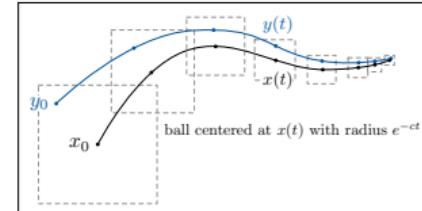
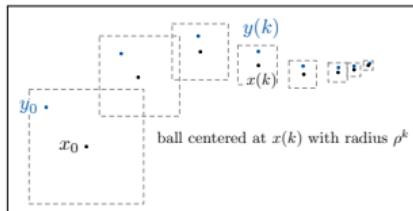


# Contracting Dynamical Systems: A Tutorial on Theory and Applications

Francesco Bullo

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

Minicourse, Focus Period "Network Dynamics and Control," University of Linköping, Sweden, 2023/9/13-15  
Tutorial (based on lectures @ SSM in Napoli Nov '22, ACC @ San Diego Jun '23). This version: 2023/09/16

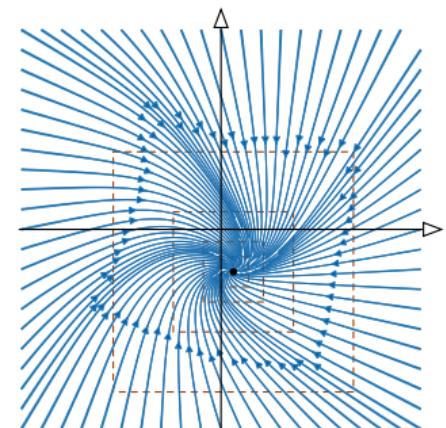


**contractivity = robust computationally-friendly stability**

fixed point theory + Lyapunov stability theory + geometry of metric spaces

**highly-ordered transient and asymptotic behavior, no anonymous constants/functions:**

- 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 periodic input, periodic output
- ④ modularity and interconnection properties
- ⑤ accurate numerical integration and equilibrium point computation



**search for** contraction properties  
**design** engineering systems to be contracting  
**verify** correct/safe behavior via known Lipschitz constants

# Acknowledgments



Veronica Centorrino  
Scuola Sup  
Meridionale



Pedro Cisneros-Velarde  
UIUC



Lily Cothren  
UC Boulder



Emiliano Dall'Anese  
UC Boulder



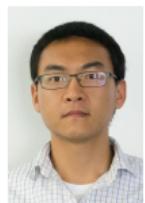
Alex Davydov  
UC Santa Barbara



Giulia De Pasquale  
ETH



Robin Delabays  
HES-SO Sion



Xiaoming Duan  
Shanghai Jiao Tong



Anand Gokhale  
UC Santa Barbara



Saber Jafarpour  
Georgia Tech



Anton Proskurnikov  
Politecnico Torino



Giovanni Russo  
Univ Salerno



John W.  
Simpson-Porco  
University of Toronto



Kevin D. Smith  
Utilidata



Elena Valcher  
Universita di Padova

# Outline

## §1. History and resources

## §2. Basic definitions: discrete and continuous-time dynamics on vector spaces

- The linear algebra of matrix norms; see CTDS Chapter 2
- Properties of induced matrix norms and Lipschitz constants

## §3. Example systems

- Constrained, distributed and proximal gradient dynamics
- Continuous-time recurrent neural networks
- Nonlinear dynamics in Lur'e form

## §4. Properties of contracting dynamics

- Equilibria, Lyapunov functions, and Euler discretization
- Incremental input-to-state stability
- Contractivity of interconnected systems
- Additional properties: entrainment, robustness wrt unmodeled dynamics and delays

## §5. Example applications

- Gradient dynamics and Nash equilibria in games
- Time-varying gradient dynamics and feedback optimization
- Recurrent and implicit neural networks

## §6. Generalizations with examples

- G1: Semicontractivity: Primal-dual gradient with redundant constraints
- G2: Contractivity on Riemannian manifolds and the Karcher mean

## §7. Conclusions and future research

## §8. Advanced Topics

- More on semicontractivity: ergodic coefficients and duality
- Network small-gain theorem for Metzler matrices
- More on the Kuramoto-Sakaguchi model and synchronization
- Proof of semicontractivity of saddle matrices
- Proof of Euler discretization theorem
- Non-Euclidean Monotone Operator Theory

- **Origins**

S. Banach. Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales. *Fundamenta Mathematicae*, 3(1):133–181, 1922. doi: 

- **Dynamics:**

G. Dahlquist. *Stability and error bounds in the numerical integration of ordinary differential equations*. PhD thesis, (Reprinted in Trans. Royal Inst. of Technology, No. 130, Stockholm, Sweden, 1959), 1958

S. M. Lozinskii. Error estimate for numerical integration of ordinary differential equations. I. *Izvestiya Vysshikh Uchebnykh Zavedenii. Matematika*, 5:52–90, 1958. URL <http://mi.mathnet.ru/eng/ivm2980>. (in Russian)



- **Computation:**

C. A. Desoer and H. Haneda. The measure of a matrix as a tool to analyze computer algorithms for circuit analysis. *IEEE Transactions on Circuit Theory*, 19(5):480–486, 1972. doi: 

- **Systems and control:**

W. Lohmiller and J.-J. E. Slotine. On contraction analysis for non-linear systems. *Automatica*, 34(6):683–696, 1998. doi: 

- **Incomplete list of scientists who influenced me**

Aminzare, Arcak, Chung, Coogan, Corless, Di Bernardo, Manchester, Margaliot, Martins, Pavel, Pavlov, Pham, Proskurnikov, Russo, Sepulchre, Slotine, Sontag, ...

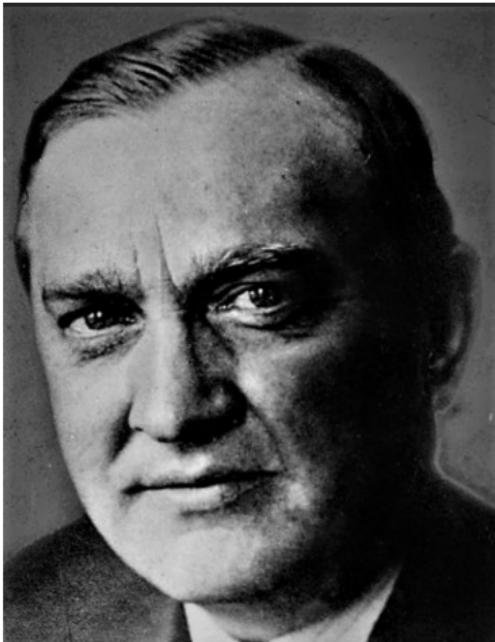
- **Surveys:**

Z. Aminzare and E. D. Sontag. Contraction methods for nonlinear systems: A brief introduction and some open problems. In *IEEE Conf. on Decision and Control*, pages 3835–3847, Dec. 2014b. doi: 

M. Di Bernardo, D. Fiore, G. Russo, and F. Scafuti. Convergence, consensus and synchronization of complex networks via contraction theory. In *Complex Systems and Networks*. Springer, 2016. doi: 

H. Tsukamoto, S.-J. Chung, and J.-J. E. Slotine. Contraction theory for nonlinear stability analysis and learning-based control: A tutorial overview. *Annual Reviews in Control*, 52:135–169, 2021. doi: 

P. Giesl, S. Hafstein, and C. Kawan. Review on contraction analysis and computation of contraction metrics. *Journal of Computational Dynamics*, 10(1):1–47, 2023. doi: 



**Figure:** Stefan Banach (Krakow, 30 Mar 1892 – Lviv, 31 Aug 1945) was a self-taught Polish mathematician

1920: doctoral thesis on Banach spaces @ University of Lviv  
1920-1922: Assistant Professor @ Lwow Polytechnic  
1922: Full Professor @ Lwow Polytechnic  
1924: Member of the Polish Academy of Arts and Sciences  
1929: Founder, Lvov School of Mathematics  
1931: first functional analysis: “Theory of Linear Operations”  
1939-45: dark years

S. Banach. Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales. *Fundamenta Mathematicae*, 3(1):133–181, 1922. doi: 

The Banach Contraction Theorem is also referred to as the *Picard-Banach-Caccioppoli*, because of the earlier work by Picard (1890) on the “method of successive approximations” and the later independent work by Renato Caccioppoli (1930).



**Figure:** Renato Caccioppoli (Napoli, 20 Jan 1904 – Napoli, 8 May 1959) was an Italian mathematician

1921-1932 student and researcher @ Napoli

1931-1934 professor @ Padova

1934-1959 professor @ Napoli

R. Caccioppoli. Un teorema generale sull'esistenza di elementi uniti in una trasformazione funzionale. *Rendiconti dell'Accademia Nazionale dei Lincei*, 11:794–799, 1930

## Contraction conditions without Jacobians

- ① **one-sided Lipschitz maps** in: G. Dahlquist. Error analysis for a class of methods for stiff non-linear initial value problems. In G. A. Watson, editor, *Numerical Analysis*, pages 60–72. Springer, 1976. doi and E. Hairer, S. P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I. Nonstiff Problems*. Springer, 1993. doi (Section 1.10, Exercise 6)
- ② **uniformly decreasing maps** in: L. Chua and D. Green. A qualitative analysis of the behavior of dynamic nonlinear networks: Stability of autonomous networks. *IEEE Transactions on Circuits and Systems*, 23(6): 355–379, 1976. doi
- ③ no-name in: A. F. Filippov. *Differential Equations with Discontinuous Righthand Sides*. Kluwer, 1988. ISBN 902772699X (Chapter 1, page 5)
- ④ **maps with negative nonlinear measure** in: H. Qiao, J. Peng, and Z.-B. Xu. Nonlinear measures: A new approach to exponential stability analysis for Hopfield-type neural networks. *IEEE Transactions on Neural Networks*, 12(2):360–370, 2001. doi
- ⑤ **dissipative Lipschitz maps** in: T. Caraballo and P. E. Kloeden. The persistence of synchronization under environmental noise. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 461(2059):2257–2267, 2005. doi
- ⑥ **maps with negative lub log Lipschitz constant** in: G. Söderlind. The logarithmic norm. History and modern theory. *BIT Numerical Mathematics*, 46(3):631–652, 2006. doi
- ⑦ **QUAD maps** in: W. Lu and T. Chen. New approach to synchronization analysis of linearly coupled ordinary differential systems. *Physica D: Nonlinear Phenomena*, 213(2):214–230, 2006. doi
- ⑧ **incremental quadratically stable maps** in: L. D'Alto and M. Corless. Incremental quadratic stability. *Numerical Algebra, Control and Optimization*, 3:175–201, 2013. doi

## Contraction conditions without Jacobians

- ① Demidovich LMI condition in: B. P. Demidovič. On the dissipativity of a certain non-linear system of differential equations. I. *Vestnik Moskovskogo Universiteta. Serija I. Matematika, Mekhanika*, 6:19–27, 1961
- ② Krasovskii's method for Lyapunov functions
- ③ common Lyapunov function approach
- ④ Pointwise quadratic constraints
- ⑤ Incremental multiplier matrices
- ⑥ Lyapunov functions for the variational system

## Links to recent related educational and research events

- 2023 ACC Workshop on "Contraction Theory for Systems, Control, and Learning"  
<http://motion.me.ucsb.edu/contraction-workshop-2023>
- Tutorial session: <https://sites.google.com/view/contractiontheory> "Contraction Theory for Machine Learning" (PDFs and youtube videos) at the 2021 IEEE CDC conference, by Soon-Jo Chung, Jean-Jacques Slotine, and Hiroyasu Tsukamoto
- Tutorial paper at CDC2021 "Contraction-Based Methods for Stable Identification and Robust Machine Learning: a Tutorial" by Ian Manchester and coauthors: <https://arxiv.org/abs/2110.00207>,  
<https://ieeexplore.ieee.org/abstract/document/9683128>
- Plenary presentation: (Slides  
<https://fbullo.github.io/talks/2022-12-FBullo-ContractionSystemsControl-CDC.pdf>) "Contraction Theory in Systems and Control" by Francesco Bullo at the 2022 IEEE CDC
- Youtube lectures: "Lectures on Nonlinear Systems" by Jean-Jacques Slotine, Fall 2013:  
<https://web.mit.edu/ns1/www/videos/lectures.html>, Lectures 14-20 (approximately 1h20min each)
- Youtube lectures: "Minicourse on Contraction Theory" by Francesco Bullo, Fall 2022. Youtube lectures  
<https://youtu.be/RvR47ZbqJjc>: 10h in 4 lectures, with chapters
- Textbook: Contraction Theory for Dynamical Systems, Francesco Bullo, rev 1.1, Mar 2023. (Book and slides freely available) <https://fbullo.github.io/ctds>

# Contraction Theory for Dynamical Systems

Francesco Bullo

**Contraction Theory for Dynamical Systems**, Francesco Bullo,  
KDP, 1.1 edition, 2023, ISBN 979-8836646806

- ① Textbook with exercises and answers. Format: textbook, slides, and paperbook
- ② Content:
  - Fixed point theory
  - Theory of contracting dynamics on vector spaces
  - Applications to nonlinear and interconnected systems
- ③ Self-Published and Print-on-Demand at:  
<https://www.amazon.com/dp/B0B4K1BTF4>
- ④ PDF Freely available at  
<https://fbullo.github.io/ctds>
- ⑤ 10h minicourse on youtube:  
<https://youtu.be/RvR47ZbqJjc>
- ⑥ Future version to include: systems on Riemannian manifolds, homogeneous spaces, and solid cones
  - "Continuous improvement is better than delayed perfection"
  - Mark Twain**

# Selected references from my group

## Contraction theory on normed spaces and Riemannian manifolds:

- A. Davydov, S. Jafarpour, and F. Bullo. Non-Euclidean contraction theory for robust nonlinear stability. *IEEE Transactions on Automatic Control*, 67(12):6667–6681, 2022a. 
- S. Jafarpour, A. Davydov, and F. Bullo. Non-Euclidean contraction theory for monotone and positive systems. *IEEE Transactions on Automatic Control*, 68(9):5653–5660, 2023. 
- J. W. Simpson-Porco and F. Bullo. Contraction theory on Riemannian manifolds. *Systems & Control Letters*, 65:74–80, 2014. 

## Contracting neural networks:

- S. Jafarpour, A. Davydov, A. V. Proskurnikov, and F. Bullo. Robust implicit networks via non-Euclidean contractions. In *Advances in Neural Information Processing Systems*, Dec. 2021. 
- A. Davydov, A. V. Proskurnikov, and F. Bullo. Non-Euclidean contractivity of recurrent neural networks. In *American Control Conference*, pages 1527–1534, Atlanta, USA, May 2022c. 
- 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, 2023b. 

## Weak and 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, 2022a. 
- 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), 2024.  To appear
- R. Delabays and F. Bullo. Semicontraction and synchronization of Kuramoto-Sakaguchi oscillator networks. *IEEE Control Systems Letters*, 7:1566–1571, 2023. 

## Optimization:

- F. Bullo, P. Cisneros-Velarde, A. Davydov, and S. Jafarpour. From contraction theory to fixed point algorithms on Riemannian and non-Euclidean spaces. In *IEEE Conf. on Decision and Control*, Dec. 2021. 
- A. Davydov, S. Jafarpour, A. V. Proskurnikov, and F. Bullo. Non-Euclidean monotone operator theory and applications. *Journal of Machine Learning Research*, June 2023b.  Submitted
- A. Davydov, V. Centorrino, A. Gokhale, G. Russo, and F. Bullo. Contracting dynamics for time-varying convex optimization. *IEEE Transactions on Automatic Control*, June 2023a.  Submitted

# Outline

## §1. History and resources

## §2. Basic definitions: discrete and continuous-time dynamics on vector spaces

- The linear algebra of matrix norms; see CTDS Chapter 2
- Properties of induced matrix norms and Lipschitz constants

## §3. Example systems

- Constrained, distributed and proximal gradient dynamics
- Continuous-time recurrent neural networks
- Nonlinear dynamics in Lur'e form

## §4. Properties of contracting dynamics

- Equilibria, Lyapunov functions, and Euler discretization
- Incremental input-to-state stability
- Contractivity of interconnected systems
- Additional properties: entrainment, robustness wrt unmodeled dynamics and delays

## §5. Example applications

- Gradient dynamics and Nash equilibria in games
- Time-varying gradient dynamics and feedback optimization
- Recurrent and implicit neural networks

## §6. Generalizations with examples

- G1: Semicontractivity: Primal-dual gradient with redundant constraints
- G2: Contractivity on Riemannian manifolds and the Karcher mean

## §7. Conclusions and future research

## §8. Advanced Topics

- More on semicontractivity: ergodic coefficients and duality
- Network small-gain theorem for Metzler matrices
- More on the Kuramoto-Sakaguchi model and synchronization
- Proof of semicontractivity of saddle matrices
- Proof of Euler discretization theorem
- Non-Euclidean Monotone Operator Theory

For a non-empty set  $\mathcal{X}$ , a map  $d : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  is a *metric* (or a *distance*) on  $\mathcal{X}$  if

(separation):

$$d(x, y) = 0 \text{ if and only if } x = y$$

(symmetry):

$$d(x, y) = d(y, x) \text{ for all } x, y \in \mathcal{X}$$

(triangle inequality):

$$d(x, y) \leq d(x, z) + d(z, y) \text{ for all } x, y, z \in \mathcal{X}$$

A map  $T : \mathcal{X} \rightarrow \mathcal{X}$  is

- ① *Lipschitz* if there exists  $\ell \geq 0$ , called *a Lipschitz constant* of  $T$ , such that

$$d(T(x), T(y)) \leq \ell d(x, y) \quad \text{for all } x, y \in \mathcal{X},$$

- ② a *contraction* if it is Lipschitz with constant  $\ell < 1$ . In this case,  $\ell$  is called the *contraction factor* of  $T$ .

## Banach Contraction Theorem

Let  $(\mathcal{X}, d)$  be a *complete metric space*

If  $T : \mathcal{X} \rightarrow \mathcal{X}$  is Lipschitz with constant  $\ell < 1$  (called the *contraction factor*), then

- ①  $T$  has a unique fixed point  $x^*$  in  $\mathcal{X}$
- ② the sequence  $\{x_k\}_{k \in \mathbb{N}}$  generated by the *Picard iteration*  $x_{k+1} = T(x_k)$  converges to  $x^*$  for all initial conditions  $x_0 \in \mathcal{X}$
- ③ the following error estimates hold for all  $k \in \mathbb{N}$ :

(geometric convergence):

$$d(x_k, x^*) \leq \ell^k d(x_0, x^*)$$

(a-priori upper bound):

$$d(x_k, x^*) \leq \frac{\ell^k}{1 - \ell} d(x_0, x_1)$$

(a-posteriori upper bound):

$$d(x_k, x^*) \leq \frac{\ell}{1 - \ell} d(x_{k-1}, x_k)$$

## Proof of Banach Contraction Theorem

For  $x_{k+1} = T(x_k)$ , note  $d(x_{k+1}, x_k) \leq \ell d(x_k, x_{k-1})$ .

- we show the sequence  $\{x_k\}_{k \in \mathbb{N}}$  is Cauchy. For all  $k$  and  $h$ ,

$$\begin{aligned} d(x_{k+h}, x_k) &\leq d(x_{k+h}, x_{k+h-1}) + \cdots + d(x_{k+1}, x_k) && \text{(triangle inequality)} \\ &\leq (\ell^{h-1} + \cdots + 1)d(x_{k+1}, x_k) && \text{(Lipschitzness)} \\ &\leq \frac{1}{1-\ell}d(x_{k+1}, x_k) && \text{(geometric series, } \ell < 1\text{)} \\ &\leq \frac{\ell^k}{1-\ell}d(x_1, x_0) && \text{(Lipschitzness)} \end{aligned}$$

- since  $\mathcal{X}$  is complete, sequence converges to a point  $x^*$
- uniqueness from  $\ell < 1$
- geometric convergence

$$d(x_k, x^*) = d(T(x_{k-1}), x^*) \leq \ell d(x_{k-1}, x^*) \leq \ell^k d(x_0, x^*)$$

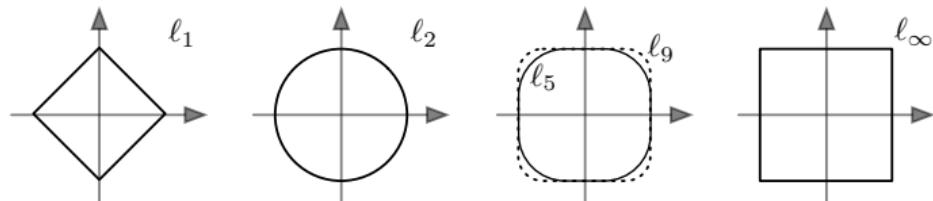
## Examples of metric spaces

- ① finite dimensional vector spaces with a norm ( $\mathbb{R}^n$  and  $d(x, y) = \|x - y\|$ )
- ② Riemannian manifolds (e.g., matrix Lie groups, Grassmannian/Stiefel ...)
- ③ infinite-dimensional Hilbert and Banach spaces
- ④ cones with the Thomson metric (e.g., positive definite matrices)
- ⑤ ...

Note: in these slides, contractivity = *contractivity on  $(\mathbb{R}^n, \|\cdot\|)$* . Available for this case: all discrete/continuous-time theorems, numerous examples, amenable to analysis.

# Linear algebra: induced 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} $ = max column "absolute sum" of $A$	$\mu_1(A) = \max_{j \in \{1, \dots, n\}} \left( a_{jj} + \sum_{i=1, i \neq j}^n  a_{ij}  \right)$ absolute value only off-diagonal
$\ x\ _2 = \sqrt{\sum_{i=1}^n x_i^2}$	$\ A\ _2 = \sqrt{\lambda_{\max}(A^\top A)}$	$\mu_2(A) = \lambda_{\max}\left(\frac{A + A^\top}{2}\right)$
$\ x\ _\infty = \max_{i \in \{1, \dots, n\}}  x_i $	$\ A\ _\infty = \max_{i \in \{1, \dots, n\}} \sum_{j=1}^n  a_{ij} $ = max row "absolute sum" of $A$	$\mu_\infty(A) = \max_{i \in \{1, \dots, n\}} \left( a_{ii} + \sum_{j=1, j \neq i}^n  a_{ij}  \right)$ absolute value only off-diagonal



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

## Lipschitz constant

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

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

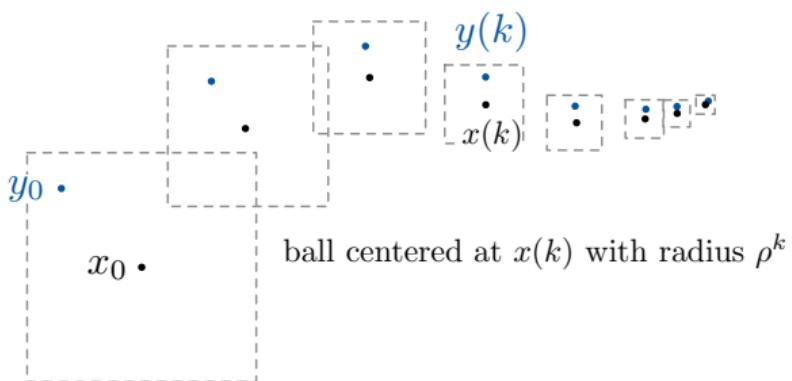
For **affine map**  $\mathsf{F}_A(x) = Ax + a$

$$\begin{array}{lll}\|x\|_{2,P} = (x^\top Px)^{1/2} & \text{Lip}_{2,P}(\mathsf{F}_A) = \|A\|_{2,P} \leq \ell & \iff A^\top PA \preceq \ell^2 P \\ \|x\|_{\infty,\eta} = \max_i |x_i|/\eta_i & \text{Lip}_{\infty,\eta}(\mathsf{F}_A) = \|A\|_{\infty,\eta} \leq \ell & \iff \eta^\top |A| \leq \ell \eta^\top\end{array}$$

## Banach contraction theorem for discrete-time dynamics:

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

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



# From induced norms to induced log norms

The **induced log norm** of  $A \in \mathbb{R}^{n \times n}$  wrt to  $\|\cdot\|$ :

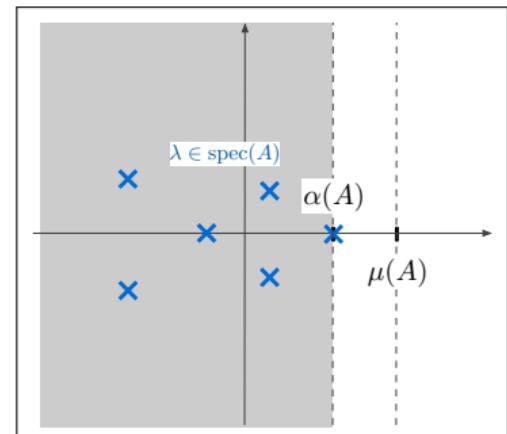
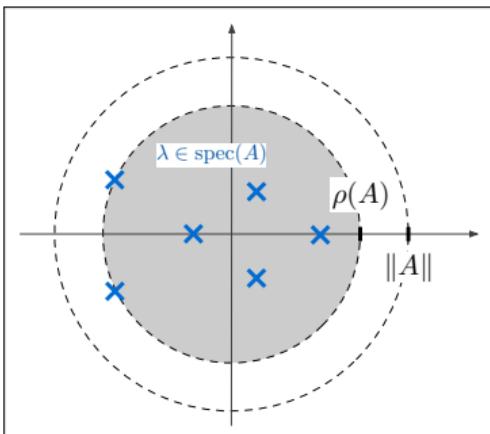
$$\mu(A) := \lim_{h \rightarrow 0^+} \frac{\|I_n + hA\| - 1}{h}$$

subadditivity:

$$\mu(A + B) \leq \mu(A) + \mu(B)$$

scaling:

$$\mu(bA) = b\mu(A), \quad \forall b \geq 0$$



## 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} $ = max column "absolute sum" of $A$	$\mu_1(A) = \max_{j \in \{1, \dots, n\}} \left( a_{jj} + \sum_{i=1, i \neq j}^n  a_{ij}  \right)$ absolute value only off-diagonal
$\ x\ _2 = \sqrt{\sum_{i=1}^n x_i^2}$	$\ A\ _2 = \sqrt{\lambda_{\max}(A^\top A)}$	$\mu_2(A) = \lambda_{\max}\left(\frac{A + A^\top}{2}\right)$
$\ x\ _\infty = \max_{i \in \{1, \dots, n\}}  x_i $	$\ A\ _\infty = \max_{i \in \{1, \dots, n\}} \sum_{j=1}^n  a_{ij} $ = max row "absolute sum" of $A$	$\mu_\infty(A) = \max_{i \in \{1, \dots, n\}} \left( a_{ii} + \sum_{j=1, j \neq i}^n  a_{ij}  \right)$ absolute value only off-diagonal

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

## One-sided Lipschitz constant

$$\begin{aligned}\text{osLip}(F) &= \inf\{b \in \mathbb{R} \text{ such that } [F(x) - F(y), x - y] \leq b\|x - y\|^2 \text{ for all } x, y\} \\ &= \sup_x \mu(DF(x))\end{aligned}$$

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

For **affine map**  $F_A(x) = Ax + a$

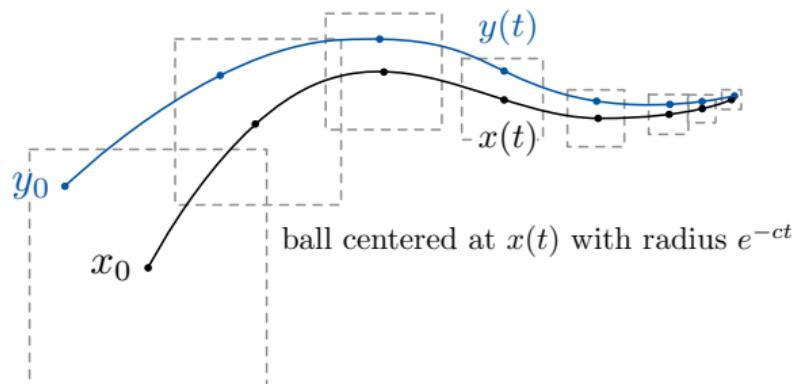
$$\text{osLip}_{2,P}(F_A) = \mu_{2,P}(A) \leq \ell \iff A^\top P + AP \preceq 2\ell P$$

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

## Banach contraction theorem for continuous-time dynamics:

If  $-c := \text{osLip}(F) < 0$ , then

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



## Key properties of inner products

Curve norm derivative formula:

$$\frac{1}{2} D^+ \|x(t)\|^2 = \langle\!\langle \dot{x}(t), x(t) \rangle\!\rangle$$

Sup of Euclidean numerical range:

$$\mu_2(A) = \lambda_{\max}\left(\frac{A+A^\top}{2}\right) = \sup_{\|x\|=1} \langle\!\langle Ax, x \rangle\!\rangle$$

An **inner product** is  $\langle\!\langle \cdot, \cdot \rangle\!\rangle : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$  satisfying

- ①  $\langle\!\langle x_1 + x_2, y \rangle\!\rangle = \langle\!\langle x_1, y \rangle\!\rangle + \langle\!\langle x_2, y \rangle\!\rangle$  and  $x \mapsto \langle\!\langle x, y \rangle\!\rangle$  is continuous, (additivity)
- ②  $\langle\!\langle bx, y \rangle\!\rangle = \langle\!\langle x, by \rangle\!\rangle = b \langle\!\langle x, y \rangle\!\rangle$  for  $b \in \mathbb{R}$ , (homogeneity)
- ③  $\langle\!\langle x, x \rangle\!\rangle > 0$ , for all  $x \neq 0_n$ , (definiteness)
- ④  $|\langle\!\langle x, y \rangle\!\rangle| \leq \langle\!\langle x, x \rangle\!\rangle^{1/2} \langle\!\langle y, y \rangle\!\rangle^{1/2}$ , (Cauchy-Schwarz)

Given norm  $\|\cdot\|$ , compatibility:  $\langle\!\langle x, x \rangle\!\rangle = \|x\|^2$  for all  $x$

# Key properties of weak pairings

Curve norm derivative formula:

$$\frac{1}{2}D^+\|x(t)\|^2 = \llbracket \dot{x}(t), x(t) \rrbracket$$

Sup of non-Euclidean numerical range (Lumer):

$$\mu(A) = \sup_{\|x\|=1} \llbracket Ax, x \rrbracket$$

A **weak pairing** is  $\llbracket \cdot, \cdot \rrbracket : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$  satisfying

- ①  $\llbracket x_1 + x_2, y \rrbracket \leq \llbracket x_1, y \rrbracket + \llbracket x_2, y \rrbracket$ , (sub-additivity)
- ②  $\llbracket bx, y \rrbracket = \llbracket x, by \rrbracket = b \llbracket x, y \rrbracket$  for  $b \geq 0$  and  $\llbracket -x, -y \rrbracket = \llbracket x, y \rrbracket$ , (positive homogeneity)
- ③  $\llbracket x, x \rrbracket > 0$ , for all  $x \neq 0_n$ , (definiteness)
- ④  $|\llbracket x, y \rrbracket| \leq \llbracket x, x \rrbracket^{1/2} \llbracket y, y \rrbracket^{1/2}$ , (Cauchy-Schwarz)

Given norm  $\|\cdot\|$ , compatibility:  $\llbracket x, x \rrbracket = \|x\|^2$  for all  $x$

A. Davydov, S. Jafarpour, and F. Bullo. Non-Euclidean contraction theory for robust nonlinear stability. *IEEE Transactions on Automatic Control*, 67(12):6667–6681, 2022a. 

## Example weak pairings

### Norms

### From inner products to sign and max pairings

### From LMIs to log norms

$$\|x\|_{2,P^{1/2}}^2 = x^\top Px$$

$$[\![x, y]\!]_{2,P^{1/2}} = x^\top Py$$

$$\mu_{2,P^{1/2}}(A) = \min\{b \mid A^\top P + PA \preceq 2bP\}$$

$$\|x\|_1 = \sum_i |x_i|$$

$$[\![x, y]\!]_1 = \|y\|_1 \text{sign}(y)^\top x$$

$$\mu_1(A) = \max_j \left( a_{jj} + \sum_{i \neq j} |a_{ij}| \right)$$

$$\|x\|_\infty = \max_i |x_i|$$

$$[\![x, y]\!]_\infty = \max_{i \in I_\infty(y)} y_i x_i$$

$$\mu_\infty(A) = \max_i \left( a_{ii} + \sum_{j \neq i} |a_{ij}| \right)$$

where  $I_\infty(x) = \{i \in \{1, \dots, n\} \text{ such that } |x_i| = \|x\|_\infty\}$

# Table of continuous-time contractivity conditions

Log Norm bound	Demidovich condition	One-sided Lipschitz condition
$\mu_{2,P}(D\mathbf{F}(x)) \leq b$	$P D\mathbf{F}(x) + D\mathbf{F}(x)^\top P \preceq 2bP$	$(x - y)^\top P(\mathbf{F}(x) - \mathbf{F}(y)) \leq b\ x - y\ _{P^{1/2}}^2$
$\mu_1(D\mathbf{F}(x)) \leq b$	$\text{sign}(v)^\top D\mathbf{F}(x)v \leq b\ v\ _1$	$\text{sign}(x - y)^\top (\mathbf{F}(x) - \mathbf{F}(y)) \leq b\ x - y\ _1$
$\mu_\infty(D\mathbf{F}(x)) \leq b$	$\max_{i \in I_\infty(v)} v_i (D\mathbf{F}(x)v)_i \leq b\ v\ _\infty^2$	$\max_{i \in I_\infty(x-y)} (x_i - y_i)(\mathbf{F}_i(x) - \mathbf{F}_i(y)) \leq b\ x - y\ _\infty^2$

## Equivalent contractivity conditions

J. A. Jacquez and C. P. Simon. Qualitative theory of compartmental systems. *SIAM Review*, 35(1):43–79, 1993. 

H. Qiao, J. Peng, and Z.-B. Xu. Nonlinear measures: A new approach to exponential stability analysis for Hopfield-type neural networks. *IEEE Transactions on Neural Networks*, 12(2):360–370, 2001. 

G. Como, E. Lovisari, and K. Savla. Throughput optimality and overload behavior of dynamical flow networks under monotone distributed routing. *IEEE Transactions on Control of Network Systems*, 2(1):57–67, 2015. 

## Advantages of non-Euclidean approaches

- ① *well suited for certain class of systems*

$\ell_1$  for monotone flow systems

- ② *computational advantages*

$\ell_1/\ell_\infty$  constraints lead to LPs, whereas  $\ell_2$  constraints leads to LMIs

- ③ *robustness to structural perturbations*

$\ell_1/\ell_\infty$  contractions are connectively robust (i.e., edge removal)

- ④ *adversarial input-output analysis*

$\ell_\infty$  better suited for the analysis of adversarial examples than  $\ell_2$

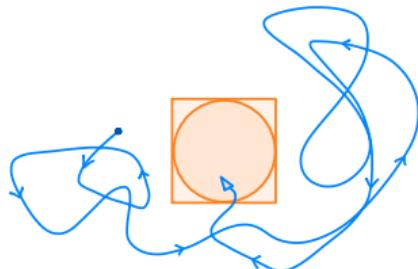
- ⑤ *asynchronous distributed computation*

$\ell_\infty$  contractions converge under fully asynchronous distributed execution

NonEuclidean contractions: biological transcriptional systems (Russo, Di Bernardo, and Sontag, 2010), Hopfield neural networks (Fang and Kincaid, 1996; Qiao, Peng, and Xu, 2001), chemical reaction networks (Al-Radhawi, Angeli, and Sontag, 2020), traffic networks (Coogan and Arcak, 2015; Como, Lovisari, and Savla, 2015; Coogan, 2019), multi-vehicle systems (Monteil, Russo, and Shorten, 2019), and coupled oscillators (Russo, Di Bernardo, and Sontag, 2013; Aminzare and Sontag, 2014a)

# Practical stability problem and the counter-intuitive nature of $\mathbb{R}^n$

Boris Polyak (1935-2023) used to say “ $\mathbb{R}^n$  contradicts our intuition”



Aim: **compute settling time inside a desired set**

- since norms on  $\mathbb{R}^n$  are equivalent, no formal difference in the choice of norm
- assume: can tolerate  $\pm 1$  error in each coordinate
  - ⇒ desired set is hypercube =  $\ell_\infty$ -ball
- assume: Lyapunov function is  $V(x) = \|x\|_2^2$ 
  - ⇒ need to wait until solution enters unit  $\ell_2$ -ball  $\subset$  unit  $\ell_\infty$ -ball
- but  $n$ -sphere inscribed in  $n$ -hypercube is very small fraction!  
as  $n \rightarrow \infty$ , the ratio of volumes decreases faster than any exponential function

**for large  $n$ , quadratic Lyap fnctns may provide exponentially conservative estimates**

Courtesy of Anton Proskurnikov, Politecnico di Torino

## Proof of Banach contraction theorem for continuous-time dynamics

For  $\dot{x} = F(x)$  with  $\text{osLip}(F) = -c < 0$  and unit-time flow map  $\phi$ :

- using the properties of the weak pairing, we compute

$$\begin{aligned}\|x - y\| D^+ \|x - y\| &= \llbracket \dot{x} - \dot{y}, x - y \rrbracket && (\text{curve norm derivative}) \\ &= \llbracket F(x) - F(y), x - y \rrbracket && (\dot{x} = F(x)) \\ &\leq -c \|x - y\|^2 && (\text{osLip}(F) = -c)\end{aligned}$$

- By the Grönwall Comparison,

$$D^+ \|x - y\| \leq -c \|x - y\| \implies \|x(t) - y(t)\| \leq e^{-ct} \|x(0) - y(0)\|$$

and  $\phi$  is a contraction with factor  $e^{-c} < 1$

- recall  $(\mathbb{R}^n, \|\cdot\|)$  is complete metric space,
- the Banach Contraction Theorem implies **existence** of a unique  $x^*$  fixed point of  $\phi$
- $\phi(x^*) = x^*$  implies that
  - either  $x^*$  is an equilibrium
  - or  $x^*$  is a point in a periodic orbit with period 1,
- by contradiction, assume a periodic orbit of period 1 exists. Then each point in the orbit is a fixed point of  $\phi$ , which violates the uniqueness of  $x^*$  as a fixed point,
- hence,  $x^*$  is the **unique** equilibrium of  $F$ .

The **upper right Dini derivative** of a continuous function  $f : ]a, b[ \rightarrow \mathbb{R}$  at a point  $t \in ]a, b[$  is

$$D^+f(t) = \limsup_{\Delta t > 0, \Delta t \rightarrow 0} \frac{f(t + \Delta t) - f(t)}{\Delta t}$$

where the limit superior of a sequence  $\{a_n\}_{n \in \mathbb{N}} \subset \mathbb{R}$  is  $\limsup_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} \sup_{m \geq n} a_m$ .

### Properties of the upper right Dini derivative

Given a continuous function  $f : ]a, b[ \rightarrow \mathbb{R}$ ,

- ① if  $f$  is differentiable at  $t \in ]a, b[$ , then  $D^+f(t) = \frac{d}{dt}f(t)$  is the usual derivative of  $f$  at  $t$ ,
- ② if  $D^+f(t) \leq 0$  for all  $t \in ]a, b[$ , then  $f$  is non-increasing on  $]a, b[$ .

### Grönwall Comparison Lemma for absolutely continuous functions

Given  $a \in \mathbb{R}$  and a continuous function  $t \mapsto \gamma(t) \in \mathbb{R}$ , assume the absolutely continuous function  $t \mapsto z(t)$  satisfies the differential inequality

$$D^+z(t) \leq az(t) + \gamma(t).$$

Then, for  $t \in [t_0, \infty)$ ,

$$z(t) \leq e^{a(t-t_0)}z(t_0) + \int_{t_0}^t e^{a(t-\tau)}\gamma(\tau)d\tau.$$

In other words,  $z(t)$  is upper bounded by the solution to the corresponding differential equality.

## Equivalence between integral and differential osLip

$$\text{Lip}(F) = \sup_x \|DF(x)\| \quad \text{and} \quad \text{osLip}(F) = \sup_x \mu(DF(x))$$

**Proof** Mean Value Theorem for vector-valued  $C^1$  function  $F(x) - F(y) = (\int_0^1 DF(y + s(x-y))ds)(x-y)$  for any  $x, y$ :

$$\begin{aligned} \text{osLip}(F) &= \sup_{x \neq y} \frac{\llbracket (\int_0^1 DF(y + s(x-y))ds)(x-y), x-y \rrbracket}{\|x-y\|^2} \\ &\leq \sup_{x \neq y} \int_0^1 \frac{\llbracket DF(y + s(x-y))(x-y), x-y \rrbracket}{\|x-y\|^2} ds \quad (\text{subadditivity of } \llbracket \cdot, \cdot \rrbracket) \\ &\leq \int_0^1 \sup_{x \neq y} \frac{\llbracket DF(y + s(x-y))(x-y), x-y \rrbracket}{\|x-y\|^2} ds = \int_0^1 \sup_{y, z \neq 0_n} \frac{\llbracket DF(y + sz)z, z \rrbracket}{\|z\|^2} ds \\ &= \int_0^1 \sup_{y, z \neq 0_n} \mu(DF(y + sz)) ds \leq \sup_{x \in \mathbb{R}^n} \mu(DF(x)) \quad (\text{Lumer's equality}) \end{aligned}$$

Vice versa, recall  $DF(y)v = \lim_{h \rightarrow 0^+} (F(y + hv) - F(y))/h$ . Pick  $x = y + hv$  for arbitrary  $v \in \mathbb{R}^n$ ,  $\|v\| = 1$ , and  $h > 0$ ,

$$\begin{aligned} \text{osLip}(F) &= \sup_{y \in \mathbb{R}^n, v \in \mathbb{R}^n, \|v\|=1, h>0} \frac{\llbracket F(x) - F(y), x-y \rrbracket}{\|x-y\|^2} \Big|_{x=y+hv} \\ &\geq \sup_{y \in \mathbb{R}^n, v \in \mathbb{R}^n, \|v\|=1} \lim_{h \rightarrow 0^+} \frac{\llbracket F(y + hv) - F(y), v \rrbracket}{h} \quad (\text{weak homogeneity}) \\ &= \sup_{y \in \mathbb{R}^n, v \in \mathbb{R}^n, \|v\|=1} \llbracket DF(y)v, v \rrbracket \quad (\text{continuity of } w \mapsto \llbracket w, v \rrbracket) \\ &= \sup_{y \in \mathbb{R}^n} \mu(DF(y)). \quad (\text{Lumer's equality}) \end{aligned}$$

# Outline

## §1. History and resources

## §2. Basic definitions: discrete and continuous-time dynamics on vector spaces

- The linear algebra of matrix norms; see CTDS Chapter 2
- Properties of induced matrix norms and Lipschitz constants

## §3. Example systems

- Constrained, distributed and proximal gradient dynamics
- Continuous-time recurrent neural networks
- Nonlinear dynamics in Lur'e form

## §4. Properties of contracting dynamics

- Equilibria, Lyapunov functions, and Euler discretization
- Incremental input-to-state stability
- Contractivity of interconnected systems
- Additional properties: entrainment, robustness wrt unmodeled dynamics and delays

## §5. Example applications

- Gradient dynamics and Nash equilibria in games
- Time-varying gradient dynamics and feedback optimization
- Recurrent and implicit neural networks

## §6. Generalizations with examples

- G1: Semicontractivity: Primal-dual gradient with redundant constraints
- G2: Contractivity on Riemannian manifolds and the Karcher mean

## §7. Conclusions and future research

## §8. Advanced Topics

- More on semicontractivity: ergodic coefficients and duality
- Network small-gain theorem for Metzler matrices
- More on the Kuramoto-Sakaguchi model and synchronization
- Proof of semicontractivity of saddle matrices
- Proof of Euler discretization theorem
- Non-Euclidean Monotone Operator Theory

For all matrices  $A, B \in \mathbb{R}^{n \times n}$ , Lipschitz maps  $F, G : \mathbb{R}^n \rightarrow \mathbb{R}^n$  and  $a \in \mathbb{R}$

### “the modulus properties”

matrix norms	Lipschitz constants
(positive definiteness) $\ A\  \geq 0$ and $\ A\  = 0 \iff A = \mathbb{0}_{n \times n}$	$\text{Lip}(F) \geq 0$ and $\text{Lip}(F) = 0 \iff F$ is constant
(homogeneity) $\ aA\  =  a  \ A\ $	$\text{Lip}(aF) =  a  \text{Lip}(F)$
(subadditivity) $\ A + B\  \leq \ A\  + \ B\ $	$\text{Lip}(F + G) \leq \text{Lip}(F) + \text{Lip}(G)$
(sub-multiplicativity) $\ AB\  \leq \ A\  \ B\ $	$\text{Lip}(F \circ G) \leq \text{Lip}(F) \text{Lip}(G)$

### “the real part properties”

matrix log norms	one-sided Lipschitz constants
(positive homogeneity) $\mu(aA) =  a  \mu(\text{sign}(a)A)$	$\text{osLip}(aF) =  a  \text{osLip}(\text{sign}(a)F)$
(subadditivity) $\mu(A + B) \leq \mu(A) + \mu(B)$	$\text{osLip}(F + G) \leq \text{osLip}(F) + \text{osLip}(G)$
(translation property) $\mu(A + aI_n) = \mu(A) + a$	$\text{osLip}(F + a \text{Id}) = \text{osLip}(F) + a$
(uniform monotonicity) $\mu(A) < 0$ $\implies A$ invertible, $\ A^{-1}\  \leq -1/\mu(A)$	$\text{osLip}(F) < 0$ $\implies F$ injective, $\text{Lip}(F^{-1}) \leq -1/\text{osLip}(F)$

# The linear algebra of matrix norms and log norms

From Chapter 2 in CTDS

**Lemma 2.12 (Weighted matrix and log norms).** Given an invertible matrix  $R$  and a norm  $\|\cdot\|$ ,  
 $\|A\|_R = \|RAR^{-1}\|$  and  $\mu_R(A) = \rho(RAR^{-1})$ . (2.34)

**Theorem 2.23 (Spectrum-norm properties).** Given a matrix  $A \in \mathbb{R}^{n \times n}$  and a norm  $\|\cdot\|$ ,

(i) for any eigenvalue  $\lambda$  of  $A$ , the spectral-radius norm property is

$$(\text{spectral-radius norm property}) \quad 0 \leq |\lambda| \leq \rho(A) \leq \|A\|. \quad (2.71)$$

and, if  $A$  is invertible,

$$0 \leq 1/\|A^{-1}\| \leq |\lambda| \leq \rho(A) \leq \|A\|. \quad (2.72)$$

(ii) for any eigenvalue  $\lambda$  of  $A$ , the spectral-abscissa log-norm property is

$$(\text{spectral-abscissa log-norm property}) \quad -\|A\| \leq -\mu(-\lambda) \leq \Re(\lambda) \leq \alpha(A) \leq \rho(A) \leq \|A\|. \quad (2.73)$$

(iii) if the norm  $\|\cdot\|$  is monotonic and  $A$  is diagonal, then

$$\|A\| = \max_{i \in \{1, \dots, n\}} \|A_{ii}\| = \rho(A), \quad (2.74)$$

$$\mu(A) = \max_{i \in \{1, \dots, n\}} A_{ii} = \alpha(A). \quad (2.75)$$

**Lemma 2.27 (Optimal weighted norms via the Jordan normal form).** Given a matrix  $A \in \mathbb{R}^{n \times n}$ , a monotonic norm  $\|\cdot\|$ , and  $\varepsilon > 0$ , define

$T \in \mathbb{C}^{n \times n}$  an invertible matrix such that  $TA\bar{T}^{-1}$  is in Jordan normal form. (2.89)

$Q \in \mathbb{C}^{n \times n}$  an unitary matrix such that  $QA\bar{Q}^{-1}$  is in Schur normal form, and (2.90)

$\delta = \varepsilon/\|A_{00}\| > 0$ , where  $A_{00}$  is a Jordan block with eigenvalue  $0$  and dimension  $n$ . (2.91)

Then

(i) the norm  $\|\cdot\|_{\text{adj}(A-Q)}$  is  $\varepsilon$ -optimal and  $\varepsilon$ -logarithmically optimal;

(ii) if  $A$  is diagonalizable, the norm  $\|\cdot\|_Y$  is optimal and logarithmically optimal and

(iii) the norm  $\|\cdot\|_{\text{adj}(A-Q)}$  is  $\varepsilon$ -optimal and  $\varepsilon$ -logarithmically optimal for sufficiently small  $\delta$ .

**Lemma 2.30 (Weighted  $\ell_1$  log norms and Lyapunov inequalities).** Given a matrix  $A \in \mathbb{R}^{n \times n}$  with spectral abscissa  $\alpha(A)$ , define for any nonnegative tolerance  $\varepsilon \geq 0$

$$P_\varepsilon := \text{any element of } \{P \in \mathbb{S}_{\geq 0}^n : AP^T + PA \preceq 2(\alpha(A) + \varepsilon)P\}. \quad (2.97)$$

Then

(i) for any  $\varepsilon > 0$ ,  $P_\varepsilon$  is well defined and  $\|\cdot\|_{\mathbb{S}_\varepsilon P_\varepsilon^{-1}}$  is logarithmically  $\varepsilon$ -optimal for  $A$ ,

(ii) if each eigenvalue  $\lambda_i(A)$  with  $\Re(\lambda_i(A)) = \alpha(A)$  is semisimple, then  $P_0$  is well defined and  $\|\cdot\|_{\mathbb{S}_0 P_0^{-1}}$  is logarithmically optimal for  $A$ .

**Theorem 2.13 (Composite induced norms and log norms).** For any set of local norms  $\|\cdot\|_i$ , and an aggregating norm  $\|\cdot\|_{\text{agg}}$  over a decomposition of  $\mathbb{R}^n$ , consider a matrix  $A \in \mathbb{R}^{n \times n}$ .

(i) the composite norm  $\|\cdot\|_{\text{agg}}$  is a well-defined, i.e., it satisfies the norm properties;

(ii) if the aggregating norm  $\|\cdot\|_{\text{agg}}$  is monotonic, then

$$\max_{i \in \{1, \dots, n\}} \|\cdot\|_{A_{ii}} \leq \|\cdot\|_{\text{agg}} \leq \|\cdot\|_{\text{agg}}. \quad (2.49)$$

$$\max_{i \in \{1, \dots, n\}} \mu_i(A_{ii}) \leq \mu_{\text{agg}}(A) \leq \mu_{\text{agg}}(\|A\|_{\text{agg}}). \quad (2.50)$$

Ref	Quadratic forms (for all $x \in \mathbb{R}^n$ )
Lemma 2.14	$\mu_{\text{Frob}}(A) = \max\left(\frac{\ PA\ ^2 + \ A^T\ ^2}{2}\right) = \max\{x^T P A x + x^T A^T P x : \ x\ _2 = 1\}$
Lemma 2.21	$\mu_{\text{Hilb}}(A) = \max\left(\eta^T K \omega \eta : \ \eta\ _2 = 1\right) = \max\{\langle x \rangle^T K \omega \langle x \rangle : \ x\ _2 = 1\}$
Lemma 2.21	$\mu_{\text{Hilb}}(A) = \max\left(\eta^T K \omega \eta : \ \eta\ _2 = 1\right) = \min\left\{\frac{\langle x \rangle^T K \omega \langle x \rangle}{\ x\ _2^2} : \ x\ _2 = 1\right\} = \min\{x^T K \omega x : \ x\ _2 = 1\}$

Table 2.3: Table of quadratic forms for weighted  $\ell_1$ ,  $\ell_\infty$  and log norms,  $P \in \mathbb{S}_{\geq 0}^n$ , and  $\eta \in \mathbb{R}_{\geq 0}^n$ . We adopt the shorthand  $I_n(x) = \{i \in \{1, \dots, n\} : |x_i| < |x|_{\infty}\}$ ,  $|x|_{\infty} = \max_i |x_i|$  as the set of indices where  $x$  takes its maximal absolute value.

**Theorem 2.24 (Monotonicity properties).** Consider a monotonic norm  $\|\cdot\|$ , a matrix  $A \in \mathbb{R}^{n \times n}$ , and a non-negative matrix  $B \in \mathbb{R}_{\geq 0}^{n \times n}$ . Then

$$(\text{monotonicity property of spectral radius}) \quad \rho(A) \leq \rho(A+B) \leq \rho(A+|B|), \quad (2.77a)$$

$$(\text{monotonicity property of induced norm}) \quad \|A\| \leq \|\|A\| + B\| \leq \|\|A\| + |B|\| \quad (2.77b)$$

and

$$(\text{monotonicity property of spectral abscissa}) \quad \alpha(A) \leq \alpha(\|A\|_M) \leq \alpha(\|A\| + B), \quad (2.78a)$$

$$(\text{monotonicity property of log norms}) \quad \mu(A) \leq \mu(\|A\|_M) \leq \mu(\|A\| + B). \quad (2.78b)$$

**Lemma 2.29 (Quasiconvex dependence upon matrix weights).** Given any  $A \in \mathbb{R}^{n \times n}$ ,

(i) the function  $P \in \mathbb{S}_{\geq 0}^n \mapsto \mu_{2,P}(A)$  is quasiconcave with sublevel sets

$$\{P \in \mathbb{S}_{\geq 0}^n : \mu_{2,P}(A) \leq k\} = \{P \in \mathbb{S}_{\geq 0}^n : A^TP + PA \preceq 2kP\}, \quad (2.93)$$

(ii) the functions  $\eta \in \mathbb{R}_{\geq 0}^n \mapsto \mu_{\infty,\eta}(A)$  and  $\eta \in \mathbb{R}_{\geq 0}^n \mapsto \mu_{\infty,|\eta|}(A)$  are quasiconcave with sublevel sets

$$\{\eta \in \mathbb{R}_{\geq 0}^n : \mu_{\infty,\eta}(A) \leq k\} = \{\eta \in \mathbb{R}_{\geq 0}^n : \eta^T A \eta \leq k\}, \quad (2.94)$$

$$\{\eta \in \mathbb{R}_{\geq 0}^n : \mu_{\infty,|\eta|}(A) \leq k\} = \{\eta \in \mathbb{R}_{\geq 0}^n : |A|_{|\eta|} \leq k\}. \quad (2.95)$$

**Lemma 2.31 (Optimal diagonally-weighted norms for non-negative and Metzler matrices).** Consider a nonnegative matrix  $A \in \mathbb{R}_{\geq 0}^{n \times n}$  and a Metzler matrix  $M \in \mathbb{R}^{n \times n}$ . For any  $p \in [1, \infty]$  and  $\delta > 0$ , define

$v$  and  $w \in \mathbb{R}_{\geq 0}^n$  to be the right and left dominant eigenvectors of  $A + \delta z_1 z_1^T$  (respectively,  $M + \delta z_1 z_1^T$ ).

$q \in [1, \infty]$  to satisfy  $1/p + 1/q = 1$  (with the convention  $1/\infty = 0$ ), and

$$u_j = \begin{pmatrix} u_1^{(j)} \\ u_2^{(j)} \\ \vdots \\ u_n^{(j)} \end{pmatrix} \in \mathbb{R}_{\geq 0}^n$$

Then

(i) for sufficiently small  $\delta$ , the norm  $\|\cdot\|_{\text{adj}(A-Q)}$  is  $\varepsilon$ -optimal for  $A$  (respectively,  $\varepsilon$ -logarithmically optimal for  $M$ ), and

(ii) if  $A$  (respectively,  $M$ ) is irreducible, then the norm  $\|\cdot\|_{\text{adj}(A-Q)}$  is optimal for  $A$  (respectively, logarithmically optimal for  $M$ ).

Specifically, for  $p \in \{1, 2, \infty\}$  and for an irreducible  $A$  with spectral radius  $\rho(A)$  and an irreducible  $M$  with spectral abscissa  $\alpha(M)$ ,

$$\rho(A) = \|A\|_{z_1,w} = \|A\|_{\infty,z_1^{-1}} = \|A\|_{2,z_1 w z_1^{-1}}. \quad (2.96)$$

$$\alpha(M) = \mu_{\infty,\eta}(M) = \mu_{\infty,|\eta|}(M) = \mu_{2,z_1 w z_1^{-1}}(M). \quad (2.97)$$

# Outline

## §1. History and resources

## §2. Basic definitions: discrete and continuous-time dynamics on vector spaces

- The linear algebra of matrix norms; see CTDS Chapter 2
- Properties of induced matrix norms and Lipschitz constants

## §3. Example systems

- Constrained, distributed and proximal gradient dynamics
- Continuous-time recurrent neural networks
- Nonlinear dynamics in Lur'e form

## §4. Properties of contracting dynamics

- Equilibria, Lyapunov functions, and Euler discretization
- Incremental input-to-state stability
- Contractivity of interconnected systems
- Additional properties: entrainment, robustness wrt unmodeled dynamics and delays

## §5. Example applications

- Gradient dynamics and Nash equilibria in games
- Time-varying gradient dynamics and feedback optimization
- Recurrent and implicit neural networks

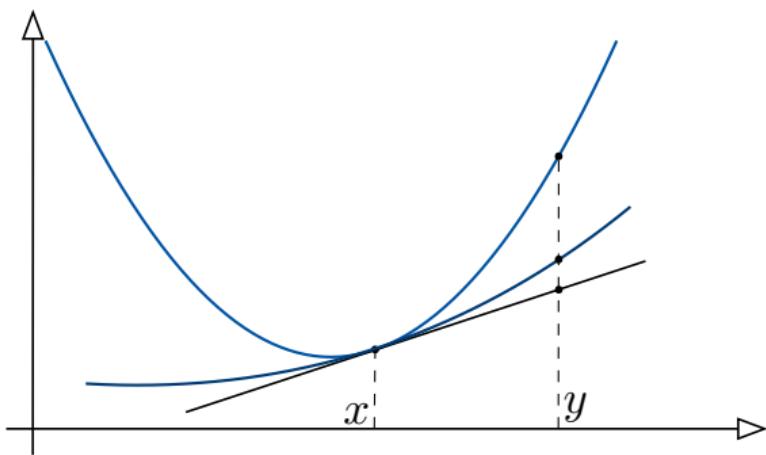
## §6. Generalizations with examples

- G1: Semicontractivity: Primal-dual gradient with redundant constraints
- G2: Contractivity on Riemannian manifolds and the Karcher mean

## §7. Conclusions and future research

## §8. Advanced Topics

- More on semicontractivity: ergodic coefficients and duality
- Network small-gain theorem for Metzler matrices
- More on the Kuramoto-Sakaguchi model and synchronization
- Proof of semicontractivity of saddle matrices
- Proof of Euler discretization theorem
- Non-Euclidean Monotone Operator Theory



$f : \mathbb{R}^n \rightarrow \mathbb{R}$  is  **$\nu$ -strongly convex** if, for all  $x, y$ ,

- ①  $f(\chi x + (1 - \chi)y) \leq \chi f(x) + (1 - \chi)f(y) - \frac{1}{2}\nu\chi(1 - \chi)\|x - y\|_2^2$  for each  $0 \leq \chi \leq 1$
- ② (if  $f$  is differentiable)  $f(y) \geq f(x) + \nabla f(x)^\top (y - x) + \frac{\nu}{2}\|y - x\|_2^2$
- ③ (if  $f$  is differentiable)  $(\nabla f(x) - \nabla f(y))^\top (x - y) \geq \nu\|x - y\|_2^2$
- ④ (if  $f$  is twice differentiable)  $\text{Hess } f(x) \succeq \nu I_n$

## Example #1: Gradient dynamics for strongly convex function

Given differentiable, strongly convex  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  with parameter  $\nu > 0$ , **gradient dynamics**

$$\dot{x} = F_G(x) := -\nabla f(x)$$

$F_G$  is infinitesimally contracting wrt  $\|\cdot\|_2$  with rate  $\nu$

unique globally exp stable point is global minimum

If  $f$  is twice-differentiable, then  $\text{Hess } f(x) \succeq \nu I_n$  for all  $x$

$$D(-\nabla f)(x) = -\text{Hess } f(x) \preceq -\nu I_n$$

$$\iff I_n D(-\nabla f)(x) + D(-\nabla f)(x)^\top I_n \preceq -2\nu I_n$$

**Kachurovskii's Theorem:** For differentiable  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , equivalent statements:

- ①  $f$  is **strongly convex** with parameter  $\nu$  (and minimum  $x^*$ )
- ②  $-\nabla f$  is  **$\nu$ -strongly infinitesimally contracting** (with equilibrium  $x^*$ ), that is

$$(-\nabla f(x) + \nabla f(y))^\top (x - y) \leq -\nu \|x - y\|_2^2$$

For map  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , equivalent statements:

- ①  $F$  is a **monotone operator<sup>a</sup>** (or a **coercive operator**) with parameter  $\nu$ ,
- ②  $-F$  is  **$\nu$ -strongly contracting** wrt  $\|\cdot\|_2$

---

<sup>a</sup> $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is a  **$\nu$ -strongly monotone operator** if  $\langle F(x) - F(y), x - y \rangle \geq \nu \|x - y\|_2^2$

## Example #2: Primal-dual gradient dynamics

strongly convex function  $f$       s.t.  $0 \prec \nu_{\min} I_n \preceq \text{Hess } f \preceq \nu_{\max} I_n$   
constraint matrix  $A$       s.t.  $0 \prec a_{\min} I_m \preceq AA^\top \preceq a_{\max} I_m$       (independent rows)

**linearly constrained optimization:**

$$\begin{aligned} & \min_{x \in \mathbb{R}^n} f(x) \\ & \text{subj. to } Ax = b \end{aligned}$$

**primal-dual gradient dynamics:**

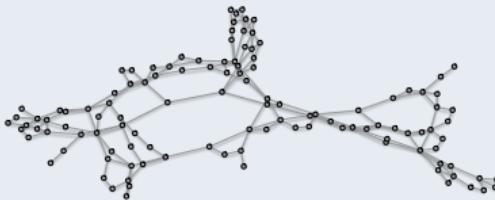
$$\begin{bmatrix} \dot{x} \\ \dot{\lambda} \end{bmatrix} = \mathsf{F}_{\text{PDG}}(x, \lambda) := \begin{bmatrix} -\nabla f(x) - A^\top \lambda \\ Ax - b \end{bmatrix}$$

$\mathsf{F}_{\text{PDG}}$  is infinitesimally contracting wrt  $\|\cdot\|_{2,P^{1/2}}$  with rate  $c$

$$P = \begin{bmatrix} I_n & \alpha A^\top \\ \alpha A & I_m \end{bmatrix} \text{ with } \alpha = \frac{1}{2} \min \left\{ \frac{1}{\nu_{\max}}, \frac{\nu_{\min}}{a_{\max}} \right\} \quad \text{and} \quad c = \frac{1}{4} \min \left\{ \frac{a_{\min}}{\nu_{\max}}, \frac{a_{\min}}{a_{\max}} \nu_{\min} \right\}$$

For each  $\nu_{\min} I_n \preceq Q \preceq \nu_{\max} I_n$ ,

$$\begin{bmatrix} -Q & -A^\top \\ A & 0 \end{bmatrix}^\top P + P \begin{bmatrix} -Q & -A^\top \\ A & 0 \end{bmatrix} \preceq -2cP$$



undirected, weighted and connected graph with  $n$  nodes and  $m$  edges  
 Laplacian  $L \in \mathbb{R}^{n \times n}$ ,  $\lambda_2$  = algebraic connectivity,  $\lambda_2/\lambda_n$  = synchronizability  
 oriented incidence matrix  $B \in \mathbb{R}^{n \times m}$

## Distributed optimization setup

cost function  $f$  is decomposable into sum of private cost function

$$f(x) = \sum_{i=1}^n f_i(x) \quad \text{where each } f_i \text{ is private to node } i$$

each node  $i$  has a local estimate  $x_{[i]}$  of global variable  $x$  and  $\mathbf{x} = [x_{[1]}, \dots, x_{[n]}]$   
 consensus constraints among estimates are imposed in two ways:

- ① incidence constraint:  $B^\top \mathbf{x} = \mathbf{0}_m$
- ② Laplacian constraint:  $L\mathbf{x} = \mathbf{0}_n$

## Example #3: Incidence-based distributed gradient

Assume graph is a tree, so that (see LNS.Exercise9.2)

$$0 \prec \lambda_2 I_{n-1} \preceq B^\top B \preceq \lambda_n I_{n-1}$$

**decomposable cost:**  $\min_{x \in \mathbb{R}} \sum_i f_i(x)$  where each  $f_i$  is  $\nu_i$ -strongly convex

$$\begin{cases} \min_{x_{[i]} \in \mathbb{R}} & \sum_{i=1}^n f_i(x_{[i]}) \\ \text{subj. to} & B^\top \mathbf{x} = \mathbb{0}_m \end{cases} \iff \begin{cases} \min_{x_{[i]} \in \mathbb{R}} & \sum_{i=1}^n f_i(x_{[i]}) \\ \text{subj. to} & x_{[i]} - x_{[j]} = 0 \quad \text{for each edge } e = (i, j) \end{cases}$$

**incidence-based distributed gradient** (primal-dual gradient,  $n + m$  vars):

$$\begin{cases} \dot{x}_{[i]} = -\nabla f_i(x_{[i]}) - \sum_{e=(i,j)} \lambda_e + \sum_{e=(j,i)} \lambda_e & \text{for each node } i \\ \dot{\lambda}_e = x_{[i]} - x_{[j]} & \text{for each edge } e = (i, j) \end{cases}$$

F<sub>Incidence-DistributedG</sub> is infinitesimally contracting with  $c = \frac{1}{4} \frac{\lambda_2}{\lambda_n} \min_i \nu_i$

## Example #4: Laplacian-based distributed gradient

Given  $\Pi_n = I_n - \mathbb{1}_n \mathbb{1}_n^\top / n$  = orthogonal projection onto  $\text{span}\{\mathbb{1}_n\}^\perp$ ,

$$0 \prec \lambda_2 \Pi_n \preceq L \preceq \lambda_n I_n$$

**decomposable cost:**  $\min_{x \in \mathbb{R}} \sum_{i=1}^n f_i(x)$  where each  $f_i$  is  $\nu_i$ -strongly convex

$$\begin{cases} \min_{x_{[i]} \in \mathbb{R}} & \sum_{i=1}^n f_i(x_{[i]}) \\ \text{subj. to} & L\mathbf{x} = \mathbb{0}_n \end{cases} \iff \begin{cases} \min_{x_{[i]} \in \mathbb{R}} & \sum_{i=1}^n f_i(x_{[i]}) \\ \text{subj. to} & \sum_{j=1}^n a_{ij}(x_i - x_j) = 0 \end{cases}$$

**Laplacian-based distributed gradient** (primal-dual gradient,  $2n$  vars):

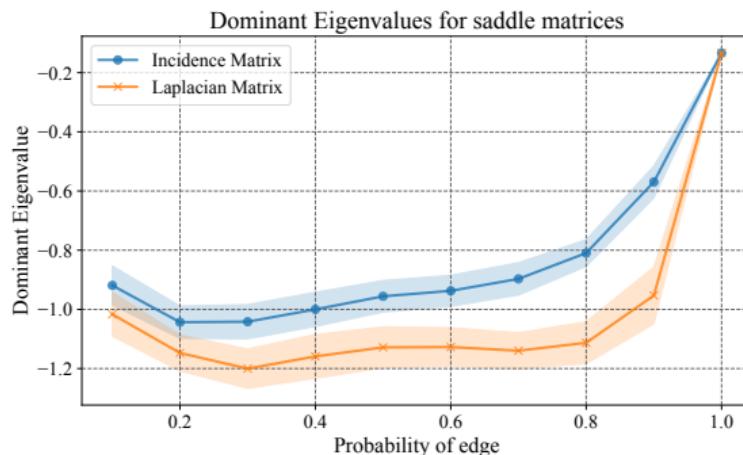
$$\begin{cases} \dot{x}_{[i]} = -\nabla f_i(x_{[i]}) - \sum_{j=1}^n a_{ij}(\lambda_i - \lambda_j) & \text{for each node } i \\ \dot{\lambda}_i = \sum_{j=1}^n a_{ij}(x_i - x_j) & \text{for each node } i \end{cases}$$

F<sub>Laplacian-DistributedG</sub> is infinitesimally contracting<sup>†</sup> with  $c = \frac{1}{4} \left( \frac{\lambda_2}{\lambda_n} \right)^2 \min_i \nu_i$

$\lambda_2/\lambda_n$  = **synchronizability** parameter from study of oscillator networks via the MSF approach

Empirically, define private functions  $q_i(x_i - v_i)^2$ , for  $x_i \in \mathbb{R}$ ,  $v_i$  and  $q_i$  uniformly sampled from  $[0, 10]$

symmetric connected Erdős-Rényi graph with  $N = 40$  nodes, with varying edge probability parameters, 50 graphs for each probability value



L. M. Pecora and T. L. Carroll. Synchronization in chaotic systems. *Physical Review Letters*, 64(8):821–824, 1990

G. Chen. Searching for best network topologies with optimal synchronizability: A brief review. *IEEE/CAA Journal of Automatica Sinica*, 9(4):573–577, 2022. doi:

# Composite optimization

**composite minimization** (cost = sum of terms with structurally different properties):

$$x^* = \operatorname{argmin}_{x \in \mathbb{R}^n} f(x) + g(x)$$

$f : \mathbb{R}^n \rightarrow \mathbb{R}$  is strongly convex and strongly smooth

$g : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is convex, closed, and proper (ccp)

**proximal operator:** for  $\gamma > 0$ , define  $\operatorname{prox}_{\gamma g} : \mathbb{R}^n \rightarrow \mathbb{R}^n$  by

$$\operatorname{prox}_{\gamma g}(z) := \operatorname{argmin}_{x \in \mathbb{R}^n} g(x) + \frac{1}{2\gamma} \|x - z\|_2^2$$

## Equivalence:

- ①  $x^*$  is minimizer for:  $\min_{x \in \mathbb{R}^n} f(x) + g(x)$
- ②  $x^*$  is fixed point for:  $x = \operatorname{prox}_{\gamma g}(x - \gamma \nabla f(x))$  for all  $\gamma$

**proximal gradient dynamics:**  $\dot{x} = F_{\operatorname{ProxG}}(x) := -x + \operatorname{prox}_{\gamma g}(x - \gamma \nabla f(x))$

$f(\mathbf{x})$	$\text{dom}(f)$	$\text{prox}_f(\mathbf{x})$	Assumptions	Reference
$\frac{1}{2}\mathbf{x}^T \mathbf{A}\mathbf{x} + \mathbf{b}^T \mathbf{x} + c$	$\mathbb{R}^n$	$(\mathbf{A} + \mathbf{I})^{-1}(\mathbf{x} - \mathbf{b})$	$\mathbf{A} \in \mathbb{S}_+^n, \mathbf{b} \in \mathbb{R}^n, c \in \mathbb{R}$	Section 6.2.3
$\lambda x^3$	$\mathbb{R}_+$	$\frac{-1 + \sqrt{1 + 12\lambda x }}{6\lambda} \mathbf{x}$	$\lambda > 0$	Lemma 6.5
$\mu x$	$[0, \alpha] \cap \mathbb{R}$	$\min\{\max\{x - \mu, 0\}, \alpha\}$	$\mu \in \mathbb{R}, \alpha \in [0, \infty]$	Example 6.14
$\lambda\ \mathbf{x}\ $	$\mathbb{E}$	$\left(1 - \frac{\lambda}{\max\{\ \mathbf{x}\ , \lambda\}}\right) \mathbf{x}$	$\ \cdot\  = \text{Euclidean norm}, \lambda > 0$	Example 6.19
$-\lambda\ \mathbf{x}\ $	$\mathbb{E}$	$\begin{cases} \left(1 + \frac{\lambda}{\ \mathbf{x}\ }\right) \mathbf{x}, & \mathbf{x} \neq \mathbf{0}, \\ \{\mathbf{u} : \ \mathbf{u}\  = \lambda\}, & \mathbf{x} = \mathbf{0}. \end{cases}$	$\ \cdot\  = \text{Euclidean norm}, \lambda > 0$	Example 6.21
$\lambda\ \mathbf{x}\ _1$	$\mathbb{R}^n$	$\mathcal{T}_\lambda(\mathbf{x}) = [\ \mathbf{x}\  - \lambda e]_+ \odot \text{sgn}(\mathbf{x})$	$\lambda > 0$	Example 6.8
$\ \omega \odot \mathbf{x}\ _1$	$\text{Box}[-\alpha, \alpha]$	$\mathcal{S}_{\omega, \alpha}(\mathbf{x})$	$\alpha \in [0, \infty]^n, \omega \in \mathbb{R}_+^n$	Example 6.23
$\lambda\ \mathbf{x}\ _\infty$	$\mathbb{R}^n$	$\mathbf{x} - \lambda P_{B_{\ \cdot\ _1}[0,1]}(\mathbf{x}/\lambda)$	$\lambda > 0$	Example 6.48
$\lambda\ \mathbf{x}\ _a$	$\mathbb{E}$	$\mathbf{x} - \lambda P_{B_{\ \cdot\ _a, *}[0,1]}(\mathbf{x}/\lambda)$	$\ \cdot\ _a = \text{arbitrary norm}, \lambda > 0$	Example 6.47
$\lambda\ \mathbf{x}\ _0$	$\mathbb{R}^n$	$\mathcal{H}_{\sqrt{2\lambda}}(x_1) \times \dots \times \mathcal{H}_{\sqrt{2\lambda}}(x_n)$	$\lambda > 0$	Example 6.10
$\lambda\ \mathbf{x}\ ^3$	$\mathbb{E}$	$\frac{2}{1 + \sqrt{1 + 12\lambda\ \mathbf{x}\ }} \mathbf{x}$	$\ \cdot\  = \text{Euclidean norm}, \lambda > 0,$	Example 6.20
$-\lambda \sum_{j=1}^n \log x_j$	$\mathbb{R}_{++}^n$	$\left( \frac{x_j + \sqrt{x_j^2 + 4\lambda}}{2} \right)_{j=1}^n$	$\lambda > 0$	Example 6.9
$\delta_C(\mathbf{x})$	$\mathbb{E}$	$P_C(\mathbf{x})$	$\emptyset \neq C \subseteq \mathbb{E}$	Theorem 6.24
$\lambda\sigma_C(\mathbf{x})$	$\mathbb{E}$	$\mathbf{x} - \lambda P_C(\mathbf{x}/\lambda)$	$\lambda > 0, C \neq \emptyset \text{ closed convex}$	Theorem 6.46
$\lambda \max\{x_i\}$	$\mathbb{R}^n$	$\mathbf{x} - \lambda P_{\Delta_n}(\mathbf{x}/\lambda)$	$\lambda > 0$	Example 6.49
$\lambda \sum_{i=1}^k x_{[i]}$	$\mathbb{R}^n$	$\mathbf{x} - \lambda P_C(\mathbf{x}/\lambda), C = H_{\mathbf{e}, k} \cap \text{Box}[\mathbf{0}, \mathbf{e}]$	$\lambda > 0$	Example 6.50
$\lambda \sum_{i=1}^k  x_{(i)} $	$\mathbb{R}^n$	$\mathbf{x} - \lambda P_C(\mathbf{x}/\lambda), C = B_{\ \cdot\ _1}[0, k] \cap \text{Box}[-\mathbf{e}, \mathbf{e}]$	$\lambda > 0$	Example 6.51
$\lambda M_f^\mu(\mathbf{x})$	$\mathbb{E}$	$\mathbf{x} + \frac{\lambda}{\mu + \lambda} (\text{prox}_{(\mu+\lambda)f}(\mathbf{x}) - \mathbf{x})$	$\lambda, \mu > 0, f \text{ proper closed convex}$	Corollary 6.64
$\lambda d_C(\mathbf{x})$	$\mathbb{E}$	$\mathbf{x} + \min \left\{ \frac{\lambda}{d_C(\mathbf{x})}, 1 \right\} (P_C(\mathbf{x}) - \mathbf{x})$	$\emptyset \neq C \text{ closed convex}, \lambda > 0$	Lemma 6.43
$\frac{\lambda}{2} d_C^2(\mathbf{x})$	$\mathbb{E}$	$\frac{\lambda}{\lambda+1} P_C(\mathbf{x}) + \frac{1}{\lambda+1} \mathbf{x}$	$\emptyset \neq C \text{ closed convex}, \lambda > 0$	Example 6.65
$\lambda H_\mu(\mathbf{x})$	$\mathbb{E}$	$(1 - \frac{\lambda}{\max\{\ \mathbf{x}\ , \mu + \lambda\}}) \mathbf{x}$	$\lambda, \mu > 0$	Example 6.66
$\rho\ \mathbf{x}\ _1^2$	$\mathbb{R}^n$	$\left[ \sqrt{\frac{v_i x_i}{v_i + 2\rho}} \right]_{i=1}^n, \mathbf{v} = \begin{cases} \frac{\mathbf{v}_1}{\sqrt{\mu}}  \mathbf{x}  - 2\rho & \text{if } \mathbf{x} \neq \mathbf{0} \\ \mathbf{0} & \text{when } \mathbf{x} = \mathbf{0} \end{cases}$	$\rho > 0$	Lemma 6.70
$\lambda\ \mathbf{Ax}\ _2$	$\mathbb{R}^n$	$\mathbf{x} - \mathbf{A}^T (\mathbf{A}\mathbf{A}^T + \alpha\mathbf{I})^{-1} \mathbf{Ax}, \alpha = 0 \text{ if } \ \mathbf{v}\ _2 \leq \lambda; \text{ otherwise, } \ \mathbf{v}_{\alpha^*}\ _2 = \lambda; \mathbf{v}_\alpha \equiv (\mathbf{A}\mathbf{A}^T + \alpha\mathbf{I})^{-1} \mathbf{Ax}$	$\mathbf{A} \in \mathbb{R}^{m \times n} \text{ with full row rank}, \lambda > 0$	Lemma 6.68

## proximal operator

well-defined for all CCP functions,  
generalized form of projection,  
non-expansive

helps generalize gradient algorithms/dynamics  
to proximal algorithms/dynamics, useful for  
nonsmooth, constrained, large-scale, and distributed optimization

evaluation of proximal operator requires small  
convex optimization,  
see [Summary of prox computations](#), Beck 2017

A. Beck. [First-Order Methods in Optimization](#). SIAM, 2017. ISBN 978-1-61197-498-0

N. Parikh and S. Boyd. Proximal algorithms. [Foundations and Trends in Optimization](#), 1(3):127–239, 2014. doi: 

## Example #5: Proximal gradient dynamics

**proximal gradient dynamics:**

$$\dot{x} = \mathsf{F}_{\text{ProxG}}(x) := -x + \text{prox}_{\gamma g}(x - \gamma \nabla f(x))$$

$f : \mathbb{R}^n \rightarrow \mathbb{R}$  is  $\nu$ -strongly convex and  $\ell$ -strongly smooth

$g : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is convex, closed, proper

①  $\mathsf{F}_{\text{ProxG}}$  is infinitesimally contracting wrt  $\|\cdot\|_2$

$$\text{for } 0 < \gamma < \frac{2}{\ell}, \quad \text{with rate} \quad c = 1 - \max\{|1 - \gamma\nu|, |1 - \gamma\ell|\},$$

$$\text{for } \gamma^* = \frac{2}{\nu + \ell}, \quad \text{with maximal rate} \quad c^* = \frac{2\nu}{\nu + \ell}$$

②  $\mathsf{F}_{\text{ProxG}}$  is infinitesimally contracting wrt  $\|\cdot\|_{2,(\gamma A - I_n)^{1/2}}$  with rate  $c = 1$

$$\text{if } f(x) = \frac{1}{2}x^\top Ax + b^\top x \quad \text{with } A \succ 0 \quad \text{and} \quad \gamma > 1/\lambda_{\min}(A)$$

# Outline

## §1. History and resources

## §2. Basic definitions: discrete and continuous-time dynamics on vector spaces

- The linear algebra of matrix norms; see CTDS Chapter 2
- Properties of induced matrix norms and Lipschitz constants

## §3. Example systems

- Constrained, distributed and proximal gradient dynamics
- Continuous-time recurrent neural networks
- Nonlinear dynamics in Lur'e form

## §4. Properties of contracting dynamics

- Equilibria, Lyapunov functions, and Euler discretization
- Incremental input-to-state stability
- Contractivity of interconnected systems
- Additional properties: entrainment, robustness wrt unmodeled dynamics and delays

## §5. Example applications

- Gradient dynamics and Nash equilibria in games
- Time-varying gradient dynamics and feedback optimization
- Recurrent and implicit neural networks

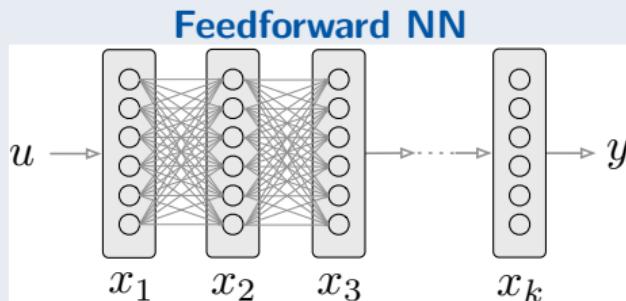
## §6. Generalizations with examples

- G1: Semicontractivity: Primal-dual gradient with redundant constraints
- G2: Contractivity on Riemannian manifolds and the Karcher mean

## §7. Conclusions and future research

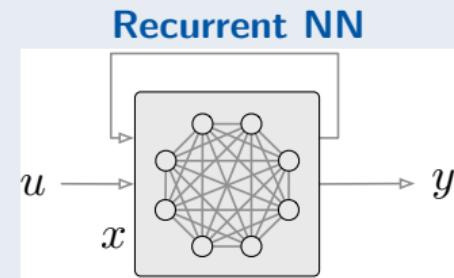
## §8. Advanced Topics

- More on semicontractivity: ergodic coefficients and duality
- Network small-gain theorem for Metzler matrices
- More on the Kuramoto-Sakaguchi model and synchronization
- Proof of semicontractivity of saddle matrices
- Proof of Euler discretization theorem
- Non-Euclidean Monotone Operator Theory



$$x_{i+1} = \Phi(W_i x_i + b_i), \quad x_0 = u,$$

$$y = C x_k + d$$



$$\dot{x} = -x + \Phi(Wx + Bu + b),$$

$$y = Cx + d$$

square matrix  $W$  = *synaptic matrix* — diagonal nonlinear  $\Phi$  = *activation function*

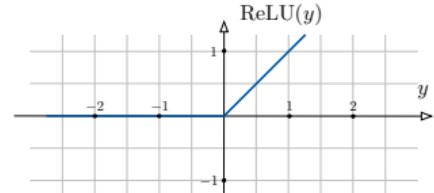
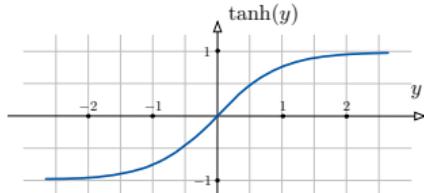
A. Davydov, A. V. Proskurnikov, and F. Bullo. Non-Euclidean contractivity of recurrent neural networks. In *American Control Conference*, pages 1527–1534, Atlanta, USA, May 2022c. [doi](#)

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, 2023b. [doi](#)

## Example #6: Firing-rate recurrent neural network

$$\dot{x} = \mathsf{F}_{\text{FR}}(x) := -x + \Phi(Wx + Bu)$$

sigmoid, hyperbolic tangent  
 $\text{ReLU} = \max\{x, 0\} = (x)_+$   
 $0 \leq \Phi'_i(y) \leq 1$



$\mathsf{F}_{\text{FR}}$  is infinitesimally contracting wrt  $\|\cdot\|_\infty$  with rate  $1 - \mu_\infty(W)_+$  if

$$\mu_\infty(W) < 1 \quad (\text{i.e., } w_{ii} + \sum_j |w_{ij}| < 1 \text{ for all } i)$$

$$\begin{aligned} \text{osLip}_\infty(\mathsf{F}_{\text{FR}}) &= \sup_{x,u} \mu_\infty(-I_n + (D\Phi(Wx + Bu))W) = -1 + \sup_{x,u} \mu_\infty(D\Phi(Wx + Bu)W) \\ &\leq -1 + \max_{d \in [0,1]^n} \mu_\infty(\text{diag}(d)W) \quad (\text{max convex polytope, } 2^n \text{ vertices}) \\ &= -1 + \max \{\mu_\infty(0), \mu_\infty(W)\} = -1 + \mu_\infty(W)_+ \end{aligned}$$

For each row  $i$ , define the  $i$ th absolute row-sum of  $A$  by

$$\mathbb{r}_i(A) = a_{ii} + \sum_{j=1, j \neq i}^n |a_{ij}|$$

and note  $\mu_\infty(A) = \max_i \mathbb{r}_i$ .

Since  $d_i \geq 0$  and  $([d]A)_{ij} = d_i a_{ij}$ , we note

$$\mathbb{r}_i([d]A) = d_i \mathbb{r}_i(A)$$

and compute

$$\begin{aligned} \max_{d \in [0,1]^n} \mu_\infty([d]A) &\stackrel{\text{(by def)}}{=} \max_{d \in [0,1]^n} \max_i d_i \mathbb{r}_i(A) \\ &\stackrel{\text{(the } n \text{ functions are decoupled)}}{=} \max_i \max_{d_i \in [0,1]} d_i \mathbb{r}_i(A) \\ (d_i \in [0,1]) \max_i (A) &\begin{cases} \mathbb{r}_i, & \text{if } \mathbb{r}_i(A) \geq 0 \\ 0, & \text{if } \mathbb{r}_i(A) < 0 \end{cases} \\ (\text{dropping the if clause}) &\leq \max_i \max \mathbb{r}_i(A), 0 \} = \max \{\mu_\infty(A), 0\}. \end{aligned}$$

## Example #7: Firing-rate network with symmetric synapses

$$\dot{x} = \mathsf{F}_{\text{FR}}(x) := -x + \Phi(Wx + Bu)$$

$$0 \leq \Phi'_i(y) \leq 1 \quad \text{and} \quad W = W^\top \text{ with } \lambda_W = \lambda_{\max}(W)$$

$\mathsf{F}_{\text{FR}}$  is infinitesimally contracting:

(for  $\lambda_W < 0$ )

**with rate 1 wrt  $\|\cdot\|_{2,(-W)^{1/2}}$**

(for  $\lambda_W = 0$ )

**with rate  $1 - \epsilon$  wrt  $\|\cdot\|_{2,Q_{\text{FR},\epsilon}}$ , for each  $\epsilon > 0$**

(for  $0 < \lambda_W < 1$ )

**with rate  $1 - \lambda_W$  wrt  $\|\cdot\|_{2,Q_{\text{FR},\lambda_W}}$**

For  $\lambda_W = 1$ ,  $\mathsf{F}_{\text{FR}}$  is weakly infinitesimally contracting wrt  $\|\cdot\|_{2,Q_{\text{FR},\lambda_W}}$

- $Q_{\text{FR},a} := Uh_a(\Lambda)U^\top \succ 0$ , where  $W = U\Lambda U^\top$  and  $h_a(z) := 2a(1 + \sqrt{1 - z/a})$
- optimal rates
- proof based upon LMI calculations and Sylvester's law of inertia

# Outline

## §1. History and resources

## §2. Basic definitions: discrete and continuous-time dynamics on vector spaces

- The linear algebra of matrix norms; see CTDS Chapter 2
- Properties of induced matrix norms and Lipschitz constants

## §3. Example systems

- Constrained, distributed and proximal gradient dynamics
- Continuous-time recurrent neural networks
- Nonlinear dynamics in Lur'e form

## §4. Properties of contracting dynamics

- Equilibria, Lyapunov functions, and Euler discretization
- Incremental input-to-state stability
- Contractivity of interconnected systems
- Additional properties: entrainment, robustness wrt unmodeled dynamics and delays

## §5. Example applications

- Gradient dynamics and Nash equilibria in games
- Time-varying gradient dynamics and feedback optimization
- Recurrent and implicit neural networks

## §6. Generalizations with examples

- G1: Semicontractivity: Primal-dual gradient with redundant constraints
- G2: Contractivity on Riemannian manifolds and the Karcher mean

## §7. Conclusions and future research

## §8. Advanced Topics

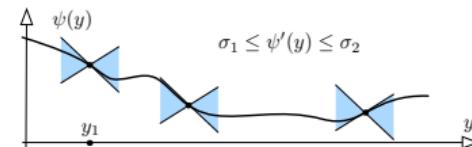
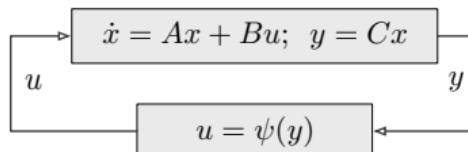
- More on semicontractivity: ergodic coefficients and duality
- Network small-gain theorem for Metzler matrices
- More on the Kuramoto-Sakaguchi model and synchronization
- Proof of semicontractivity of saddle matrices
- Proof of Euler discretization theorem
- Non-Euclidean Monotone Operator Theory

# Systems in Lur'e form

**nonlinear system in Lur'e form**  $x \in \mathbb{R}^n, u \in \mathbb{R}, y \in \mathbb{R}$ :

$$\dot{x} = Ax + Bu \quad y = Cx$$

$$u = \psi(y) \quad \psi : \mathbb{R} \rightarrow \mathbb{R}$$



$M = M^\top \in \mathbb{R}^{2 \times 2}$  is an *incremental multiplier matrix* for  $\psi$  if

$$\begin{bmatrix} y_1 - y_2 \\ \psi(y_1) - \psi(y_2) \end{bmatrix}^\top M \begin{bmatrix} y_1 - y_2 \\ \psi(y_1) - \psi(y_2) \end{bmatrix} \geq 0 \quad \text{for all } y_1, y_2 \in \mathbb{R}$$

Eg, *slope constraint*  $\sigma_1 \leq \psi'(y) \leq \sigma_2$  is described by  $M_{\sigma_1, \sigma_2} = \begin{bmatrix} -\sigma_1 \sigma_2 & (\sigma_1 + \sigma_2)/2 \\ (\sigma_1 + \sigma_2)/2 & -1 \end{bmatrix}$

## Example #8: Systems in Lur'e form

$$F_{\text{Lur'e}}(x) = Ax + B\psi(Cx)$$

assume

- ① nonlinearity  $\psi : \mathbb{R} \rightarrow \mathbb{R}$  described by incremental multiplier  $M$
- ② there exist an  $n \times n$  matrix  $P = P^\top \succ 0$  and a scalar  $c > 0$  satisfying LMI

$$\begin{bmatrix} PA + A^\top P + 2cP & PB \\ B^\top P & 0 \end{bmatrix} + \begin{bmatrix} C & 0 \\ 0_{1 \times n} & 1 \end{bmatrix}^\top M \begin{bmatrix} C & 0 \\ 0_{1 \times n} & 1 \end{bmatrix} \preceq 0$$

$F_{\text{Lur'e}}(x)$  is infinitesimally contracting wrt  $\|\cdot\|_{2,P^{1/2}}$  with rate  $c$

- proof based upon S-lemma
- LMIs defining  $P$  and  $M$  together imply contractivity LMI
- typical vector valued constraints: monotonic or sector bound

L. D'Alto and M. Corless. Incremental quadratic stability. *Numerical Algebra, Control and Optimization*, 3:175–201, 2013. 

M. Giaccagli, V. Andrieu, S. Tarbouriech, and D. Astolfi. Infinite gain margin, contraction and optimality: An LMI-based design. *European Journal of Control*, 68:100685, 2022. 

# Outline

## §1. History and resources

## §2. Basic definitions: discrete and continuous-time dynamics on vector spaces

- The linear algebra of matrix norms; see CTDS Chapter 2
- Properties of induced matrix norms and Lipschitz constants

## §3. Example systems

- Constrained, distributed and proximal gradient dynamics
- Continuous-time recurrent neural networks
- Nonlinear dynamics in Lur'e form

## §4. Properties of contracting dynamics

- Equilibria, Lyapunov functions, and Euler discretization
- Incremental input-to-state stability
- Contractivity of interconnected systems
- Additional properties: entrainment, robustness wrt unmodeled dynamics and delays

## §5. Example applications

- Gradient dynamics and Nash equilibria in games
- Time-varying gradient dynamics and feedback optimization
- Recurrent and implicit neural networks

## §6. Generalizations with examples

- G1: Semicontractivity: Primal-dual gradient with redundant constraints
- G2: Contractivity on Riemannian manifolds and the Karcher mean

## §7. Conclusions and future research

## §8. Advanced Topics

- More on semicontractivity: ergodic coefficients and duality
- Network small-gain theorem for Metzler matrices
- More on the Kuramoto-Sakaguchi model and synchronization
- Proof of semicontractivity of saddle matrices
- Proof of Euler discretization theorem
- Non-Euclidean Monotone Operator Theory

## Equilibrium and Lyapunov functions for a contracting vector field

For a time-invariant  $\mathbf{F}$ ,  $c$ -strongly contracting wrt  $\|\cdot\|$

- ① for each  $t > 0$ , flow at time  $t$  of  $\mathbf{F}$  is a contraction with factor  $e^{-ct}$ ,  
i.e., distance between solutions exponentially decreases with rate  $c$
- ② there exists an equilibrium  $x^*$ , that is unique, globally exponentially stable with global  
Lyapunov functions

$$V_1(x) = \|x - x^*\|^2 \quad \text{and} \quad V_2(x) = \|\mathbf{F}(x)\|^2$$

- ③ if additionally  $D\mathbf{F}(x) = D\mathbf{F}(x)^\top$  for all  $x$ , then another global Lyapunov function is

$$V_3(x) = - \int_0^1 x^\top \mathbf{F}(tx) dt + w \quad \text{for each scalar } w$$

Also,  $V_3$  is  $c$ -strongly convex and  $\mathbf{F} = -\nabla V_3$

## Proof of global Lyapunov functions

Regarding  $V_1(x) = \|x - x^*\|^2$ , from  $D^+ \|x - y\| \leq -c \|x - y\|$ , we immediately have

$$\|x(t) - x^*\| \leq e^{-ct} \|x(0) - x^*\|$$

Regarding  $V_2(x) = \|\mathbf{F}(x)\|^2$ , note  $\frac{d}{dt} \mathbf{F}(x(t)) = D\mathbf{F}(x(t))\dot{x}(t) = D\mathbf{F}(x(t))\mathbf{F}(x(t))$  and

$$\begin{aligned}\|\mathbf{F}(x(t))\| D^+ \|\mathbf{F}(x(t))\| &= \left[ \left[ \frac{d}{dt} \mathbf{F}(x(t)), \mathbf{F}(x(t)) \right] \right] && \text{(curve norm derivative)} \\ &= [D\mathbf{F}(x(t))\mathbf{F}(x(t)), \mathbf{F}(x(t))] \\ &= \mu(D\mathbf{F}(x(t))) [\mathbf{F}(x(t)), \mathbf{F}(x(t))] && \text{(Lumer equality)} \\ &\leq \sup_{z \in \mathbb{R}^n} \mu(D\mathbf{F}(z)) \|\mathbf{F}(x(t))\|^2 = -c \|\mathbf{F}(x(t))\|^2\end{aligned}$$

Regarding  $V_3$ , see M. Fitzsimmons and J. Liu. A note on the equivalence of a strongly convex function and its induced contractive differential equation. *Automatica*, page 110349, 2022. doi. URL

<https://doi.org/10.1016/j.automatica.2022.110349>

## Euler discretization theorem for contracting dynamics

Given arbitrary norm  $\|\cdot\|$  and differentiable  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , equivalent statements

- ①  $\dot{x} = F(x)$  is infinitesimally contracting
- ② there exists  $\alpha > 0$  such that  $x_{k+1} = x_k + \alpha F(x_k)$  is contracting

### Optimal\* contractivity of Euler discretization $\text{Id} + \alpha F$

Given  $c := -\text{osLip}(F) > 0$  and  $\ell := \text{Lip}(F)$ , define *condition number*  $\kappa = \ell/c \geq 1$ :

$$\textcircled{3} \quad 0 < \alpha < \frac{1}{c\kappa(1+\kappa)} \implies \text{Lip}(\text{Id} + \alpha F) \leq \left(1 + \alpha c - \frac{\alpha^2 \ell^2}{1 - \alpha \ell}\right)^{-1} < 1$$

- ④ the optimal\* step size and contraction factor are

$$\alpha^* = \frac{1}{c} \left( \frac{1}{2\kappa^2} - \frac{3}{8\kappa^3} + \mathcal{O}\left(\frac{1}{\kappa^4}\right) \right), \quad \text{Lip}(\text{Id} + \alpha^* F) = 1 - \frac{1}{4\kappa^2} + \frac{1}{8\kappa^3} + \mathcal{O}\left(\frac{1}{\kappa^4}\right)$$

# Outline

## §1. History and resources

## §2. Basic definitions: discrete and continuous-time dynamics on vector spaces

- The linear algebra of matrix norms; see CTDS Chapter 2
- Properties of induced matrix norms and Lipschitz constants

## §3. Example systems

- Constrained, distributed and proximal gradient dynamics
- Continuous-time recurrent neural networks
- Nonlinear dynamics in Lur'e form

## §4. Properties of contracting dynamics

- Equilibria, Lyapunov functions, and Euler discretization
- **Incremental input-to-state stability**
- Contractivity of interconnected systems
- Additional properties: entrainment, robustness wrt unmodeled dynamics and delays

## §5. Example applications

- Gradient dynamics and Nash equilibria in games
- Time-varying gradient dynamics and feedback optimization
- Recurrent and implicit neural networks

## §6. Generalizations with examples

- G1: Semicontractivity: Primal-dual gradient with redundant constraints
- G2: Contractivity on Riemannian manifolds and the Karcher mean

## §7. Conclusions and future research

## §8. Advanced Topics

- More on semicontractivity: ergodic coefficients and duality
- Network small-gain theorem for Metzler matrices
- More on the Kuramoto-Sakaguchi model and synchronization
- Proof of semicontractivity of saddle matrices
- Proof of Euler discretization theorem
- Non-Euclidean Monotone Operator Theory

For time and input-dependent vector  $\mathbf{F}$ ,

$$\dot{x} = \mathbf{F}(t, x, u(t)), \quad x(0) = x_0 \in \mathcal{X}, \quad u(t) \in \mathcal{U}$$

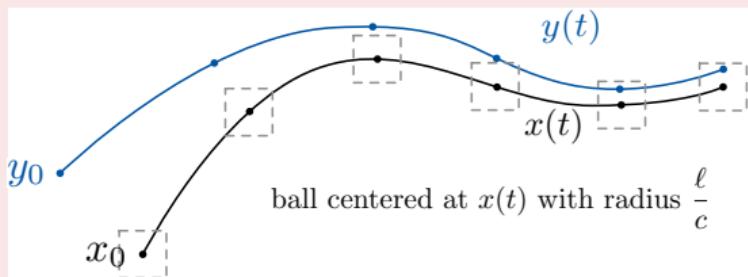
Given norms  $\|\cdot\|_{\mathcal{X}}$  and  $\|\cdot\|_{\mathcal{U}}$ , assume

- **contractivity wrt  $x$ :**  $\text{osLip}_x(\mathbf{F}) \leq -c < 0$ , uniformly in  $t, u$
- **Lipschitz wrt  $u$ :**  $\text{Lip}_u(\mathbf{F}) \leq \ell$ , uniformly in  $t, x$

Then

- ① any soltns:  $x(t)$  with input  $u_x$  and  $y(t)$  with input  $u_y$

$$D^+ \|x(t) - y(t)\|_{\mathcal{X}} \leq -c \|x(t) - y(t)\|_{\mathcal{X}} + \ell \|u_x(t) - u_y(t)\|_{\mathcal{U}}$$



- ②  $F$  is **incrementally ISS**, that is, for all  $x_0, y_0$

$$\|x(t) - y(t)\|_{\mathcal{X}} \leq e^{-ct} \|x_0 - y_0\|_{\mathcal{X}} + \frac{\ell(1 - e^{-ct})}{c} \sup_{\tau \in [0, t]} \|u_x(\tau) - u_y(\tau)\|_{\mathcal{U}}$$

## Proof of iISS property

Using the properties of the weak pairing, we compute

$$\begin{aligned} \|x(t) - y(t)\| D^+ \|x(t) - y(t)\| &= \llbracket \dot{x}(t) - \dot{y}(t), x - y \rrbracket && (\text{curve norm derivative}) \\ &= \llbracket \mathbb{F}(t, x, u_x) - \mathbb{F}(t, y, u_y), x - y \rrbracket \\ &\leq \llbracket \mathbb{F}(t, x, u_x) - \mathbb{F}(t, y, u_x), x - y \rrbracket \\ &\quad + \llbracket \mathbb{F}(t, y, u_x) - \mathbb{F}(t, y, u_y), x - y \rrbracket && (\text{subadditivity}) \\ &\leq -c \|x - y\|^2 + \llbracket \mathbb{F}(t, y, u_x) - \mathbb{F}(t, y, u_y), x - y \rrbracket && (\text{contractivity}) \\ &\leq -c \|x - y\|^2 + \|\mathbb{F}(t, y, u_x) - \mathbb{F}(t, y, u_y)\| \|x - y\| && (\text{Cauchy-Schwartz}) \\ &\leq -c \|x - y\|^2 + \ell \|u_x - u_y\|_{\mathcal{U}} \|x - y\|. && (\text{Lipschitzness}) \end{aligned}$$

## Signal norms and system gains

Given norm  $\|\cdot\|_{\mathcal{X}}$  on  $\mathbb{R}^n$  (or  $\|\cdot\|_{\mathcal{U}}$  on  $\mathbb{R}^k$ ),

- $\mathcal{L}_{\mathcal{X}}^q$ ,  $q \in [1, \infty]$ , is vector space of continuous signals,  $x : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}^n$ , with well-defined bounded norm

$$\|x(\cdot)\|_{\mathcal{X},q} = \begin{cases} \left( \int_0^\infty \|x(t)\|_{\mathcal{X}}^q dt \right)^{1/q} & \text{if } q \in [1, \infty[ \\ \sup_{t \geq 0} \|x(t)\|_{\mathcal{X}} & \text{if } q = \infty \end{cases}$$

- Input-state system has  $\mathcal{L}_{\mathcal{X},\mathcal{U}}^q$ -induced gain upper bounded by  $\gamma > 0$  if, for all  $u \in \mathcal{L}_{\mathcal{U}}^q$ , the state  $x$  from zero initial state satisfies

$$\|x(\cdot)\|_{\mathcal{X},q} \leq \gamma \|u(\cdot)\|_{\mathcal{U},q}$$

- ③ F has incremental  $\mathcal{L}_{\mathcal{X},\mathcal{U}}^q$  gain equal to  $\ell/c$ , for  $q \in [1, \infty]$ ,

$$\|x(\cdot) - y(\cdot)\|_{\mathcal{X},q} \leq \frac{\ell}{c} \|u_x(\cdot) - u_y(\cdot)\|_{\mathcal{U},q} \quad (\text{for } x_0 = y_0)$$

## Application: Parametrized fixed point problem

$$\emptyset_n = F(x, u), \quad x \in \mathcal{X}, u \in \mathcal{U}$$

Given norms  $\|\cdot\|_{\mathcal{X}}$  and  $\|\cdot\|_{\mathcal{U}}$ , assume

- **contractivity wrt  $x$ :**  $\text{osLip}_x(F) \leq -c < 0$ , uniformly in  $u$
- **Lipschitz wrt  $u$ :**  $\text{Lip}_u(F) \leq \ell$ , uniformly in  $x$

① for each fixed  $u$ , there exists a unique equilibrium  $x^*(u)$

② the equilibrium map  $x^* : \mathcal{U} \rightarrow \mathcal{X}$  is Lipschitz with constant  $\frac{\ell}{c}$

### Sensitivity analysis in convex optimization

If  $f(x, u)$  is  $\nu$ -strongly convex and differentiable wrt  $x$ ,

$\nabla_x f$  is  $\ell$ -Lipschitz wrt  $u$ ,

then global minimum  $u \mapsto x^*(u)$  is Lipschitz with constant  $\frac{\ell}{\nu}$

## Proof of Parametrized continuous-time Banach Contraction Theorem

Recall iISS: any soltns  $x_1(t)$  with input  $u_1(t)$  and  $x_2(t)$  with input  $u_2(t)$

$$D^+ \|x_1(t) - x_2(t)\|_{\mathcal{X}} \leq -c \|x_1(t) - x_2(t)\|_{\mathcal{X}} + \ell \|u_1(t) - u_2(t)\|_{\mathcal{U}}$$

For constant  $u_1(t) = u_1$  and  $u_2(t) = u_2$ , as  $t \rightarrow +\infty$ ,

$$x_1(t) \rightarrow x^*(u_1) \quad \text{and} \quad x_2(t) \rightarrow x^*(u_2)$$

Taking the limit, we obtain

$$0 \leq -c \|x^*(u_1) - x^*(u_2)\|_{\mathcal{X}} + \ell \|u_1 - u_2\|_{\mathcal{U}}$$

that is,  $\|x^*(u_1) - x^*(u_2)\|_{\mathcal{X}} \leq \frac{\ell}{c} \|u_1 - u_2\|_{\mathcal{U}}$

# Outline

## §1. History and resources

## §2. Basic definitions: discrete and continuous-time dynamics on vector spaces

- The linear algebra of matrix norms; see CTDS Chapter 2
- Properties of induced matrix norms and Lipschitz constants

## §3. Example systems

- Constrained, distributed and proximal gradient dynamics
- Continuous-time recurrent neural networks
- Nonlinear dynamics in Lur'e form

## §4. Properties of contracting dynamics

- Equilibria, Lyapunov functions, and Euler discretization
- Incremental input-to-state stability
- **Contractivity of interconnected systems**
- Additional properties: entrainment, robustness wrt unmodeled dynamics and delays

## §5. Example applications

- Gradient dynamics and Nash equilibria in games
- Time-varying gradient dynamics and feedback optimization
- Recurrent and implicit neural networks

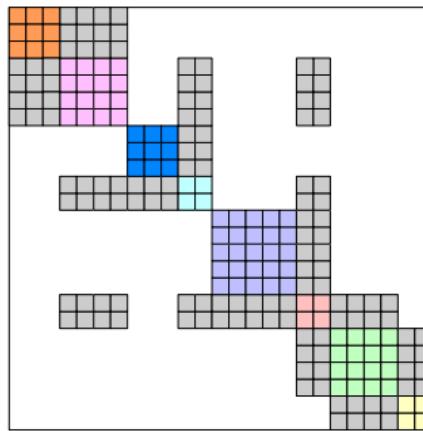
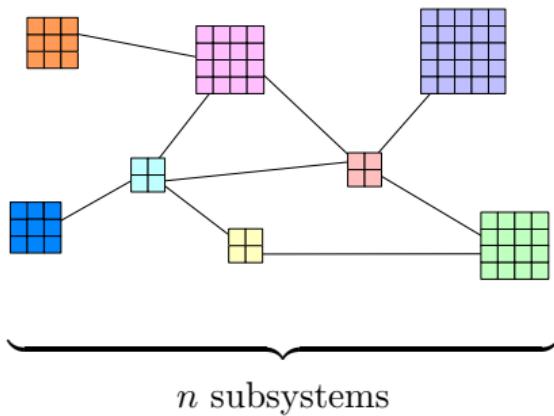
## §6. Generalizations with examples

- G1: Semicontractivity: Primal-dual gradient with redundant constraints
- G2: Contractivity on Riemannian manifolds and the Karcher mean

## §7. Conclusions and future research

## §8. Advanced Topics

- More on semicontractivity: ergodic coefficients and duality
- Network small-gain theorem for Metzler matrices
- More on the Kuramoto-Sakaguchi model and synchronization
- Proof of semicontractivity of saddle matrices
- Proof of Euler discretization theorem
- Non-Euclidean Monotone Operator Theory



- ➊  $n$  local norms  $\|\cdot\|_i$  on  $\mathbb{R}^{N_i}$ ,  $i \in \{1, \dots, n\}$
- ➋ a aggregating norm  $\|\cdot\|_{\text{agg}}$  on  $\mathbb{R}^n$
- ➌  $\implies$  a composite norm on  $\mathbb{R}^N$ ,  $N = N_1 + \dots + N_n$

T. Ström. On logarithmic norms. *SIAM Journal on Numerical Analysis*, 12(5):741–753, 1975. [doi](#)

O. Pastravanu and M. Voicu. Generalized matrix diagonal stability and linear dynamical systems. *Linear Algebra and its Applications*, 419(2):299–310, 2006. [doi](#)

G. Russo, M. Di Bernardo, and E. D. Sontag. A contraction approach to the hierarchical analysis and design of networked systems. *IEEE Transactions on Automatic Control*, 58(5):1328–1331, 2013. [doi](#)

Interconnected subsystems:  $x_i \in \mathbb{R}^{N_i}$  and  $x_{-i} \in \mathbb{R}^{N-N_i}$ :

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

**Network contraction theorem.** Given local norms, assume

- **contractivity wrt  $x_i$ :**  $\text{osLip}_{x_i}(F_i) \leq -c_i < 0$ , uniformly in  $x_{-i}$
- **Lipschitz wrt  $x_j$ ,  $j \neq i$ :**  $\text{Lip}_{x_j}(F_i) \leq \ell_{ij}$ , uniformly in  $x_{-j}$

- the Lipschitz constants matrix  $\Gamma = \begin{bmatrix} -c_1 & \dots & \ell_{1n} \\ \vdots & & \vdots \\ \ell_{n1} & \dots & -c_n \end{bmatrix}$  is **Hurwitz**

$\implies$  the **interconnected system** is infinitesimally contracting  
wrt composite  $\|\cdot\|$  generated by log optimal norm for  $\Gamma$  and  $c = |\alpha(\Gamma)|$

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

(see LNS.Section10.4)

**Hurwitzness depends upon both topology and edge weights**

- $M$  Hurwitz iff there exists a positive  $\xi$  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**

## Proof of Network Contraction Theorem

First, design a log optimal norm for  $\Gamma = \begin{bmatrix} -c_1 & \dots & \ell_{1r} \\ \vdots & & \vdots \\ \ell_{r1} & \dots & -c_r \end{bmatrix} \in \mathbb{R}^{r \times r}$

From Lemma 3.21 on Metzler matrices in CTDS, for arbitrarily small  $\epsilon$ , one can compute  $\eta \in \mathbb{R}_{>0}^n$  such that  $\|\cdot\|_{2,\text{diag}(\eta)^{1/2}}$  is  $\epsilon$ -log optimal for  $\Gamma$ :

$$\mu_{2,\text{diag}(\eta)^{1/2}}(\Gamma) \leq \alpha(\Gamma) + \epsilon \iff \text{diag}(\eta)\Gamma + \Gamma^\top \text{diag}(\eta) \preceq 2(\alpha(\Gamma) + \epsilon) \text{diag}(\eta)$$

Next, define the composite norm  $\|\cdot\|_\eta$  on  $\mathbb{R}^N$  by

$$\|(x_1, \dots, x_r)\|_\eta^2 = \sum_{i=1}^r \eta_i \|x_i\|_i^2$$

with weak pairing

$$[(x_1, \dots, x_r), (y_1, \dots, y_r)]_\eta = \sum_{i=1}^r \eta_i [x_i, y_i]_i$$

For each  $i$ , compute

$$\begin{aligned}
 & \| \mathbb{F}_i(t, x_i, x_{-i}) - \mathbb{F}_i(t, y_i, y_{-i}), x_i - y_i \|_i \\
 & \leq \| \mathbb{F}_i(t, x_i, x_{-i}) - \mathbb{F}_i(t, y_i, x_{-i}), x_i - y_i \|_i + \| \mathbb{F}_i(t, y_i, x_{-i}) - \mathbb{F}_i(t, y_i, y_{-i}), x_i - y_i \|_i \\
 & \leq -c_i \|x_i - y_i\|_i^2 + \sum_{j=1, j \neq i}^r \ell_{ij} \|x_j - y_j\|_j \|x_i - y_i\|_i
 \end{aligned}$$

Next, we check the one-sided Lipschitz condition for the vector field on  $\mathbb{R}^N$ :

$$\begin{aligned}
 & \sum_{i=1}^r \eta_i \| \mathbb{F}_i(t, x_i, x_{-i}) - \mathbb{F}_i(t, y_i, y_{-i}), x_i - y_i \|_i \\
 & \leq - \sum_{i=1}^r \eta_i c_i \|x_i - y_i\|_i^2 + \sum_{i,j=1, j \neq i}^r \eta_i \ell_{ij} \|x_j - y_j\|_j \|x_i - y_i\|_i \\
 & = \begin{bmatrix} \|x_1 - y_1\|_1 \\ \vdots \\ \|x_r - y_r\|_r \end{bmatrix}^\top \text{diag}(\eta) \Gamma \begin{bmatrix} \|x_1 - y_1\|_1 \\ \vdots \\ \|x_r - y_r\|_r \end{bmatrix} \\
 & = \begin{bmatrix} \|x_1 - y_1\|_1 \\ \vdots \\ \|x_r - y_r\|_r \end{bmatrix}^\top \frac{\text{diag}(\eta) \Gamma + \Gamma^\top \text{diag}(\eta)}{2} \begin{bmatrix} \|x_1 - y_1\|_1 \\ \vdots \\ \|x_r - y_r\|_r \end{bmatrix}
 \end{aligned}$$

so that the interconnected system is contracting if the gain matrix  $\Gamma$  is diagonally stable.

## Application: Singularly perturbed matrices

Given a constant  $\epsilon > 0$ , consider block matrix

$$\mathcal{A}_\epsilon = \begin{bmatrix} \epsilon A & \epsilon B \\ C & D \end{bmatrix} \in \mathbb{R}^{(n+m) \times (n+m)}.$$

$$\begin{array}{c} \mu(A) < 0, \mu(D) < 0, \text{ and} \\ \mu(A)\mu(D) > \|B\|\|C\| \end{array} \implies \mathcal{A}_\epsilon \text{ is Hurwitz for all } \epsilon > 0$$

$$\begin{array}{ccc} & & \Downarrow \\ \Downarrow & & \Downarrow \\ D \text{ and } A - BD^{-1}C \text{ are Hurwitz} & \implies & \exists \epsilon^* \text{ s.t. } \mathcal{A}_\epsilon \text{ is Hurwitz for each } \epsilon < \epsilon^* \end{array}$$

Additionally, a valid choice of  $\epsilon^*$  is:

$$\epsilon^* := \frac{|\mu(A - BD^{-1}C)| \cdot |\mu(D)|}{\|B\|\|D^{-1}C(A - BD^{-1}C)\| + |\mu(A - BD^{-1}C)|\|D^{-1}CB\|}$$

# Outline

## §1. History and resources

## §2. Basic definitions: discrete and continuous-time dynamics on vector spaces

- The linear algebra of matrix norms; see CTDS Chapter 2
- Properties of induced matrix norms and Lipschitz constants

## §3. Example systems

- Constrained, distributed and proximal gradient dynamics
- Continuous-time recurrent neural networks
- Nonlinear dynamics in Lur'e form

## §4. Properties of contracting dynamics

- Equilibria, Lyapunov functions, and Euler discretization
- Incremental input-to-state stability
- Contractivity of interconnected systems
- Additional properties: entrainment, robustness wrt unmodeled dynamics and delays

## §5. Example applications

- Gradient dynamics and Nash equilibria in games
- Time-varying gradient dynamics and feedback optimization
- Recurrent and implicit neural networks

## §6. Generalizations with examples

- G1: Semicontractivity: Primal-dual gradient with redundant constraints
- G2: Contractivity on Riemannian manifolds and the Karcher mean

## §7. Conclusions and future research

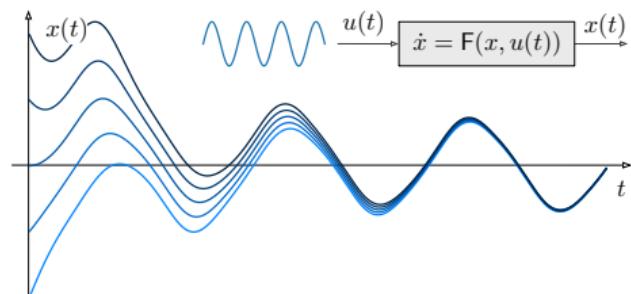
## §8. Advanced Topics

- More on semicontractivity: ergodic coefficients and duality
- Network small-gain theorem for Metzler matrices
- More on the Kuramoto-Sakaguchi model and synchronization
- Proof of semicontractivity of saddle matrices
- Proof of Euler discretization theorem
- Non-Euclidean Monotone Operator Theory

# Entrainment in systems with periodic time-dependence

For time-varying vector field  $\mathbf{F}$  and norm  $\|\cdot\|$

- ①  $\text{osLip}_x(\mathbf{F}) \leq -c < 0$
- ②  $\mathbf{F}$  is  $T$ -periodic



Then

- ① there exists a unique periodic solution  $x^* : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}^n$  with period  $T$
- ② for every initial condition  $x_0$ ,

$$\|x(t, x_0) - x^*(t)\| \leq e^{-ct} \|x_0 - x^*(0)\|$$

Given a norm  $\|\cdot\|$ , consider

$$\dot{x} = F(x) + \Delta(x)$$

Assume:

- **contractivity:**  $\text{osLip}(F) \leq -c < 0$
- **bounded disturbance:**  $\text{osLip}(\Delta) \leq d < c$

Then

- ①  $F + \Delta$  is strongly contracting with rate  $c - d$
- ② the unique equilibria  $x_F^*$  of  $F$  and  $x_{F+\Delta}^*$  of  $F + \Delta$  satisfy

$$\|x_F^* - x_{F+\Delta}^*\| \leq \frac{\|\Delta(x_F^*)\|}{c - d}$$

$$\dot{x}(t) = F(x(t), x(t-s), u(t)), 0 \leq s \leq S, \quad \| \cdot \|_{\mathcal{X}}, \| \cdot \|_{\mathcal{U}} \quad (2)$$

assume there exist positive constants  $c, \ell_{\mathcal{U}}, \ell_{\mathcal{X}}$  such that, for all variables,

$$\text{osL } x : \quad \|F(x, d, u) - F(y, d, u), x - y\|_{\mathcal{X}} \leq -c\|x - y\|_{\mathcal{X}}^2 \quad (3)$$

$$\text{Lip } x(t-s) : \quad \|F(x, x_1, u) - F(x, x_2, u)\|_{\mathcal{X}} \leq \ell_{\mathcal{X}}\|x_1 - x_2\|_{\mathcal{X}} \quad (4)$$

$$\text{Lip } u : \quad \|F(x, d, u) - F(x, d, v)\|_{\mathcal{X}} \leq \ell_{\mathcal{U}}\|u - v\|_{\mathcal{U}} \quad (5)$$

By the curve norm derivative formula, subadditivity, and Cauchy-Schwarz inequality,

$$\begin{aligned} \|x(t) - y(t)\|_{\mathcal{X}} D^+ \|x(t) - y(t)\|_{\mathcal{X}} &= \|F(x(t), x(t-s), u_x(t)) - F(y(t), y(t-s), u_y(t)), x(t) - y(t)\|_{\mathcal{X}} \\ &\leq \|F(x(t), x(t-s), u_x(t)) - F(y(t), x(t-s), u_x(t)), x(t) - y(t)\|_{\mathcal{X}} \\ &\quad + \|F(y(t), x(t-s), u_x(t)) - F(y(t), y(t-s), u_x(t)), x(t) - y(t)\|_{\mathcal{X}} \\ &\quad + \|F(y(t), y(t-s), u_x(t)) - F(y(t), y(t-s), u_y(t)), x(t) - y(t)\|_{\mathcal{X}} \\ &\leq -c\|x(t) - y(t)\|_{\mathcal{X}}^2 + \ell_{\mathcal{X}}\|x(t-s) - y(t-s)\|_{\mathcal{U}}\|x(t) - y(t)\|_{\mathcal{X}} \\ &\quad + \ell_{\mathcal{U}}\|u_x(t) - u_y(t)\|_{\mathcal{U}}\|x(t) - y(t)\|_{\mathcal{X}}. \end{aligned}$$

Thus, with  $m(t) = \|x(t) - y(t)\|_{\mathcal{X}}$ , delay differential inequality:

$$D^+ m(t) \leq -cm(t) + \ell_{\mathcal{X}} \sup_{0 \leq s \leq S} m(t-s) + \ell_{\mathcal{U}}\|u_x(t) - u_y(t)\|_{\mathcal{U}}, \quad (6)$$

Halanay inequality is applicable. If  $c > \ell_{\mathcal{X}}$ , then

$$m(t) \leq m_0 e^{-\rho(t-t_0)} + \ell_{\mathcal{U}} \int_{t_0}^t e^{-\rho(t-\tau)} \|u_x(\tau) - u_y(\tau)\|_{\mathcal{U}} d\tau, \quad (7)$$

where  $\rho > 0$  is the unique positive root of  $\rho = c - \ell_{\mathcal{X}} e^{\rho S}$  and  $m_0 = \sup_{0 \leq s \leq S} m(t_0 - s)$ .

Interconnected subsystems  $i \in \{1, \dots, n\}$

$$\dot{x}_i = F_i(x_i, x_{-i}, x_{-i}(t-s), u_i), \quad 0 \leq s \leq S, \quad \|\cdot\|_i, \|\cdot\|_{i,\mathcal{U}} \quad (8)$$

Assume there exist positive constants st

**osL**  $x_i$ :  $\llbracket F_i(x_i, \dots) - F_i(y_i, \dots), x_i - y_i \rrbracket_i \leq -c_i \|x_i - y_i\|_i^2$

**Lip**  $x_{-i}$ :  $\|F_i(\dots, x_{-i}, \dots) - F_i(\dots, y_{-i}, \dots)\|_i \leq \sum_{j=1, j \neq i}^n \gamma_{ij} \|x_j - y_j\|_j$

**Lip**  $x_{-1}^{-s}$ :  $\|F_i(\dots, x_{-i}^{-s}, \dots) - F_i(\dots, y_{-i}^{-s}, \dots)\|_i \leq \sum_{j=1, j \neq i}^n \hat{\gamma}_{ij} \|x_j^{-s} - y_j^{-s}\|_j$

**Lip**  $u_i$ :  $\|F_i(\dots, u_i) - F_i(\dots, v_i)\|_i \leq \ell_{i,\mathcal{U}} \|u_i - v_i\|_{i,\mathcal{U}}$

With  $m_i(t) = \|x_i(t) - y_i(t)\|_i$ , delay differential inequality:

$$D^+ m(t) \leq -Cm(t) + \Gamma m(t) + \widehat{\Gamma} \sup_{0 \leq s \leq S} m(t-s) + \ell_{i,\mathcal{U}} \|u_x(t) - u_y(t)\|_{i,\mathcal{U}}$$

and, if the Metzler matrix  $-C + \Gamma + \widehat{\Gamma}$  is Hurwitz, then (8) is incremental ISS

# Outline

## §1. History and resources

## §2. Basic definitions: discrete and continuous-time dynamics on vector spaces

- The linear algebra of matrix norms; see CTDS Chapter 2
- Properties of induced matrix norms and Lipschitz constants

## §3. Example systems

- Constrained, distributed and proximal gradient dynamics
- Continuous-time recurrent neural networks
- Nonlinear dynamics in Lur'e form

## §4. Properties of contracting dynamics

- Equilibria, Lyapunov functions, and Euler discretization
- Incremental input-to-state stability
- Contractivity of interconnected systems
- Additional properties: entrainment, robustness wrt unmodeled dynamics and delays

## §5. Example applications

- Gradient dynamics and Nash equilibria in games
- Time-varying gradient dynamics and feedback optimization
- Recurrent and implicit neural networks

## §6. Generalizations with examples

- G1: Semicontractivity: Primal-dual gradient with redundant constraints
- G2: Contractivity on Riemannian manifolds and the Karcher mean

## §7. Conclusions and future research

## §8. Advanced Topics

- More on semicontractivity: ergodic coefficients and duality
- Network small-gain theorem for Metzler matrices
- More on the Kuramoto-Sakaguchi model and synchronization
- Proof of semicontractivity of saddle matrices
- Proof of Euler discretization theorem
- Non-Euclidean Monotone Operator Theory

- Nash equilibria: existence, uniqueness, computation, convergence for gradient-like dynamics, robustness
- games with partial information
- aggregative games: demand-side management in the smart grid, charging control for plug-in electric vehicles, spectrum sharing in wireless networks, and network congestion control

S. Li and T. Başar. Distributed algorithms for the computation of noncooperative equilibria. *Automatica*, 23(4):523–533, 1987. 

D. Gadjov and L. Pavel. A passivity-based approach to Nash equilibrium seeking over networks. *IEEE Transactions on Automatic Control*, 64(3):1077–1092, 2019. 

M. Arcak and N. C. Martins. Dissipativity tools for convergence to Nash equilibria in population games. *IEEE Transactions on Control of Network Systems*, 8(1):39–50, 2021. 

G. Belgioioso, P. Yi, S. Grammatico, and L. Pavel. Distributed generalized Nash equilibrium seeking: An operator-theoretic perspective. *IEEE Control Systems*, 42(4):87–102, 2022. 

A. Gokhale, A. Davydov, and F. Bullo. Contractivity of distributed optimization and Nash seeking dynamics. *IEEE Control Systems Letters*, Sept. 2023.  Submitted

## Example #9: Saddle dynamics

Assume  $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$

- $x \mapsto f(x, y)$  is  $\nu_x$ -strongly convex, uniformly in  $y$
- $y \mapsto f(x, y)$  is  $\nu_y$ -strongly concave, uniformly in  $x$

**saddle dynamics (primal-descent / dual-ascent):**

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \mathsf{F}_S(x, y) := \begin{bmatrix} -\nabla_x f(x, y) \\ \nabla_y f(x, y) \end{bmatrix}$$

$\mathsf{F}_S$  is infinitesimally contracting wrt  $\|\cdot\|_2$  with rate  $\min\{\nu_x, \nu_y\}$

unique globally exp stable point is saddle point (min in  $x$ , max in  $y$ )

If  $f$  is twice-differentiable, then

$$\begin{aligned} \mu_2(D\mathsf{F}_S(x, y)) &= \mu_2 \left( \begin{bmatrix} -\text{Hess}_x f(x, y) & -D_y \nabla_x f(x, y) \\ D_x \nabla_y f(x, y) & \text{Hess}_y f(x, y) \end{bmatrix} \right) \\ &\stackrel{\mu_2(A)=\mu_2(\frac{A+A^\top}{2})}{=} \mu_2 \left( \begin{bmatrix} -\text{Hess}_x f(x, y) & 0 \\ 0 & \text{Hess}_y f(x, y) \end{bmatrix} \right) = -\min\{\nu_x, \nu_y\} \end{aligned}$$

## Example #10: Pseudogradient play

Each player  $i$  aims to minimize its own cost function  $J_i(x_i, x_{-i})$  (not a potential game)

**pseudogradient dynamics (aka gradient play in game theory):**

$$\begin{aligned}\dot{x} &= \mathsf{F}_{\text{PseudoG}}(x) = -(\nabla_1 J_1(x_1, x_{-1}), \dots, \nabla_n J_n(x_n, x_{-n})) \quad (\text{stacked vector}) \\ \iff \quad \dot{x}_i &= -\nabla_i J_i(x_i, x_{-i})\end{aligned}$$

- **strong convexity wrt  $x_i$ :**  $J_i$  is  $\mu_i$  strongly convex wrt  $x_i$ , uniformly in  $x_{-i}$
  - **Lipschitz wrt  $x_{-i}$ :**  $\text{Lip}_{x_j}(\nabla_i J_i) \leq \ell_{ij}$ , uniformly in  $x_{-j}$
  - $\mathsf{F}_{\text{PseudoG}}$  gain matrix is Hurwitz
- ⇒  $\mathsf{F}_{\text{PseudoG}}$  is infinitesimally contracting wrt appropriate diag-weighted  $\|\cdot\|_2$

if  $\mathsf{F}_{\text{PseudoG}}$  is infinitesimally contracting (wrt any norm)  
then unique globally exp stable Nash equilibrium  $J_i(x_i^*, x_{-i}^*) \leq J_i(y_i, x_{-i}^*)$  for all  $y_i$

## Example #11: Best response play

Each player  $i$  aims to minimize its own cost function  $J_i(x_i, x_{-i})$

$\text{BR}_i : x_{-i} \rightarrow \operatorname{argmin}_{x_i} J_i(x_i, x_{-i})$  best response of player  $i$  wrt other decisions  $x_{-i}$

**best response dynamics:**

$$\begin{aligned}\dot{x} &= F_{\text{BR}}(x) := \text{BR}(x) - x \\ \iff \quad \dot{x}_i &= \text{BR}_i(x_{-i}) - x_i\end{aligned}$$

- **strong convexity wrt  $x_i$ :**  $J_i$  is  $\mu_i$  strongly convex wrt  $x_i$ , uniformly in  $x_{-i}$
- **Lipschitz wrt  $x_{-i}$ :**  $\text{Lip}_{x_j}(\nabla_i J_i) \leq \ell_{ij}$ , uniformly in  $x_{-j}$   
 $\implies \text{BR}_i$  is Lipschitz wrt  $x_j$  with constant  $\ell_{ij}/\mu_i$
- $F_{\text{BR}}$  gain matrix is Hurwitz  $\iff$  BR is a discrete-time contraction  
 $\implies \text{BR} - \text{Id}$  is infinitesimally contracting wrt appropriate diag-weighted  $\|\cdot\|_2$

if  $F_{\text{BR}}$  is infinitesimally contracting (wrt any norm)  
then unique globally exp stable Nash equilibrium (fixed point of BR)

## Equivalent statements:

①  $F_{\text{PseudoG}}$  gain matrix:

$$\begin{bmatrix} -\mu_1 & \dots & \ell_{1n} \\ \vdots & & \vdots \\ \ell_{n1} & \dots & -\mu_n \end{bmatrix} \text{ is Hurwitz}$$

②  $F_{\text{BR}}$  gain matrix:

$$\begin{bmatrix} -1 & \dots & \ell_{1n}/\mu_1 \\ \vdots & & \vdots \\ \ell_{n1}/\mu_n & \dots & -1 \end{bmatrix} \text{ is Hurwitz}$$

③ discrete-time  $F_{\text{BR}}$  gain matrix:

$$\begin{bmatrix} 0 & \dots & \ell_{1n}/\mu_1 \\ \vdots & & \vdots \\ \ell_{n1}/\mu_n & \dots & 0 \end{bmatrix} \text{ is Schur}$$

**Aggregative games:**  $J_i(x_i, x_{-i}) = f_i(x_i, \frac{1}{n} \sum_{j=1}^n x_j)$

assume  $f_i$  is  $\mu_i$ -strongly convex wrt  $x_i$  and  $\ell_i = \text{Lip}_y(\nabla_{x_i} f_i(x_i, y))$

$$\mu_i > \ell_i \text{ for each agent } i \quad \implies \quad \text{Hurwitz}$$

# Outline

## §1. History and resources

## §2. Basic definitions: discrete and continuous-time dynamics on vector spaces

- The linear algebra of matrix norms; see CTDS Chapter 2
- Properties of induced matrix norms and Lipschitz constants

## §3. Example systems

- Constrained, distributed and proximal gradient dynamics
- Continuous-time recurrent neural networks
- Nonlinear dynamics in Lur'e form

## §4. Properties of contracting dynamics

- Equilibria, Lyapunov functions, and Euler discretization
- Incremental input-to-state stability
- Contractivity of interconnected systems
- Additional properties: entrainment, robustness wrt unmodeled dynamics and delays

## §5. Example applications

- Gradient dynamics and Nash equilibria in games
- **Time-varying gradient dynamics and feedback optimization**
- Recurrent and implicit neural networks

## §6. Generalizations with examples

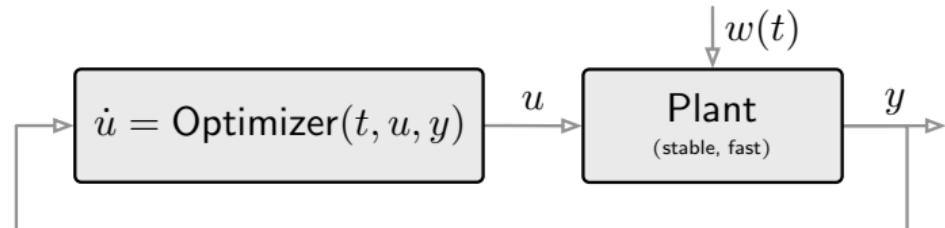
- G1: Semicontractivity: Primal-dual gradient with redundant constraints
- G2: Contractivity on Riemannian manifolds and the Karcher mean

## §7. Conclusions and future research

## §8. Advanced Topics

- More on semicontractivity: ergodic coefficients and duality
- Network small-gain theorem for Metzler matrices
- More on the Kuramoto-Sakaguchi model and synchronization
- Proof of semicontractivity of saddle matrices
- Proof of Euler discretization theorem
- Non-Euclidean Monotone Operator Theory

## Solving optimization problems via dynamical systems



- studies in linear and nonlinear programming (Arrow, Hurwicz, and Uzawa 1958)
- neural networks (Hopfield and Tank 1985) and analog circuits (Kennedy and Chua 1988)
- optimization on manifolds (Brockett 1991)
- ...
- online and dynamic feedback optimization (Dall'Anese, Dörfler, Simonetto, ...)

A. Davydov, V. Centorrino, A. Gokhale, G. Russo, and F. Bullo. Contracting dynamics for time-varying convex optimization. *IEEE Transactions on Automatic Control*, June 2023a. Submitted

L. Cothren, F. Bullo, and E. Dall'Anese. Singular perturbation via contraction theory. *Technical Report*, Sept. 2023

# From convex optimization to contracting dynamics – time-varying

Many convex optimization problems can be solved with contracting dynamics

$$\dot{x} = \mathsf{F}(x, \theta)$$

	Convex Optimization	Contracting Dynamics
Unconstrained	$\min_{x \in \mathbb{R}^n} f(x, \theta)$	$\dot{x} = -\nabla_x f(x, \theta)$
Constrained	$\min_{\substack{x \in \mathbb{R}^n \\ \text{s.t.} \\ x \in \mathcal{X}(\theta)}} f(x, \theta)$	$\dot{x} = -x + \text{Proj}_{\mathcal{X}(\theta)}(x - \gamma \nabla_x f(x, \theta))$
Composite	$\min_{x \in \mathbb{R}^n} f(x, \theta) + g(x, \theta)$	$\dot{x} = -x + \text{prox}_{\gamma g_\theta}(x - \gamma \nabla_x f(x, \theta))$
Equality	$\min_{\substack{x \in \mathbb{R}^n \\ \text{s.t.} \\ Ax = b(\theta)}} f(x, \theta)$	$\begin{aligned} \dot{x} &= -\nabla_x f(x, \theta) - A^\top \lambda, \\ \dot{\lambda} &= Ax - b(\theta) \end{aligned}$
Inequality	$\min_{\substack{x \in \mathbb{R}^n \\ \text{s.t.} \\ Ax \leq b(\theta)}} f(x, \theta)$	$\begin{aligned} \dot{x} &= -\nabla f(x, \theta) - A^\top \nabla M_{\gamma, b(\theta)}(Ax + \gamma \lambda), \\ \dot{\lambda} &= \gamma(-\lambda + \nabla M_{\gamma, b(\theta)}(Ax + \gamma \lambda)) \end{aligned}$

# Tracking equilibrium trajectories

For parameter-dependent vector field  $\mathbf{F} : \mathbb{R}^n \times \mathbb{R}^d \rightarrow \mathbb{R}^n$  and differentiable  $\theta : \mathbb{R}_{\geq 0} \rightarrow \Theta \subset \mathbb{R}^d$

$$\dot{x}(t) = \mathbf{F}(x(t), \theta(t))$$

Assume there exist norms  $\|\cdot\|_{\mathcal{X}}$  and  $\|\cdot\|_{\Theta}$  s.t.

- **contractivity wrt  $x$ :**  $\text{osLip}_x(\mathbf{F}) \leq -c < 0$ , uniformly in  $\theta$
- **Lipschitz wrt  $\theta$ :**  $\text{Lip}_{\theta}(\mathbf{F}) \leq \ell$ , uniformly in  $x$

**Theorem: Incremental ISS** any two soltns:  $x(t)$  with input  $\theta_x$  and  $y(t)$  with input  $\theta_y$

$$D^+ \|x(t) - y(t)\|_{\mathcal{X}} \leq -c \|x(t) - y(t)\|_{\mathcal{X}} + \ell \|\theta_x(t) - \theta_y(t)\|_{\Theta}$$

# Tracking equilibrium trajectories

For parameter-dependent vector field  $\mathbf{F} : \mathbb{R}^n \times \mathbb{R}^d \rightarrow \mathbb{R}^n$  and differentiable  $\theta : \mathbb{R}_{\geq 0} \rightarrow \Theta \subset \mathbb{R}^d$

$$\dot{x}(t) = \mathbf{F}(x(t), \theta(t))$$

Assume there exist norms  $\|\cdot\|_{\mathcal{X}}$  and  $\|\cdot\|_{\Theta}$  s.t.

- **contractivity wrt  $x$ :**  $\text{osLip}_x(\mathbf{F}) \leq -c < 0$ , uniformly in  $\theta$
- **Lipschitz wrt  $\theta$ :**  $\text{Lip}_{\theta}(\mathbf{F}) \leq \ell$ , uniformly in  $x$

## Theorem: Equilibrium tracking for contracting dynamics

- ① for each fixed  $\theta$ , there exists a unique equilibrium  $x^*(\theta)$
- ② the equilibrium map  $x^*(\cdot)$  is Lipschitz with constant  $\frac{\ell}{c}$
- ③  $D^+ \|x(t) - x^*(\theta(t))\|_{\mathcal{X}} \leq -c \|x(t) - x^*(\theta(t))\|_{\mathcal{X}} + \frac{\ell}{c} \|\dot{\theta}(t)\|_{\Theta}$

## Proof of equilibrium tracking

Given:  $\dot{x} = F(x, \theta(t))$  with  $\text{osLip}_x(F) \leq -c$  and  $\text{Lip}_u(F) \leq \ell$

Task: compare **traj**  $x(t)$  with **equilibrium traj**  $x^*(\theta(t))$  implicitly defined by  $F(x, \theta(t)) = 0$

Consider **auxiliary dynamics** with two trajectories:

$$\dot{x} = F(x, \theta(t)) + v(t) =: F_{\text{aux}}(x, \theta, v)$$

①  $v(t) = 0 \implies \text{trajectory } x(t)$

②  $v(t) = \dot{x}^*(\theta(t)) \implies \text{equilibrium trajectory } x^*(\theta(t))$

$F_{\text{aux}}$  is contracting with  $\text{osLip}_x(F_{\text{aux}}) \leq -c$  and  $\text{Lip}_v(F_{\text{aux}}) = 1$ . Hence, iISS:

$$\begin{aligned} D^+ \|x(t) - x^*(\theta(t))\|_{\mathcal{X}} &\leq -c \cdot \|x(t) - x^*(\theta(t))\|_{\mathcal{X}} + 1 \cdot \|0 - \dot{x}^*(\theta(t))\|_{\mathcal{X}} \\ &\leq -c \cdot \|x(t) - x^*(\theta(t))\|_{\mathcal{X}} + \frac{\ell}{c} \cdot \|\dot{\theta}(t)\|_{\Theta} \quad \left(\text{since } \text{Lip}(x^*) = \frac{\ell}{c}\right) \end{aligned}$$

$$D^+ \|x(t) - x^*(\theta(t))\|_{\mathcal{X}} \leq -c \|x(t) - x^*(\theta(t))\|_{\mathcal{X}} + \frac{\ell}{c} \|\dot{\theta}(t)\|_{\Theta}$$

- bounded input, bounded error  
with asymptotic bound:

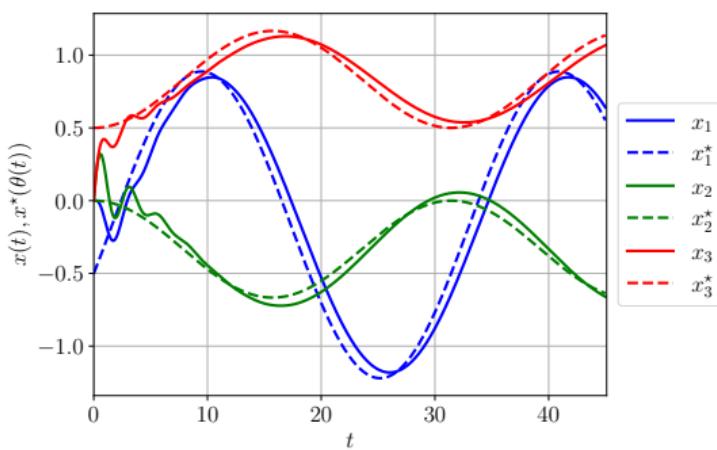
$$\limsup_{t \rightarrow \infty} \|x(t) - x^*(\theta(t))\|_{\mathcal{X}} \leq \frac{\ell}{c^2} \limsup_{t \rightarrow \infty} \|\dot{\theta}(t)\|_{\Theta}$$

- bounded energy input, bounded energy error
- vanishing input, vanishing error
- exponentially vanishing input  $\sim e^{-ht}$ , exponentially vanishing error  $\sim e^{-\min\{c,h\}t}$
- periodic input, periodic error

# Numerical simulations

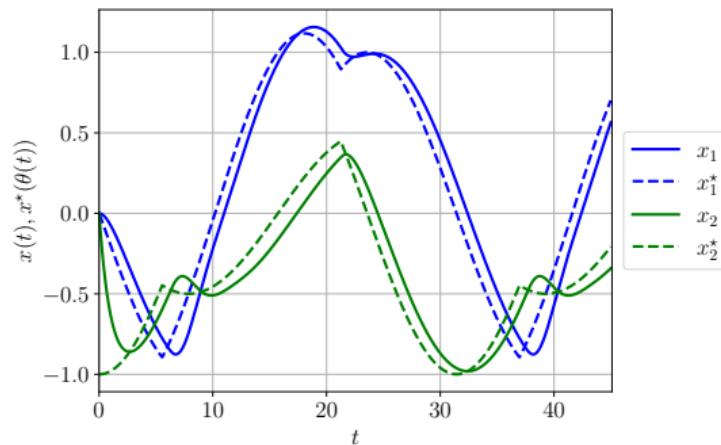
$$\begin{array}{ll} \min_{x \in \mathbb{R}^3} & \frac{1}{2} \|x - r(t)\|_2^2 \\ \text{subj. to} & x_1 + 2x_2 + x_3 = \sin(\omega t), \end{array}$$

$$r(t) = (\sin(\omega t), \cos(\omega t), 1), \omega = 0.2$$

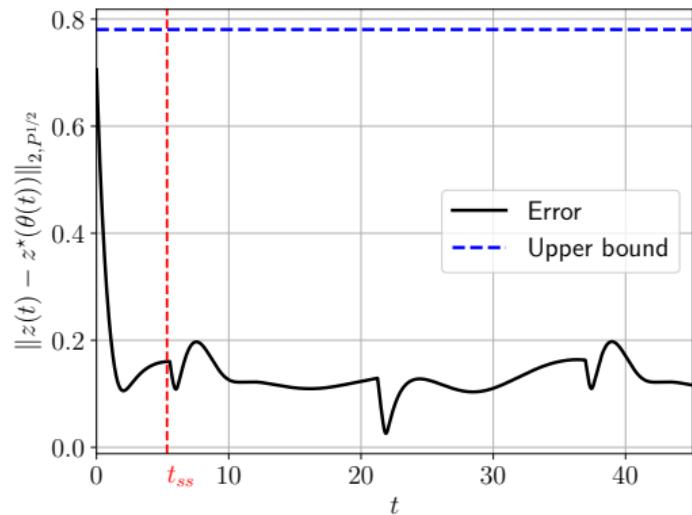
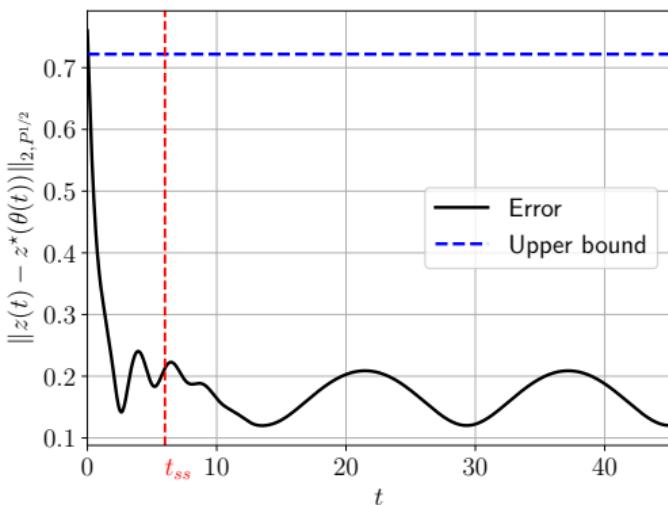


$$\begin{array}{ll} \min_{x \in \mathbb{R}^2} & \frac{1}{2} \|x + r(t)\|_2^2 \\ \text{subj. to} & -x_1 + x_2 \leq \cos(\omega t), \end{array}$$

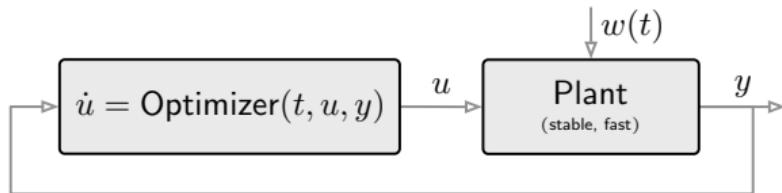
$$r(t) = (\sin(\omega t), \cos(\omega t)), \omega = 0.2$$



## Empirical error versus theoretical upper bound



# Application: Dynamic feedback optimization



## dynamic feedback optimization

online optimization, optimization-based feedback, input/output regulation . . .

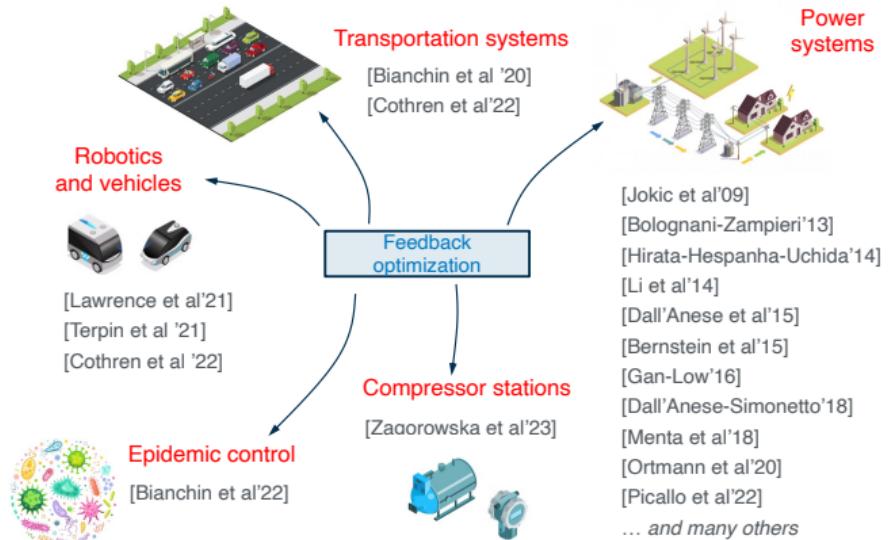
$$\begin{cases} \min & \text{cost}_1(u) + \text{cost}_2(y) \\ \text{subj. to} & y = \text{Plant}(u, w(t)) \end{cases} \implies \begin{cases} \dot{u} = \text{Optimizer}(t, u, y) \\ y = \text{Plant}(u, w(t)) \end{cases}$$

A. Jokic, M. Lazar, and P. van den Bosch. On constrained steady-state regulation: Dynamic KKT controllers. *IEEE Transactions on Automatic Control*, 54(9):2250–2254, 2009. [doi](#)

A. Hauswirth, S. Bolognani, G. Hug, and F. Dorfler. Timescale separation in autonomous optimization. *IEEE Transactions on Automatic Control*, 66(2):611–624, 2021. [doi](#)

G. Bianchin, J. Cortés, J. I. Poveda, and E. Dall'Anese. Time-varying optimization of LTI systems via projected primal-dual gradient flows. *IEEE Transactions on Control of Network Systems*, 9(1):474–486, 2022. [doi](#)

## Some works on feedback optimization



## Example #12: Gradient controller

Fast/stable LTI plant with control input  $u$  and state/measurement disturbance  $w(t)$ :

$$\begin{aligned}\epsilon \dot{x} &= Ax + Bu + Ew(t) && A \text{ Hurwitz} \\ y &= Cx + Dw(t)\end{aligned}$$

In singular perturbation limit as  $\epsilon \rightarrow 0^+$ , steady state map ( $Y_u$  and  $Y_w$ )

$$y = \underbrace{-CA^{-1}B}_{=: Y_u} u + \underbrace{(D - CA^{-1}E)}_{=: Y_w} w$$

### Feedback optimization

equilibrium trajectory  $u^*(t)$  is solution to

$$\min_u \phi(u) + \psi(y(t)) \quad (\nu\text{-strongly convex } \phi, \text{ convex } \psi)$$

$$\text{subj to } y(t) = Y_u u + Y_w w(t)$$

### Gradient controller (as function of measured output):

$$\begin{cases} \dot{u}(t) = -\nabla \phi(u(t)) - Y_u^\top \nabla \psi(y(t)), & u(0) = u_0 \\ \text{fast/stable LTI} \end{cases}$$

## Example #12: Gradient controller

In singular perturbation limit as  $\epsilon \rightarrow 0^+$ ,

$$\mathcal{E}(u, w) = \phi(u) + \psi(Y_u u + Y_w w), \quad (\nu\text{-strongly convex in } u)$$

$$\begin{aligned}\nabla_u \mathcal{E}(u, w) &= \nabla \phi(u) + Y_u^\top \nabla \psi(Y_u u + Y_w w) \\ &= \nabla \phi(u) + Y_u^\top \nabla \psi(y) \quad (\text{no need to measure } w(t) \text{ to compute } \dot{u}(t))\end{aligned}$$

Hence, **gradient controller** is equivalently defined by

$$\dot{u} = F_{\text{GradCtrl}}(u, w) := -\nabla_u \mathcal{E}(u, w) = -\nabla \phi(u) - Y_u^\top \nabla \psi(Y_u u + Y_w w)$$

### Equilibrium tracking for the gradient controller

①  $\text{osLip}_u(F_{\text{GradCtrl}}) \leq -\nu$  (gradient of  $\nu$ -strongly convex function)

②  $\text{Lip}_w(F_{\text{GradCtrl}}) = \ell_w := \|Y_u^\top\| \text{Lip}(\nabla \psi) \|Y_w\|$

$$\limsup_{t \rightarrow \infty} \|u(t) - u^*(t)\| \leq \frac{\ell_w}{\nu^2} \limsup_{t \rightarrow \infty} \|\dot{w}(t)\|$$

## Example #13: Projected gradient controller

### Constrained feedback optimization:

$$\begin{aligned} \min_u \quad & \mathcal{E}(u, w) = \phi(u) + \psi(Y_u u + Y_w w) \quad (\nu \text{ strongly convex}, \ell_u \text{ strongly smooth}, \ell_w) \\ \text{subj. to} \quad & u \in \mathcal{U} \quad (\text{nonempty, closed, convex. } P_{\mathcal{U}} = \text{orthogonal projection}) \end{aligned}$$

### Projected gradient controller (example of proximal gradient dynamics):

$$\dot{u} = F_{\text{PGC}}(u, w) := -u + P_{\mathcal{U}}(u - \gamma \nabla_u \mathcal{E}(u, w))$$

**Equilibrium tracking for projected gradient controller** At  $\gamma = \frac{2}{\nu + \ell_u}$ ,

①  $\text{osLip}_u(F_{\text{PGC}}) \leq -c_{\text{PGC}} := -\frac{2\nu}{\nu + \ell_u}$  (contractivity prox gradient)

②  $\text{Lip}_w(F_{\text{PGC}}) = \ell_{\text{PGC}} := \frac{2}{\nu + \ell_u} \ell_w$

$$\limsup_{t \rightarrow \infty} \|u(t) - u^*(t)\| \leq \frac{\ell_{\text{PGC}}}{c_{\text{PGC}}^2} \limsup_{t \rightarrow \infty} \|\dot{w}(t)\| \quad (\text{eq tracking})$$

## Summary:

- ① from convex optimization to contracting dynamics
- ② tracking-bounds for time-varying contracting systems
- ③ applications to convex optimization and feedback optimization

## Ongoing work and open problems:

- ① contracting predictor-corrector methods
- ② tracking bounds in time-varying norms
- ③ convex but not strongly convex problems

# Outline

## §1. History and resources

## §2. Basic definitions: discrete and continuous-time dynamics on vector spaces

- The linear algebra of matrix norms; see CTDS Chapter 2
- Properties of induced matrix norms and Lipschitz constants

## §3. Example systems

- Constrained, distributed and proximal gradient dynamics
- Continuous-time recurrent neural networks
- Nonlinear dynamics in Lur'e form

## §4. Properties of contracting dynamics

- Equilibria, Lyapunov functions, and Euler discretization
- Incremental input-to-state stability
- Contractivity of interconnected systems
- Additional properties: entrainment, robustness wrt unmodeled dynamics and delays

## §5. Example applications

- Gradient dynamics and Nash equilibria in games
- Time-varying gradient dynamics and feedback optimization
- Recurrent and implicit neural networks

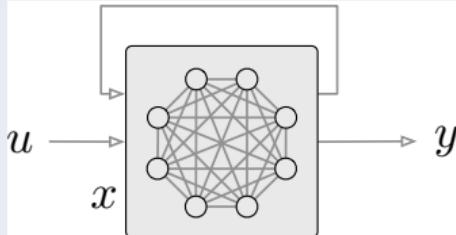
## §6. Generalizations with examples

- G1: Semicontractivity: Primal-dual gradient with redundant constraints
- G2: Contractivity on Riemannian manifolds and the Karcher mean

## §7. Conclusions and future research

## §8. Advanced Topics

- More on semicontractivity: ergodic coefficients and duality
- Network small-gain theorem for Metzler matrices
- More on the Kuramoto-Sakaguchi model and synchronization
- Proof of semicontractivity of saddle matrices
- Proof of Euler discretization theorem
- Non-Euclidean Monotone Operator Theory



$$\dot{x} = -x + \Phi(Ax + Bu + b)$$

(*recurrent NN*)

$$x = \Phi(Ax + Bu + b)$$

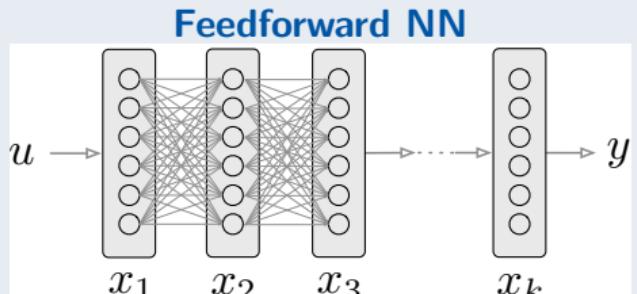
(*implicit NN*)

$$x_{k+1} = (1 - \alpha)x_k + \alpha\Phi(Ax_k + Bu + b) \quad (\text{Euler discretization})$$

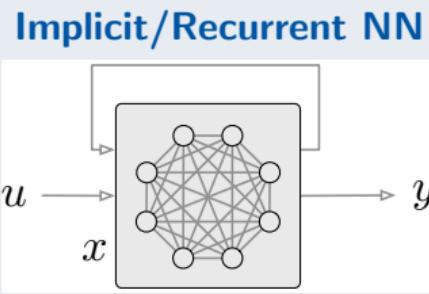
If

$$\mu_\infty(A) < 1 \quad \left( \text{i.e., } a_{ii} + \sum_j |a_{ij}| < 1 \text{ for all } i \right)$$

- recurrent NN is contracting with rate  $1 - \mu_\infty(A)_+$
- implicit NN is well posed
- Euler discretization is contracting with factor  $1 - \frac{1 - \mu_\infty(A)_+}{1 - \min_i(a_{ii})_-}$  at  $\alpha^* = \frac{1}{1 - \min_i(a_{ii})_-}$
- input-state Lipschitz constant  $\text{Lip}_{u \rightarrow x} = \frac{\|B\|_\infty}{1 - \mu_\infty(A)_+}$



$$x_{i+1} = \Phi(A_i x_i + b_i), \quad x_0 = u,$$
$$y = C x_k + d$$



$$x = \Phi(Ax + Bu + b),$$
$$y = Cx + d$$

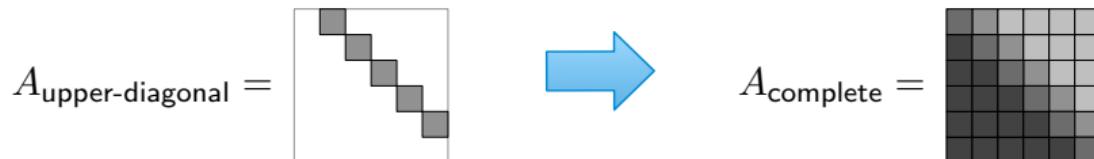
## ML advantages of implicit/equilibrium/fixed point formulation:

- ① bio-inspired
- ② expressivity and ability to model I/O behavior, instead of modalities
- ③ simplicity and memory efficiency
- ④ input-output robustness

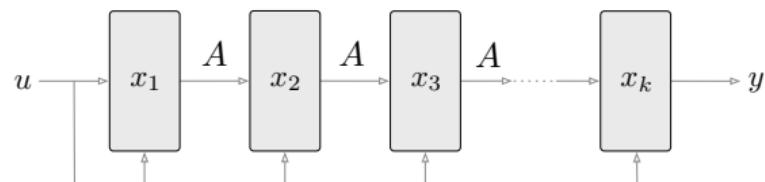
S. Jafarpour, A. Davydov, A. V. Proskurnikov, and F. Bullo. Robust implicit networks via non-Euclidean contractions. In *Advances in Neural Information Processing Systems*, Dec. 2021.

**Motivation #1:** Generalizing FF to fully-connected synaptic matrices

$$x^{i+1} = \Phi(A_i x^i + B_i u + b_i) \iff x = \Phi(Ax + Bu + b), \text{ where } A \text{ has upper diagonal structure.}$$



**Motivation #2:** Weight-tied infinite-depth NN  $\rightarrow$  fixed-point of INN



$$x^{i+1} = \Phi(Ax^i + Bu + b) \implies \lim_{i \rightarrow \infty} x^i = x^* \text{ solution to the INN}$$

- ① S. Bai, J. Z. Kolter, and V. Koltun. Deep equilibrium models. In *Advances in Neural Information Processing Systems*, 2019. URL <https://arxiv.org/abs/1909.01377>
- ② L. El Ghaoui, F. Gu, B. Travacca, A. Askari, and A. Tsai. Implicit deep learning. *SIAM Journal on Mathematics of Data Science*, 3(3):930–958, 2021. doi: 
- ③ E. Winston and J. Z. Kolter. Monotone operator equilibrium networks. In *Advances in Neural Information Processing Systems*, 2020. URL <https://arxiv.org/abs/2006.08591>
- ④ M. Revay, R. Wang, and I. R. Manchester. Lipschitz bounded equilibrium networks. 2020. URL <https://arxiv.org/abs/2010.01732>
- ⑤ A. Kag, Z. Zhang, and V. Saligrama. RNNs incrementally evolving on an equilibrium manifold: A panacea for vanishing and exploding gradients? In *International Conference on Learning Representations*, 2020. URL <https://openreview.net/forum?id=HylpqA4FwS>
- ⑥ K. Kawaguchi. On the theory of implicit deep learning: Global convergence with implicit layers. In *International Conference on Learning Representations*, 2021. URL <https://openreview.net/forum?id=p-NZIuwqhl4>
- ⑦ S. W. Fung, H. Heaton, Q. Li, D. McKenzie, S. Osher, and W. Yin. Fixed point networks: Implicit depth models with Jacobian-free backprop, 2021. URL <https://arxiv.org/abs/2103.12803>. ArXiv e-print

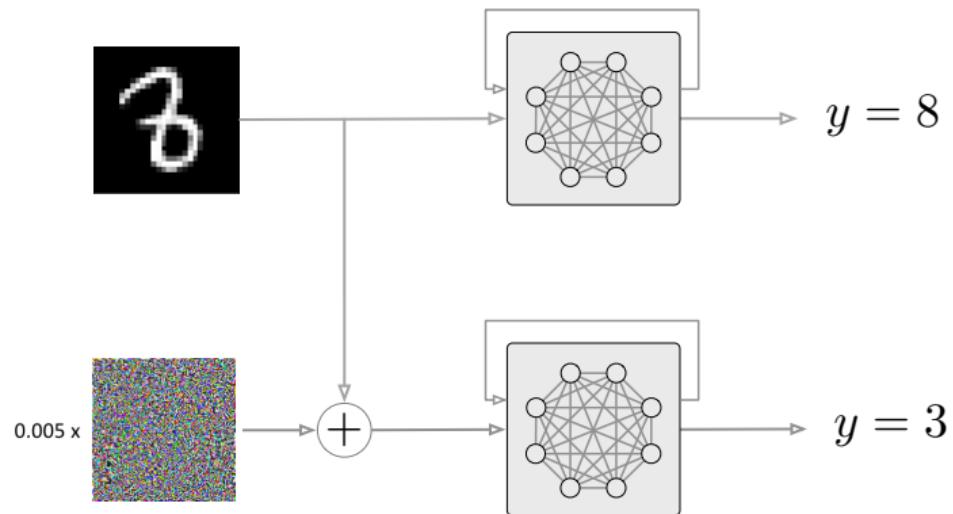
- Training INNs:

- ① loss function  $\mathcal{L}$
- ② training data  $(\hat{u}_i, \hat{y}_i)_{i=1}^N$
- ③ **training optimization problem**

$$\min_{A, B, C, b, x} \sum_{i=1}^N \mathcal{L}(\hat{y}_i, Cx_i + c)$$
$$x_i = \Phi(Ax_i + B\hat{u}_i + b)$$

- Efficient back-propagation through implicit differentiation
- Stochastic gradient descent: at each step solve  $x = \Phi(Ax + Bu + b)$ .

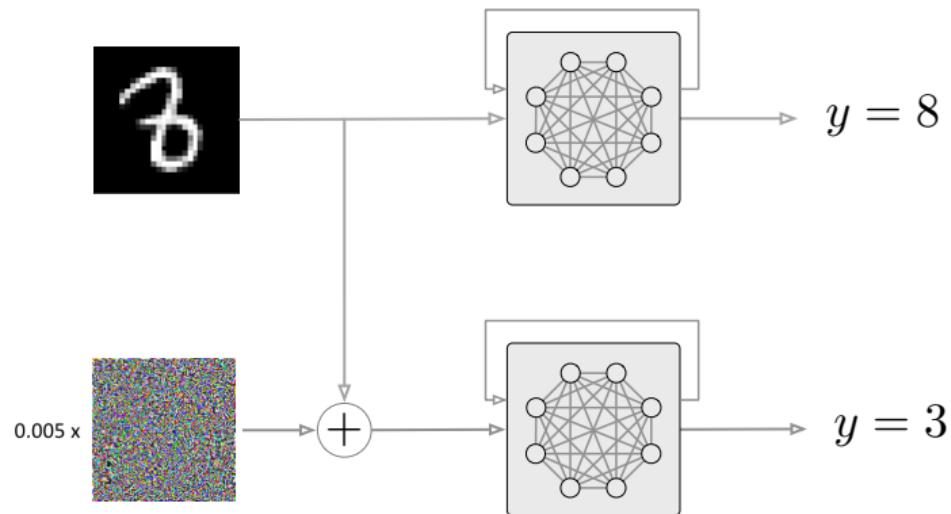
**Adversarial examples:** small input change can cause large output change!



Robustness measures: **input-output Lipschitz constant**

- ①  **$\ell_2$ -norm Lipschitz constant**: not informative in many scenarios
- ②  **$\ell_\infty$ -norm Lipschitz constant**: large-scale input wrt wide-spread perturbations

**Adversarial examples:** small input change can cause large output change!



Robustness measures: **input-output Lipschitz constants**

- ① NP-hard to compute exactly
- ② Approximations provide only coarse certified robustness guarantees

Training optimization problem:

$$\begin{aligned} \min_{A,B,C,b} \quad & \sum_{i=1}^N \mathcal{L}(\hat{y}_i, Cx_i + c) + \lambda \text{Lip}_{u \rightarrow y} \\ & x_i = \Phi(Ax_i + Bu_i + b) \\ & \mu_\infty(A) \leq \gamma \end{aligned}$$

- $\lambda \geq 0$  is a regularization parameter
- $\gamma < 1$  is a hyperparameter

**Parametrization of  $\mu_\infty$  constraint:**

$$\mu_\infty(A) \leq \gamma \iff \exists T \text{ s.t. } A = T - \text{diag}(|T| \mathbf{1}_n) + \gamma I_n.$$

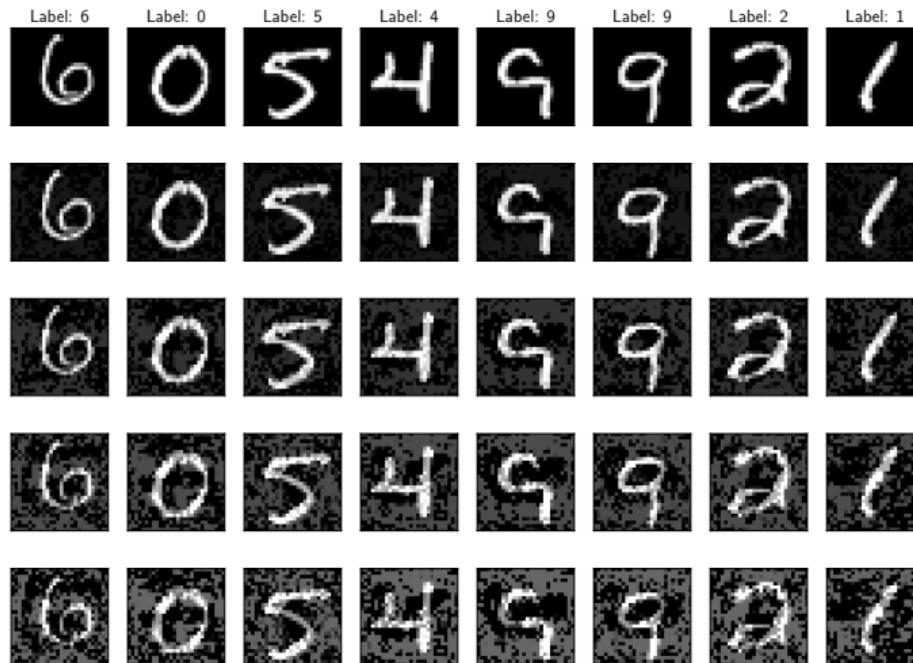
Synaptic matrix  $A$  encodes interactions between neurons



- $A_{\text{dropout}}$  is a principal submatrix of  $A_{\text{complete}}$
- $\mu_\infty(A_{\text{dropout}}) \leq \mu_\infty(A_{\text{complete}})$ 
  - Well-posedness of original INN implies well-posedness of INN with subset of neurons
  - Promotes *compression* and *sparsity* of overparametrized models

## Numerical Experiments

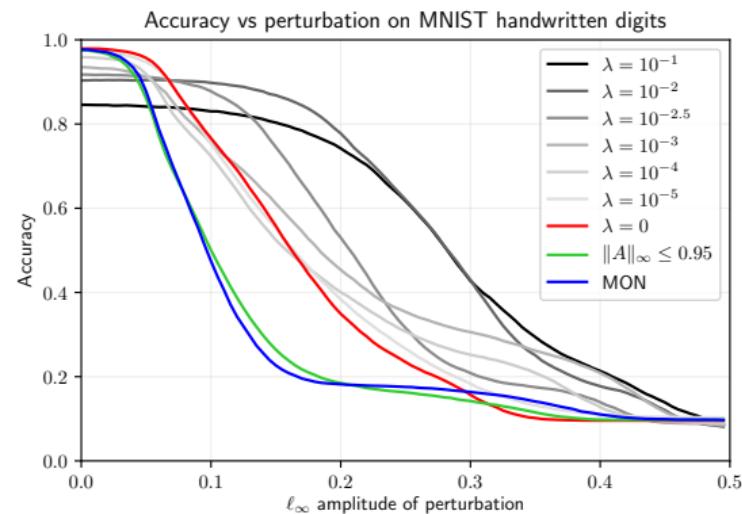
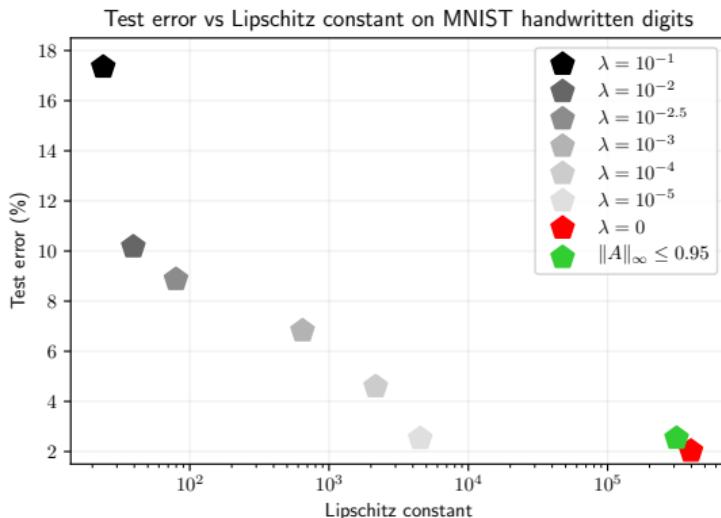
- MNIST handwritten digit dataset (60K+10K, 28x28, grayscale)
- implicit neural network order:  $n = 100$



# Numerical Experiments

## Robustness of INNs

### Tradeoff between **accuracy** and **robustness**



- Pareto-optimal curve

- Clean performance vs. robustness

# Outline

## §1. History and resources

## §2. Basic definitions: discrete and continuous-time dynamics on vector spaces

- The linear algebra of matrix norms; see CTDS Chapter 2
- Properties of induced matrix norms and Lipschitz constants

## §3. Example systems

- Constrained, distributed and proximal gradient dynamics
- Continuous-time recurrent neural networks
- Nonlinear dynamics in Lur'e form

## §4. Properties of contracting dynamics

- Equilibria, Lyapunov functions, and Euler discretization
- Incremental input-to-state stability
- Contractivity of interconnected systems
- Additional properties: entrainment, robustness wrt unmodeled dynamics and delays

## §5. Example applications

- Gradient dynamics and Nash equilibria in games
- Time-varying gradient dynamics and feedback optimization
- Recurrent and implicit neural networks

## §6. Generalizations with examples

- G1: Semicontractivity: Primal-dual gradient with redundant constraints
- G2: Contractivity on Riemannian manifolds and the Karcher mean

## §7. Conclusions and future research

## §8. Advanced Topics

- More on semicontractivity: ergodic coefficients and duality
- Network small-gain theorem for Metzler matrices
- More on the Kuramoto-Sakaguchi model and synchronization
- Proof of semicontractivity of saddle matrices
- Proof of Euler discretization theorem
- Non-Euclidean Monotone Operator Theory



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

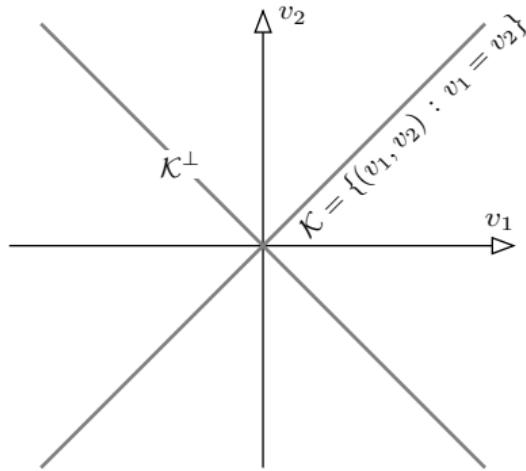
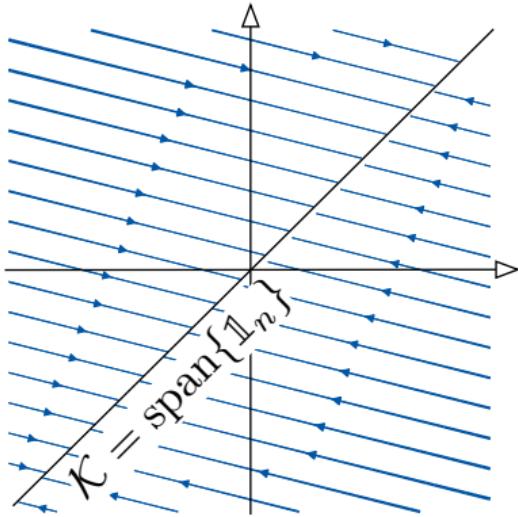
- **Invariance property:** for all  $x, y \in \mathbb{R}^n$  and  $\alpha \in \mathbb{R}$ ,

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

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

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

**systems with invariance or conservation properties are not strongly contracting**



For  $\dot{x} = -Lx$

- ①  $\mathcal{K} = \text{span}\{\mathbf{1}_n\}$
- ②  $x_{\text{avg}} = \frac{1}{n} \mathbf{1}_n^\top x$  along  $\mathcal{K}$
- ③  $x_\perp = x - x_{\text{avg}} \mathbf{1}_n \in \mathcal{K}^\perp$

decomposition: perpendicular dynamics + reconstruction equation:

$$\dot{x}_\perp := -\Pi_n L x_\perp \quad \in \mathbb{1}_m^\perp$$

$$\dot{x}_{\text{avg}} = -\frac{1}{n} \mathbf{1}_n^\top L x_\perp \quad \in \mathbb{R}$$

# Systems with symmetry and their reduced dynamics

Model	Symmetry	Reduced space
Laplacian	$\dot{x} = \mathsf{F}_{\text{Laplacian}}(x) := -Lx$ $\mathsf{F}_{\text{Laplacian}}(x + \alpha \mathbb{1}_n) = \mathsf{F}_{\text{Laplacian}}(x)$	$\mathbb{R}^n / \mathbb{R}$
Kuramoto-Sakaguchi	$\dot{\theta} = \mathsf{F}_{\text{KS}}(\theta) := \omega + B\mathcal{A}(\sin(B^\top \theta - \varphi) + \sin(\varphi))$ $\mathsf{F}_{\text{KS}}(\theta + \alpha \mathbb{1}_n) = \mathsf{F}_{\text{KS}}(\theta)$	$\mathbb{T}^n / \mathbb{S} \rightarrow \mathbb{R}^n / \mathbb{R}$
Primal-dual gradient with $k$ redundant constraints	$\begin{bmatrix} \dot{x} \\ \dot{\lambda} \end{bmatrix} = \mathsf{F}_{\text{PDG}} \left( \begin{bmatrix} x \\ \lambda \end{bmatrix} \right) := \begin{bmatrix} -\nabla f(x) - A^\top \lambda \\ Ax - b \end{bmatrix}$ $\mathsf{F}_{\text{PDG}} \left( \begin{bmatrix} x \\ \lambda \end{bmatrix} + \begin{bmatrix} 0 \\ \alpha \nu \end{bmatrix} \right) = \mathsf{F}_{\text{PDG}} \left( \begin{bmatrix} x \\ \lambda \end{bmatrix} \right) \quad \text{for all } \nu \in \ker(A^\top)$	$\mathbb{R}^{n+m} / \mathbb{R}^k$

If  $\mathsf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is invariant under  $\mathbb{R}^k$  translations, then

**perpendicular dynamics**  $\mathsf{F}_\perp : \mathbb{R}^n / \mathbb{R}^k \rightarrow \mathbb{R}^n / \mathbb{R}^k$  is well defined  
full solution obtained via **reconstruction equation**

A **seminorm** is a function  $\|\cdot\| : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$  such that  $\forall a \in \mathbb{R}$  and  $\forall x, y \in \mathbb{R}^n$

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

**kernel is a subspace**  $\mathcal{K} = \{x \in \mathbb{R}^n \text{ such that } \|x\| = 0\}$

**seminorm is invariant**  $\|x + \kappa\| = \|x\| \text{ for all } \kappa \in \mathcal{K}$

**seminorm on  $\mathbb{R}^n$  with kernel  $\mathcal{K} \sim \mathbb{R}^k$**   $\iff$  **norm on  $\mathcal{K}^\perp \sim \mathbb{R}^n / \mathbb{R}^k$**

- **matrix seminorm** is  $\|A\| = \max_{\substack{\|v\|=1 \\ v \perp \mathcal{K}}} \|Av\|$
- **matrix log seminorm**  $\mu_{\|\cdot\|}(A) = \lim_{h \rightarrow 0^+} \frac{\|I_n + hA\| - 1}{h}$
- $F$  is **infinitesimally semicontracting** if  $\sup_x \mu_{\|\cdot\|}(DF(x)) \leq -c$

**$F$  is inf semicontracting on  $\mathbb{R}^n$**   $\iff$   **$F_\perp$  is inf contracting on  $\mathbb{R}^n / \mathbb{R}^k$**

$\ell_2$  seminorm with kernel  $\mathcal{K}$   $\iff P = P^\top \succeq 0$  and  $\ker(P) = \mathcal{K}$

$$\|x\|_{2,P^{1/2}}^2 := x^\top Px$$

consensus  $\ell_2$  seminorm with  $\mathcal{K} = \text{span}\{\mathbb{1}_n\}$

$$\|x\|_{2,\Pi_n}^2 := \sum_{i,j} (x_i - x_j)^2,$$

$$\Pi_n = I_n - \mathbb{1}_n \mathbb{1}_n^\top / n = \text{orthogonal projection onto } \mathcal{K}^\perp = \text{span}\{\mathbb{1}_n\}^\perp$$

Given  $\ell_2$  seminorm defined by  $P = P^\top \succeq 0$  and  $\ker(P) = \mathcal{K}$ ,

**semicontractivity LMIs** for  $A\mathcal{K} \subset \mathcal{K}$

$$\|A\|_{2,P} \leq \ell \iff A^\top PA \preceq \ell^2 P$$

$$\mu_{2,P}(A) \leq \ell \iff A^\top P + AP \preceq 2\ell P$$

## Example #14: Laplacian flow

Laplacian flow

$$\dot{x} = F_{\text{Laplacian}}(x) := -Lx$$

where  $L$  is the Laplacian of a weighted undirected graph

$F_{\text{Laplacian}}$  is semicontracting wrt  $\|\cdot\|_{2,\Pi_n}$  with rate  $\lambda_2$

- $L \succeq \lambda_2 \Pi_n$
- $\Pi_n L = L \Pi_n = L$
- $\Pi_n (-L) + (-L) \Pi_n \preceq -2\lambda_2 \Pi_n$
- $\text{osLip}_{2,\Pi_n}(F_{\text{Laplacian}}) := \mu_{2,\Pi_n}(-L) = -\lambda_2$

## Example #15: Kuramoto-Sakaguchi model and synchronization

graph: incidence matrix  $B$ , weight matrix  $A$ , max degree  $d_{\max}$  and algebraic connectivity  $\lambda_2$   
natural frequency  $\omega$ , frustration parameter  $\varphi$

$$\dot{\theta}_i = \omega_i + \sum_j a_{ij} \sin(\theta_i - \theta_j + \varphi_{ij})$$

$F_{KS}$  is locally infinitesimally semicontracting wrt  $\|\cdot\|_{2,\Pi_n}$

### Proof sketch

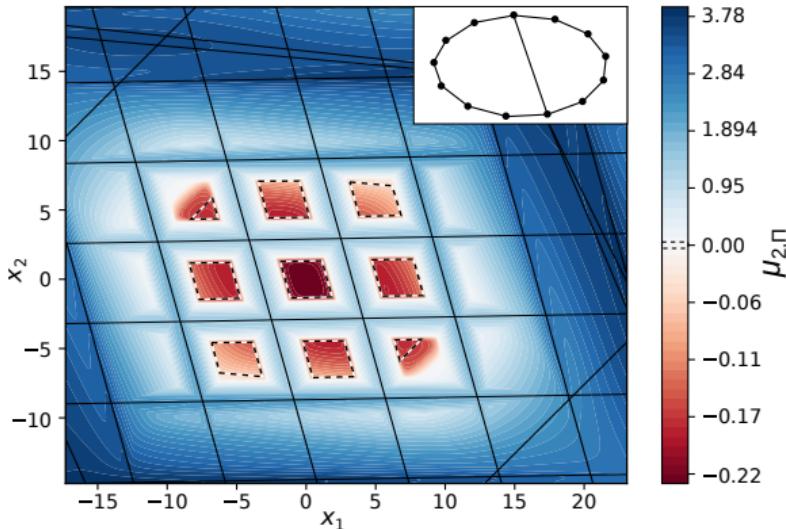
$$\dot{\theta} = \omega + \cos(\varphi) \underbrace{BA \sin(B^\top \theta)}_{F_{\text{odd}}(\theta)} - \sin(\varphi) \underbrace{BA(1 - \cos(B^\top \theta))}_{F_{\text{even}}(\theta)}$$

For  $\theta \in \mathbb{T}^n$ , define  $\gamma(\theta) = \max_{(i,j)} |\theta_i - \theta_j|$

$$\begin{aligned} \mu_{2,\Pi_n}(DF_{\text{odd}}(\theta)) &= \mu_{2,\Pi_n}(-L(\theta)) \leq -\lambda_2 \cos(\gamma(\theta)) \quad (\text{Jacobian} = -\text{Laplacian}, L = BAB^\top) \\ \mu_{2,\Pi_n}(DF_{\text{even}}(\theta)) &\leq d_{\max} \sin(\gamma(\theta)) \end{aligned}$$

$$\implies \mu_{2,\Pi_n}(DF_{KS}(\theta)) < 0 \text{ locally in } \left\{ \theta \in \mathbb{T}^n \mid \gamma(\theta) < \arctan \frac{\lambda_2}{d_{\max} \tan(\varphi)} \right\}$$

## Local semicontractivity of KS system, inside cells



$\mu_{2,II_n}(DF_{KS}(\theta))$  for  $\theta$  in two-dimensional slice of  $\mathbb{R}^{13}$   
model parameters: frustration  $\varphi = 0.01$ , graph in inset

## Example #16: Primal-dual gradient dynamics with redundant constraints

strongly convex function  $f$

constraint matrix  $A$

$$\text{s.t. } 0 \prec \nu_{\min} I_n \preceq \text{Hess } f \preceq \nu_{\max} I_n$$

$$\text{s.t. } 0 \preceq a_{\min} \Pi_A \preceq AA^\top \preceq a_{\max} I_m$$

where  $\Pi_A$  is the orthogonal projection onto  $\text{Im}(A)$   
i.e., redundant constraints are allowed

**primal-dual gradient dynamics:**

$$\begin{bmatrix} \dot{x} \\ \dot{\lambda} \end{bmatrix} = \mathsf{F}_{\text{PDG}}(x, \lambda) := \begin{bmatrix} -\nabla f(x) - A^\top \lambda \\ Ax - b \end{bmatrix}$$

$\mathsf{F}_{\text{PDG}}$  is infinitesimally semicontracting wrt  $\|\cdot\|_{2,P^{1/2}}$  with rate  $c$

$$P = \begin{bmatrix} I_n & \alpha A^\top \\ \alpha A & \Pi_A \end{bmatrix} \text{ and } \alpha = \frac{1}{2} \min \left\{ \frac{1}{\nu_{\max}}, \frac{\nu_{\min}}{a_{\max}} \right\}, \quad \text{and} \quad c = \frac{1}{4} \min \left\{ \frac{a_{\min}}{\nu_{\max}}, \frac{a_{\min}}{a_{\max}} \nu_{\min} \right\}$$

$$\text{For each } \nu_{\min} I_n \preceq Q \preceq \nu_{\max} I_n, \quad \begin{bmatrix} -Q & -A^\top \\ A & 0 \end{bmatrix}^\top P + P \begin{bmatrix} -Q & -A^\top \\ A & 0 \end{bmatrix} \preceq -2cP$$

# Outline

## §1. History and resources

## §2. Basic definitions: discrete and continuous-time dynamics on vector spaces

- The linear algebra of matrix norms; see CTDS Chapter 2
- Properties of induced matrix norms and Lipschitz constants

## §3. Example systems

- Constrained, distributed and proximal gradient dynamics
- Continuous-time recurrent neural networks
- Nonlinear dynamics in Lur'e form

## §4. Properties of contracting dynamics

- Equilibria, Lyapunov functions, and Euler discretization
- Incremental input-to-state stability
- Contractivity of interconnected systems
- Additional properties: entrainment, robustness wrt unmodeled dynamics and delays

## §5. Example applications

- Gradient dynamics and Nash equilibria in games
- Time-varying gradient dynamics and feedback optimization
- Recurrent and implicit neural networks

## §6. Generalizations with examples

- G1: Semicontractivity: Primal-dual gradient with redundant constraints
- G2: Contractivity on Riemannian manifolds and the Karcher mean

## §7. Conclusions and future research

## §8. Advanced Topics

- More on semicontractivity: ergodic coefficients and duality
- Network small-gain theorem for Metzler matrices
- More on the Kuramoto-Sakaguchi model and synchronization
- Proof of semicontractivity of saddle matrices
- Proof of Euler discretization theorem
- Non-Euclidean Monotone Operator Theory

Contraction theory on Riemannian manifolds originates in

W. Lohmiller and J.-J. E. Slotine. On contraction analysis for non-linear systems. *Automatica*, 34(6):683–696, 1998. 

A formal coordinate-free analysis (with connection to monotone operators) is given in

J. W. Simpson-Porco and F. Bullo. Contraction theory on Riemannian manifolds. *Systems & Control Letters*, 65:74–80, 2014. 

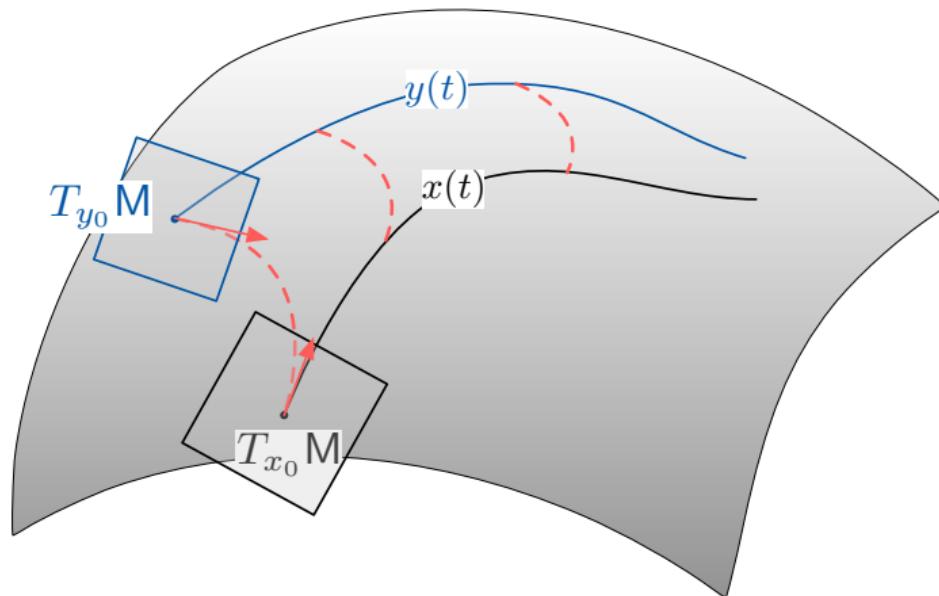
In the differential geometry literature, geodesically monotonic vector fields are studied by

S. Z. Németh. Geodesic monotone vector fields. *Lobachevskii Journal of Mathematics*, 5:13–28, 1999. URL <http://mi.mathnet.ru/eng/ljm145>

J. X. Da Cruz Neto, O. P. Ferreira, and L. R. Lucambio Pérez. Contributions to the study of monotone vector fields. *Acta Mathematica Hungarica*, 94(4):307–320, 2002. 

J. H. Wang, G. López, V. Martín-Márquez, and C. Li. Monotone and accretive vector fields on Riemannian manifolds. *Journal of Optimization Theory and Applications*, 146(3):691–708, 2010. 

Assume: existence and uniqueness of geodesic curve  $\gamma(t) = x \#_t y$  between each  $(x, y)$   
F **contracting** if geodesic distances from  $x$  to  $y$  diminishes along the flow of F



**integral test:** the inner product between  $F$  and the geodesic velocity vector  $\gamma'$  at  $x$  and  $y$

**differential test:** condition on covariant differential of  $F$

Given vector field  $F$  on a Riemannian manifold  $(M, \mathbb{G})$  and  $c > 0$ , equivalent statements:

- ① **integral condition:** for each  $x, y \in M$  and geodesic  $\gamma : [0, 1] \rightarrow M$  with  $\gamma(0) = x, \gamma(1) = y$ ,

$$\langle\langle F(y), \gamma'(1) \rangle\rangle_{\mathbb{G}} - \langle\langle F(x), \gamma'(0) \rangle\rangle_{\mathbb{G}} \leq -c d_{\mathbb{G}}(x, y)^2$$

or, equivalently, using the parallel transport map  $P_{y \rightarrow x} : T_y M \rightarrow T_x M$ ,

$$\langle\langle P_{y \rightarrow x} F(y) - F(x), \gamma'(0) \rangle\rangle_{\mathbb{G}} \leq -c d_{\mathbb{G}}(x, y)^2$$

- ② **differential condition:** for all  $v_x \in T_x M$

$$\langle\langle \nabla_{v_x} F(x), v_x \rangle\rangle_{\mathbb{G}} \leq -c \|v_x\|_{\mathbb{G}}^2,$$

where  $\nabla F$  is covariant derivative. In components, generalized Demidovich condition:

$$\mathbb{G}(x) D F(x) + D F(x)^\top \mathbb{G}(x) + \mathcal{L}_F \mathbb{G}(x) \preceq -2c \mathbb{G}(x)$$

- ③ **trajectory condition:** for all solutions  $x(\cdot), y(\cdot)$

$$D^+ d_{\mathbb{G}}(x(t), y(t)) \leq -c d_{\mathbb{G}}(x(t), y(t))$$

## Example #17: Natural gradient dynamics on Riemannian manifolds

Given Riemannian manifold  $(M, \mathbb{G})$ ,

a function  $f : M \rightarrow \mathbb{R}$  is  **$\nu$ -strongly geodesically convex** if, for each  $x, y$ ,

- ①  $f(x\#_t y) \leq (1 - \chi)f(x) + \chi f(y) - \frac{1}{2}\nu\chi(1 - \chi)d_{\mathbb{G}}(x, y)^2$
- ② (if  $f$  is twice differentiable)  $\text{Hess } f(x) \succeq \nu\mathbb{G}(x)$

### natural gradient dynamics

$$\dot{x} = F_{\mathbb{G}}(x) := -\mathbb{G}(x)^{-1}\nabla f(x)$$

$F_{\mathbb{G}}$  is infinitesimally contracting wrt  $\mathbb{G}$  with rate  $\nu$

unique globally exp stable point is global minimum

Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  smooth, strongly convex

**natural gradient on  $(\mathbb{R}^n, \text{Hess}(f))$  = Newton's continuous-time method  
infinitesimally contracting with rate 1**

## Example #18: Rosenbrock function

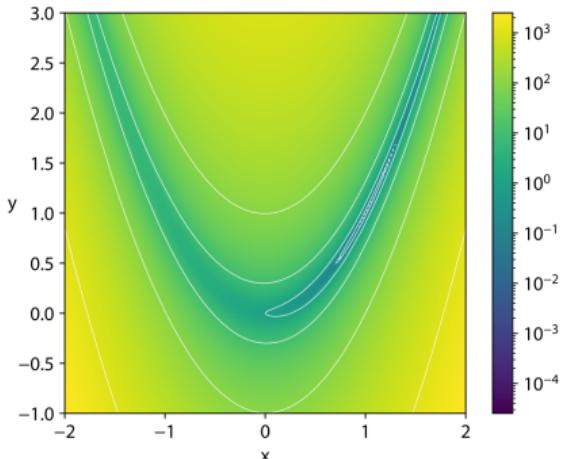
$$f_{\text{Rsnbrck}}(x_1, x_2) = 100(x_1^2 - x_2)^2 + (x_1 - 1)^2$$

is 2-strongly geodesically convex wrt

$$\mathbb{G}(x) = \begin{bmatrix} 400x_1^2 + 1 & -200x_1 \\ -200x_1 & 100 \end{bmatrix}$$

and natural gradient is 2-strongly contracting

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = -\mathbb{G}(x)^{-1} \nabla f_{\text{Rsnbrck}} = -2 \begin{bmatrix} x_1 - 1 \\ x_1^2 - 2x_1 + x_2 \end{bmatrix}$$



contour plot for  $f_{\text{Rsnbrck}}$   
long, shallow parabolic valley  
global minimum (1, 1)

## Example #19: Karcher mean on manifold of positive-definite matrices

$\mathbb{S}_{>0}^n$  = manifold of symmetric positive-definite matrices with

$$\mathbb{G}(X)(\xi, \eta) = \text{trace}(X^{-1}\xi X^{-1}\eta) \quad (\text{Riemannian metric})$$

$$X \#_t Y = X^{1/2} (X^{-1/2} Y X^{-1/2})^t X^{1/2} \quad (\text{geodesic})$$

$$d_{\mathbb{G}}(X, Y) = \|\log(X^{-1/2} Y X^{-1/2})\|_{\text{F}} \quad (\text{geodesic distance})$$

Given dataset  $\{A_i \in \mathbb{S}_{>0}^n\}_{i \in \{1, \dots, N\}}$ , define **Karcher loss function**

$$f_{\text{Karcher}}(X) = \sum_{i=1}^N d_{\mathbb{G}}(X, A_i)^2$$

$f_{\text{Karcher}}$  is  $2N$ -strongly geodesically convex on  $\mathbb{S}_{>0}^n$

**Karcher mean** = global minimizer = globally exp stable point of natural gradient

# Outline

## §1. History and resources

## §2. Basic definitions: discrete and continuous-time dynamics on vector spaces

- The linear algebra of matrix norms; see CTDS Chapter 2
- Properties of induced matrix norms and Lipschitz constants

## §3. Example systems

- Constrained, distributed and proximal gradient dynamics
- Continuous-time recurrent neural networks
- Nonlinear dynamics in Lur'e form

## §4. Properties of contracting dynamics

- Equilibria, Lyapunov functions, and Euler discretization
- Incremental input-to-state stability
- Contractivity of interconnected systems
- Additional properties: entrainment, robustness wrt unmodeled dynamics and delays

## §5. Example applications

- Gradient dynamics and Nash equilibria in games
- Time-varying gradient dynamics and feedback optimization
- Recurrent and implicit neural networks

## §6. Generalizations with examples

- G1: Semicontractivity: Primal-dual gradient with redundant constraints
- G2: Contractivity on Riemannian manifolds and the Karcher mean

## §7. Conclusions and future research

## §8. Advanced Topics

- More on semicontractivity: ergodic coefficients and duality
- Network small-gain theorem for Metzler matrices
- More on the Kuramoto-Sakaguchi model and synchronization
- Proof of semicontractivity of saddle matrices
- Proof of Euler discretization theorem
- Non-Euclidean Monotone Operator Theory

- ① **Lotka-Volterra population dynamics** (Lotka, 1920; Volterra, 1928):  
 $\ell_1$ - semiglobally strongly contracting (after a rescaling change of coordinates)
- ② **Matrosov-Bellman interconnected stable systems** (Bellman, 1962; Matrosov, 1962):  
strongly contracting wrt composite norm
- ③ **Kuramoto coupled oscillators** (Kuramoto, 1975):  
strongly semicontracting wrt  $(\ell_2, \Pi_n)$  norm, in neightb'd of each phase-cohesive equilibrium
- ④ **Yorke multigroup SIS epidemic model** (Lajmanovich and Yorke, 1976):  
equilibrium contracting wrt weighted  $\ell_1/\ell_\infty$  norms (at disease-free and endemic eq.)
- ⑤ **Hopfield and cellular neural networks** (Hopfield, 1982):  
 $\ell_1/\ell_\infty$ -strongly contracting
- ⑥ **Daganzo cell transmission model for traffic networks** (Daganzo, 1994):  
 $\ell_1$ -weakly contracting, when the dynamics is monotone
- ⑦ **Chua's diffusively-coupled dynamical systems** (Wu and Chua, 1995):  
strongly semi-contracting wrt  $(2, p)$  tensor norm on  $\mathbb{R}^n \otimes \mathbb{R}^k$
- ⑧ ...

**contractivity = robust computationally-friendly stability**

fixed point theory + Lyapunov stability theory + geometry of metric spaces

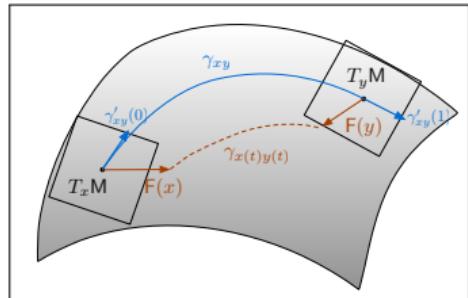


	Lyapunov Theory	Contraction Theory for Dynamical Systems
existence of equilibrium Lyapunov function inputs	F admits global Lyapunov function assumed arbitrary ISS via $\mathcal{KL}$ and $\mathcal{L}$ functions	F is strongly contracting implied + computational methods $\ x - x^*\ $ and $\ F(x)\ $ iISS via explicit constants

search for contraction properties  
design engineering systems to be contracting

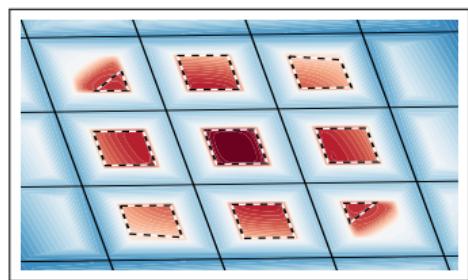
## Theoretical frontiers

- higher order contraction
- relationship with monotone operator theory
- metric spaces
- computational methods



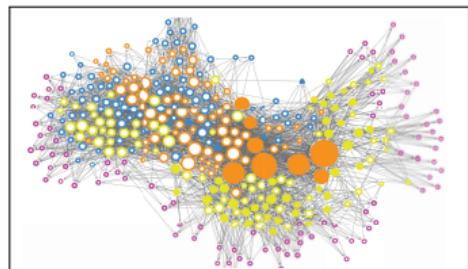
**Limitations:** not all stable systems are contractive:

- Lyapunov-diagonally-stable networks
- multistable and locally contracting systems
- biochemical networks
- control contraction design



## Application to control and learning

- ① control: optimization-based control design
- ② ML: implicit models and energy-based learning
- ③ neuroscience: robust dynamical modeling



# Outline

## §1. History and resources

## §2. Basic definitions: discrete and continuous-time dynamics on vector spaces

- The linear algebra of matrix norms; see CTDS Chapter 2
- Properties of induced matrix norms and Lipschitz constants

## §3. Example systems

- Constrained, distributed and proximal gradient dynamics
- Continuous-time recurrent neural networks
- Nonlinear dynamics in Lur'e form

## §4. Properties of contracting dynamics

- Equilibria, Lyapunov functions, and Euler discretization
- Incremental input-to-state stability
- Contractivity of interconnected systems
- Additional properties: entrainment, robustness wrt unmodeled dynamics and delays

## §5. Example applications

- Gradient dynamics and Nash equilibria in games
- Time-varying gradient dynamics and feedback optimization
- Recurrent and implicit neural networks

## §6. Generalizations with examples

- G1: Semicontractivity: Primal-dual gradient with redundant constraints
- G2: Contractivity on Riemannian manifolds and the Karcher mean

## §7. Conclusions and future research

## §8. Advanced Topics

- More on semicontractivity: ergodic coefficients and duality
- Network small-gain theorem for Metzler matrices
- More on the Kuramoto-Sakaguchi model and synchronization
- Proof of semicontractivity of saddle matrices
- Proof of Euler discretization theorem
- Non-Euclidean Monotone Operator Theory



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

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

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

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

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

## Example #20: Averaging and dynamical flow systems

### Prototypical dynamics with invariance and conservation

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  $\text{span}\{\mathbb{1}_n\}$

**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

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

$\tau_1(A) < 1$  under mild connectivity conditions

$\tau_p(A)$  also defined for general  $p \in [1, \infty]$

How is  $\tau_1$  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. doi:10.1007/BF01043417

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

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

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

$$d_{\max\text{-}\min}(Ax) \leq \tau_1(A) d_{\max\text{-}\min}(x), \quad \text{where } d_{\max\text{-}\min}(x) = \max_i \{x_i\} - \min_j \{x_j\}$$

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

Given  $\pi, \sigma$  in the simplex  $\Delta_n$ , total variation distance:

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

Why is the same  $\tau_1$  relevant in both cases?

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

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

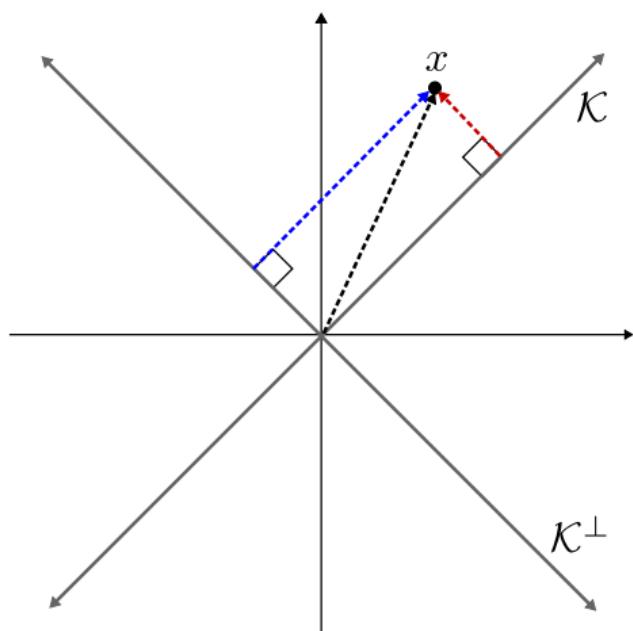
The *kernel* is the vector space:

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

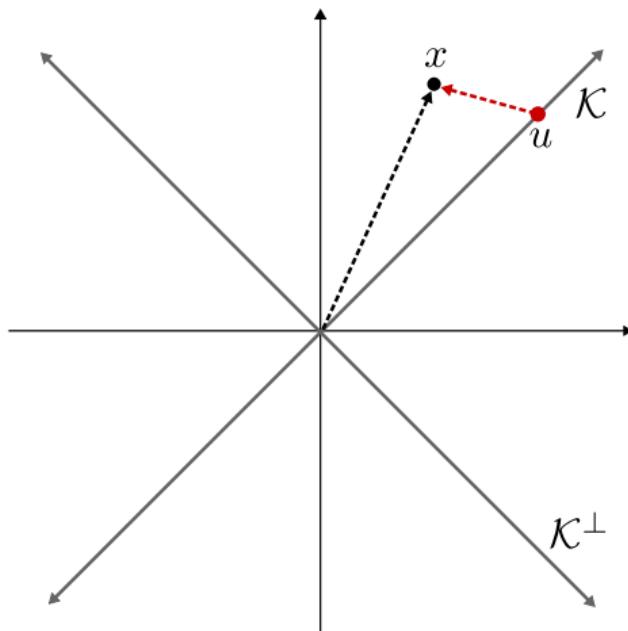
We focus on *consensus seminorms*, where  $\mathcal{K} = \text{span}\{\mathbf{1}_n\}$ .

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

## Projection seminorms



## Distance seminorms



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

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

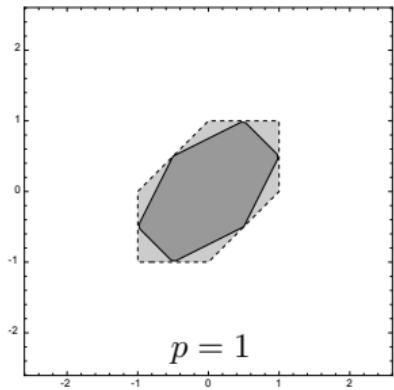
## Projection and distance-based consensus seminorms ( $\mathcal{K} = \text{span}\{\mathbb{1}_n\}$ )

	$\ x\ _{\text{proj},p}$	$\ x\ _{\text{dist},p}$
$\ell_1$	$\sum_{i=1}^n  x_i - x_{\text{avg}} $	$\sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} x_{(i)} - \sum_{j=\lceil \frac{n}{2} \rceil + 1}^n x_{(j)}$
$\ell_2$	$\sqrt{\frac{1}{n} \sum_{i,j} (x_i - x_j)^2}$	$\sqrt{\frac{1}{n} \sum_{i,j} (x_i - x_j)^2}$
$\ell_\infty$	$\max_i  x_i - x_{\text{avg}} $	$\frac{1}{2} (x_{(1)} - x_{(n)})$

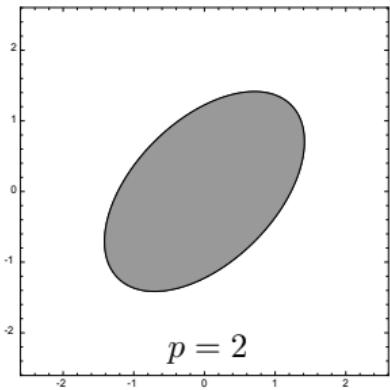
where we have sorted  $x_{(1)} \geq x_{(2)} \geq \dots \geq x_{(n)}$

Therefore

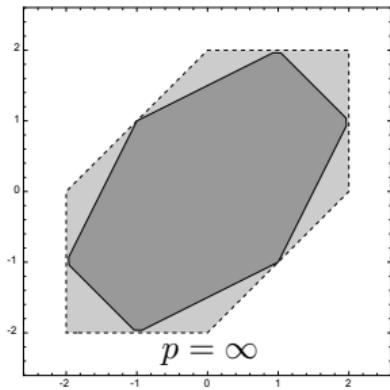
$$d_{\text{max-min}}(x) = 2\|x\|_{\text{dist},\infty} \quad \text{and} \quad d_{\text{TV}}(\pi, \sigma) = \|\pi - \sigma\|_{\text{proj},1}$$



$$p = 1$$



$$p = 2$$



$$p = \infty$$

**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  $(x_1, x_2, x_3 = 0)$ .

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

**Induced matrix seminorm:** function  $\|\cdot\| : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}_{\geq 0}$  where

$$\|A\| = \max_{\substack{\|x\| \leq 1 \\ x \perp \mathcal{K}}} \|Ax\|, \quad \forall A \in \mathbb{R}^{n \times n}$$



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

## Properties of dual and induced norms

- ①  $\ell_p$  and  $\ell_q$  norms are dual, for  $1/p + 1/q = 1$

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

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

## Properties of dual and induced seminorms

- ①  $\ell_p$ -distance and  $\ell_q$ -projection seminorms are dual, for  $1/p + 1/q = 1$

$$\|\cdot\|_{\text{dist},p} = (\|\cdot\|_{\text{proj},q})_* \quad \|\cdot\|_{\text{proj},q} = (\|\cdot\|_{\text{dist},p})_*$$

- ② dual seminorm satisfies (sharp) *Markov inequality*:  $x^\top \Pi_\perp y \leq \|x\|_{\text{dist},p} \|y\|_{\text{proj},q}$
- ③ equality between dual induced seminorms:  $\|A\|_{\text{dist},p} = \|A^\top\|_{\text{proj},q}$
- 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

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

# How Markov and Banach's results meet

## Classical Property of Averaging Systems

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

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

## Classical Property of Markov Chains

Given row-stochastic  $A \in \mathbb{R}^{n \times n}$  and  $\pi, \sigma$  in the simplex  $\Delta_n$ :

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

## 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 primal-dual dynamics with redundant constraints
  - ④ local contractivity of Kuramoto-Sakaguchi models

### Future work

consider the set of undirected, unweighted connected graphs + selfloops

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

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

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

or **prove** that it does not exist

# Continuous-time semicontraction theory

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

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

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

$$\begin{aligned}\mu_{\text{dist},1}(-L) &= -\min_j \left\{ (d_{\text{out}})_j - \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor - 1} a_{(i),j} + \sum_{i=\lceil \frac{n}{2} \rceil}^{n-1} a_{(i),j} \right\}, \quad d_{\text{out}} = A\mathbb{1}_n \\ \mu_{\text{dist},2}(-L) &= \min \left\{ b : \Pi_{\perp} L + L^{\top} \Pi_{\perp} \succeq -2b\Pi_{\perp} \right\}, \quad \Pi_{\perp} = I_n - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n^{\top}\end{aligned}$$

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

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  $\mathcal{K}$ ,

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

# Outline

## §1. History and resources

## §2. Basic definitions: discrete and continuous-time dynamics on vector spaces

- The linear algebra of matrix norms; see CTDS Chapter 2
- Properties of induced matrix norms and Lipschitz constants

## §3. Example systems

- Constrained, distributed and proximal gradient dynamics
- Continuous-time recurrent neural networks
- Nonlinear dynamics in Lur'e form

## §4. Properties of contracting dynamics

- Equilibria, Lyapunov functions, and Euler discretization
- Incremental input-to-state stability
- Contractivity of interconnected systems
- Additional properties: entrainment, robustness wrt unmodeled dynamics and delays

## §5. Example applications

- Gradient dynamics and Nash equilibria in games
- Time-varying gradient dynamics and feedback optimization
- Recurrent and implicit neural networks

## §6. Generalizations with examples

- G1: Semicontractivity: Primal-dual gradient with redundant constraints
- G2: Contractivity on Riemannian manifolds and the Karcher mean

## §7. Conclusions and future research

## §8. Advanced Topics

- More on semicontractivity: ergodic coefficients and duality
- **Network small-gain theorem for Metzler matrices**
- More on the Kuramoto-Sakaguchi model and synchronization
- Proof of semicontractivity of saddle matrices
- Proof of Euler discretization theorem
- Non-Euclidean Monotone Operator Theory

**Hurwitz Metzler Theorem** (see LNS.Section10.4)

- ①  $M$  is Hurwitz,
- ② there exists  $\eta \in \mathbb{R}_{>0}^n$  such that  $\eta^\top M < 0_n^\top$  or, equivalently,  $\mu_{1,\text{diag}(\eta)}(M) < 0$ ,
- ③ there exists  $\xi \in \mathbb{R}_{>0}^n$  such that  $M\xi < 0_n$  or, equivalently,  $\mu_{\infty,\text{diag}(\xi)^{-1}}(M) < 0$ , and
- ④ there exists a diagonal  $P = P^\top \succ 0$  satisfying  $M^\top P + PM \prec 0$  or, equivalently,  
 $\mu_{2,P^{1/2}}(M) < 0$ .

**Input:** a Metzler matrix  $M \in \mathbb{R}^{n \times n}$

**Output:** polynomials  $\{\gamma_{\mathcal{C}_2}, \dots, \gamma_{\mathcal{C}_n}\}$  in entries of  $M$

- 1:  $\mathcal{C} :=$  set of simple cycles of digraph associated to  $M$
- 2:  $\gamma_\phi :=$  gain of cycle  $\phi \in \mathcal{C}$
- 3: **for**  $i$  from 2 to  $n$
- 4:    $\mathcal{C}_i :=$  cycles in  $\mathcal{C}$  passing through only nodes  $1, \dots, i$
- 5:    $\gamma_{\mathcal{C}_i} := \sum_{\phi \in \mathcal{C}_i} \gamma_\phi - \sum_{\substack{\phi, \psi \in \mathcal{C}_i \\ \phi \perp \psi}} \gamma_\phi \gamma_\psi + \sum_{\substack{\phi, \psi, \rho \in \mathcal{C}_i \\ \phi \perp \psi, \phi \perp \rho, \psi \perp \rho}} \gamma_\phi \gamma_\psi \gamma_\rho - \dots$

### Network small-gain theorem for Metzler matrices

$$\text{Metzler } M \text{ is Hurwitz} \iff \gamma_{\mathcal{C}_2} < 1, \dots, \gamma_{\mathcal{C}_n} < 1$$

- not unique: distinct/equivalent conditions after renumbering, redundancy
- computational efficiency: after precomputation of simple cycles

$$M = \begin{bmatrix} -c_1 & 0 & 0 & \ell_{14} \\ 0 & -c_2 & \ell_{23} & \ell_{24} \\ 0 & \ell_{32} & -c_3 & \ell_{34} \\ \ell_{41} & \ell_{42} & \ell_{43} & -c_4 \end{bmatrix}$$

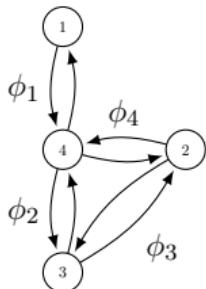


Figure: associated digraph and simple cycles

- $\gamma_{\phi_1} = \frac{\ell_{14}\ell_{41}}{c_1 c_4}$ ,  $\gamma_{\phi_2} = \frac{\ell_{34}\ell_{43}}{c_3 c_4}$ ,  $\gamma_{\phi_3} = \frac{\ell_{23}\ell_{32}}{c_2 c_3}$ , and  $\gamma_{\phi_4} = \frac{\ell_{24}\ell_{42}}{c_2 c_4}$
- $\mathcal{C}_2 = \emptyset$
- $\mathcal{C}_3 = \{\phi_3\}$ :  $\gamma_{\mathcal{C}_3} = \gamma_{\phi_3} < 1$  (redundant)
- $\mathcal{C}_4 = \{\phi_1, \dots, \phi_4\}$ :  $\gamma_{\mathcal{C}_4} = \sum_{i=1}^4 \gamma_{\phi_i} - \gamma_{\phi_1} \gamma_{\phi_3} < 1$

$$\begin{bmatrix} -c_1 & 0 & 0 & 0 & \ell_{15} & \ell_{16} \\ 0 & -c_2 & 0 & \ell_{24} & \ell_{25} & 0 \\ 0 & 0 & -c_3 & \ell_{34} & 0 & \ell_{36} \\ 0 & \ell_{42} & \ell_{43} & -c_4 & 0 & 0 \\ \ell_{51} & \ell_{52} & 0 & 0 & -c_5 & 0 \\ \ell_{61} & 0 & \ell_{63} & 0 & 0 & -c_6 \end{bmatrix}$$

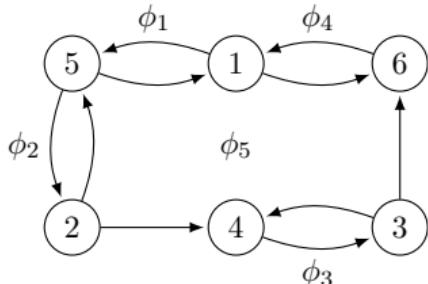


Figure: associated digraph and simple cycles

- $\mathcal{C}_2, \mathcal{C}_3$  empty
- $\mathcal{C}_4 = \{\phi_3\}$ :  $\gamma_3 < 1$  (redundant)
- $\mathcal{C}_5 = \{\phi_1, \phi_2, \phi_3\}$ :  $\gamma_{\mathcal{C}_5} = \gamma_1 + \gamma_2 + \gamma_3 - \gamma_1\gamma_3 - \gamma_2\gamma_3 < 1$
- $\mathcal{C}_6 = \{\phi_1, \dots, \phi_5\}$ :  $\gamma_{\mathcal{C}_6} = \sum_{i=1}^5 \gamma_i - \gamma_1\gamma_3 - \gamma_2\gamma_3 - \gamma_3\gamma_4 - \gamma_2\gamma_4 + \gamma_2\gamma_3\gamma_4 < 1$

# Outline

## §1. History and resources

## §2. Basic definitions: discrete and continuous-time dynamics on vector spaces

- The linear algebra of matrix norms; see CTDS Chapter 2
- Properties of induced matrix norms and Lipschitz constants

## §3. Example systems

- Constrained, distributed and proximal gradient dynamics
- Continuous-time recurrent neural networks
- Nonlinear dynamics in Lur'e form

## §4. Properties of contracting dynamics

- Equilibria, Lyapunov functions, and Euler discretization
- Incremental input-to-state stability
- Contractivity of interconnected systems
- Additional properties: entrainment, robustness wrt unmodeled dynamics and delays

## §5. Example applications

- Gradient dynamics and Nash equilibria in games
- Time-varying gradient dynamics and feedback optimization
- Recurrent and implicit neural networks

## §6. Generalizations with examples

- G1: Semicontractivity: Primal-dual gradient with redundant constraints
- G2: Contractivity on Riemannian manifolds and the Karcher mean

## §7. Conclusions and future research

## §8. Advanced Topics

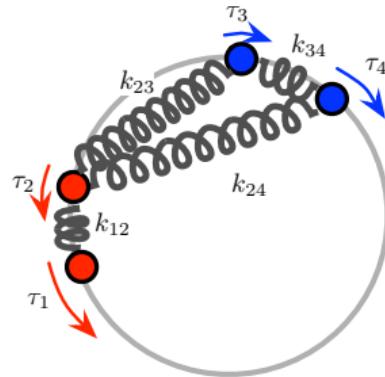
- More on semicontractivity: ergodic coefficients and duality
- Network small-gain theorem for Metzler matrices
- **More on the Kuramoto-Sakaguchi model and synchronization**
- Proof of semicontractivity of saddle matrices
- Proof of Euler discretization theorem
- Non-Euclidean Monotone Operator Theory

$$\omega_i = \sum_{j=1}^n a_{ij} \sin(\theta_i - \theta_j)$$

## Spring network

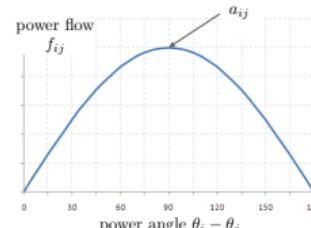
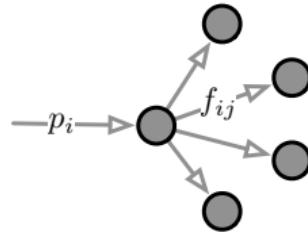
- $\omega_i = \tau_i$  : torque at  $i$
- $a_{ij} = k_{ij}$  : spring stiffness  $i, j$
- $\sin(\theta_i - \theta_j)$  : modulation
- elastic energy

$$\mathcal{E} = \sum_{ij} (1 - \cos(\theta_i - \theta_j))$$



## Power network

- $\omega_i = p_i$  : injected power
- $a_{ij}$  : max power flow  $i, j$
- $\sin(\theta_i - \theta_j)$  : modulation
- KCL flow conservation and Ohm's law

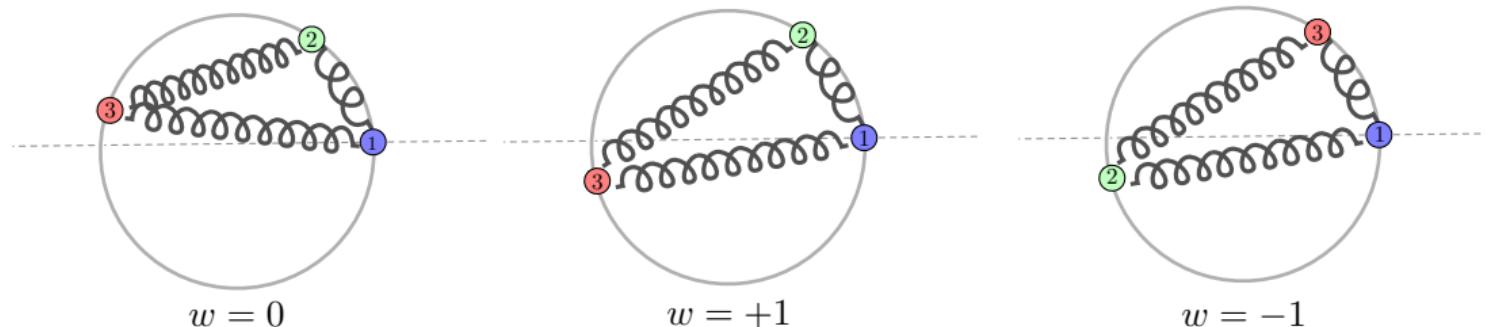


# Winding numbers and partitions

Given a cycle  $\sigma = (1, \dots, n_\sigma)$  and orientation

- ① **winding number of  $\theta \in \mathbb{T}^n$  along  $\sigma$**

= number of times the **shortest-arc path wraps around torus**



- ② given basis  $\sigma_1, \dots, \sigma_r$  for cycles, **winding vector of  $\theta$**  is

$$w(\theta) = (w_{\sigma_1}(\theta), \dots, w_{\sigma_r}(\theta))$$

**Theorem: Kirchhoff angle law on  $\mathbb{T}^n$**

winding number is at most  $\pm \lfloor n_\sigma / 2 \rfloor - 1$

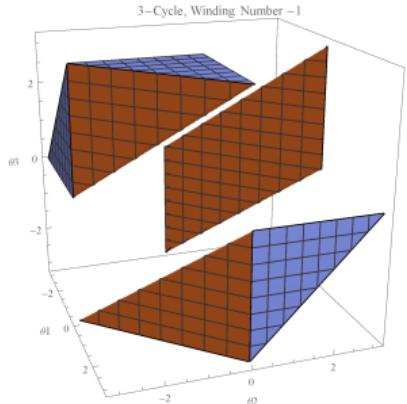


**Theorem: Winding partition** For each possible winding vector  $u$ , define

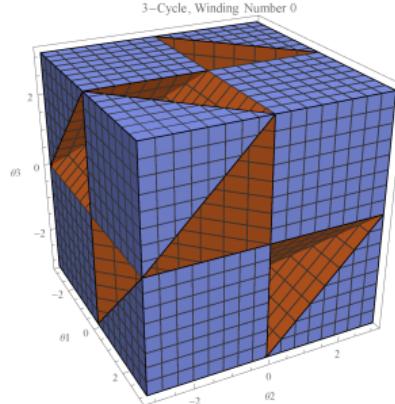
$$\text{WindingCell}(u) := \{\theta \in \mathbb{T}^n \text{ such that } w(\theta) = u\}$$

Then

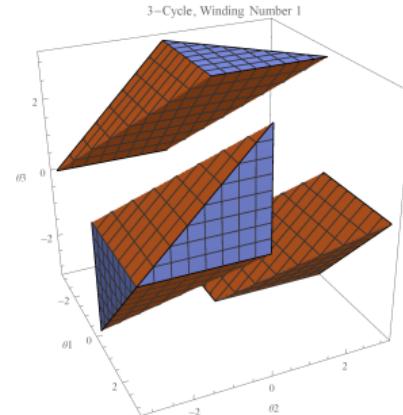
$$\mathbb{T}^n = \cup_u \text{WindingCell}(u)$$



$w = -1$



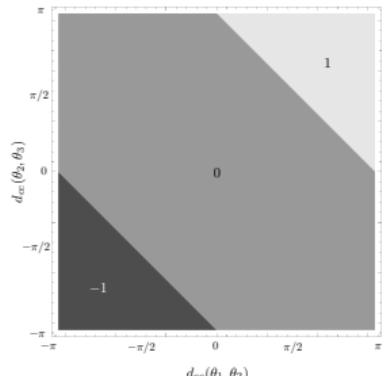
$w = 0$

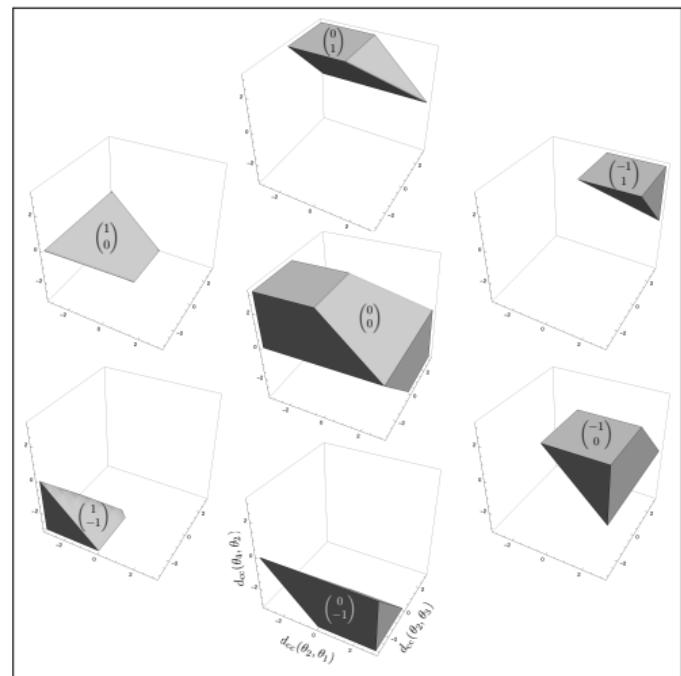
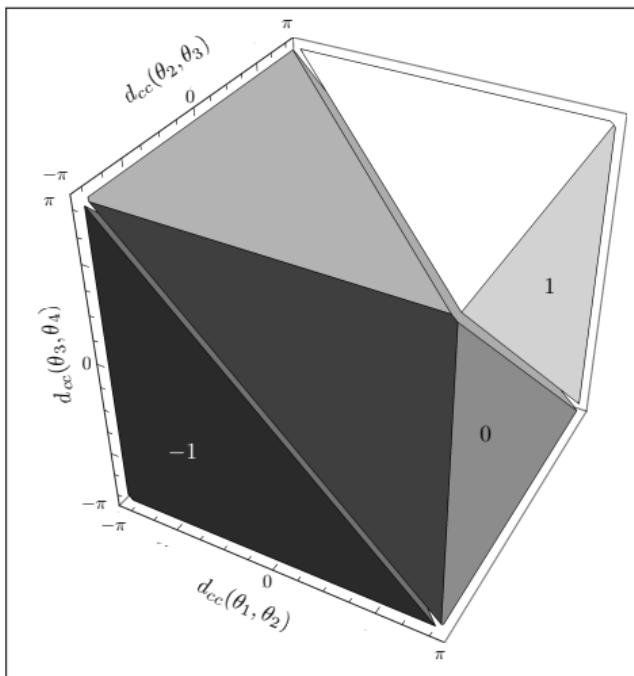
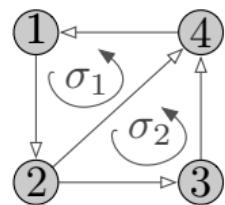
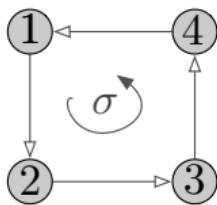


$w = +1$

## Theorem: Reduced cell is convex polytope

- each winding cell is connected and invariant under rotation
- bijection:**  
reduced winding cell  $\longleftrightarrow$  open convex polytope





$$\dot{\theta}_i = \omega_i - \sum_{j=1}^n a_{ij} \sin(\theta_i - \theta_j)$$

in each winding cell

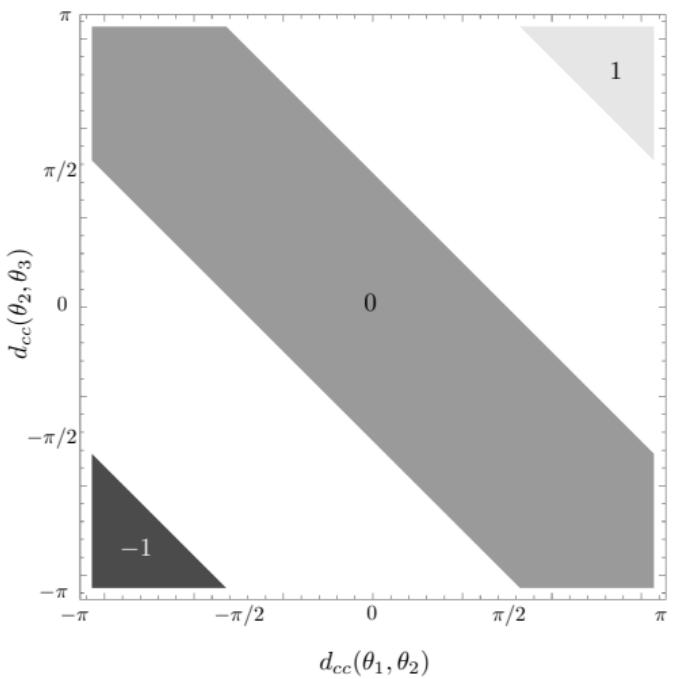
- ①  $\dot{\theta} = -\nabla \mathcal{E}(\theta)$ , where (well-posed smooth)

$$\mathcal{E}(\theta) = \sum_{ij} (1 - \cos(\theta_i - \theta_j)) + \omega^\top \theta$$

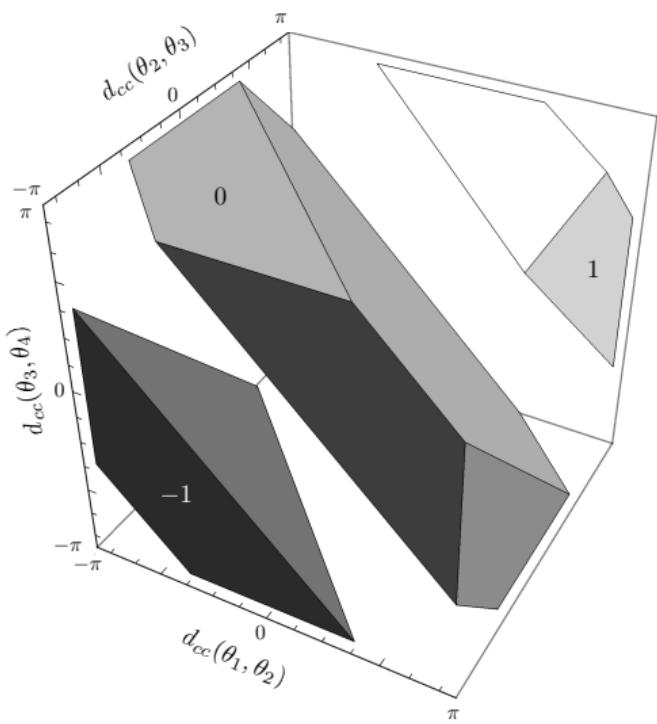
- ② Hessian  $\mathcal{E}(\theta) = -\text{Cosine-Laplacian}(\theta)$  (with possibly negative weights)
- ③ Hessian  $\mathcal{E}(\theta) \preceq 0$  on the **cohesive subset**  $|\theta_i - \theta_j| \leq \pi/2$
- ④ modulo the symmetry, the dynamics is strongly contracting (on strictly cohesive subset)

## Theorem:

- ① each winding cell has at most one cohesive equilibrium
- ② contraction algorithm to decide/compute in each winding cell



(a)



(b)

$$\dot{\theta}_i = \omega_i - \sum_{j=1}^n a_{ij} \sin(\theta_i - \theta_j + \phi_{ij})$$

same properties, by robustness of contracting dynamics

R. Delabays and F. Bullo. Semicontraction and synchronization of Kuramoto-Sakaguchi oscillator networks. *IEEE Control Systems Letters*, 7:1566–1571, 2023. 

# Outline

## §1. History and resources

## §2. Basic definitions: discrete and continuous-time dynamics on vector spaces

- The linear algebra of matrix norms; see CTDS Chapter 2
- Properties of induced matrix norms and Lipschitz constants

## §3. Example systems

- Constrained, distributed and proximal gradient dynamics
- Continuous-time recurrent neural networks
- Nonlinear dynamics in Lur'e form

## §4. Properties of contracting dynamics

- Equilibria, Lyapunov functions, and Euler discretization
- Incremental input-to-state stability
- Contractivity of interconnected systems
- Additional properties: entrainment, robustness wrt unmodeled dynamics and delays

## §5. Example applications

- Gradient dynamics and Nash equilibria in games
- Time-varying gradient dynamics and feedback optimization
- Recurrent and implicit neural networks

## §6. Generalizations with examples

- G1: Semicontractivity: Primal-dual gradient with redundant constraints
- G2: Contractivity on Riemannian manifolds and the Karcher mean

## §7. Conclusions and future research

## §8. Advanced Topics

- More on semicontractivity: ergodic coefficients and duality
- Network small-gain theorem for Metzler matrices
- More on the Kuramoto-Sakaguchi model and synchronization
- Proof of semicontractivity of saddle matrices
- Proof of Euler discretization theorem
- Non-Euclidean Monotone Operator Theory

# Outline

## §1. History and resources

## §2. Basic definitions: discrete and continuous-time dynamics on vector spaces

- The linear algebra of matrix norms; see CTDS Chapter 2
- Properties of induced matrix norms and Lipschitz constants

## §3. Example systems

- Constrained, distributed and proximal gradient dynamics
- Continuous-time recurrent neural networks
- Nonlinear dynamics in Lur'e form

## §4. Properties of contracting dynamics

- Equilibria, Lyapunov functions, and Euler discretization
- Incremental input-to-state stability
- Contractivity of interconnected systems
- Additional properties: entrainment, robustness wrt unmodeled dynamics and delays

## §5. Example applications

- Gradient dynamics and Nash equilibria in games
- Time-varying gradient dynamics and feedback optimization
- Recurrent and implicit neural networks

## §6. Generalizations with examples

- G1: Semicontractivity: Primal-dual gradient with redundant constraints
- G2: Contractivity on Riemannian manifolds and the Karcher mean

## §7. Conclusions and future research

## §8. Advanced Topics

- More on semicontractivity: ergodic coefficients and duality
- Network small-gain theorem for Metzler matrices
- More on the Kuramoto-Sakaguchi model and synchronization
- **Proof of semicontractivity of saddle matrices**
- Proof of Euler discretization theorem
- Non-Euclidean Monotone Operator Theory

# Semicontractivity of saddle matrices

Given  $Q \in \mathbb{R}^{n \times n}$ ,  $A \in \mathbb{R}^{m \times n}$ , and a time-scale parameter  $\tau > 0$ , define

saddle matrix

$$\mathcal{S} = \begin{bmatrix} -Q & -A^\top \\ \tau^{-1}A & 0 \end{bmatrix} \in \mathbb{R}^{(m+n) \times (m+n)}$$

$$q_{\min} := \lambda_{\min}(Q + Q^\top)/2 > 0$$

$$q_{\max} := \min\{q \text{ such that } Q^\top Q \preceq q(Q + Q^\top)/2\} \leq \sigma_{\max}^2(Q)/q_{\min}$$

$a_{\min}\Pi_A \preceq AA^\top \preceq a_{\max}I_m$ , where  $\Pi_A \in \mathbb{R}^{m \times m}$  is orthogonal projection onto image of  $A$

## Semi-contractivity LMI

$$\mathcal{S}^\top P + P\mathcal{S} \preceq -2cP$$

where

$$P = \begin{bmatrix} I_n & \alpha A^\top \\ \alpha A & \tau \Pi_A \end{bmatrix} \succeq 0 \quad \text{with} \quad \alpha = \frac{1}{2} \min \left\{ \frac{1}{\nu_{\max}}, \tau \frac{\nu_{\min}}{a_{\max}} \right\}$$

$$c = \frac{1}{2} \tau^{-1} \alpha a_{\min} = \frac{1}{4} \min \left\{ \frac{a_{\min}}{\tau q_{\max}}, \frac{a_{\min}}{a_{\max}} q_{\min} \right\}$$

### Proof of saddle matrix semicontractivity I: $P \succeq 0$

Use Schur complement to show that  $P \succeq 0$ . Clearly the  $(1, 1)$  block is positive definite. Therefore,

$$P \succeq 0 \iff \tau \Pi_A - \alpha^2 A A^\top \succ 0 \iff \tau - \alpha^2 a_{\max} > 0 \iff \alpha^2 < \tau/a_{\max}.$$

The inequality  $\alpha^2 < \tau/a_{\max}$  follows from the stronger inequality  $(2\alpha)^2 < \frac{\tau}{a_{\max}}$  with the following argument:

$$\min \left\{ \frac{1}{q_{\max}}, \tau \frac{q_{\min}}{a_{\max}} \right\}^2 \leq \min \left\{ \frac{1}{q_{\max}}, \tau \frac{q_{\min}}{a_{\max}} \right\} \cdot \max \left\{ \frac{1}{q_{\max}}, \tau \frac{q_{\min}}{a_{\max}} \right\} = \frac{q_{\min}}{q_{\max}} \cdot \frac{\tau}{a_{\max}} \leq \frac{\tau}{a_{\max}}.$$

## Proof of saddle matrix semicontractivity II: factorization of LMI

Next, we aim to show that  $-\mathcal{S}^\top P - PS - 2cP \succeq 0$ . After some bookkeeping, we compute

$$\begin{aligned}-\mathcal{S}^\top P - PS - 2cP &= \begin{bmatrix} Q^\top & -\tau^{-1}A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} I_n & \alpha A^\top \\ \alpha A & \tau \Pi_A \end{bmatrix} + \begin{bmatrix} I_n & \alpha A^\top \\ \alpha A & \tau \Pi_A \end{bmatrix} \begin{bmatrix} Q & A^\top \\ -\tau^{-1}A & 0 \end{bmatrix} - 2c \begin{bmatrix} I_n & \alpha A^\top \\ \alpha A & \tau \Pi_A \end{bmatrix} \\ &= \begin{bmatrix} Q + Q^\top - 2\tau^{-1}\alpha A^\top A - 2cI_n & \alpha Q^\top A^\top - A^\top + A^\top \Pi_A^\top - 2c\alpha A^\top \\ A + \alpha A Q - \Pi_A A - 2c\alpha A & 2\alpha A A^\top - 2c\tau \Pi_A \end{bmatrix}.\end{aligned}$$

The (2,2) block satisfies the lower bound

$$2\alpha A A^\top - 2c\tau \Pi_A = 2\left(\frac{1}{2}\alpha A A^\top - c\tau \Pi_A\right) + \alpha A A^\top \succeq 2\left(\frac{1}{2}\alpha a_{\min} - c\tau\right)\Pi_A + \alpha A A^\top = \alpha A A^\top \succ 0.$$

Given this lower bound and the equality  $\Pi_A A = A$ , we can factorize the resulting matrix as follows:

$$-\mathcal{S}^\top P - PS - cP \succeq \begin{bmatrix} I_n & 0 \\ 0 & A \end{bmatrix} \underbrace{\begin{bmatrix} Q + Q^\top - 2(\tau^{-1}\alpha A^\top A + cI_n) & \alpha Q^\top - 2c\alpha I_n \\ \alpha Q - 2c\alpha I_n & \alpha I_n \end{bmatrix}}_{n \times n} \begin{bmatrix} I_n & 0 \\ 0 & A^\top \end{bmatrix}.$$

### Proof of saddle matrix semicontractivity III: Schur complement and final bounds

Since  $\alpha I_n \succ 0$ , it suffices to show that the Schur complement of the (2,2) block is positive semidefinite:

$$Q + Q^\top - 2(\tau^{-1}\alpha A^\top A + cI_n) - \alpha(Q^\top - 2cI_n)(Q - 2cI_n) \succeq 0 \quad (9)$$

$$\iff (Q + Q^\top - \alpha Q^\top Q) + 2\alpha c(Q + Q^\top) \succeq 2(\tau^{-1}\alpha A^\top A + cI_n) + 4\alpha c^2 I_n \quad (10)$$

$$\iff Q + Q^\top - \alpha Q^\top Q \succeq 2(\tau^{-1}\alpha A^\top A + cI_n) \quad \text{and} \quad 2\alpha c(Q + Q^\top) \succeq 4\alpha c^2 I_n. \quad (11)$$

To prove the first inequality in (11), we upper bound the right hand side as follows:

$$\begin{aligned} 2(\tau^{-1}\alpha A^\top A + cI_n) &\preceq 2(\tau^{-1}\alpha a_{\max} + c)I_n \stackrel{c=\frac{1}{2}\tau^{-1}\alpha a_{\min}}{=} \tau^{-1}\alpha(2a_{\max} + a_{\min})I_n \\ &\stackrel{\alpha \leq \frac{1}{2}\tau q_{\min}/a_{\max}}{\preceq} \frac{1}{2} \frac{q_{\min}}{a_{\max}} (2a_{\max} + a_{\min})I_n \preceq \frac{3}{2}q_{\min}I_n. \end{aligned}$$

Next, since  $\alpha \leq \frac{1}{2q_{\max}}$ , we know  $-\alpha q_{\max} \geq -\frac{1}{2}$ . We then lower bound the left hand side as follows:

$$Q + Q^\top - \alpha Q^\top Q \stackrel{\text{by definition}}{\succeq} Q + Q^\top - \alpha q_{\max}(Q + Q^\top)/2 \succeq (2 - \frac{1}{2})(Q + Q^\top) \succeq \frac{3}{2}q_{\min}I_n.$$

Finally, we prove the second inequality in (11) that is,  $2\alpha c(Q + Q^\top) \succeq 4\alpha c^2 I_n$ . This is equivalent to  $Q + Q^\top \succeq 2cI_n$  and follows from noting  $c \leq \frac{1}{2} \frac{a_{\min}}{a_{\max}} q_{\min} < q_{\min}$ .

# Outline

## §1. History and resources

## §2. Basic definitions: discrete and continuous-time dynamics on vector spaces

- The linear algebra of matrix norms; see CTDS Chapter 2
- Properties of induced matrix norms and Lipschitz constants

## §3. Example systems

- Constrained, distributed and proximal gradient dynamics
- Continuous-time recurrent neural networks
- Nonlinear dynamics in Lur'e form

## §4. Properties of contracting dynamics

- Equilibria, Lyapunov functions, and Euler discretization
- Incremental input-to-state stability
- Contractivity of interconnected systems
- Additional properties: entrainment, robustness wrt unmodeled dynamics and delays

## §5. Example applications

- Gradient dynamics and Nash equilibria in games
- Time-varying gradient dynamics and feedback optimization
- Recurrent and implicit neural networks

## §6. Generalizations with examples

- G1: Semicontractivity: Primal-dual gradient with redundant constraints
- G2: Contractivity on Riemannian manifolds and the Karcher mean

## §7. Conclusions and future research

## §8. Advanced Topics

- More on semicontractivity: ergodic coefficients and duality
- Network small-gain theorem for Metzler matrices
- More on the Kuramoto-Sakaguchi model and synchronization
- Proof of semicontractivity of saddle matrices
- **Proof of Euler discretization theorem**
- Non-Euclidean Monotone Operator Theory

## Euler discretization theorem for contracting dynamics

Given arbitrary norm  $\|\cdot\|$  and differentiable  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , equivalent statements

- ①  $\dot{x} = F(x)$  is infinitesimally contracting
- ② there exists  $\alpha > 0$  such that  $x_{k+1} = x_k + \alpha F(x_k)$  is contracting

### Optimal\* contractivity of Euler discretization $\text{Id} + \alpha F$

Given  $c := -\text{osLip}(F) > 0$  and  $\ell := \text{Lip}(F)$ , define *condition number*  $\kappa = \ell/c \geq 1$ :

$$\textcircled{3} \quad 0 < \alpha < \frac{1}{c\kappa(1+\kappa)} \implies \text{Lip}(\text{Id} + \alpha F) \leq \left(1 + \alpha c - \frac{\alpha^2 \ell^2}{1 - \alpha \ell}\right)^{-1} < 1$$

- ④ the optimal\* step size and contraction factor are

$$\alpha^* = \frac{1}{c} \left( \frac{1}{2\kappa^2} - \frac{3}{8\kappa^3} + \mathcal{O}\left(\frac{1}{\kappa^4}\right) \right), \quad \text{Lip}(\text{Id} + \alpha^* F) = 1 - \frac{1}{4\kappa^2} + \frac{1}{8\kappa^3} + \mathcal{O}\left(\frac{1}{\kappa^4}\right)$$

## Euler discretization theorem: Additional equivalences

Given  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  and  $\alpha \geq 0$ , define

- **shifted map**  $G := \text{Id} + F \iff F = -\text{Id} + G$
- $F_\alpha := \underbrace{\text{Id} + \alpha F}_{\text{Euler discretization of } F} = \underbrace{(1 - \alpha) \text{Id} + \alpha G}_{\text{average map of } G} =: G_\alpha$

For a differentiable  $F$  and  $x \in \mathbb{R}^n$

$$\begin{array}{ccc} F(x) = 0 & \iff & F_\alpha(x) = G_\alpha(x) = G(x) = x \\ \text{equilibrium point} & & \text{fixed point} \end{array}$$

$$\begin{array}{ccc} \begin{array}{c} \text{osLip}(F) < 0 \\ F \text{ is infinitesimally contracting} \end{array} & \iff & \begin{array}{c} \text{osLip}(G) = \text{osLip}(F) + 1 < 1 \\ \uparrow \\ \exists \alpha^* \text{ s.t. } \text{Lip}(F_{\alpha^*}) < 1 \\ F_{\alpha^*} \text{ is contracting} \end{array} \\ \iff & & \iff \\ \begin{array}{c} \exists \alpha^* \text{ s.t. } \text{Lip}(G_{\alpha^*}) < 1 \\ G_{\alpha^*} \text{ is contracting} \end{array} & & \end{array}$$

## Optimal\* contractivity of Euler discretization $\text{Id} + \alpha F$ : inner-product norms $\|\cdot\|_{2,P^{1/2}}$

Given  $c := -\text{osLip}(F) > 0$  and  $\ell := \text{Lip}(F)$ , define *condition number*  $\kappa = \ell/c \geq 1$ :

①  $0 < \alpha < \frac{2}{c\kappa^2} \implies \text{Lip}(\text{Id} + \alpha F) \leq 1 - 2\alpha c + \alpha^2 \ell^2 < 1$

② the optimal\* step size and contraction factor are

$$\alpha^* = \frac{1}{c\kappa^2}, \quad \text{Lip}(\text{Id} + \alpha^* F) = 1 - \frac{1}{2\kappa^2} + \mathcal{O}\left(\frac{1}{\kappa^3}\right)$$

Standard proof from monotone operator theory. For  $\alpha > 0$ , compute

$$\begin{aligned} \|(\text{Id} + \alpha F)x - (\text{Id} + \alpha F)y\|^2 &= \|x - y + \alpha(F(x) - F(y))\|^2 \\ &= \|x - y\|^2 + 2\alpha \langle F(x) - F(y), x - y \rangle + \alpha^2 \|F(x) - F(y)\|^2 \\ &\leq (1 - 2\alpha c + \alpha^2 \ell^2) \|x - y\|^2 \end{aligned}$$

Next, study convex parabola  $\alpha \mapsto 1 - 2\alpha c + \alpha^2 \ell^2$ . Eg,  $1 - 2\alpha c + \alpha^2 \ell^2 < 1$  iff  $0 < \alpha < 2c/\ell^2$

**Optimal\* contractivity of Euler discretization  $\text{Id} + \alpha F$ : nonEuclidean  $\|\cdot\|_{\infty, \text{diag}(\eta)^{-1}}$ ,**

$\|\cdot\|_{1, \text{diag}(\eta)}$

Let  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be differentiable and Lipschitz

define *contraction rate*  $c := -\text{osLip}(F) > 0$

define *diagonal Lipschitz constant*  $\ell_{\text{diag}} = \max_{i \in \{1, \dots, n\}} \sup_{x \in \mathbb{R}^n} |DF_{ii}(x)|$ ; can show  $\ell_{\text{diag}} \geq c$

$$\textcircled{1} \quad 0 < \alpha \leq \frac{1}{\ell_{\text{diag}}} \implies \text{Lip}(\text{Id} + \alpha F) \leq 1 - \alpha c < 1$$

\textcircled{2} the optimal\* step size and contraction factor are

$$\alpha^* = \frac{1}{\ell_{\text{diag}}}, \quad \text{Lip}(\text{Id} + \alpha^* F) = 1 - \frac{c}{\ell_{\text{diag}}}$$

**Acceleration:** (i) the condition number improves/diminishes  $\kappa \geq \kappa_\infty := \frac{c}{\ell_{\text{diag}}}$ , and  
(ii)  $\text{Lip}(\text{Id} + \alpha^* F) = 1 - \frac{1}{4\kappa^2} + \mathcal{O}\left(\frac{1}{\kappa^4}\right)$  improves/decreases to  $\text{Lip}(\text{Id} + \alpha^* F) = 1 - \frac{1}{\kappa_\infty}$ .

S. Jafarpour, A. Davydov, A. V. Proskurnikov, and F. Bullo. Robust implicit networks via non-Euclidean contractions. In *Advances in Neural Information Processing Systems*, Dec. 2021. 

### Proof of $\ell_\infty/\ell_1$ Euler discretization theorem

For every  $A = [a_{ij}] \in \mathbb{R}^{n \times n}$ ,  $\eta \in \mathbb{R}_{>0}^n$ , and  $\alpha \in \mathbb{R}$  such that  $|\alpha| \leq (\max_i |a_{ii}|)^{-1}$ , **norm=lognorm identity**:

$$\|I_n + \alpha A\|_{1,\text{diag}(\eta)} = 1 + \alpha \mu_{1,\text{diag}(\eta)}(A), \quad \|I_n + \alpha A\|_{\infty,\text{diag}(\eta)^{-1}} = 1 + \alpha \mu_{\infty,\text{diag}(\eta)^{-1}}(A), \quad (12)$$

whose proof is an algebraic exercise (hint: diagonal of  $I_n + \alpha A$  is nonnegative).

Next, consider  $\|\cdot\|_{\infty,\text{diag}(\eta)^{-1}}$ ; the proof for  $\|\cdot\|_{1,\text{diag}(\eta)}$  is omitted. Regarding part 1, for each  $i \in \{1, \dots, n\}$  and  $x \in \mathbb{R}^n$

$$\begin{aligned} \ell_{\text{diag}} &= \max_{i \in \{1, \dots, n\}} \sup_{x \in \mathbb{R}^n} |D\mathbf{F}_{ii}(x)| \stackrel{(\text{osLip}(\mathbf{F}) < 0 \implies D\mathbf{F}_{ii}(x) < 0)}{=} \max_{i \in \{1, \dots, n\}} \sup_{x \in \mathbb{R}^n} (-D\mathbf{F}_{ii}(x)), \\ &\geq \max_{i \in \{1, \dots, n\}} \sup_{x \in \mathbb{R}^n} \left( -D\mathbf{F}_{ii}(x) - \sum_{j \neq i} |D\mathbf{F}_{ij}(x)| \frac{\eta_j}{\eta_i} \right) \\ &= - \max_{i \in \{1, \dots, n\}} \sup_{x \in \mathbb{R}^n} \mu_{\infty,\text{diag}(\eta)^{-1}}(D\mathbf{F}(x)) = -\text{osLip}(\mathbf{F}) = c. \end{aligned}$$

Since  $\ell_{\text{diag}} = \sup_x \max_i |D\mathbf{F}_{ii}(x)| \geq \max_i |D\mathbf{F}_{ii}(x)|$  for all  $x$  and  $\alpha \leq \frac{1}{\ell_{\text{diag}}} \leq \frac{1}{\max_i |D\mathbf{F}_{ii}(x)|}$ , equation (12) implies

$$\|I_n + \alpha D\mathbf{F}(x)\|_{\infty,\text{diag}(\eta)^{-1}} = 1 + \alpha \mu_{\infty,\text{diag}(\eta)^{-1}}(D\mathbf{F}(x)) \leq 1 + \alpha \text{osLip}(\mathbf{F}) = 1 - \alpha c.$$

Finally,  $\text{Lip}(\text{Id} + \alpha \mathbf{F}) \leq \sup_x \|I_n + \alpha D\mathbf{F}(x)\|_{\infty,\text{diag}(\eta)^{-1}} \leq 1 - \alpha c$ .

Regarding part 2,  $\alpha \rightarrow \text{Lip}(\text{Id} + \alpha \mathbf{F})$  is decreasing and therefore minimum at the maximum of allowable value of  $\alpha$ . Note that  $\alpha^* = \ell_{\text{diag}}^{-1}$  is the maximum value of  $\alpha$  and  $\text{Lip}(\text{Id} + \alpha^* \mathbf{F}) = 1 - c/\ell_{\text{diag}} > 0$  since  $c/\ell_{\text{diag}} \leq 1$ .

# Outline

## §1. History and resources

## §2. Basic definitions: discrete and continuous-time dynamics on vector spaces

- The linear algebra of matrix norms; see CTDS Chapter 2
- Properties of induced matrix norms and Lipschitz constants

## §3. Example systems

- Constrained, distributed and proximal gradient dynamics
- Continuous-time recurrent neural networks
- Nonlinear dynamics in Lur'e form

## §4. Properties of contracting dynamics

- Equilibria, Lyapunov functions, and Euler discretization
- Incremental input-to-state stability
- Contractivity of interconnected systems
- Additional properties: entrainment, robustness wrt unmodeled dynamics and delays

## §5. Example applications

- Gradient dynamics and Nash equilibria in games
- Time-varying gradient dynamics and feedback optimization
- Recurrent and implicit neural networks

## §6. Generalizations with examples

- G1: Semicontractivity: Primal-dual gradient with redundant constraints
- G2: Contractivity on Riemannian manifolds and the Karcher mean

## §7. Conclusions and future research

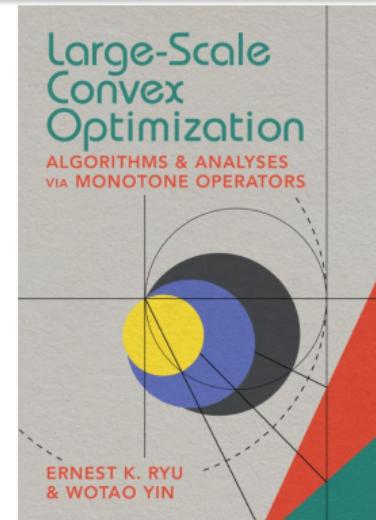
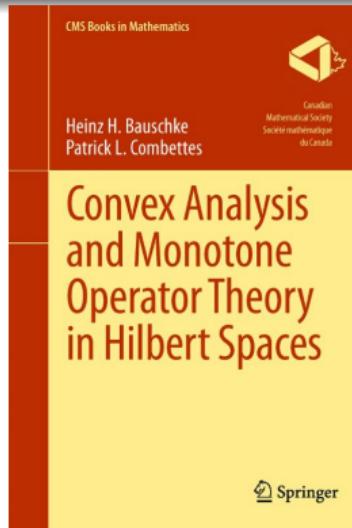
## §8. Advanced Topics

- More on semicontractivity: ergodic coefficients and duality
- Network small-gain theorem for Metzler matrices
- More on the Kuramoto-Sakaguchi model and synchronization
- Proof of semicontractivity of saddle matrices
- Proof of Euler discretization theorem
- Non-Euclidean Monotone Operator Theory

# Monotone operator methods

Success in many disparate fields

- ① Optimization and control
  - Subdifferentials are monotone
- ② Game theory
  - Monotone games
- ③ Systems analysis
  - Input-output behavior
- ④ Machine learning



L. Pavel. Distributed GNE seeking under partial-decision information over networks via a doubly-augmented operator splitting approach. *IEEE Transactions on Automatic Control*, 65(4):1584–1597, 2020. doi: [https://doi.org/10.1109/TAC.2019.2932750](#)

P. L. Combettes and J.-C. Pesquet. Fixed point strategies in data science. *IEEE Transactions on Signal Processing*, 2021. doi: [https://doi.org/10.1109/TSP.2021.3057440](#)

E. Winston and J. Z. Kolter. Monotone operator equilibrium networks. In *Advances in Neural Information Processing Systems*, 2020. URL <https://arxiv.org/abs/2006.08591>

A. Davydov, S. Jafarpour, A. V. Proskurnikov, and F. Bullo. Non-Euclidean monotone operator theory with applications to recurrent neural networks. In *IEEE Conf. on Decision and Control*, Cancún, México, Dec. 2022b. doi: [https://doi.org/10.1109/CDC46503.2022.9805000](#)

A. Davydov, S. Jafarpour, A. V. Proskurnikov, and F. Bullo. Non-Euclidean monotone operator theory and applications. *Journal of Machine Learning Research*, June 2023b. doi: [https://doi.org/10.4236/jmlr.v24i21.166000](#). Submitted

## Background on monotone operators

operator  $A : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is **monotone** with parameter  $m \geq 0$  if

$$\langle\langle A(x) - A(y), x - y \rangle\rangle \geq m \|x - y\|_2^2 \quad (\text{osLip}_2(-A) \leq -m)$$

A **monotone inclusion problem** is of the form

$$\text{find } x \in \mathbb{R}^n \text{ s.t. } 0 \in A(x)$$

A **monotone splitting problem** is of the form

$$\text{find } x \in \mathbb{R}^n \text{ s.t. } 0 \in (A + B)(x)$$

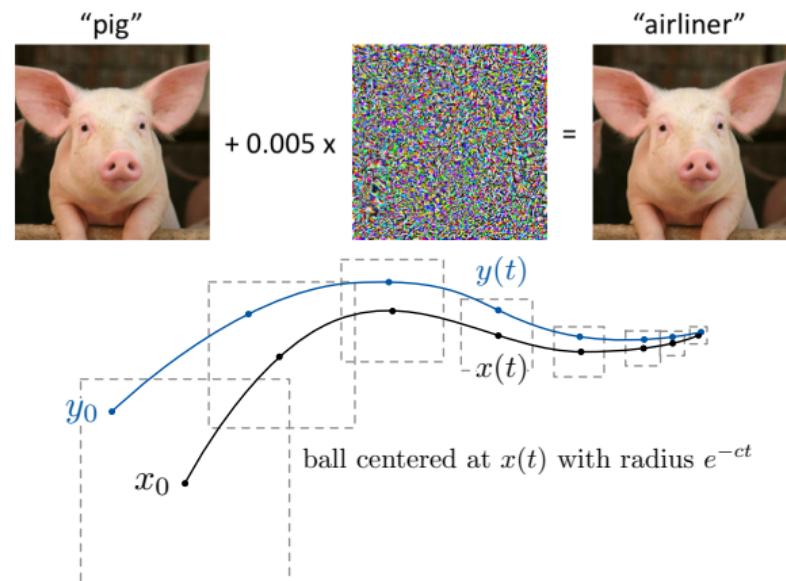
Existing algorithms based on Banach contractions or Krasnosel'skii–Mann iterations:

- Forward step method, proximal-point algorithm, etc.
- Forward-backward splitting, Peaceman-Rachford splitting, etc.

# Why non-Euclidean?

Algorithms for inclusions and splittings are limited to Hilbert settings  
Many problems are better stated in **Banach spaces!**

- ①  $\ell_\infty$  robustness analysis of neural networks
- ②  $L_\infty$  norm systems analysis
- ③ Non-Euclidean contracting dynamics
- ④ Totally asynchronous distributed optimization



A differentiable  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is **strongly monotone w.r.t  $\|\cdot\|$  with parameter  $m$**  if

$$-\mu(-DF(x)) \geq m, \quad \forall x \in \mathbb{R}^n. \quad (\text{osLip}(-F) \leq -m)$$

The **resolvent** and **reflected resolvent** of  $F$  with parameter  $\alpha > 0$  are given by:

$$J_{\alpha F} := (\text{Id} + \alpha F)^{-1}, \quad R_{\alpha F} := 2J_{\alpha F} - \text{Id}$$

Fixed points of  $J_{\alpha F}$  and  $R_{\alpha F}$  correspond to zeros of  $F$

**Lipschitz constants:** Suppose  $F$  is monotone w.r.t. a diagonally-weighted  $\ell_1/\ell_\infty$  norm

$$\text{Lip}(J_{\alpha F}) = \frac{1}{1 + \alpha m}, \quad \forall \alpha > 0$$

$$\text{Lip}(R_{\alpha F}) = \frac{1 - \alpha m}{1 + \alpha m}, \quad \forall \alpha \in ]0, \text{diagL}(F)^{-1}]$$

$$\text{diagL}(F) = \sup_{x \in \mathbb{R}^n} \max_{i \in \{1, \dots, n\}} (DF(x))_{ii}$$

Monotone inclusion problem  $F(x) = 0$

The **forward step method** of  $F$  ( $\ell_1/\ell_\infty$  monotone) is the iteration

$$x_{k+1} = (\text{Id} - \alpha F)(x_k)$$

- ① if  $m > 0$ ,  $\|x_{k+1} - x^*\| \leq (1 - \alpha m) \|x_k - x^*\|$ ,  $\forall \alpha \in ]0, \text{diagL}(F)^{-1}]$
- ② if  $m = 0$  and  $\text{zero}(F) \neq \emptyset$ , then convergence to an element of  $\text{zero}(F)$  with rate  $\mathcal{O}(1/\sqrt{k})$

The **proximal point method** of  $F$  ( $\ell_1/\ell_\infty$  monotone) is the iteration

$$x_{k+1} = J_{\alpha F}(x_k)$$

- ① if  $m > 0$ ,  $\|x_{k+1} - x^*\| \leq \frac{1}{1 + \alpha m} \|x_k - x^*\|$ ,  $\forall \alpha > 0$
- ② if  $m = 0$  and  $\text{zero}(F) \neq \emptyset$ , then convergence to an element of  $\text{zero}(F)$  with rate  $\mathcal{O}(1/\sqrt{k})$

# Comparison to standard convergence rates

Algorithm	F strongly monotone and globally Lipschitz			
	$\ell_2$		Diagonally weighted $\ell_1$ or $\ell_\infty$	
	$\alpha$ range	Optimal Lip	$\alpha$ range	Optimal Lip
Forward step	$]0, \frac{2m}{\ell^2} [$	$1 - \frac{1}{2\kappa^2} + \mathcal{O}\left(\frac{1}{\kappa^3}\right)$	$]0, \frac{1}{\text{diagL}(F)} [$	$1 - \frac{1}{\kappa_\infty}$
Proximal point	$]0, \infty[$	N/A	$]0, \infty[$	N/A
Cayley method	$]0, \infty[$	$1 - \frac{1}{2\kappa} + \mathcal{O}\left(\frac{1}{\kappa^2}\right)$	$]0, \frac{1}{\text{diagL}(F)} [$	$1 - \frac{2}{\kappa_\infty} + \mathcal{O}\left(\frac{1}{\kappa_\infty^2}\right)$

Step size ranges and Lipschitz constants for algorithms for finding zeros of monotone operators.  $\kappa := \ell/m \geq 1$  and  $\kappa_\infty := \text{diagL}(F)/m \in [1, \kappa]$

# Non-Euclidean operator splitting

Monotone splitting problem  $(F + G)(x) = 0$

The **forward-backward splitting method** of  $F$  and  $G$  ( $\ell_1/\ell_\infty$  monotone) is

$$x_{k+1} = (J_{\alpha G} \circ (\text{Id} - \alpha F))(x_k)$$

- ① if  $F$  s.m.,  $m > 0$ ,  $\|x_{k+1} - x^*\| \leq (1 - \alpha m) \|x_k - x^*\|$ ,  $\forall \alpha \in ]0, \text{diagL}(F)^{-1}]$
- ② if  $m = 0$  and  $\text{zero}(F + G) \neq \emptyset$ , then iteration converges to an element of  $\text{zero}(F + G)$  with rate  $\mathcal{O}(1/\sqrt{k})$

The **Peaceman-Rachford splitting method** of  $F$  and  $G$  ( $\ell_1/\ell_\infty$  monotone) is

$$x_{k+1} = J_{\alpha G}(z_k),$$

$$z_{k+1} = z_k + 2J_{\alpha F}(2x_{k+1} - z_k) - 2x_{k+1}.$$

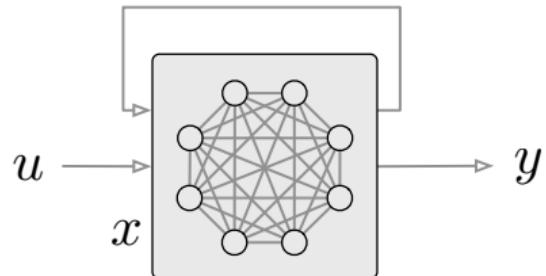
- ① If  $F$  s.m.,  $m > 0$ ,

$$\|x_{k+1} - x^*\| \leq \frac{1 - \alpha m}{1 + \alpha m} \|x_k - x^*\|, \quad \forall \alpha \in ]0, \min\{\text{diagL}(F)^{-1}, \text{diagL}(G)^{-1}\}]$$

## Equilibrium computation of RNN

$$\dot{x} = -x + \Phi(Ax + Bu + b) =: F(x, u)$$

$$\Phi(x) = \text{LeakyReLU}(x) = \max\{x, ax\}$$



A sufficient condition for contractivity is  $\mu_\infty(A) = \gamma < 1$ .

In this case,  $-F(x, u)$  is strongly monotone and can apply **forward step method**

$$x_{k+1} = (1 - \alpha)x_k + \alpha\Phi(Ax_k + Bu + b),$$

converges for  $\alpha \in ]0, \alpha^*]$  with linear convergence rate  $1 - \alpha(1 - \Phi(\gamma))$

$$\alpha^* = (1 - \min_{i \in \{1, \dots, n\}} \min\{a \cdot (A)_{ii}, (A)_{ii}\})^{-1}$$

A. Davydov, A. V. Proskurnikov, and F. Bullo. Non-Euclidean contractivity of recurrent neural networks. In *American Control Conference*, pages 1527–1534, Atlanta, USA, May 2022c. doi: [10.23915/ACC22.1534](#)

# Splitting methods for equilibrium computation

Finding an equilibrium point  $x^*(u)$  is equivalent to  $(F + G)(x^*(u)) = 0$  where

$$F(z) = (I_n - A)z - Bu - b, \quad G(z) = \frac{1-a}{a} \min\{z, 0\}$$

Apply **forward-backward splitting**

$$x_{k+1} = J_{\alpha G}((1-\alpha)x_k + \alpha(Ax_k + Bu + b)).$$

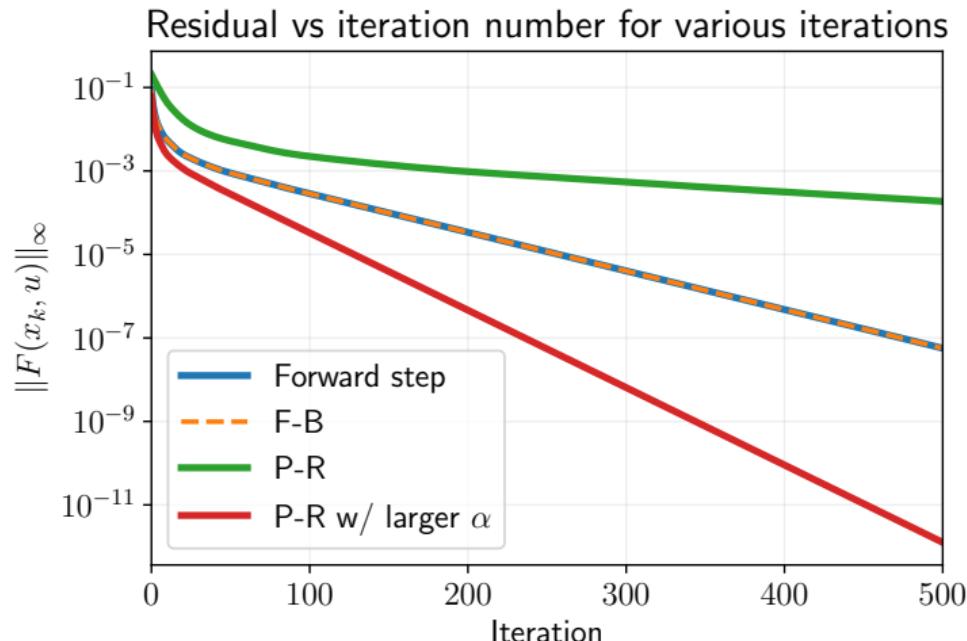
Converges with rate  $1 - \alpha(1 - \gamma)$  for  $\alpha \in ]0, (1 - \min_i(A)_{ii})^{-1}]$

Apply **Peaceman-Rachford splitting**

$$x_{k+1} = (I_n + \alpha(I_n - A))^{-1}(z_k + \alpha(Bu + b)),$$

$$z_{k+1} = z_k + 2J_{\alpha G}(2x_{k+1} - z_k) - 2x_{k+1}.$$

Converges with rate  $\frac{1 - \alpha(1 - \gamma)}{1 + \alpha(1 - \gamma)}$  for  $\alpha \in ]0, \min\{(1 - \min_i(A)_{ii})^{-1}, \frac{a}{1-a}\}[,$



We generate  $A \in \mathbb{R}^{200 \times 200}$ ,  $B \in \mathbb{R}^{200 \times 50}$ ,  $b \in \mathbb{R}^{200}$ ,  $u \in \mathbb{R}^{50}$  with entries according to a normal distribution and then project  $A$  so that  $\mu_\infty(A) \leq 0.99$

## Summary:

- ① provide a transcription of monotone operator theory for non-Euclidean norms
- ② provable convergence of classical iterations for monotone inclusions and splittings
- ③ application to continuous-time recurrent neural network

## Extensions and open problems:

- ① tightening Lipschitz estimates for operator splittings
- ② infinite-dimensional Banach spaces and set-valued  $F$
- ③ further applications to systems analysis and neural networks

**Thank you for reading!**

**For any questions, please do not hesitate to email me**

Let  $F$  denote a  $\ell_1$  weakly-contracting analytic vector field on a subset  $C$  of  $\mathbb{R}^n$ . Assume there exists a bounded solution  $x(\cdot)$  in  $C$  of  $\dot{x} = F(x)$  defined for all  $t \in \mathbb{R}_{\geq 0}$ . If the function  $t \mapsto \|F(x(t))\|_1$  is constant, then the solution  $x(\cdot)$  is an equilibrium of  $F$ , that is,  $x(t) = x^*$  for all  $t$  and  $F(x^*) \equiv 0$ .

**Proof** For simplicity take  $n = 2$ . By analyticity, and unless  $\|f(x(t))\|_1$  is identically zero (in case we are done), we can pick an interval  $J$  where both  $f_i(x(t))$  have no zeroes, and hence a constant sign. (If one is identically zero, the proof is the same ignoring that variable.) Without loss of generality (take  $-f_i$  if necessary), assume that both have positive sign, so  $\|f(x(t))\|_1 = f_1(x(t)) + f_2(x(t)) = \frac{dx_1}{dt} + \frac{dx_2}{dt} = \frac{d(x_1+x_2)}{dt}$ . Since  $\|f(x)\|_1$  is constant, this means that  $\frac{d(x_1+x_2)}{dt} \equiv c$  on the interval, and therefore  $x_1(t) + x_2(t) = ct + b$  on the interval  $J$ . By analytic continuation, this is true for all  $t \in \mathbb{R}_{\geq 0}$ , contradicting boundedness of  $x(\cdot)$  unless  $c = 0$ . So we have that  $\frac{d(x_1+x_2)}{dt} \equiv 0$ , that is,  $\|f(x(t))\|_1 \equiv 0$ , as desired. The proof for  $\ell_\infty$  norm is even easier - just take an interval where one of the two terms is max.