Perspectives on Contraction Theory and Neural Networks

Francesco Bullo

Center for Control,
Dynamical Systems & Computation
University of California at Santa Barbara

http://motion.me.ucsb.edu

ECE Seminar Series, NYU, Feb 10, 2022

Acknowledgments

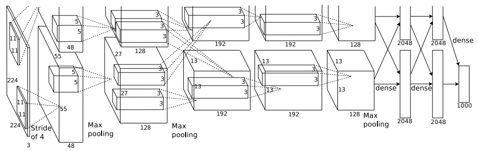
Alex Davydov
PhD student
UC Santa Barbara

Saber Jafarpour
Postdoc
GeorgiaTech

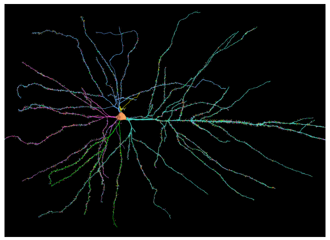
Anton Proskurnikov
Politecnico Torino & Russian
Academy of Sciences

- 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. 2021b. URL **http://arxiv.org/abs/2106.03194**
- A. Davydov, S. Jafarpour, and F. Bullo. Non-Euclidean contraction theory for robust nonlinear stability. *IEEE Transactions on Automatic Control*, July 2021. URL **https://arxiv.org/abs/2103.12263**. Submitted
- CDC 2021 tutorial (**https://arxiv.org/abs/2110.03623**), ACC 2022 (**https://arxiv.org/abs/2110.08298**), L4DC 2022 (**https://arxiv.org/abs/2112.05310**)

Biological and Artificial Neural Networks



artificial neural network (AlexNet '12)



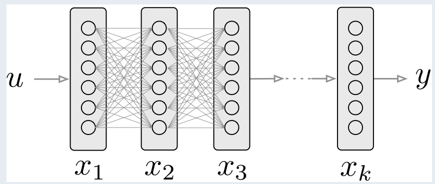
human neocortical neuron

Aim: understand the dynamics of neural networks, so that

- **reproducible behavior, i.e., equilibrium response as function of stimula**
- robust behavior in face of uncertain stimuli
- robust behavior in face of uncertain dynamics
- learning models, efficient computational tools, periodic behaviors ...

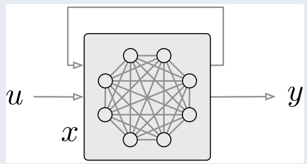
Fixed point computation

Feedforward NN



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

Implicit/Recurrent NN



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

Fixed point strategies in data science = simplifying and unifying framework to model, analyze, and solve advanced convex optimization methods, Nash equilibria, monotone inclusions, etc.

P. L. Combettes and J.-C. Pesquet. Fixed point strategies in data science. *IEEE Transactions on Signal Processing*, 2021.

Outline

1. Scientific and engineering problems from neural networks
2. Contraction theory
   - Banach contractions and infinitesimal counterparts
   - Contraction on Euclidean and inner product spaces
   - Contraction on non-Euclidean normed vector spaces
3. Detour: Network systems
4. Application to recurrent neural networks and implicit ML models
   - Contractivity of recurrent neural networks
   - Implicit neural networks in machine learning
5. Conclusions and future research

Contraction theory: historical notes

- **Origins**

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

  S. M. Lozinskii. Error estimate for numerical integration of ordinary differential equations. I. *Izvestiya Vysshikh Uchebnykh Zavedenii. Matematika*, 5:52–90, 1958

  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.

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

- **Reviews**:

  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. 2014.

  M. Di Bernardo, D. Fiore, G. Russo, and F. Scafuti. Convergence, consensus and synchronization of complex networks via contraction theory. In J. Lü, X. Yu, G. Chen, and W. Yu, editors, *Complex Systems and Networks*, pages 313–339. Springer, 2016. ISBN 978-3-662-47824-0.

  H. Tsukamotoa, S.-J. Chung, and J.-J. E. Slotine. Contraction theory for nonlinear stability analysis and learning-based control: A tutorial overview, 2021. URL **https://arxiv.org/abs/2110.00675**

- contraction conditions without Jacobians have been studied under many different names:
  1. **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.
  2. **one-sided Lipschitz maps** in: E. Hairer, S. P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I. Nonstiff Problems*. Springer, 1993. (Section 1.10, Exercise 6)
  3. **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.
  4. **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.
  5. **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.
  6. **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.
  7. **incremental quadratically stable maps** in: L. D'Alto and M. Corless. Incremental quadratic stability. *Numerical Algebra, Control and Optimization*, 3:175–201, 2013.

- deep connections: infinitesimal contraction, fixed point and monotone operator theory
  1. V. Berinde. *Iterative Approximation of Fixed Points*. Springer, 2007. ISBN 3-540-72233-5
  2. H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, 2 edition, 2017. ISBN 978-3-319-48310-8
  3. E. K. Ryu and W. Yin. *Large-Scale Convex Optimization via Monotone Operators*. Cambridge, 2022

On fixed point algorithms and Banach contractions

$$x = \mathsf{G}(x)$$

Banach Contraction Theorem

If $\mathrm{Lip}(\mathsf{G}) < 1$ that is $\|\mathsf{G}(u) - \mathsf{G}(v)\| \leq \mathrm{Lip}(\mathsf{G})\|u - v\|$,
then *Picard iteration* $x_{k+1} = \mathsf{G}(x_k)$ is a Banach contraction

For $\mathrm{Lip}(\mathsf{G}) \geq 1$, define the *average/damped/Mann-Krasnosel'skii iteration*

$$x_{k+1} = (1 - \alpha)x_k + \alpha\mathsf{G}(x_k)$$

Infinitesimal Contraction Theorem

1. there exists $0 < \alpha < 1$ such that the average iteration is a Banach contraction
2. the map G satisfies $\mathrm{osLip}(\mathsf{G}) < 1$
3. the dynamics $\dot{x} = -x + \mathsf{G}(x)$ is infinitesimally strongly contracting

Robustness of fixed point algorithms

Robustness via Lipschitz constants (Lim's Lemma)

x_u^* is a fixed point of $x = \mathsf{G}(x,u)$ and $\mathsf{Lip}_x\, \mathsf{G} < 1$, then

$$\|x_u^* - x_v^*\| \leq \frac{\mathsf{Lip}_u\, \mathsf{G}}{1 - \mathsf{Lip}_x\, \mathsf{G}} \|u - v\|$$

Robustness via one-sided Lipschitz constants

x_u^* is a fixed point of $x = \mathsf{G}(x,u)$

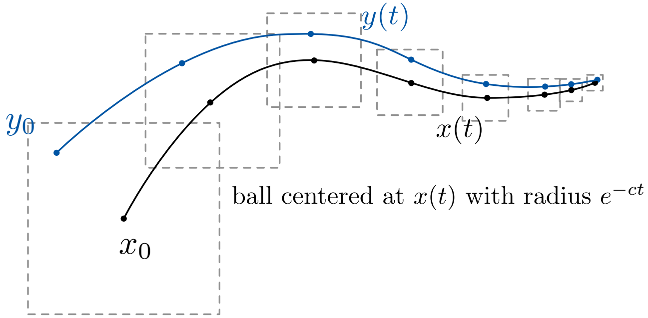
x_v^* is a fixed point of $x = \mathsf{G}(x,v) + \mathsf{D}(x,v)$, and

$\mathsf{osLip}_x(\mathsf{G}+\mathsf{D}) < 1$, then

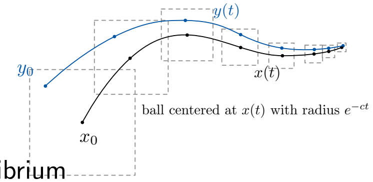
$$\|x_u^* - x_v^*\| \leq \frac{1}{1 - \mathsf{osLip}_x(\mathsf{G}+\mathsf{D})} \Big(\mathsf{Lip}_u(\mathsf{G}+\mathsf{D})\|u - v\| + \|\mathsf{D}(x_u^*, u)\| \Big)$$

On infinitesimal contraction theory

Given $\dot{x} = \mathsf{F}(t,x)$, F is *infinitesimally strongly contractive* if its flow is a Banach contraction



Properties of contracting dynamical systems



Highly ordered **transient** and **asymptotic** behavior:

1. time-invariant F: unique globally exponential stable equilibrium
   two natural Lyapunov functions
2. periodic F: contracting system entrain to periodic inputs
3. contractivity rate is natural measure/indicator of robust stability
4. accurate numerical integration, and
5. there exist efficient methods for their **equilibrium computation**

Scalar maps and vector field

$F : \mathbb{R} \to \mathbb{R}$ is **one-sided Lipschitz** with $\mathsf{osLip}(F) = b$ if

$$\begin{aligned} & F'(x) \leq b, && \forall x \\ \iff \quad & F(x) - F(y) \leq b(x-y), && \forall x > y \\ \iff \quad & (x-y)(F(x) - F(y)) \leq b(x-y)^2, && \forall x, y \end{aligned}$$

- F is osL with $b = 0$ iff F weakly decreasing
- if F is Lipschitz with bound ℓ, then F is osL with $b \leq \ell$
- For
  $$\dot{x} = F(x)$$
  the Grönwall lemma implies $|x(t) - y(t)| \leq \mathrm{e}^{bt}|x(0) - y(0)|$

For $x \in \mathbb{R}^n$ and differentiable time-dep

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

For $P = P^\top \succ 0$, define $\|x\|_{2,P^{1/2}}^2 = x^\top P x$

Main equivalences: For $c > 0$, map **F is c-strongly contracting** (i.e., $\mathrm{osLip}(\mathsf{F}) \leq -c$) if

1. **osL** : $(\mathsf{F}(x) - \mathsf{F}(y))^\top P(x-y) \leq -c\|x-y\|_{2,P^{1/2}}^2$ for all x, y
2. **d-osL** : $P D\mathsf{F}(x) + D\mathsf{F}(x)^\top P \preceq -2cP$ for all x
3. **d-IS** : $D^+\|x(t) - y(t)\|_{2,P^{1/2}} \leq -c\|x(t) - y(t)\|_{2,P^{1/2}}$ for all soltns $x(\cdot), y(\cdot)$

For differentiable $V : \mathbb{R}^n \to \mathbb{R}$, equivalent statements:

1. V is **strongly convex** with parameter m
2. $-\mathrm{grad}V$ is **m-strongly contracting**, that is

$$\big(-\mathrm{grad}V(x) + \mathrm{grad}V(y)\big)^\top (x-y) \leq -m\|x-y\|_2^2$$

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

1. F is a **monotone operator** (or a **coercive operator**) with parameter m,
2. $-\mathsf{F}$ is **m-strongly contracting**

Equilibria of contracting vector fields:
For a time-invariant F, c-strongly contracting with respect to $\|\cdot\|_{2,P^{1/2}}$

1. flow of F is a contraction,
   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

$$x \mapsto \|x - x^*\|_{2,P^{1/2}}^2 \quad \text{and} \quad x \mapsto \|\mathsf{F}(x)\|_{2,P^{1/2}}^2$$

Given $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$

$$x^* \in \mathrm{zero}(\mathsf{F}) \quad \Longleftrightarrow \quad x^* \in \mathrm{fixed}(G), \text{ where } \mathsf{G} = \mathsf{Id} + \mathsf{F}$$

consider **forward step = Euler integration** for F = averaged iteration for G:

$$x_{k+1} = (\mathsf{Id} + \alpha\mathsf{F})x_k = x_k + \alpha\mathsf{F}(x_k) \quad = (1-\alpha)\,\mathsf{Id} + \alpha\mathsf{G}$$

Given *contraction rate c* and *Lipschitz constant ℓ*, define *condition number $\kappa = \ell/c \geq 1$*

1. the map $\mathsf{Id} + \alpha\mathsf{F}$ is a contraction map with respect to $\|\cdot\|_{2,P^{1/2}}$ for

$$0 < \alpha < \frac{2}{c\kappa^2}$$

2. the optimal step size minimizing and minimum contraction factor:

$$\alpha_{\mathsf{E}}^* = \frac{1}{c\kappa^2}$$
$$\ell_{\mathsf{E}}^* = 1 - \frac{1}{2\kappa^2} + \mathcal{O}\Big(\frac{1}{\kappa^4}\Big)$$

Outline

1. Scientific and engineering problems from neural networks
2. Contraction theory
   - Banach contractions and infinitesimal counterparts
   - Contraction on Euclidean and inner product spaces
   - Contraction on non-Euclidean normed vector spaces
3. Detour: Network systems
4. Application to recurrent neural networks and implicit ML models
   - Contractivity of recurrent neural networks
   - Implicit neural networks in machine learning
5. Conclusions and future research

Contraction theory on the normed vector spaces $(\mathbb{R}^n, \ell_1/\ell_\infty)$ 1/5

Norms	From inner products to sign and max pairings	From LMIs to log norms
$\|x\|^2_{2,P^{1/2}} = x^\top P x$	$[\![x,y]\!]_{2,P^{1/2}} = x^\top P y$	$\mu_{2,P^{1/2}}(A) = \min\{b \mid A^\top P + PA \preceq 2bP\}$
$\|x\|_1 = \sum_i \|x_i\|$	$[\![x,y]\!]_1 = \|y\|_1 \operatorname{sign}(y)^\top x$	$\mu_1(A) = \max_j \left(a_{jj} + \sum_{i \neq j} \|a_{ij}\|\right)$
$\|x\|_\infty = \max_i \|x_i\|$	$[\![x,y]\!]_\infty = \max_{i \in I_\infty(y)} y_i x_i$	$\mu_\infty(A) = \max_i \left(a_{ii} + \sum_{j \neq i} \|a_{ij}\|\right)$

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

Generalizing inner products: weak pairings 2/5

A **weak pairing** is $[\![\cdot,\cdot]\!] : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ satisfying

1. $[\![x_1 + x_2, y]\!] \leq [\![x_1, y]\!] + [\![x_2, y]\!]$ and $x \mapsto [\![x, y]\!]$ is continuous,
2. $[\![bx, y]\!] = [\![x, by]\!] = b\,[\![x, y]\!]$ for $b \geq 0$ and $[\![-x, -y]\!] = [\![x, y]\!]$,
3. $[\![x, x]\!] > 0$, for all $x \neq \mathbb{0}_n$,
4. $|\,[\![x, y]\!]\,| \leq [\![x, x]\!]^{1/2}\, [\![y, y]\!]^{1/2}$,

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

Sup of non-Euclidean numerical range: $\mu(A) = \sup_{\|x\|=1} [\![Ax, x]\!]$

Norm derivative formula: $\frac{1}{2} D^+ \|x(t)\|^2 = [\![\dot{x}(t), x(t)]\!]$

Generalizing LMIs: log norms conditions 3/5

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

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

Basic properties:

subadditivity:	$\mu(A+B) \leq \mu(A) + \mu(B)$	
scaling:	$\mu(bA) = b\mu(A),$	$\forall b \geq 0$
convexity:	$\mu(\theta A + (1-\theta)B) \leq \theta\mu(A) + (1-\theta)\mu(B),$	$\forall \theta \in [0,1]$

$$\mu_2(A) \leq -c \iff A + A^\top \preceq -2cI_n$$

$$\mu_\infty(A) \leq -c \iff a_{ii} + \sum_{j \neq i} |a_{ij}| \leq -c \text{ for all } i$$

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

Contraction theory on the normed vector spaces $(\mathbb{R}^n, \ell_1/\ell_\infty)$ 4/5

For $x \in \mathbb{R}^n$ and differentiable time-dep

$$\dot{x} = \mathsf{F}(x) \qquad (1)$$

For norm $\|\cdot\|$ with log norm $\mu(\cdot)$ and compatible weak pairing $[\![\cdot,\cdot]\!]$

Main equivalences: for $c > 0$

1. **osL** : $[\![\mathsf{F}(x) - \mathsf{F}(y), x - y]\!] \leq -c\|x - y\|^2$ for all x, y
2. **d-osL** : $\mu(D\mathsf{F}(x)) \leq -c$ for all x
3. **d-IS** : $D^+\|x(t) - y(t)\| \leq -c\|x(t) - y(t)\|$ for soltns $x(\cdot), y(\cdot)$

A. Davydov, S. Jafarpour, and F. Bullo. Non-Euclidean contraction theory for robust nonlinear stability. *IEEE Transactions on Automatic Control*, July 2021. URL **https://arxiv.org/abs/2103.12263**. Submitted

Contraction theory on the normed vector spaces $(\mathbb{R}^n, \ell_1/\ell_\infty)$ 5/5

Consider a norm $\|\cdot\|$ with compatible weak pairing $[\![\cdot,\cdot]\!]$
Recall **forward step method** $x_{k+1} = (\mathsf{Id} + \alpha\mathsf{F})x_k = x_k + \alpha\mathsf{F}(x_k)$

Given *contraction rate c* and *Lipschitz constant ℓ*, define *condition number $\kappa = \ell/c \geq 1$*

1. the map $\mathsf{Id} + \alpha\mathsf{F}$ is a contraction map with respect to $\|\cdot\|$ for

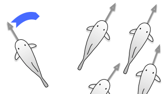
$$0 < \alpha < \frac{1}{c\kappa(1+\kappa)}$$

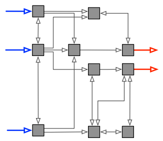
2. the optimal step size minimizing and minimum contraction factor:

$$\alpha^*_{\mathsf{nE}} = \frac{1}{c}\Big(\frac{1}{2\kappa^2} - \frac{3}{8\kappa^3} + \mathcal{O}\Big(\frac{1}{\kappa^4}\Big)\Big)$$
$$\ell^*_{\mathsf{nE}} = 1 - \frac{1}{4\kappa^2} + \frac{1}{8\kappa^3} + \mathcal{O}\Big(\frac{1}{\kappa^4}\Big)$$

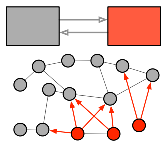
Outline

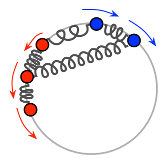
1. Scientific and engineering problems from neural networks
2. Contraction theory
   - Banach contractions and infinitesimal counterparts
   - Contraction on Euclidean and inner product spaces
   - Contraction on non-Euclidean normed vector spaces
3. Detour: Network systems
4. Application to recurrent neural networks and implicit ML models
   - Contractivity of recurrent neural networks
   - Implicit neural networks in machine learning
5. Conclusions and future research

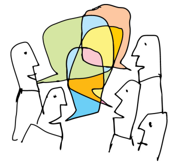

averaging

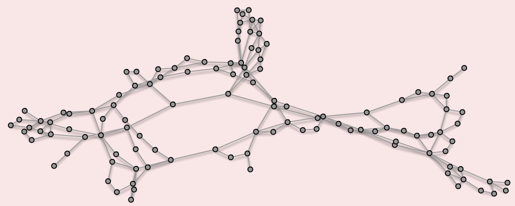

compartmental flows


mutualism


virus spread


coupled oscillators


social power



network structure $\Longleftrightarrow$ **function = dynamic behavior**

Control theories: general Lyapunov theory, passivity/dissipativity, monotone dynamics ...

Networks of contracting systems

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

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

- **osL**: $x_i \mapsto f_i(x_i, x_{-i})$ is infinitesimally strongly contracting with rate c_i
- **Lip**: $x_{-i} \mapsto f_i(x_i, x_{-i})$ is Lipschitz: $\|f_i(x_i,x_{-i}) - f_i(x_i,y_{-i})\|_i \le \sum_{j\neq i} \gamma_{ij}\|x_j - y_j\|_j$
- the gain matrix $\begin{bmatrix} -c_1 & \dots & \gamma_{1n} \\ \vdots & & \vdots \\ \gamma_{n1} & \dots & -c_n \end{bmatrix}$ is **Metzler Hurwitz**

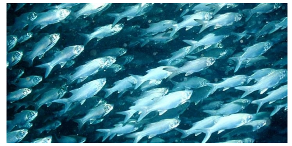
$\implies$ the **interconnected system** is infinitesimally strongly contracting

A. Davydov, S. Jafarpour, and F. Bullo. Non-Euclidean contraction theory for robust nonlinear stability. *IEEE Transactions on Automatic Control*, July 2021. URL **https://arxiv.org/abs/2103.12263**. Submitted

Contraction theory for networks

Challenge: many real-world networks are not contracting.





For a vector field F and positive vectors $\eta, \xi \in \mathbb{R}^n_{>0}$,

conservation law	$\eta^\top f(x) = \eta^\top f(y)\ \forall x,y$	$\iff$	$\eta^\top DF(x) = 0\ \forall x$
translation invariance	$f(x+\alpha\xi) = f(x)\ \forall x,\alpha$	$\iff$	$DF(x)\xi = 0\ \forall x$

If F satisfies a conservation law or translation invariance, then

1. $\mathrm{osLip}(f) \ge 0$
2. if additionally F is monotone, then $\mathrm{osLip}_{1,[\eta]}(f) = 0$ or $\mathrm{osLip}_{\infty,[\xi]^{-1}}(f) = 0$

Weakly contracting systems

$\dot{x} = f(x)$ is **weakly contracting** wrt $\|\cdot\|$:

$$\mathrm{osLip}(f) \le 0$$

1. Lotka-Volterra population dynamics (Lotka, 1920; Volterra, 1928) (**ℓ_1-norm for mutualistic**)
2. Kuramoto oscillators (Kuramoto, 1975) and coupled swing equations (Bergen and Hill, 1981) (**ℓ_1-norm and ℓ_∞-norm**)
3. Daganzo's cell transmission model for traffic networks (Daganzo, 1994), (**ℓ_1-norm for non-FIFO intersection**)
4. compartmental systems in biology, medicine, and ecology (Sandberg, 1978; Maeda et al., 1978). (**ℓ_1-norm**)
5. saddle-point dynamics for optimization of weakly-convex functions (Arrow et al., 1958). (**ℓ_2-norm**)

Semi-contracting systems

$\dot{x} = f(x)$ is **semi-contracting** wrt the semi-norm $|||\cdot|||$ with rate $c > 0$:

$$\mathrm{osLip}_{|||\cdot|||}(f) \le -c$$

or, for differentiable systems, $\mu_{|||\cdot|||}(DF(x)) \le -c$

1. Kuramoto oscillators (Kuramoto, 1975) and coupled swing equations (Bergen and Hill, 1981), (**ℓ_1-norm**)
2. Chua's diffusively-coupled circuits (Wu and Chua, 1995), (**ℓ_2-norm**)
3. morphogenesis in developmental biology (Turing, 1952), (**ℓ_1-norm, over some param ranges**)
4. Goodwin model for oscillating auto-regulated gene (Goodwin, 1965). (**ℓ_1-norm, over some param ranges**)

S. Jafarpour, P. Cisneros-Velarde, and F. Bullo. Weak and semi-contraction for network systems and diffusively-coupled oscillators. *IEEE Transactions on Automatic Control*, 2021a. . To appear

k and α-contracting systems

- M. Y. Li and J. S. Muldowney. A geometric approach to global-stability problems. *SIAM Journal on Mathematical Analysis*, 27(4):1070–1083, 1996.
- C. Wu, I. Kanevskiy, and M. Margaliot. k-contraction: Theory and applications. *Automatica*, 136:110048, 2022.
- C. Wu, R. Pines, M. Margaliot, and J.-J. E. Slotine. Generalization of the multiplicative and additive compounds of square matrices and contraction in the Hausdorff dimension, Dec. 2020. Available at **https://arxiv.org/abs/2012.13441**

Outline

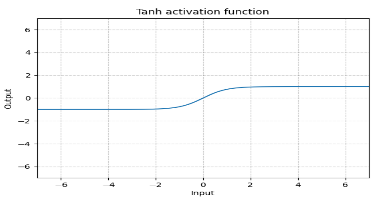
1. Scientific and engineering problems from neural networks
2. Contraction theory
   - Banach contractions and infinitesimal counterparts
   - Contraction on Euclidean and inner product spaces
   - Contraction on non-Euclidean normed vector spaces
3. Detour: Network systems
4. Application to recurrent neural networks and implicit ML models
   - Contractivity of recurrent neural networks
   - Implicit neural networks in machine learning
5. Conclusions and future research

Applications to recurrent neural networks

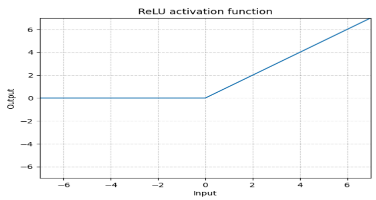
Continuous-time recurrent neural networks:

$$\dot{x} = -x + A\Phi(x) + u \qquad \text{(*Hopfield*)}$$
$$\dot{x} = -x + \Phi(Ax + u) =: f_{\mathsf{FR}}(x) \qquad \text{(*Firing rate* $\sim$ *Implicit NNs*)}$$
$$\dot{x} = A\Phi(x) \qquad \text{(*Persidskii-type*)}$$
$$\dot{x} = Ax - \Phi(x) \qquad (\dots)$$

sigmoid, hyperbolic tangent



ReLU $= \max\{x, 0\} = (x)_+$



activation functions are locally-Lip and slope-restricted: for all i

$$d_{\mathsf{min}} := \operatorname{ess\,inf}_{y\in\mathbb{R}} \frac{\partial \Phi_i(y)}{\partial y} \geq 0 \quad \text{and} \quad d_{\mathsf{max}} := \operatorname{ess\,sup}_{y\in\mathbb{R}} \frac{\partial \Phi_i(y)}{\partial y} < \infty$$

Tight transcription. $Df_{\mathsf{FR}}(x) = -I_n + (D\Phi(x))A$ a.e., and so

$$\mathsf{osLip}_\infty(f_{\mathsf{FR}}) = \operatorname*{ess\,sup}_{x\in\mathbb{R}^n} \mu_\infty\big(-I_n + (D\Phi(x))A\big) = -1 + \max_{d\in[d_{\mathsf{min}},d_{\mathsf{max}}]^n} \mu_\infty(\mathsf{dg}(d)A)$$

Max log norms over hypercubes. For $A \in \mathbb{R}^{n\times n}$ and $0 \leq d_{\mathsf{min}} \leq d_{\mathsf{max}}$

$$\max_{d\in[d_{\mathsf{min}},d_{\mathsf{max}}]^n} \mu_1(\mathsf{dg}(d)A) = \max\{\mu_1(d_{\mathsf{max}}A), \mu_1(d_{\mathsf{max}}A - (d_{\mathsf{max}} - d_{\mathsf{min}})(I_n \circ A))\}$$
$$\max_{d\in[d_{\mathsf{min}},d_{\mathsf{max}}]^n} \mu_\infty(\mathsf{dg}(d)A) = \max\big\{\mu_\infty(d_{\mathsf{min}}A), \mu_\infty(d_{\mathsf{max}}A)\big\}$$

Recall: max convex function over polytope achieved at a vertex; here $2^n \to 2$ vertices only.

NonEuclidean contractivity of firing rate model

$$\dot{x} = -Cx + \Phi(Ax+u) =: f_{\mathsf{FR}}(x)$$

1. for arbitrary $\eta \in \mathbb{R}^n_{>0}$

$$\mathsf{osLip}_{\infty,[\eta]^{-1}}(f_{\mathsf{FR}}) = \max\{\mu_{\infty,[\eta]^{-1}}(-C + d_{\mathsf{min}}A), \mu_{\infty,[\eta]^{-1}}(-C + d_{\mathsf{max}}A)\}$$

2. optimal weight η and minimim value of $\mathsf{osLip}_{\infty,[\eta]^{-1}}(f_{\mathsf{FR}})$ from quasiconvex opt:

$$\begin{aligned} \inf_{b\in\mathbb{R},\eta\in\mathbb{R}^n_{>0}} \quad & b \\ \text{s.t.} \quad & (-C + d_{\mathsf{min}}|A|_{\mathsf{M}})\eta \leq b\eta \\ & (-C + d_{\mathsf{max}}|A|_{\mathsf{M}})\eta \leq b\eta \end{aligned}$$

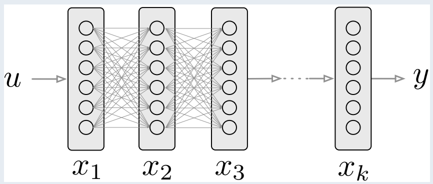
3. if $d_{\mathsf{min}} = 0$ and $C \succ 0$, let $v_* \in \mathbb{R}^n_{>0}$ be right eigenvector of $-C + d_{\mathsf{max}}|A|_{\mathsf{M}}$,

$$\inf_{\eta\in\mathbb{R}^n_{>0}} \mathsf{osLip}_{\infty,[\eta]}(f_{\mathsf{FR}}) = \mathsf{osLip}_{\infty,[v_*]^{-1}}(f_{\mathsf{FR}}) = \max\left\{\alpha(-C), \alpha(-C + d_{\mathsf{max}}|A|_{\mathsf{M}})\right\}.$$

A. Davydov, A. V. Proskurnikov, and F. Bullo. Non-Euclidean contractivity of recurrent neural networks. In *American Control Conference*, 2022. URL **https://arxiv.org/abs/2110.08298**. To appear

Implicit neural networks in machine learning

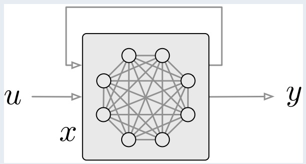
Feedforward NN



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

$y = Cx_k + d$

Implicit/Recurrent NN



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

$y = Cx + d$

ML advantages of implicit/equilibrium/fixed point formulation:

bio-inspired, simplicity, accuracy, memory efficiency, input-output robustness

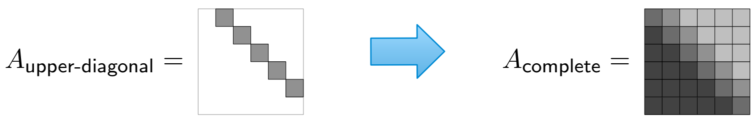
S. Jafarpour, A. Davydov, A. V. Proskurnikov, and F. Bullo. Robust implicit networks via non-Euclidean contractions. In *Advances in Neural Information Processing Systems*, Dec. 2021b. URL **http://arxiv.org/abs/2106.03194**

Outline

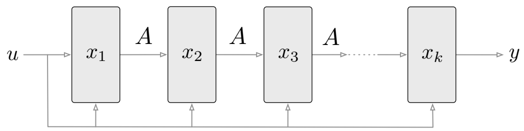
1. Scientific and engineering problems from neural networks
2. Contraction theory
   - Banach contractions and infinitesimal counterparts
   - Contraction on Euclidean and inner product spaces
   - Contraction on non-Euclidean normed vector spaces
3. Detour: Network systems
4. Application to recurrent neural networks and implicit ML models
   - Contractivity of recurrent neural networks
   - Implicit neural networks in machine learning
5. Conclusions and future research

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

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



Motivation #2: Weight-tied infinite-depth NN → fixed-point of INN



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

Motivation #3: Neural ODE model (infinite time) → fixed-point of INN

$\dot{x} = -x + \Phi(Ax + Bu + b) \quad \Longrightarrow \quad \lim_{t\to\infty} x(t) = x^*$ solution to INN

Recent literature on implicit NNs

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.
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-NZIuwqhI4**
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

Implicit Neural Networks (INNs)

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

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

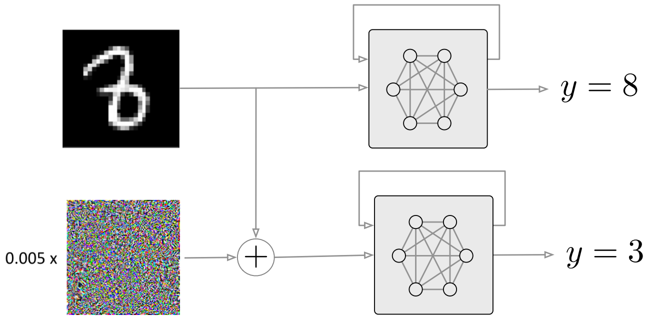
- Efficient back-propagation through implicit differentiation
- Stochastic gradient descent: at each step solve $x = \Phi(Ax + Bu + b)$.

Challenge #1: well-posedness of fixed-point equation

Challenge #2: algorithm for fixed-point equation

Robustness of INNs

Adversarial examples: small input change causes large output change!



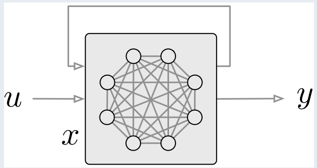
Robustness measures: input-to-output Lipschitz constant

1. ℓ_2-**norm Lipschitz constant**: not informative in many scenarios
2. ℓ_∞-**norm Lipschitz constant**: large-scale input wrt wide-spread perturbations

Challenge #3: compute robustness margins

Challenge #4: implement robustness in training

Well-posedness and robustness of ℓ_∞-contracting INNs



$$x = \Phi(Ax + Bu + b) \qquad (\textit{INN fixed point})$$
$$\dot{x} = -x + \Phi(Ax + Bu + b) \qquad (\textit{Recurrent NN})$$
$$x_{k+1} = (1-\alpha)x_k + \alpha\Phi(Ax_k + Bu + b) \qquad (\textit{Average iter.n})$$

If

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

- dynamics is contracting with rate $1 - \mu_\infty(A)_+$
- average iteration is Banach with factor $1 - \dfrac{1-\mu_\infty(A)_+}{1-\min_i(a_{ii})_-}$ at $\alpha = \dfrac{1}{1-\min_i(a_{ii})_-}$
- input-output Lipschitz constant $\mathrm{Lip}_{u\to y} = \dfrac{\|B\|_\infty \|C\|_\infty}{1-\mu_\infty(A)_+}$

Training INNs

Training optimization problem:

$$\min_{A,B,C,b} \quad \sum_{i=1}^{N} \mathcal{L}(\widehat{y}_i, Cx_i + c) + \quad \lambda \quad \mathsf{Lip}_{u \to y}$$
$$x_i = \Phi(Ax_i + B\widehat{u}_i + b)$$
$$\mu_\infty(A) \le \gamma$$

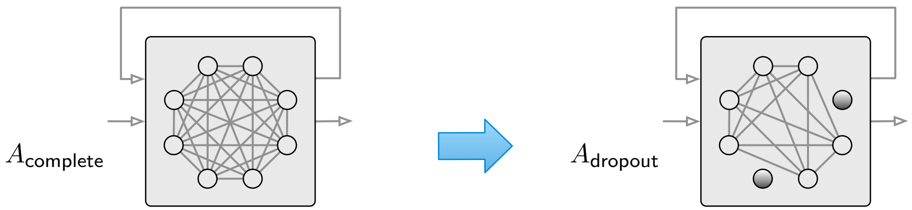
- $\lambda \ge 0$ is a regularization parameter
- $\gamma < 1$ is a hyperparameter

Parametrization of μ_∞ constraint:

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

Graph-Theoretic Regularization

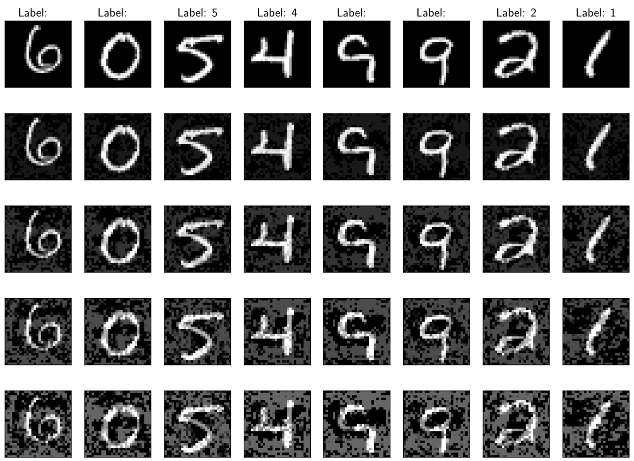
Synaptic matrix A encodes interactions between neurons



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

Numerical Experiments

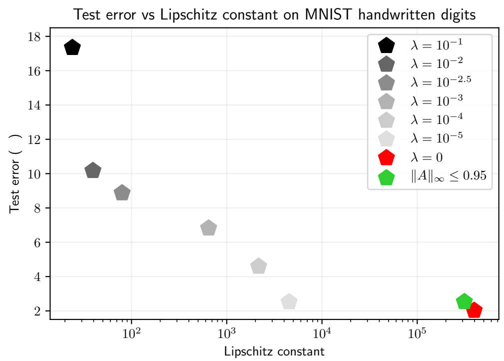
- MNIST handwritten digit dataset (60K+10K, 28x28, grayscale)
- implicit neural network order: $n = 100$

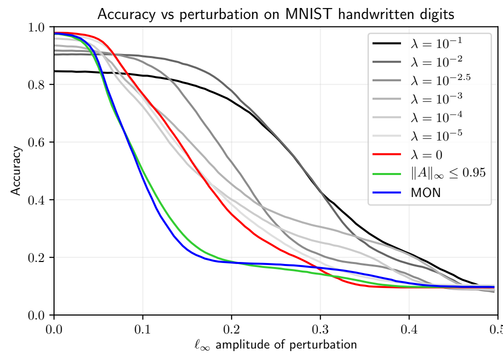


Numerical Experiments

Robustness of INNs

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

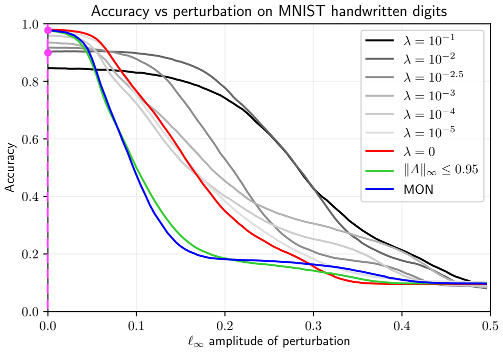


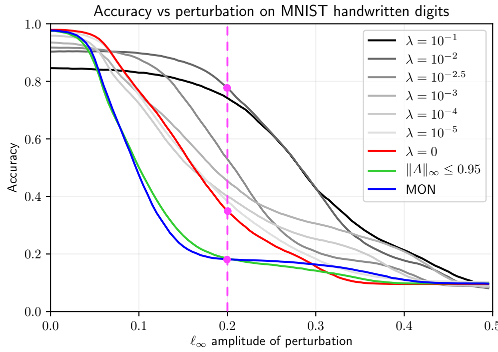


- Pareto-optimal curve
- Clean performance vs. robustness

Numerical Experiments
Robustness of INNs

Clean performance vs. robustness





Outline

1. Scientific and engineering problems from neural networks
2. Contraction theory
   - Banach contractions and infinitesimal counterparts
   - Contraction on Euclidean and inner product spaces
   - Contraction on non-Euclidean normed vector spaces
3. Detour: Network systems
4. Application to recurrent neural networks and implicit ML models
   - Contractivity of recurrent neural networks
   - Implicit neural networks in machine learning
5. Conclusions and future research

Conclusions

From Contracting Dynamics to Contracting Algorithms:

1. contraction theory, monotone operator theory, convex optimization
   - effective methodologies to tackle control, optimization and learning problems
   - extensions to network dynamics
2. from Euclidean to non-Euclidean norms
3. application to recurrent and implicit neural networks
   - existence, uniqueness, and computation of fixed-points
   - robustness analysis and robust training via Lipschitz bounds
   - **https://github.com/davydovalexander/Non-Euclidean_Mon_Op_Net**

From Contracting Dynamics to Contracting Algorithms:

1. mixed-monotone contraction theory (**https://arxiv.org/abs/2112.05310**)
2. implicit graph neural architectures
3. bio-inspired Hebbian learning
4. robustness of implicit models

Supplementary slides

Background on Infinitesimal Contraction Theorem

1. there exists $0 < \alpha < 1$ such that the average iteration is a Banach contraction
2. the map G satisfies $\mathrm{osLip}(\mathsf{G}) < 1$
3. the dynamics $\dot{x} = \mathsf{F}(x) := -x + \mathsf{G}(x)$ is infinitesimally contracting

- the equivalence (2) $\iff$ (3) is just a transcription:
  - $\mathsf{F} = -\mathrm{Id} + \mathsf{G}$ contracting with rate $c \iff \mathrm{osLip}(\mathsf{F}) < -c \iff \mathrm{osLip}(\mathsf{G}) < 1 - c$, for $c > 0$
  - in (ℓ_2, P), $\mathrm{osLip}(\mathsf{F}) < -c$ is usual Krasovskii: $PJ(x) + J(x)^\top P \preceq -2cP$ for all x and $J = D\mathsf{F}$
- (2) $\implies$ (1): known in monotone operator theory (page 15 "forward step method" in[1])
  - vector field F is contracting with rate $c \iff -\mathsf{F}$ is strongly monotone with parameter c
- Theorem 1 in[2] proves the equivalence (1) $\iff$ (2) for any norm, i.e., the implication (2) $\implies$ (1) for any norm (with proper osLip definitions) and the converse direction (1) $\implies$ (2) for ℓ_2, P. Theorem 3 in[2] proves the one-sided Lim Lemma (see next slide).

[1] E. K. Ryu and S. Boyd. Primer on monotone operator methods. *Applied Computational Mathematics*, 15(1):3–43, 2016

[2] 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. 2021b. URL http://arxiv.org/abs/2106.03194

Euclidean vs. non-Euclidean contractions

Most foundational results in systems theory are based on ℓ_2 linear-quadratic theory; their ℓ_1/ℓ_∞ analogs are yet to be worked out.

Advantages of non-Euclidean approach

1. *computational advantages*: non-Euclidean log-norm constraints lead to LPs, whereas ℓ_2 constraints leads to LMIs. Parametrization of log-norm constrained matrices is polytopic.

   A. Rantzer. Scalable control of positive systems. *European Journal of Control*, 24:72–80, 2015.

2. *guaranteed robustness to structural perturbations*: ℓ_∞ contractivity ensures:
   1. absolute contractivity = with respect to a class of activation functions
   2. total contractivity = remove any node and all its incident connections
   3. connective contractivity = remove any set of edges

3. *adversarial input-output analysis*
   ℓ_∞ better suited for the analysis of adversarial examples than ℓ_2: in high dimensions, large inner product between two vectors is possible even when one vector has small ℓ_∞ norm

   I. J. Goodfellow, J. Shlens, and C. Szegedy. Explaining and harnessing adversarial examples. In *International Conference on Learn Representations (ICLR)*, 2015. URL https://arxiv.org/abs/1412.6572

Literature on recurrent NN ODEs

1. J. J. Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences*, 81(10):3088–3092, 1984.
2. E. Kaszkurewicz and A. Bhaya. On a class of globally stable neural circuits. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 41(2):171–174, 1994.
3. M. Forti, S. Manetti, and M. Marini. Necessary and sufficient condition for absolute stability of neural networks. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 41(7):491–494, 1994.
4. Y. Fang and T. G. Kincaid. Stability analysis of dynamical neural networks. *IEEE Transactions on Neural Networks*, 7(4):996–1006, 1996.
5. 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.
6. W. He and J. Cao. Exponential synchronization of chaotic neural networks: a matrix measure approach. *Nonlinear Dynamics*, 55:55–65, 2009.
7. H. Zhang, Z. Wang, and D. Liu. A comprehensive review of stability analysis of continuous-time recurrent neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 25(7): 1229–1262, 2014.

Contractivity conditions with respect to arbitrary norms

Log norm bound	Demidovich condition	One-sided Lipschitz condition
$\mu_{2,P}(D\mathsf{F}(x)) \leq b$	$PD\mathsf{F}(x) + D\mathsf{F}(x)^\top P \preceq 2bP$	$(x-y)^\top P\big(\mathsf{F}(x) - \mathsf{F}(y)\big) \leq b\lVert x-y\rVert^2_{P^{1/2}}$
$\mu_p(D\mathsf{F}(x)) \leq b$	$(v \circ \lvert v\rvert^{p-2})^\top D\mathsf{F}(x)v \leq b\lVert v\rVert_p^p$	$((x-y)\circ\lvert x-y\rvert^{p-2})^\top(\mathsf{F}(x) - \mathsf{F}(y)) \leq b\lVert x-y\rVert_p^p$
$\mu_1(D\mathsf{F}(x)) \leq b$	$\mathrm{sign}(v)^\top D\mathsf{F}(x)v \leq b\lVert v\rVert_1$	$\mathrm{sign}(x-y)^\top(\mathsf{F}(x) - \mathsf{F}(y)) \leq b\lVert x-y\rVert_1$
$\mu_\infty(D\mathsf{F}(x)) \leq b$	$\max_{i\in I_\infty(v)} v_i\,(D\mathsf{F}(x)v)_i \leq b\lVert v\rVert_\infty^2$	$\max_{i\in I_\infty(x-y)}(x_i - y_i)(f_i(x) - f_i(y)) \leq b\lVert x-y\rVert_\infty^2$

Table of equivalences between measure bounded Jacobians, differential Demidovich and one-sided Lipschitz conditions. Note: $I_\infty(v) = \{i \in \{1,\dots,n\} \mid \lvert v_i\rvert = \lVert v\rVert_\infty\}$.

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.

Robustness to unmodeled dynamics

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

$$\dot{x} = f(x) + g(x) \tag{2}$$

If F has one-sided Lipschitz constant $-c < 0$ and
g has one-sided Lipschitz constant $d > 0$, then

1. **(contractivity under perturbations)** if $d < c$, then $f + g$ is strongly contracting with rate $c - d$,
2. **(equilibrium point under perturbations)** if additionally F and g are time-invariant, then the unique equilibrium point x^* of F and x^{**} of $f + g$ satisfy

$$\|x^* - x^{**}\| \leq \frac{\|g(x^*)\|}{c - d} \tag{3}$$

Metzler matrices and monotone systems

- For Metzler M and monotonic $\|\cdot\|$, $\mu(M) = \sup_{x \geq \mathbb{0}_n} \frac{[\![Ax, x]\!]}{\|x\|}$.
- For $\eta, \xi \in \mathbb{R}^n_{>0}$,

$$\mu_{1,[\eta]}(M) = \max(\eta^\top M[\eta]^{-1}) = \min\{b \in \mathbb{R} \mid \eta^\top M \leq b\eta^\top\}$$
$$\mu_{\infty,[\xi]^{-1}}(M) = \max([\xi]^{-1} M\xi) = \min\{b \in \mathbb{R} \mid M\xi \leq b\xi\}$$

F **monotone** if $DF(x)$ Metzler for all x

1. **osL** : $[\![f(x) - f(y), x - y]\!] \leq b\|x - y\|^2$ for all $x \geq y$
2. **d-osL** : $[\![DF(x)v, v]\!] \leq b\|v\|^2$, for all $v \geq 0$ and x

$\mu_{1,[\eta]}(DF(x)) \leq b$	$\eta^\top DF(x) \leq b\eta^\top$	$\eta^\top(f(x) - f(y)) \leq b\eta^\top(x - y)$ for all $x \geq y$
$\mu_{\infty,[\xi]^{-1}}(DF(x)) \leq b$	$DF(x)\xi \leq b\xi$	$f(x) - f(y) \leq b(x - y)$ for all $x = y + \beta\xi, \beta > 0$

Input-state stability and gain of contracting systems 1/3

For a time and input-dependent vector F,

$$\dot{x} = f(x, u(t)), \qquad x(0) = x_0 \in \mathbb{R}^n, u(t) \in \mathbb{R}^k \tag{4}$$

Assume $\|\cdot\|_{\mathcal{X}}$ with compatible $[\![\cdot,\cdot]\!]_{\mathcal{X}}$, a norm $\|\cdot\|_{\mathcal{U}}$, and $c, \ell > 0$ such that

- **osL**: $[\![f(x,u) - f(y,u), x - y]\!]_{\mathcal{X}} \leq -c\|x - y\|^2_{\mathcal{X}}$, for all x, y, u,
- **Lip**: $\|f(x,u) - f(x,v)\|_{\mathcal{X}} \leq \ell\|u - v\|_{\mathcal{U}}$, for all x, u, v.

Input-state stability and gain of contracting systems 2/3

Then

1. any two soltns $x(t)$ and $y(t)$ to (4) with inputs u_x, 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 input-to-state stable**, i.e., for all x_0, y_0

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

3. F has **incremental $\mathcal{L}^q_{\mathcal{X},\mathcal{U}}$ gain equal to ℓ/c, for $q \in [1, \infty]$**,

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

Signal norms and system gains 3/3

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

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

$$\|x(\cdot)\|_{\mathcal{X},q} = \begin{cases} \Big(\int_0^\infty \|x(t)\|^q_{\mathcal{X}} dt \Big)^{1/q} & \text{if } q \in [1,\infty[\\ \sup_{t\geq 0} \|x(t)\|_{\mathcal{X}} & \text{if } q = \infty \end{cases} \qquad (5)$$

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

$$\|x(\cdot)\|_{\mathcal{X},q} \leq \gamma \, \|u(\cdot)\|_{\mathcal{U},q} \qquad (6)$$

Incremental ISS for strongly contracting delay ODEs

$$\dot{x}(t) = f\big(x(t), x(t-s), u(t)\big), 0 \leq s \leq S, \qquad \|\cdot\|_{\mathcal{X}}, \|\cdot\|_{\mathcal{U}} \qquad (7)$$

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

osL x : $\quad [\![f(x,d,u) - f(y,d,u), x-y]\!]_{\mathcal{X}} \leq -c\|x-y\|^2_{\mathcal{X}} \qquad (8)$

Lip $x(t-s)$: $\quad \|f(x,x_1,u) - f(x,x_2,u)\|_{\mathcal{X}} \leq \ell_{\mathcal{X}} \|x_1 - x_2\|_{\mathcal{X}} \qquad (9)$

Lip u : $\quad \|f(x,d,u) - f(x,d,v)\|_{\mathcal{X}} \leq \ell_{\mathcal{U}} \|u - v\|_{\mathcal{U}} \qquad (10)$

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

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

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

$$D^+ m(t) \leq -cm(t) + \ell_{\mathcal{X}} \sup_{0\leq s\leq S} m(t-s) + \ell_{\mathcal{U}} \|u_x(t) - u_y(t)\|_{\mathcal{U}}, \qquad (11)$$

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

$$m(t) \leq m_0 \mathrm{e}^{-\rho(t-t_0)} + \ell_{\mathcal{U}} \int_{t_0}^t \mathrm{e}^{-\rho(t-\tau)} \|u_x(\tau) - u_y(\tau)\|_{\mathcal{U}} d\tau, \qquad (12)$$

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

Networks of contracting systems with time delays

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

$$\dot{x}_i = f_i(x_i, x_{-i}, x_{-i}(t-s), u_i), \qquad 0 \leq s \leq S, \qquad \|\cdot\|_i, \|\cdot\|_{i,\mathcal{U}} \qquad (13)$$

Assume there exist positive constants st

osL x_i : $\quad [\![f_i(x_i,\dots) - f_i(y_i,\dots), x_i - y_i]\!]_i \leq -c_i \|x_i - y_i\|_i^2$

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

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

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

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

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

and, if the Metzler matrix $-C + \Gamma + \widehat{\Gamma}$ is Hurwitz, then (13) is incremental ISS

F. Mazenc, M. Malisoff, and M. Krstic. Vector extensions of Halanay's inequality. *IEEE Transactions on Automatic Control*, 2021. to appear

Networks of ISS systems

Interconnections scalar ISS subsystems

$$\dot{x}_i = -a_i(x_i) + \sum_{j\neq i} \gamma_{ij}(x_j) + u_i, \qquad \text{for } i \in \{1,\dots,n\}. \qquad (14)$$

where a_i are of class $\mathcal{K}_\infty$ and γ_{ij} are of class $\mathcal{K}$. Define

$$A_i(x) = a_i(x_i), \qquad \text{and } \Gamma_i(x) = \sum_{j\neq i} \gamma_{ij}(x_j)$$

If there exist $\eta \in \mathbb{R}^n_{>0}$ and $c > 0$ satisfying

$$\eta^\top \big(A(v) - A(w)\big) \geq \eta^\top \big(\Gamma(v) - \Gamma(w) + c(v-w)\big), \qquad \text{for all } v \geq w \geq \mathbb{0}_n$$

then the **interconnected system** is strongly contracting with respect to $\|\cdot\|_{1,[\eta]}$ and with rate c

Proof: $\mathrm{osLip}_{1,[\eta]}(f) \leq b$ if and only if $\eta^\top\big(f(x) - f(y)\big) \leq b\eta^\top(x-y)$