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

Prospective Centre of Excellence of
Network Systems Learning, Control, and Evolution.
IIT Madras, March 15, 2022

Acknowledgments

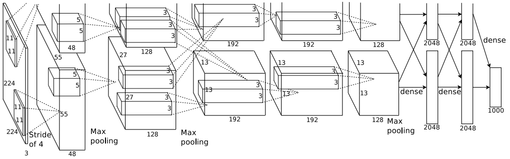
Alex Davydov
PhD student
UC Santa Barbara

Saber Jafarpour
Postdoc
GeorgiaTech

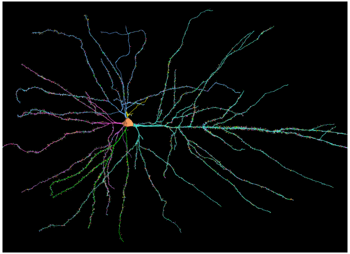
Anton Proskurnikov
Politecnico Torino & Russian
Academy of Sciences

- 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. 2021. 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**. Conditionally accepted as Paper
- 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)



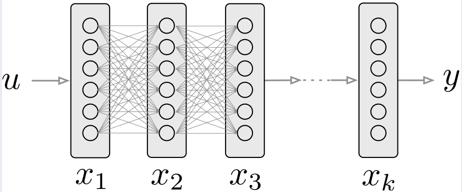
human neocortical neuron

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 ...

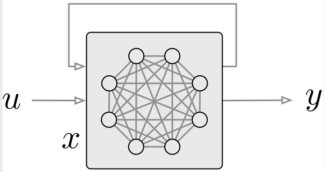
Fixed point computation

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

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. URL **http://mi.mathnet.ru/eng/ivm2980**. (in Russian)

  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. 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.

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

For $\mathsf{Lip}(\mathsf{G}) \geq 1$, 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 $\mathsf{osLip}(\mathsf{G}) < 1$
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 $\operatorname{Lip}_x \mathsf{G} < 1$, then

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

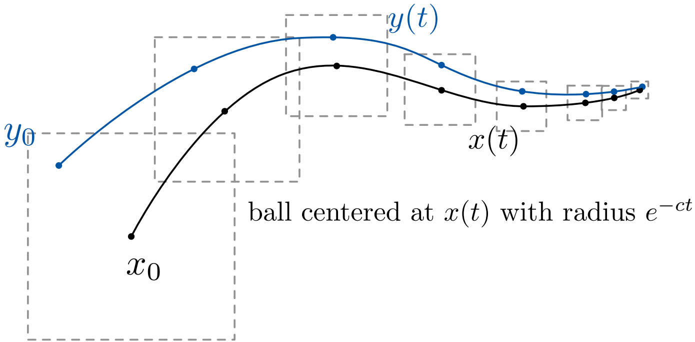
Robustness via one-sided Lipschitz constants

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

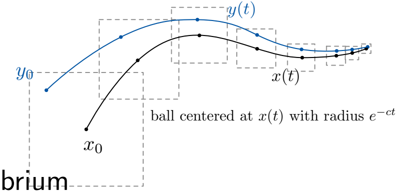
$$\|x_u^* - x_v^*\| \le \frac{1}{1 - \operatorname{osLip}_x(\mathsf{G}+\mathsf{D})} \Big(\operatorname{Lip}_u(\mathsf{G}+\mathsf{D}) \|u - v\| + \|\mathsf{D}(x_u^*, u)\| \Big)$$

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 $\text{osLip}(F) = b$ 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., osLip(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 $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$

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

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

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

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

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

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

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.

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

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

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

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**. Conditionally accepted as Paper

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

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

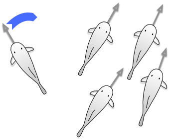
1. the map $\mathsf{Id} + \alpha \mathsf{F}$ is a contraction map with respect to $\|\cdot\|$ for

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

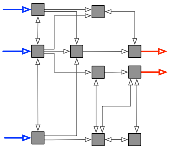
2. the optimal step size minimizing and minimum contraction factor:

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

Outline

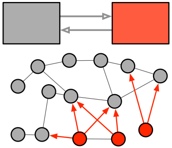


averaging

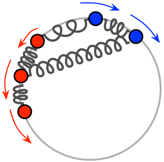


compartmental flows

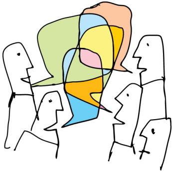
mutualism



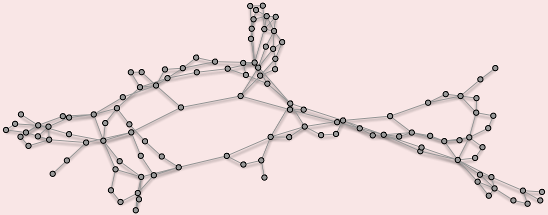
virus spread



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 $x_{-i} \in \mathbb{R}^{N-N_i}$:

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

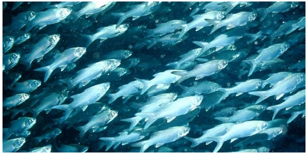
- **osL**: $x_i \mapsto f_i(x_i, x_{-i})$ is infinitesimally strongly contracting with rate c_i
- **Lip**: $x_{-i} \mapsto f_i(x_i, x_{-i})$ is Lipschitz: $\|f_i(x_i, x_{-i}) - f_i(x_i, y_{-i})\|_i \leq \sum_{j \neq i} \gamma_{ij} \|x_j - y_j\|_j$
- the gain matrix $\begin{bmatrix} -c_1 & \dots & \gamma_{1n}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

$\Longrightarrow$ 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**. Conditionally accepted as Paper

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	$\eta^\top f(x) = \eta^\top f(y) \;\; \forall x, y$	$\iff$	$\eta^\top D\mathsf{F}(x) = 0 \;\; \forall x$
translation invariance	$f(x + \alpha\xi) = f(x) \;\; \forall x, \alpha$	$\iff$	$D\mathsf{F}(x)\xi = 0 \;\; \forall x$

If F satisfies a conservation law or translation invariance, then

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

Weakly contracting systems

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

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

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 $c > 0$:

$$\mathsf{osLip}_{|||\cdot|||}(f) \leq -c$$

or, for differentiable systems, $\mu_{|||\cdot|||}(D\mathsf{F}(x)) \leq -c$

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*, 67(3):1285–1300, 2022. doi

- M. Y. Li and J. S. Muldowney. A geometric approach to global-stability problems. *SIAM Journal on Mathematical Analysis*, 27(4):1070–1083, 1996. doi
- C. Wu, I. Kanevskiy, and M. Margaliot. k-contraction: Theory and applications. *Automatica*, 136:110048, 2022a. doi
- 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. *IEEE Transactions on Automatic Control*, 2022b. doi

Outline

Continuous-time recurrent neural networks:

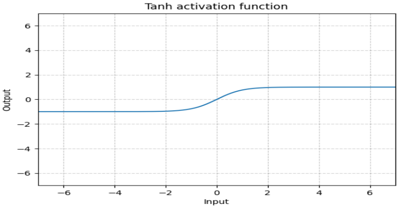
$$\dot{x} = -x + A\Phi(x) + u \qquad \text{(\textit{Hopfield})}$$

$$\dot{x} = -x + \Phi(Ax + u) =: f_{\mathsf{FR}}(x) \qquad \text{(\textit{Firing rate} $\sim$ \textit{Implicit NNs})}$$

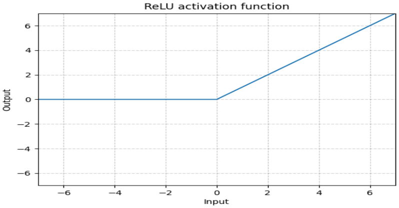
$$\dot{x} = A\Phi(x) \qquad \text{(\textit{Persidskii-type})}$$

$$\dot{x} = Ax - \Phi(x) \qquad (\dots)$$

sigmoid, hyperbolic tangent



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



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

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

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

Tight transcription.

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

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

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

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

$$\max_{d\in[d_{\mathsf{min}},d_{\mathsf{max}}]^n} \mu_\infty(A\mathsf{dg}(d)) = \dots$$

$$\max_{d\in[d_{\mathsf{min}},d_{\mathsf{max}}]^n} \mu_1(A\mathsf{dg}(d)) = \dots$$

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:

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

Specifically, if $d_{\mathsf{min}} = 0$ and $C \succ 0$,

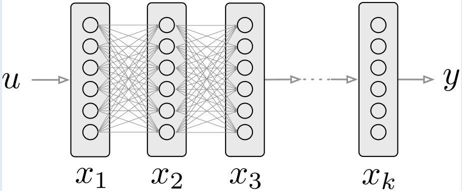
$$\inf_{\eta\in\mathbb{R}^n_{>0}} \mathsf{osLip}_{\infty,[\eta]}(f_{\mathsf{FR}}) = \max\left\{\alpha(-C), \alpha(-C + d_{\mathsf{max}}|A|_{\mathsf{M}})\right\}$$

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

Outline

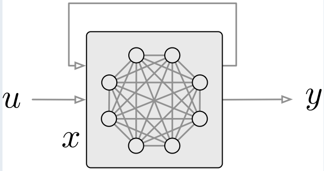
Implicit neural networks in machine learning

Feedforward NN



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

Implicit/Recurrent NN



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

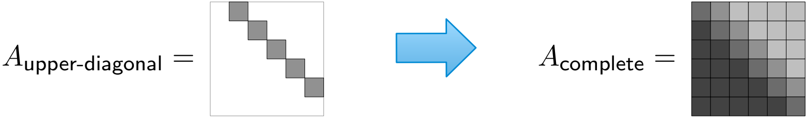
ML advantages of implicit/equilibrium/fixed point formulation:

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

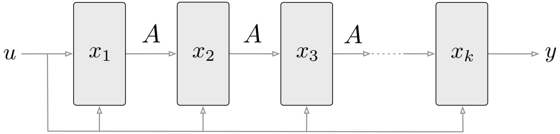
S. Jafarpour, A. Davydov, A. V. Proskurnikov, and F. Bullo. Robust implicit networks via non-Euclidean contractions. In *Advances in Neural Information Processing Systems*, Dec. 2021. URL **http://arxiv.org/abs/2106.03194**

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

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



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



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

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

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

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. 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-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**

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

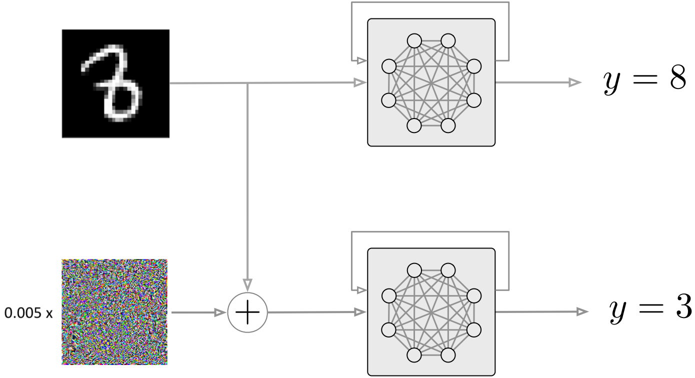
- Efficient back-propagation through implicit differentiation
- Stochastic gradient descent: at each step solve $x = \Phi(Ax + Bu + b)$.

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!



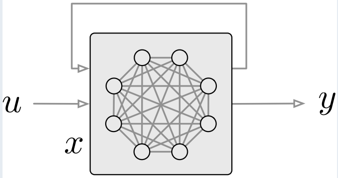
Robustness measures: input-to-output Lipschitz constant

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:

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

$$x_i = \Phi(Ax_i + B\widehat{u}_i + b)$$

$$\mu_\infty(A) \le \gamma$$

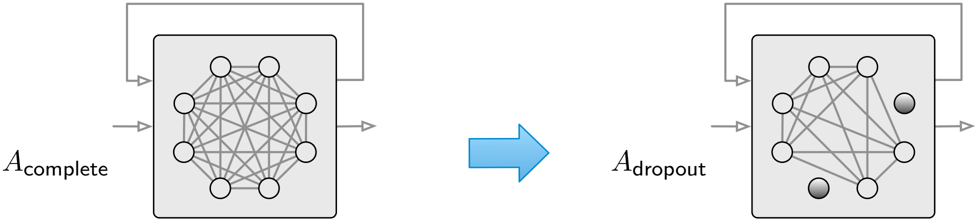
- $\lambda \ge 0$ is a regularization parameter
- $\gamma < 1$ is a hyperparameter

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

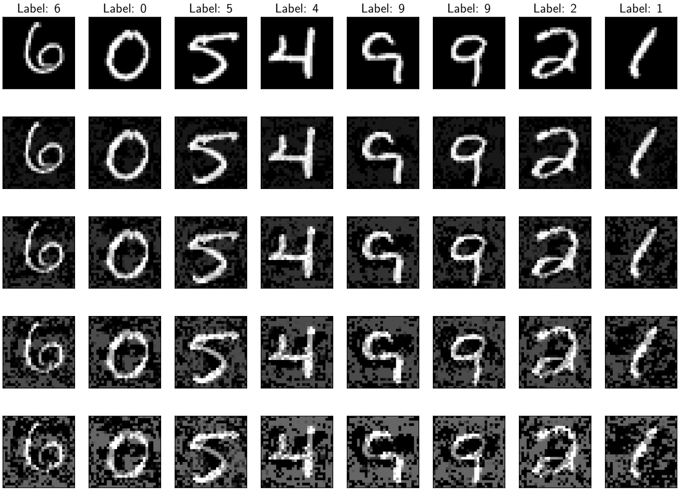
Synaptic matrix A encodes interactions between neurons



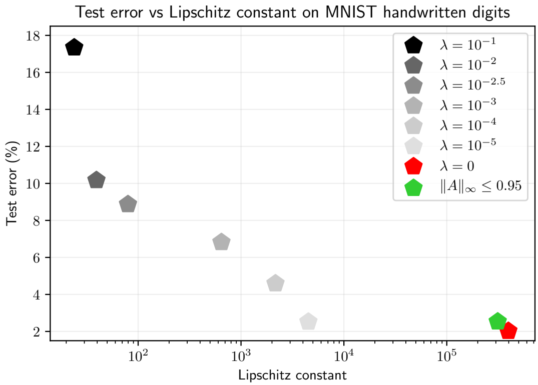
- A_{dropout} is a principal submatrix of A_{complete}
- $\mu_\infty(A_{\text{dropout}}) \leq \mu_\infty(A_{\text{complete}})$
    - Well-posedness of original INN implies well-posedness of INN with subset of neurons
    - Promotes compression and sparsity of overparametrized models

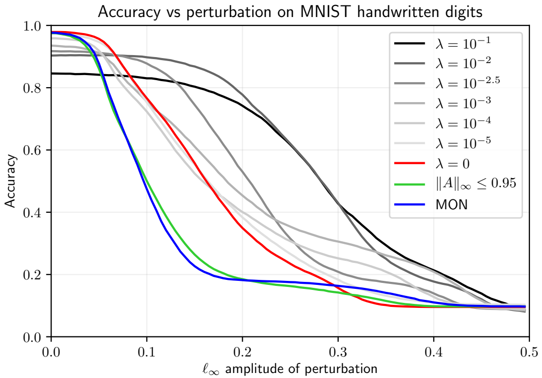
Numerical Experiments

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



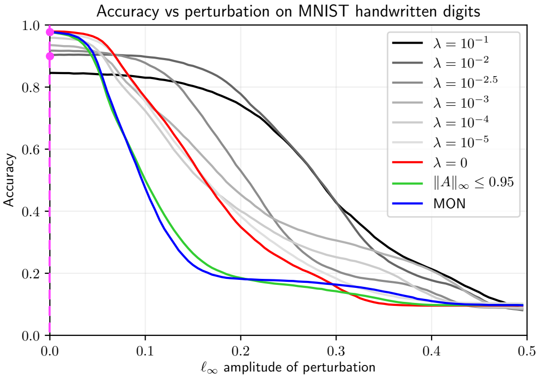
Tradeoff between **accuracy** and **robustness**

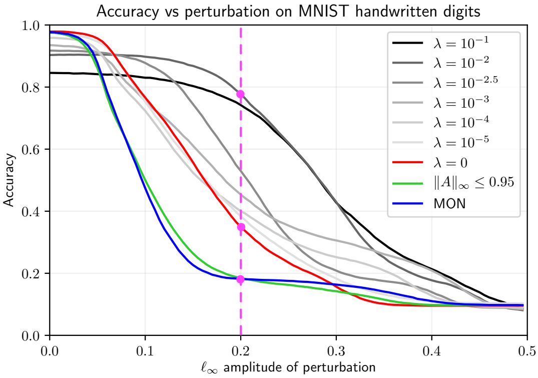




- Pareto-optimal curve
- Clean performance vs. robustness

Clean performance vs. robustness





Outline

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
   (L4DC **https://arxiv.org/abs/2112.05310**, oral presentation)
2. implicit graph neural architectures
3. bio-inspired Hebbian learning
4. robustness of implicit models