Contracting and Semicontracting Dynamics on Networks

Francesco Bullo

Center for Control,
Dynamical Systems & Computation
University of California at Santa Barbara
https://fbullo.github.io

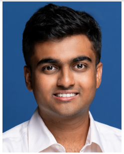
Workshop on Emergent Behavior in Complex Systems of Interacting Agents
Institute for Mathematical and Statistical Innovation (IMSI), Chicago, March 21, 2025

Acknowledgments

Veronica Centorrino
Scuola Sup Meridionale / ETH

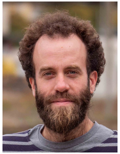
Alexander Davydov
UC Santa Barbara / Rice University

Giulia De Pasquale
TU Eindhoven



Anand Gokhale
UC Santa Barbara

Kevin D. Smith
Utilidata



Robin Delabays
HES-SO Sion

Anton Proskurnikov
Politecnico Torino

Giovanni Russo
Univ Salerno

Elena Valcher
Univ Padova

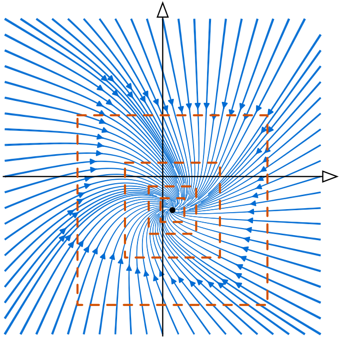
AFOSR ARO DTRA/ERDC

Frederick Leve @AFOSR FA9550-22-1-0059
Paul Tandy @DTRA W912HZ-22-2-0010
Donald Wagner @AFOSR FA9550-21-1-0203
Derya Cansever @ARO W911NF-24-1-0228

contractivity = robust computationally-friendly stability
fixed point theory + Lyapunov stability theory + geometry of metric spaces

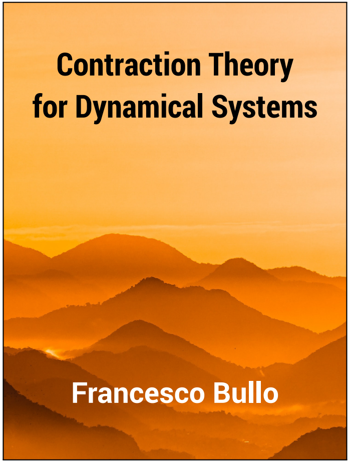
highly-ordered transient and asymptotic behavior:

1. unique globally exponential stable equilibrium
   & two natural Lyapunov functions
2. robustness properties
   - bounded input, bounded output (iss)
   - finite input-state gain
   - robustness margin wrt unmodeled dynamics
   - robustness margin wrt delayed dynamics
3. ...



search for contraction properties
design engineering systems to be contracting

Ongoing education and research on Contraction Theory



"Continuous improvement is better than delayed perfection"
Mark Twain

- *Textbook*: Contraction Theory for Dynamical Systems, Francesco Bullo, rev 1.2, Aug 2024. (PDF freely available)
  https://fbullo.github.io/ctds
- *Tutorial slides*: **https://fbullo.github.io/ctds**
- *Youtube lectures*: "Minicourse on Contraction Theory"
  https://youtu.be/FQV5PrRHks8 6 lectures, total 12h

Outline

Discrete-time dynamics and Lipschitz constants

$$x_{k+1} = \mathsf{F}(x_k) \qquad \text{on } \mathbb{R}^n \text{ with norm } \|\cdot\| \text{ and induced norm } \|\cdot\|$$

Lipschitz constant (max expansion factor)

$$\begin{aligned}\mathsf{Lip}(\mathsf{F}) &= \inf\{\ell > 0 \text{ such that } \|\mathsf{F}(x) - \mathsf{F}(y)\| \leq \ell\|x - y\| \quad \text{for all } x, y\}\\ &= \sup_x \|\mathsf{J}_{\mathsf{F}}(x)\|\end{aligned}$$

For **scalar map** f, $\quad \mathsf{Lip}(f) = \sup_x |f'(x)|$

For **affine map** $\mathsf{F}_A(x) = Ax + a$

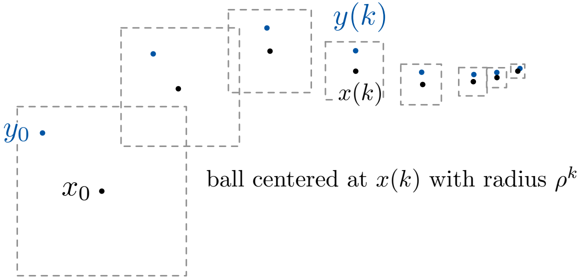
$$\|x\|_{2,P} = (x^\top P x)^{1/2} \qquad \mathsf{Lip}_{2,P}(\mathsf{F}_A) = \|A\|_{2,P} \leq \ell \quad \iff \quad A^\top P A \preceq \ell^2 P$$

$$\|x\|_{\infty,\eta} = \max_i |x_i|/\eta_i \qquad \mathsf{Lip}_{\infty,\eta}(\mathsf{F}_A) = \|A\|_{\infty,\eta} \leq \ell \quad \iff \quad \eta^\top |A| \leq \ell \eta^\top$$

Banach contraction theorem for discrete-time dynamics:

If $\rho := \text{Lip}(\mathsf{F}) < 1$, then

1. F is **contracting** = distance between trajectories decreases exp fast (ρ^k)
2. F has a unique, glob exp stable equilibrium x^*



Example induced log norms

Vector norm	Induced matrix norm	Induced matrix log norm
$\|x\|_1 = \sum_{i=1}^n \|x_i\|$	$\|A\|_1 = \max_{j\in\{1,\dots,n\}} \sum_{i=1}^n \|a_{ij}\|$	$\mu_1(A) = \max_{j\in\{1,\dots,n\}} \Big(a_{jj} + \sum_{i=1,i\neq j}^n \|a_{ij}\|\Big)$ $=$ max column "absolute sum" of A
$\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$	$\|A\|_2 = \sqrt{\lambda_{\mathsf{max}}(A^\top A)}$	$\mu_2(A) = \lambda_{\mathsf{max}}\Big(\frac{A+A^\top}{2}\Big)$
$\|x\|_\infty = \max_{i\in\{1,\dots,n\}} \|x_i\|$	$\|A\|_\infty = \max_{i\in\{1,\dots,n\}} \sum_{j=1}^n \|a_{ij}\|$	$\mu_\infty(A) = \max_{i\in\{1,\dots,n\}} \Big(a_{ii} + \sum_{j=1,j\neq i}^n \|a_{ij}\|\Big)$ $=$ max row "absolute sum" of A

$$\dot{x} = \mathsf{F}(x) \qquad \text{on } \mathbb{R}^n \text{ with norm } \|\cdot\| \text{ and induced log norm } \mu(\cdot)$$

One-sided Lipschitz constant (max expansion rate)

$$\begin{aligned} \mathsf{osLip}(\mathsf{F}) &= \inf\{b \in \mathbb{R} \text{ such that } \langle\!\langle \mathsf{F}(x) - \mathsf{F}(y), x - y \rangle\!\rangle \leq b\|x-y\|^2 \quad \text{for all } x, y\} \\ &= \sup_x \mu(\mathsf{J}_{\mathsf{F}}(x)) \end{aligned}$$

For **scalar map** f, $\quad \mathsf{osLip}(f) = \sup_x f'(x)$

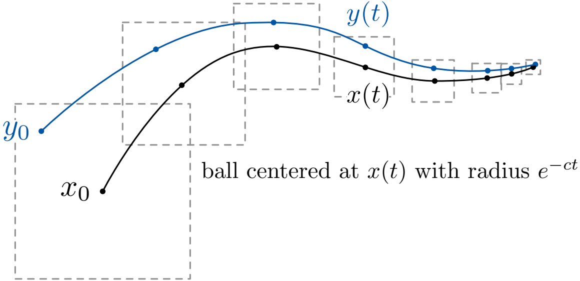
For **affine map** $\mathsf{F}_A(x) = Ax + a$

$$\mathsf{osLip}_{2,P}(\mathsf{F}_A) = \mu_{2,P}(A) \leq \ell \qquad \iff \qquad A^\top P + AP \preceq 2\ell P$$

$$\mathsf{osLip}_{\infty,\eta}(\mathsf{F}_A) = \mu_{\infty,\eta}(A) \leq \ell \qquad \iff \qquad a_{ii} + \sum_{j \neq i} |a_{ij}| \eta_i / \eta_j \leq \ell$$

Banach contraction theorem for continuous-time dynamics:
If $-c := \text{osLip}(\mathsf{F}) < 0$, then

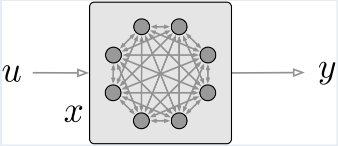
1. F is **infinitesimally contracting** = distance between trajectories decreases exp fast (e^{-ct})
2. F has a unique, glob exp stable equilibrium x^*



Example contracting systems

1. **neural network dynamics** under assumptions on synaptic matrix
   (recurrent, implicit, reservoir computing, etc)
2. **interconnected systems** under contractivity and small-gain assumptions
   (Hurwitz Metzler matrices, network small-gain theorem, etc)
3. Lur'e-type systems under assumptions on nonlinearity and LMI conditions
   (Lipschitz, incrementally passive, monotone, conic, etc)
4. gradient descent flows under strong convexity assumptions
   (proximal, primal-dual, distributed, Hamiltonian, saddle, pseudo, best response, etc)
5. data-driven learned models (imitation learning)
6. feedback linearizable systems with stabilizing controllers
7. incremental ISS systems
8. nonlinear systems with a locally exponentially stable equilibrium
   are contracting with respect to appropriate Riemannian metric

Firing-rate neural networks



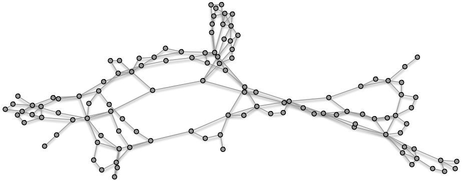
$$\dot{x} = -x + \Phi(Ax + Bu + b) \qquad \text{(recurrent NN)}$$
$$x = \Phi(Ax + Bu + b) \qquad \text{(implicit NN)}$$
$$x_{k+1} = (1-\alpha)x_k + \alpha\Phi(Ax_k + Bu + b) \qquad \text{(Euler discrt.)}$$

If
$$\mu_\infty(A) < 1 \qquad \Big(\text{i.e., } a_{ii} + \sum\nolimits_{j\neq i} |a_{ij}| < 1 \text{ for all } i\Big)$$

- **recurrent NN is infinitesimally contracting** with rate $1 - \mu_\infty(A)_+$
- **implicit NN is well posed**
- **Euler discretization is contracting** at $\alpha^* = (1 - \min_i (a_{ii})_-)^{-1}$

Outline

1. A brief review of contractivity concepts
   - From discrete-time to continuous-time dynamics
   - Examples and selected properties
2. Network contraction theorem
3. Semicontractivity, ergodic coefficients, and duality
   - Systems with invariance/conservation properties
   - Induced seminorms and duality
4. Conclusions and future research

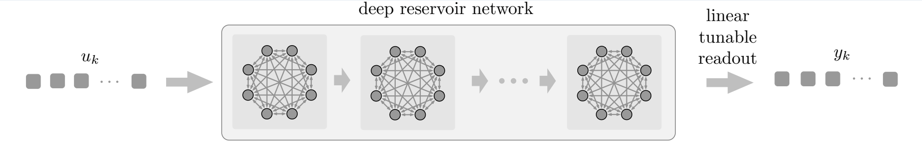


Network Contraction Theorem. Consider interconnected subsystems

$$\dot{x}_i = \mathsf{F}_i(x_i, x_{-i}), \qquad \text{for } i \in \{1, \dots, n\}$$

- **contractivity wrt x_i**: $\quad \mathrm{osLip}_{x_i}(\mathsf{F}_i) \leq -c_i < 0$
- **Lipschitz wrt x_j, $j \neq i$**: $\quad \mathrm{Lip}_{x_j}(\mathsf{F}_i) \leq \ell_{ij}$
- gain matrix $\begin{bmatrix} -c_1 & \dots & \ell_{1n} \\ \vdots & \ddots & \vdots \\ \ell_{n1} & \dots & -c_n \end{bmatrix}$ is **Hurwitz**

$\Longrightarrow$ **interconnected system** is contracting with rate $|\alpha(\text{gain matrix})|$



$$x_{k+1}^{(1)} = (1-\alpha)x_k^{(1)} + \alpha\Phi\big(A^{(1)}x_k^{(1)} + B^{(1)}u_k + b^{(1)}\big)$$
$$x_{k+1}^{(i)} = (1-\alpha)x_k^{(i)} + \alpha\Phi\big(A^{(i)}x_k^{(i)} + B^{(i)}x_k^{(i-1)} + b^{(i)}\big)$$

(leaky integrator reservoirs)

Deep reservoir network is contracting (and "echo state property") if

$$\mu_\infty(A^{(i)}) < 1 \qquad \text{for each} \quad i \quad \text{and} \quad \alpha \leq \alpha^{**}$$

H. Jaeger. The "echo state" approach to analysing and training recurrent neural networks. Technical report, German National Research Center for Information Technology, 2001

Metzler Hurwitz matrices and the small gain theorem

$$\begin{bmatrix} -c_1 & \dots & \ell_{1n} \\ \vdots & & \vdots \\ \ell_{n1} & \dots & -c_n \end{bmatrix}$$ is **Metzler** (Perron-Frobenius Theorem applies)

Hurwitzness depends upon both topology and edge weights

- M Hurwitz iff there exists a positive ξ such that $M\xi < \mathbb{0}_n$ (power method)
- For $n = 2$, Hurwitz if and only if **small gain condition**

$$\text{cycle gain} := \frac{\ell_{12}}{c_1}\frac{\ell_{21}}{c_2} < 1$$

- For $n \geq 3$, Hurwitz if **network small gain condition**

see **network small-gain theorem for Metzler matrices**

Outline

Consider a vector field $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$, and let $\xi, \eta \in \mathbb{R}^n$.

- Invariance property: for all $x, y \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$,

$$\mathsf{F}(x + \alpha\xi) = \mathsf{F}(x) \qquad \text{or equivalently} \qquad D\mathsf{F}(x)\xi = \mathbb{0}_n$$

- Conservation property: for all $x, y \in \mathbb{R}^n$,

$$\eta^\top \mathsf{F}(x) = \eta^\top \mathsf{F}(y) \qquad \text{or equivalently} \qquad \eta^\top D\mathsf{F}(x) = \mathbb{0}_n^\top$$

Prototypical dynamics

Let $A \in \mathbb{R}^{n\times n}$ be row-stochastic: $A\mathbb{1}_n = \mathbb{1}_n$ and $A \geq 0$

Averaging Systems

$$x_{k+1} = Ax_k$$

Invariance: dynamics unaffected by translations in $\operatorname{span}\{\mathbb{1}_n\}$

Examples: distributed optimization, robotic coordination, frequency synchronization, ...

Dynamical Flow Systems

$$x_{k+1} = A^\top x_k$$

Conservation: quantity $\mathbb{1}_n^\top x$ is constant

Examples: compartmental models, Markov chains

Given row-stochastic $A \in \mathbb{R}^{n \times n}$,
Markov-Dobrushin ergodic coefficient

$$\tau_1(A) = \max_{\|z\|_1 = 1, \mathbb{1}_n^\top z = 0} \|A^\top z\|_1$$

$\tau_1(A) < 1$ under mild connectivity conditions
$\tau_p(A)$ also defined for general $p \in [1, \infty]$

How is τ_1 an induced norm?

A. A. Markov. Extensions of the law of large numbers to dependent quantities. *Izvestiya Fiziko-matematicheskogo obschestva pri Kazanskom universitete*, 15, 1906. (in Russian)
R. L. Dobrushin. Central limit theorem for nonstationary Markov chains. I. *Theory of Probability & Its Applications*, 1(1):65–80, 1956.

$A \in \mathbb{R}^{n\times n}$ row-stochastic

Classical Property of Averaging Systems $x_{k+1} = Ax_k$

Given $x \in \mathbb{R}^n$, max-min disagreement:

$$s(Ax) \;\leq\; \tau_1(A)\; s(x), \qquad \text{where} \quad s(x) = \max_i\{x_i\} - \min_j\{x_j\}$$

Classical Property of Markov Chains $x_{k+1} = A^\top x_k$

Given π, σ in the simplex Δ_n, total variation distance:

$$d_{\mathsf{TV}}(A^\top\pi, A^\top\sigma) \;\leq\; \tau_1(A)\; d_{\mathsf{TV}}(\pi,\sigma), \qquad \text{where} \quad d_{\mathsf{TV}}(\pi,\sigma) = \frac{1}{2}\sum_i |\pi_i - \sigma_i|$$

Why is the same τ_1 relevant in both cases?

Seminorms

A **seminorm** is a function $|||\cdot||| : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ s.t., $\forall a \in \mathbb{R}$ and $\forall x, y \in \mathbb{R}^n$:

1. (*homogeneity*): $|||ax||| = |a|\,|||x|||$
2. (*subadditivity*): $|||x+y||| \leq |||x||| + |||y|||$

The *kernel* is the vector space:

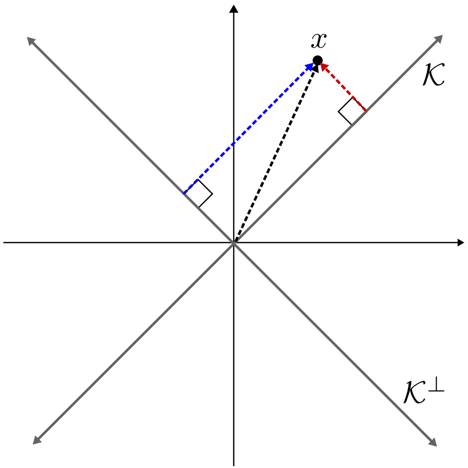
$$\mathcal{K} = \{x \in \mathbb{R}^n : |||x||| = 0\}$$

We focus on *consensus seminorms*, where $\mathcal{K} = \operatorname{span}\{\mathbb{1}_n\}$.

Note: $|||\cdot|||$ is invariant under translations in $\mathcal{K}$

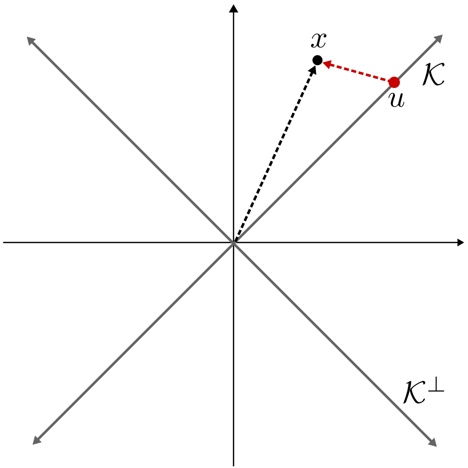
Projection and distance-based seminorms: graphical definitions

Projection seminorms



$$|||x|||_{\mathsf{proj},p} \triangleq \|\Pi_\perp x\|_p, \quad \Pi_\perp = \Pi_\perp^\top$$

Distance seminorms



$$|||x|||_{\mathsf{dist},p} \triangleq \min_{u\in\mathcal{K}} \|x-u\|_p$$

Consensus seminorms

When $\mathcal{K} = \operatorname{span}\{\mathbb{1}_n\}$, **consensus seminorms**

	$\|\|\|x\|\|\|_{\mathsf{proj},p}$	$\|\|\|x\|\|\|_{\mathsf{dist},p}$
ℓ_1	$\sum_{i=1}^{n} \lvert x_i - x_{\mathrm{avg}} \rvert$	$\sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} x_{(i)} - \sum_{j=\lceil \frac{n}{2} \rceil + 1}^{n} x_{(j)}$
ℓ_2	$\sqrt{\frac{1}{n} \sum_{i,j} (x_i - x_j)^2}$	$\sqrt{\frac{1}{n} \sum_{i,j} (x_i - x_j)^2}$
ℓ_∞	$\max_i \lvert x_i - x_{\mathrm{avg}} \rvert$	$\frac{1}{2} \big(x_{(1)} - x_{(n)} \big)$

where we have sorted $x_{(1)} \geq x_{(2)} \geq \cdots \geq x_{(n)}$

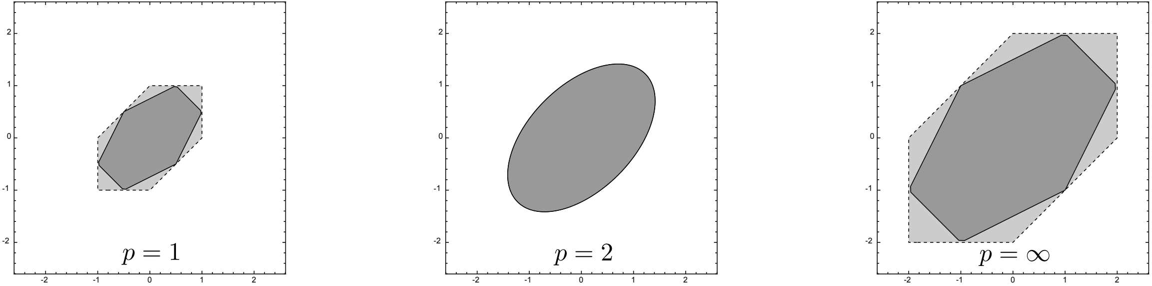


Figure: Two-dimensional sections of three-dimensional unit disks of projection (solid contours) and distance (dashed contours) consensus seminorms. We plot the sections corresponding to $(x_1, x_2, x_3 = 0)$.

Induced matrix seminorms

Consider a seminorm $|||\cdot|||$ on $\mathbb{R}^n$ with kernel $\mathcal{K}$.

Induced matrix seminorm: function $|||\cdot||| : \mathbb{R}^{n\times n} \to \mathbb{R}_{\geq 0}$ defined by

$$|||A||| \;=\; \max_{\substack{|||x|||\leq 1\\ x\perp\mathcal{K}}} |||Ax|||$$

⚠ In general, $|||Ax||| \not\leq |||A|||\,|||x|||$
Inequality is true if $x \in \mathcal{K}^\perp$ or $A\mathcal{K} \subseteq \mathcal{K}$

Key facts about dual and induced norms

Properties of dual and induced norms

1. **ℓ_p and ℓ_q norms are dual**, for $1/p + 1/q = 1$

$$\|\cdot\|_p = (\|\cdot\|_q)_\star \qquad\qquad \|\cdot\|_q = (\|\cdot\|_p)_\star$$

2. dual norm satisfies (sharp) *Hölder inequality*: $\quad x^\top y \leq \|x\|_p \|y\|_q$
3. equality between dual induced norms: $\quad \|A\|_p = \|A^\top\|_q$
4. induced norm is submultiplicative: $\quad \|AB\| \leq \|A\| \|B\|$

Key facts about dual and induced seminorms

Properties of dual and induced seminorms

1. **ℓ_p-distance and ℓ_q-projection seminorms are dual**, for $1/p + 1/q = 1$

$$|||\cdot|||_{\mathrm{dist},p} = (|||\cdot|||_{\mathrm{proj},q})_\star \qquad\qquad |||\cdot|||_{\mathrm{proj},q} = (|||\cdot|||_{\mathrm{dist},p})_\star$$

2. dual seminorm satisfies (sharp) *Markov inequality*: $x^\top \Pi_\perp y \leq |||x|||_{\mathrm{dist},p}\, |||y|||_{\mathrm{proj},q}$
3. equality between dual induced seminorms: $|||A|||_{\mathrm{dist},p} = |||A^\top|||_{\mathrm{proj},q}$
4. induced seminorm is submultiplicative: $|||AB||| \leq |||A|||\,|||B|||$ if $A\mathcal{K} \subseteq \mathcal{K}$ or $B\mathcal{K}^\top \subseteq \mathcal{K}^\top$

Ergodic coefficients are induced seminorms

$$|||A|||_{\mathrm{dist},p} = |||A^\top|||_{\mathrm{proj},q} = \tau_q(A) := \max_{\|z\|_q=1,\; z\perp\mathbb{1}_n} \|A^\top z\|_q$$

How Markov and Banach's results meet

Classical Property of Averaging Systems

Given row-stochastic $A \in \mathbb{R}^{n\times n}$ and $x, y \in \mathbb{R}^n$:

$$\begin{aligned} |||A(x-y)|||_{\mathsf{dist},\infty} &\leq \tau_1(A)|||x-y|||_{\mathsf{dist},\infty} \\ &= |||A|||_{\mathsf{dist},\infty}|||x-y|||_{\mathsf{dist},\infty} \end{aligned}$$

Classical Property of Markov Chains

Given row-stochastic $A \in \mathbb{R}^{n\times n}$ and π, σ in the simplex Δ_n:

$$\begin{aligned} |||A^\top(\pi-\sigma)|||_{\mathsf{proj},1} &\leq \tau_1(A)|||\pi-\sigma|||_{\mathsf{proj},1} \\ &= |||A^\top|||_{\mathsf{proj},1}|||\pi-\sigma|||_{\mathsf{proj},1} \end{aligned}$$

Outline

1. A brief review of contractivity concepts
   - From discrete-time to continuous-time dynamics
   - Examples and selected properties
2. Network contraction theorem
3. Semicontractivity, ergodic coefficients, and duality
   - Systems with invariance/conservation properties
   - Induced seminorms and duality
4. Conclusions and future research

Summary and future work

1. ergodic coefficients are contraction factors
2. duality explains their roles in both averaging and flow systems
3. nonEuclidean norms play a key role
4. **semicontraction theory**
    1. discrete/continuous-time Markov chains
    2. discrete/continuous-time nonlinear consensus algorithms
    3. local contractivity of Kuramoto and Kuramoto-Sakaguchi models

References

Contraction theory, the network contraction theorem, and neural networks:

- A. Davydov, S. Jafarpour, and F. Bullo. Non-Euclidean contraction theory for robust nonlinear stability. *IEEE Transactions on Automatic Control*, 67(12):6667–6681, 2022. doi
- A. Davydov, A. V. Proskurnikov, and F. Bullo. Non-Euclidean contraction analysis of continuous-time neural networks. *IEEE Transactions on Automatic Control*, 70(1), 2025. doi
- V. Centorrino, A. Gokhale, A. Davydov, G. Russo, and F. Bullo. Euclidean contractivity of neural networks with symmetric weights. *IEEE Control Systems Letters*, 7:1724–1729, 2023. doi

Semicontraction theory

- S. Jafarpour, P. Cisneros-Velarde, and F. Bullo. Weak and semi-contraction for network systems and diffusively-coupled oscillators. *IEEE Transactions on Automatic Control*, 67(3):1285–1300, 2022. doi
- G. De Pasquale, K. D. Smith, F. Bullo, and M. E. Valcher. Dual seminorms, ergodic coefficients, and semicontraction theory. *IEEE Transactions on Automatic Control*, 69(5):3040–3053, 2024. doi
- R. Delabays and F. Bullo. Semicontraction and synchronization of Kuramoto-Sakaguchi oscillator networks. *IEEE Control Systems Letters*, 7:1566–1571, 2023. doi

Continuous-time semicontraction theory

The *induced log seminorm* of $A \in \mathbb{R}^{n\times n}$ is

$$\mu_{|||\cdot|||}(A) \triangleq \lim_{h\to 0^+} \frac{|||I_n + hA||| - 1}{h}$$

Laplacian L, corresponding to weighted digraph with adj. matrix A:

$$\mu_{\mathsf{dist},1}(-L) = -\min_j \left\{ (d_{\mathrm{out}})_j - \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor - 1} a_{(i),j} + \sum_{i=\lceil \frac{n}{2} \rceil}^{n-1} a_{(i),j} \right\}, \quad d_{\mathrm{out}} = A\mathbb{1}_n$$

$$\mu_{\mathsf{dist},2}(-L) = \min \left\{ b : \Pi_\perp L + L^\top \Pi_\perp \succeq -2b\Pi_\perp \right\}, \quad \Pi_\perp = I_n - \tfrac{1}{n}\mathbb{1}_n\mathbb{1}_n^\top$$

$$\mu_{\mathsf{dist},\infty}(-L) = -\min_{i\neq j} \left\{ a_{ij} + a_{ji} + \sum_{k\neq i,j} \min\{a_{ik}, a_{jk}\} \right\}$$

Let $p, q \in [1,\infty]$ such that $p^{-1} + q^{-1} = 1$. For any matrix $M \in \mathbb{R}^{n\times n}$, and any kernel $\mathcal{K}$,

$$\mu_{\mathrm{dist},p}(M) = \mu_{\mathrm{proj},q}(M^\top)$$

Open problem

consider the set of undirected, unweighted connected graphs + selfloops

for each adjacency A_i, define row-stochastic $\mathcal{A}_i = \operatorname{diag}(A_i \mathbb{1}_n)^{-1} A_i$ (equal neighbor)

find a consensus seminorm $|||\cdot|||$ such that, for each i,

$$|||\mathcal{A}_i||| < 1$$

or **prove** that it does not exist