tfrac{1}{N}\textstyle\sum_{i=1}^{N} x_i(0)$$





[2] Shim, H. and Trenn, S. Proc. 54th IEEE Conf. Decision Control (CDC) 2015.

Application to decentralized optimization

$$\dot{s}(t) = \frac{1}{N}\sum_{i=1}^{N} f_i(t, s(t))$$

Limit trajectory

In fact: $\limsup_{t\to\infty} |x_i(t) - s(t)| < \varepsilon \quad \forall i \in V$

Average dynamics for local gradient descent

For $f_i(t,s) = -\nabla F_i(s)$ we have

$$\dot{s}(t) = -\frac{1}{N}\sum_{i=1}^{N} \nabla F_i(s(t)) = -\frac{1}{N}\nabla F(s(t))$$

$\hookrightarrow$ Average dynamics = global gradient descent

Corollary

For sufficiently large diffusive coupling gain: $x_i(t) \to \mathbb{B}_\varepsilon(x^*)$

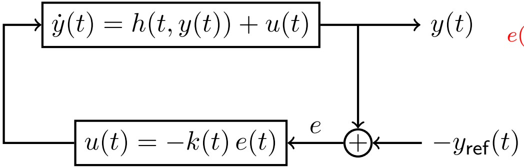
Remaining problems

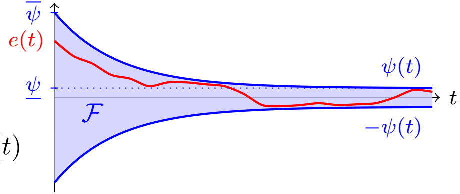
Problems

- necessary gain depends on global network structure
- all agents must use the same gain
- no direct control over convergence rate
- only practical synchronization, no asymptotic convergence

Contents

Reminder Funnel Controller





Theorem (Practical tracking[3])

Funnel Control

$$k(t) = \frac{1}{\psi(t) - |e(t)|}$$

works, in particular, errors remains within funnel for all times.

Basic idea for funnel coupling

$$u = -k\ \mathcal{L}\ x \quad \longrightarrow \quad u = -k(t)\ \mathcal{L}\ x$$

[3] A. Ilchmann, E.P. Ryan and C.J. Sangwin. ESAIM: Control, Opt. Calc. Variations, 2002

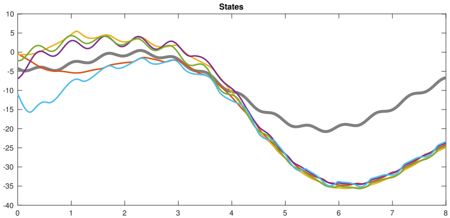
Node-wise funnel coupling approach[2]

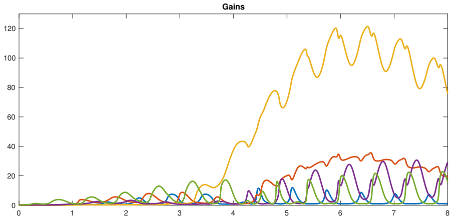
Local error

$$u_i = -k \sum_{j \in \mathcal{N}_i} (x_i - x_j) = -k \left(d_i x_i - \sum_{j \in \mathcal{N}_i} x_j \right) =: -k\, d_i\, (x_i - \overline{x}_i) =: -k_i\ e_i$$

Funnel coupling rule

$$u_i(t) = -k_i(t) e_i(t) \quad \text{with} \quad k_i(t) = \frac{1}{\psi(t) - |e_i(t)|}$$





Edgewise Funnel coupling

Diffusive coupling → edgewise Funnel coupling

$$u_i = -k_i(t) \sum_{j \in \mathcal{N}_i} (x_i - x_j) \quad \longrightarrow \quad u_i = -\sum_{j \in \mathcal{N}_i} k_{ij}(t) \cdot (x_i - x_j)$$

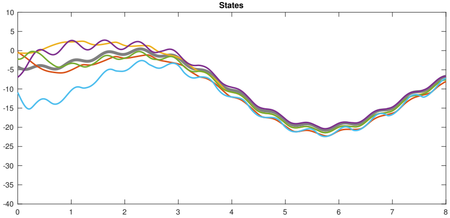
Edgewise funnel gain

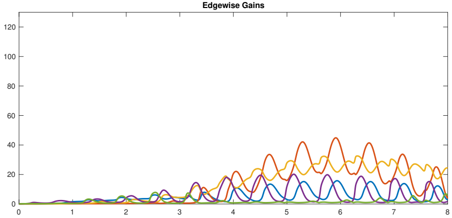
$$k_{ij}(t) = \frac{1}{\psi(t) - |e_{ij}|}, \quad \text{with} \quad e_{ij} := x_i - x_j$$

Properties:

- Decentralized, i.e. u_i only depends on state of neighbors
- Symmetry, $k_{ij} = k_{ji}$
- Laplacian feedback, $u = -\mathcal{L}_K(t,x)x$

Simulation

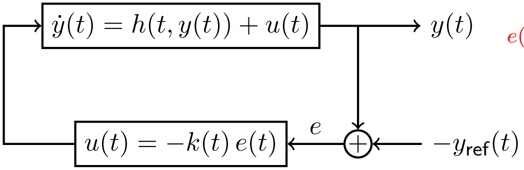


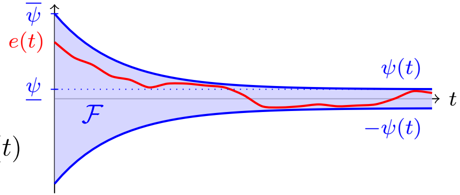


Properties

+ Synchronization occurs
+ Predictable limit trajectory (given by average dynamics)
+ Local feedback law
– No asymptotic convergence towards limit trajectory

Funnel control and asymptotic tracking





Asymptotic tracking only with unbounded gain

Asymptotic tracking $\Leftrightarrow$ $\psi(t) \underset{t\to\infty}{\to} 0$ $\Rightarrow$ $\psi(t) - \|e(t)\| \underset{t\to\infty}{\to} 0$

$\Leftrightarrow$ $k(t) = \frac{1}{\psi(t) - \|e(t)\|} \underset{t\to\infty}{\to} \infty$

Conclusion: Funnel control and asymptotic tracking not compatible **?**

Rewrite rule for funnel control

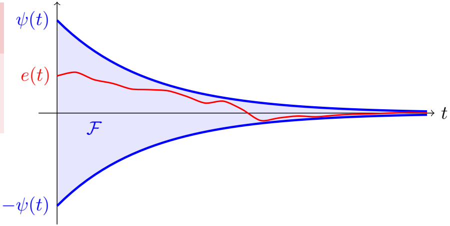
Important observation: $\psi(t) \to 0 \quad \Longrightarrow \quad \|e(t)\| \to 0$

Hence $\quad u(t) = -k(t)e(t) \to \infty \cdot 0 \quad$ not necessarily unbounded!

$$u(t) = -\frac{1}{\psi(t) - \|e(t)\|} e(t) = -\frac{1}{1 - \left\| \frac{e(t)}{\psi(t)} \right\|} \frac{e(t)}{\psi(t)} =: -\alpha(\eta(t)) \cdot \eta(t) \text{ where } \eta(t) := \frac{e(t)}{\psi(t)}$$

Asymptotic funnel control possible[5]

Rewriting classical funnel control rule
$\hookrightarrow \psi(t) \to 0$ allowed!



[5] Lee, J.G. and Trenn, S.: Asymptotic tracking via funnel control. Proc. 58th IEEE Conf. Decision Control (CDC) 2019, pp. 4228–4233, Nice, France, 2019.

Contents

Back to optimization problem

$$\boxed{\min_{x\in\mathbb{R}^n} F(x)} \qquad F \text{ convex \& differentiable} \qquad F(x) = \sum_{i=1}^{N} F_i(x)$$

Gradient descent = average dynamics $\boxed{\dot{s} = -\frac{1}{N}\sum_{i=1}^{N} \nabla F_i(s)}$

Theorem (Optimization via high gain coupling[6])

$$\dot{x}_i = -\nabla F_i(x_i) + u_i \qquad u_i = -\bar{k} \sum_{j\in\mathcal{N}_i} (x_i - x_j)$$

Then $\limsup_{t\to\infty} \|x_i(t) - s(t)\| < \varepsilon(\bar{k})$ *and* $s(t) \to x^*$

[6] J.G. Lee and H. Shim. Automatica 2020.

Replacing constant gain by edge-wise funnel gain

Replace $u_i = -\bar{k} \sum_{j \in \mathcal{N}_i} (x_i - x_j)$ by edgewise funnel coupling

$$u_i(t) = -\sum_{j \in \mathcal{N}_i} \boldsymbol{\alpha}\left(\frac{x_i(t) - x_j(t)}{\psi(t)}\right) \frac{x_i - x_j}{\psi(t)}$$

with

$$\boldsymbol{\alpha} : \mathbb{R}^n \to \mathbb{R}^{n \times n}, \eta = (\eta^1, \dots, \eta^n) \mapsto \begin{bmatrix} \frac{1}{1-|\eta^1|} & & & \\ & \frac{1}{1-|\eta^2|} & & \\ & & \ddots & \\ & & & \frac{1}{1-|\eta^n|} \end{bmatrix}$$

or, by introducing $(e_{ij}^1, e_{ij}^2, \dots, e_{ij}^n)^\top = e_{ij} := x_i - x_j$

$$u_i(t) = -\sum_{j \in \mathcal{N}_i} \left(\frac{e_{ij}^1}{\psi - |e_{ij}^1|}, \frac{e_{ij}^2}{\psi - |e_{ij}^2|}, \dots, \frac{e_{ij}^n}{\psi - |e_{ij}^n|} \right)^\top$$

Decentralized optimization via funnel coupling

$$\boxed{\min_{x\in\mathbb{R}^n} F(x)} \qquad F(x) = \sum_{i=1}^{N} F_i(x) \qquad \text{unique minimizer } x^* \in \mathbb{R}^n$$

$$\dot{x}_i = -\nabla F_i(x) - \sum_{j\in\mathcal{N}_i} \left(\frac{e_{ij}^1}{\psi - |e_{ij}^1|}, \frac{e_{ij}^2}{\psi - |e_{ij}^2|}, \ldots, \frac{e_{ij}^n}{\psi - |e_{ij}^n|} \right)^\top, \quad x_i(0) = x_{i,0} \in \mathbb{R}^n, \; i \in \mathcal{V}$$

Theorem (Decentralized optimization via funnel coupling[7])

Assume ∇F_i is globally Lipschitz and coupling graph is undirected and connected, funnel boundary $\psi \in W^{1,\infty}(\mathbb{R}_{\geq 0} \to \mathbb{R}_{>0})$ with $\psi(t) \to 0$ as $t \to \infty$ and $\|x_{i,0} - x_{j,0}\| < \psi(0)$ $\forall (i,j) \in \mathcal{E}$. Then edge-wise funnel coupling yields

$$\lim x_i(t) = x^*$$

[7] J.G. Lee, T. Berger, S. Trenn and H. Shim. Proc. European Control Conference (ECC) 2020.

Main steps of proof

Skip Proof

Step 1: Show that solution of closed loop exists on $[0, \infty)$, in particular

$$\|x_i(t) - x_j(t)\|_\infty < \psi(t) \quad \forall t \geq 0$$

Step 2: Write $x_i = x_{\text{avg}} + (\boldsymbol{t}_i^\top \otimes I_n)\widetilde{x}$, where $\boldsymbol{t}_i^\top$ is the i-th row of $T(T^\top T)^{-1}$, T is the incidence matrix of a spanning tree and

$$x_{\text{avg}} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad \text{and} \quad \widetilde{x} = (T^\top \otimes I_n)(x_1^\top, \ldots, x_N^\top)^\top$$

Then $\|\tilde{x}(t)\|_\infty < \psi(t)$, in particular, $x_i(t) \to x_{\text{avg}}(t)$

Step 3: Use Lyapunov function $V(x_{\text{avg}}) := F(x_{\text{avg}}) - F(x^*)$ to show convergence of x_{avg} towards x^*.

Convergence rate

Average dynamics = gradient descent

$$\dot{s} = -\frac{1}{N}\sum_{i=1}^{N} F_i(s)$$

Convergence towards x^* not influenced by choice of coupling rule

Edge-wise funnel coupling

$$x_i(t) - s(t) \to 0 \quad \text{as } t \to \infty$$

Convergence rate (and transient behavior) directly influenceable by choice of ψ

Example: Distributed Least-square solver

Problem

Find least-square solution of $Ax = b$, where $A \in \mathbb{R}^{M \times n}$ with $M = \sum_{i=1}^{N} m_i$. Then

$$\frac{1}{2}\|Ax - b\|_2^2 = \sum_{i=1}^{N} \frac{1}{2}\|A_i x - b_i\|_2^2$$

where $A_i \in \mathbb{R}^{m_i \times n}$, $b_i \in \mathbb{R}^{m_i}$ are the corresponding block rows of A and b.

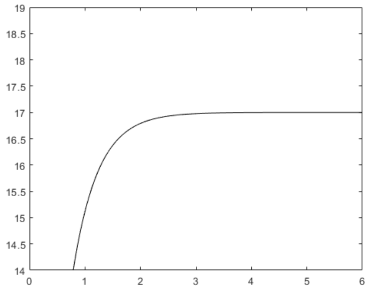
Local gradient descent with funnel coupling

$$\dot{x}_i = -A_i^\top (A_i x_i - b_i) - \sum_{j \in \mathcal{N}_i} \left(\frac{e_{ij}^1}{\psi - |e_{ij}^1|}, \frac{e_{ij}^2}{\psi - |e_{ij}^2|}, \ldots, \frac{e_{ij}^n}{\psi - |e_{ij}^n|} \right)^\top$$

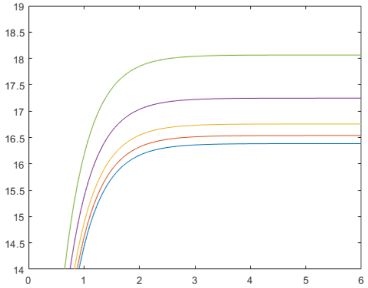
Simulation

For $A = [1,1,2,2,1]^\top$ and $b = [1,10,20,18,100]^\top$,
$x_1(0) = 0$, $x_2(0) = -x_3(0) = 0.1$, $x_4(0) = -x_5(0) = 0.2$,
$\psi(t) = \exp(-0.8t)$ and a line coupling we obtain

Average dynamics
= global gradient descent



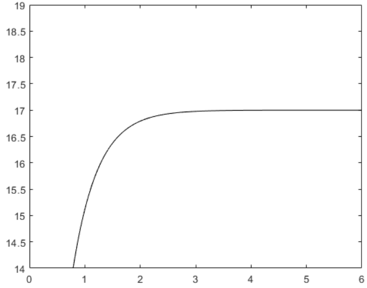
Constant gain
$k = 100$



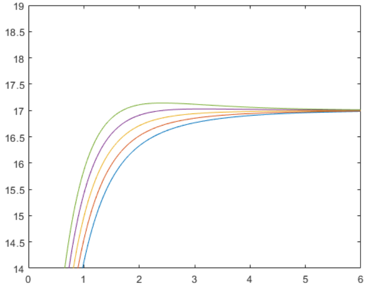
Simulation

For $A = [1, 1, 2, 2, 1]^\top$ and $b = [1, 10, 20, 18, 100]^\top$,
$x_1(0) = 0$, $x_2(0) = -x_3(0) = 0.1$, $x_4(0) = -x_5(0) = 0.2$,
$\psi(t) = \exp(-0.8t)$ and a line coupling we obtain

Average dynamics
$=$ global gradient descent



Edgewise Funnel coupling



Conclusion

Optimzation problem

$\min_x F(x)$ with convex $F = \sum_i F_i$ can be solved

- completely decentralized
- no derivatives need to be communicated
- prespecified convergence rate (funnel)
- asymptotic convergence

Future work

- Decentralized optimization
  - Constrained optimization
  - Nonconvex cost function
- Funnel coupling
  - Other applications, e.g. decentralized observation/control
  - Directed graphs