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

2021/2022 Sackler Lecturer
Mortimer and Raymond Sackler Institute of Advanced Studies

Seminar of the Electrical Engineering-Systems Department
Tel Aviv University, May 19, 2022

Acknowledgments

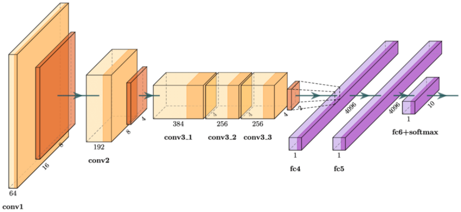
Alex Davydov
PhD student
UC Santa Barbara

Saber Jafarpour
Postdoc
GeorgiaTech

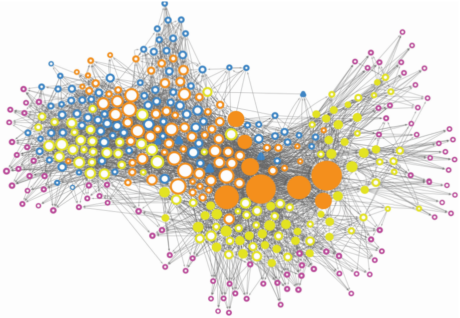
Anton Proskurnikov
Politecnico Torino & Russian
Academy of Sciences

- 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
- 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**), LCSS 2022 (**https://arxiv.org/abs/2204.00187**)

Artificial and Biological Neural Networks



artificial neural network AlexNet '12



C. elegans connectome '17

Aim: understand the dynamics of neural networks, so that

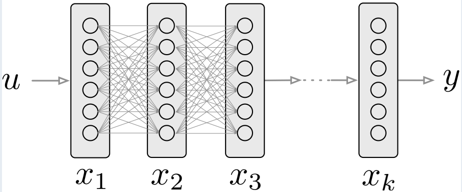
- **reproducible behavior, i.e., equilibrium response as function of stimula**
- robust behavior in face of uncertain stimuli and dynamics
- learning models, efficient computational tools, periodic behaviors ...

A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 2012

G. Yan, P. E. Vértes, E. K. Towlson, Y. L. Chew, D. S. Walker, W. R. Schafer, and A.-L. Barabási. Network control principles predict neuron function in the Caenorhabditis elegans connectome. *Nature*, 550(7677):519–523, 2017.

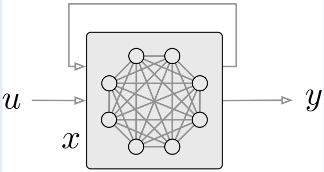
Fixed point computation

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. doi

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. doi

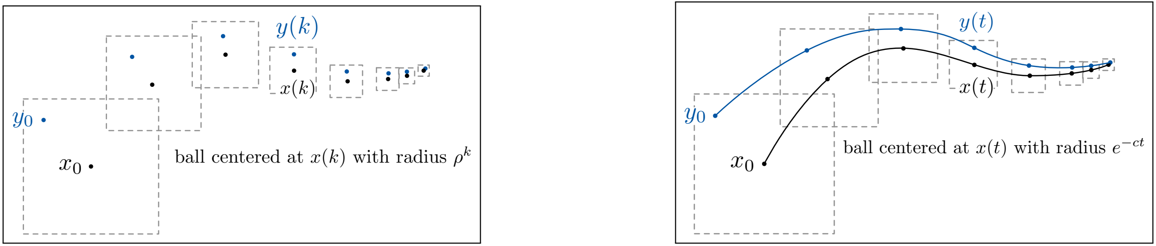
- **Application in dynamics and control**: W. Lohmiller and J.-J. E. Slotine. On contraction analysis for non-linear systems. *Automatica*, 34(6):683–696, 1998. doi

- **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. doi

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. 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



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

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 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 based upon Contraction

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

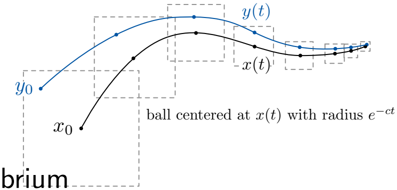
$$\|x_u^* - x_v^*\| \leq \frac{\mathrm{Lip}_u \mathsf{G}}{1 - \mathrm{Lip}_x \mathsf{G}} \|u - v\|$$

Robustness based upon Infinitesimal Contraction

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

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

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. modularity and interconnection properties,
5. accurate numerical integration, and

6. there exist efficient methods for their **equilibrium computation**

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:	$\mu(bA) = b\mu(A),$	$\forall b \geq 0$
norm/spectrum:	$\mathrm{Re}(\lambda) \leq \mu(A) \leq \|A\|,$	$\forall \lambda \in \mathrm{spec}(A)$

Convexity and quasi-convexity:

$$\mu_{2,P^{1/2}(A)} \leq -c \qquad \Longleftrightarrow \qquad PA + A^\top P \preceq -2cP$$

$$\mu_{\infty,\eta}(A) \leq -c \qquad \Longleftrightarrow \qquad a_{ii}\eta_i + \sum_{j\neq i} |a_{ij}|\eta_j \leq -c\eta_i \ \text{ for all } i$$

Contraction equivalences on normed vector spaces

For $x \in \mathbb{R}^n$ and continuously differentiable

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

For norm $\|\cdot\|$ with log norm $\mu(\cdot)$

$$\mathsf{osLip}(\mathsf{F}) := \sup_{x \in \mathbb{R}^n} \mu(D\mathsf{F}(x))$$

Main equivalences: for $c > 0$

1. **d-osL** : $\mathsf{osLip}(\mathsf{F}) \leq -c$
2. **d-IS** : $D^+\|x(t) - y(t)\| \leq -c\|x(t) - y(t)\|$ for soltns $x(\cdot), y(\cdot)$
3. **IS** : $\|x(t) - y(t)\| \leq e^{-c(t-t_0)}\|x(t_0) - y(t_0)\|$, for all soltns $x(\cdot), y(\cdot)$

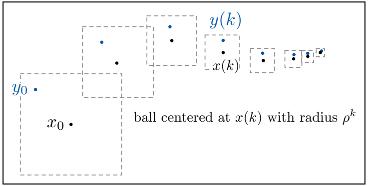
Equilibria of contracting vector fields:

For a time-invariant F, c-strongly contracting with respect to $\|\cdot\|$

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

$$x \mapsto \|x - x^*\| \quad \text{and} \quad x \mapsto \|\mathsf{F}(x)\|$$



Computing equilibria

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

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

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* $\kappa = \ell/c \geq 1$

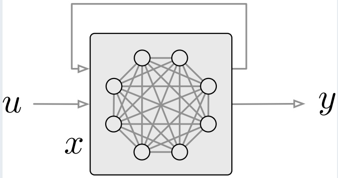
1. the map $\mathsf{Id} + \alpha \mathsf{F}$ is a contraction map with respect to $\|\cdot\|$ for

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

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

$$\alpha^* \approx \frac{1}{2c\kappa^2}, \qquad \ell^* \approx 1 - \frac{1}{4\kappa^2}$$

Application: ℓ_∞-contracting neural networks



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

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 $\mathrm{Lip}_{u\to y} = \dfrac{\|B\|_\infty \|C\|_\infty}{1-\mu_\infty(A)_+}$

Outline

Detour: a bit more detail

Continuous-time recurrent neural networks:

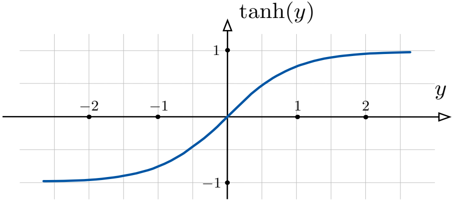
$$\dot{x} = -x + A\Phi(x) + u \qquad \text{(\textit{Hopfield})}$$

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

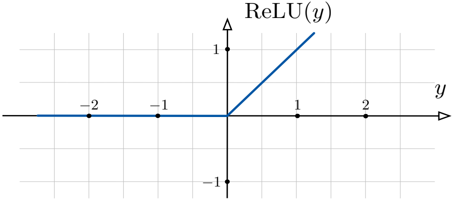
$$\dot{x} = A\Phi(x) \qquad \text{(\textit{Persidskii-type})}$$

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

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.

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

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

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

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

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

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

NonEuclidean contractivity of firing rate model

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

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

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

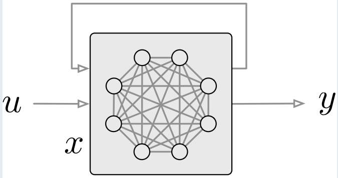
2. *optimal weight* η and minimim value of $\mathsf{osLip}_{\infty,[\eta]^{-1}}(f_{\mathsf{FR}})$ from quasiconvex opt:

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

Explicit solution (from PF theory) when $d_{\mathsf{min}} = 0$

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

Example: ℓ_∞-contracting neural networks



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

If

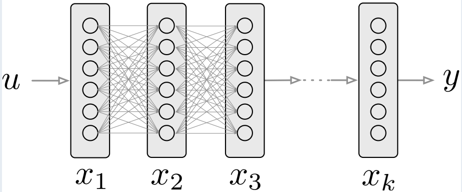
$$\mu_\infty(A) < 1 \qquad \Big(\text{i.e., } a_{ii} + \sum_j |a_{ij}| < 1 \text{ for all } i\Big)$$

- 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 $\mathrm{Lip}_{u\to y} = \dfrac{\|B\|_\infty \|C\|_\infty}{1-\mu_\infty(A)_+}$

Outline

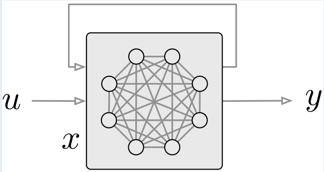
Implicit neural networks in machine learning

Feedforward NN



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

Implicit/Recurrent NN



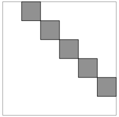
$$x = \Phi(Ax + Bu + b),$$
$$y = Cx + d$$

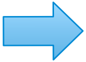
ML advantages of implicit/equilibrium/fixed point formulation:

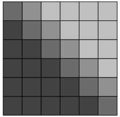
1. bio-inspired
2. expressivity and ability to model I/O behavior, instead of modalities
3. simplicity and memory efficiency
4. accuracy
5. input-output robustness

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

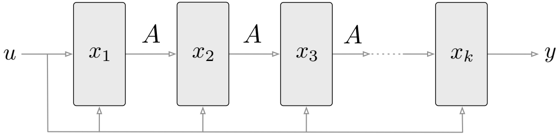
$x^{i+1} = \Phi(A_i x^i + B_i u + b_i) \quad \iff \quad x = \Phi(Ax + Bu + b)$, where A has upper diagonal structure.

$A_{\text{upper-diagonal}} =$ 



$A_{\text{complete}} =$ 

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



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

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)

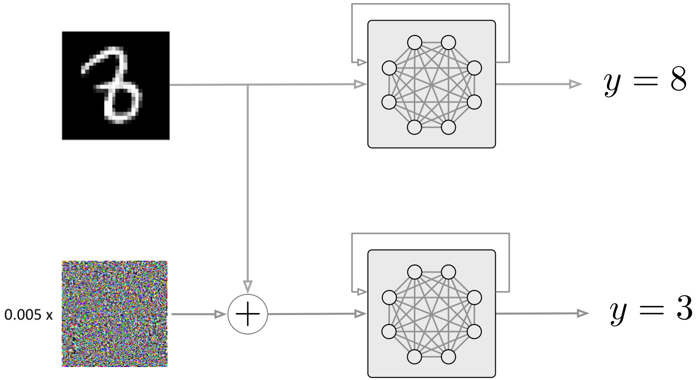
- Training INNs:
  1. loss function $\mathcal{L}$
  2. training data $(\widehat{u}_i, \widehat{y}_i)_{i=1}^N$
  3. **training optimization problem**

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

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

Robustness of INNs

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

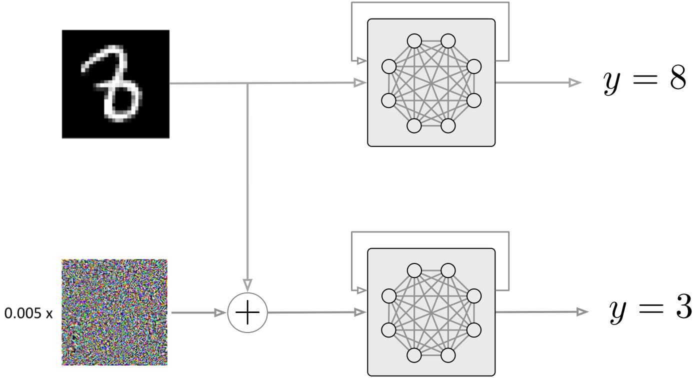


Robustness measures: **input-output Lipschitz constant**

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

Robustness of INNs

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 INNs

Training optimization problem:

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

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

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

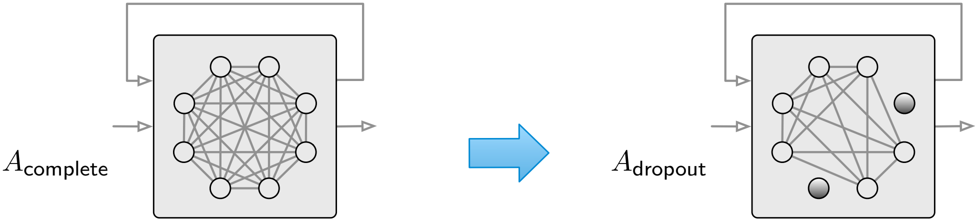
- $\lambda \geq 0$ is a regularization parameter
- $\gamma < 1$ is a hyperparameter

Parametrization of μ_∞ constraint:

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

Graph-Theoretic Regularization

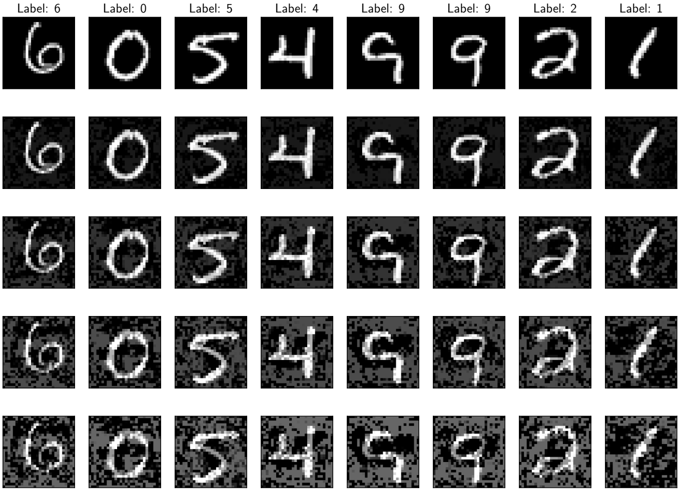
Synaptic matrix A encodes interactions between neurons



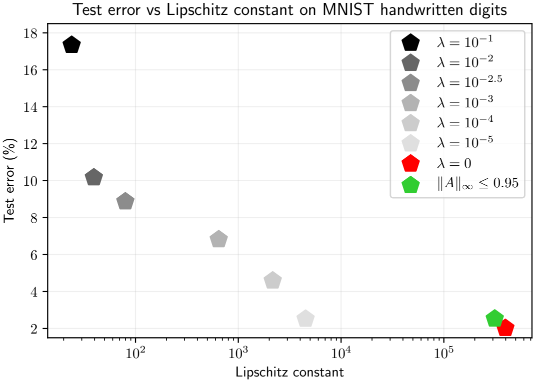
- A_{dropout} is a principal submatrix of A_{complete}
- $\mu_\infty(A_{\text{dropout}}) \leq \mu_\infty(A_{\text{complete}})$
  - Well-posedness of original INN implies well-posedness of INN with subset of neurons
  - Promotes compression and sparsity of overparametrized models

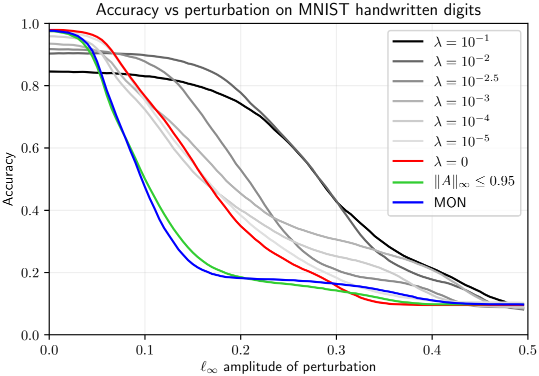
Numerical Experiments

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



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

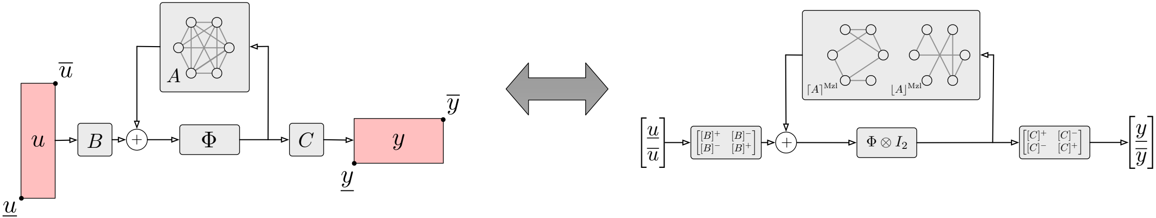




- Pareto-optimal curve
- Clean performance vs. robustness

Mixed monotonicity for INN reachability

Idea: mixed monotone systems theory for reachability analysis of RNN dynamics



INN with $u \in [\underline{u}, \overline{u}]$

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

Embedded INN

$$\begin{bmatrix} \underline{x} \\ \overline{x} \end{bmatrix} = \phi\left(\begin{bmatrix} \lceil A \rceil^{\mathrm{Mzr}} & \lfloor A \rfloor^{\mathrm{Mzr}} \\ \lfloor A \rfloor^{\mathrm{Mzr}} & \lceil A \rceil^{\mathrm{Mzr}} \end{bmatrix} \begin{bmatrix} \underline{x} \\ \overline{x} \end{bmatrix} + \begin{bmatrix} [B]^+ & [B]^- \\ [B]^- & [B]^+ \end{bmatrix} \begin{bmatrix} \underline{u} \\ \overline{u} \end{bmatrix} + \begin{bmatrix} b \\ b \end{bmatrix}\right),$$

$$\begin{bmatrix} \underline{y} \\ \overline{y} \end{bmatrix} = \begin{bmatrix} [C]^+ & [C]^- \\ [C]^- & [C]^+ \end{bmatrix} \begin{bmatrix} \underline{x} \\ \overline{x} \end{bmatrix} + \begin{bmatrix} d \\ d \end{bmatrix}.$$

Training provably robust INNs

Training INNs

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

$$\min_{\Theta} \quad \sum_{i=1}^{N} (1-\kappa)\mathcal{L}(y^i, \widehat{y}^i) + \kappa\mathcal{L}(\underline{y}^i, \widehat{y}^i),$$

$$x^i = \phi(Ax^i + Bu^i + b), \quad y^i = Cx^i + d,$$

$$\underline{y}^i = [T^i C]^+ \underline{x}^i + [T^i C]^- \overline{x}^i + T^i c, \quad \mu_{\infty,[\eta]^{-1}}(A) \leq \gamma.$$

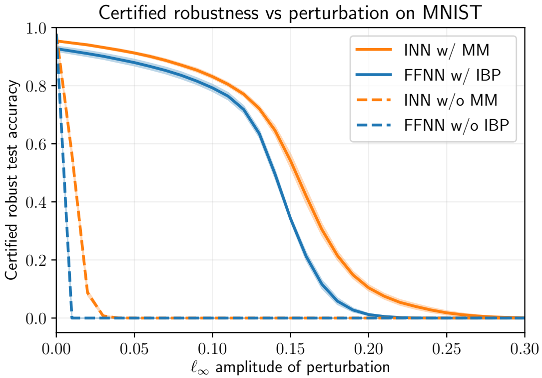
where $\kappa \in [0,1]$ and $\gamma < 1$ are hyperparameters.

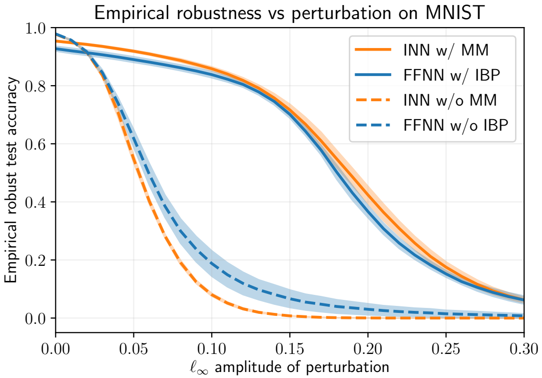
- Stochastic gradient descent: at each step solve 2 fixed-point problems
- Backpropagation through fixed-point equations

Numerical Experiments

Robustness of INNs

- Tradeoff between **accuracy** and **robustness**
- implicit neural network order: $n = 100$ vs 5-layer feedforward network





Outline

1. Scientific and engineering problems from neural networks
2. Contraction theory
   - Banach contractions and infinitesimal counterparts
3. Application to recurrent neural networks and implicit ML models
   - Implicit neural networks in machine learning
4. 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. implicit graph neural architectures
2. bio-inspired Hebbian learning
3. 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 $\mathsf{osLip}(\mathsf{G}) < 1$
3. the dynamics $\dot{x} = \mathsf{F}(x) := -x + \mathsf{G}(x)$ is infinitesimally contracting

- the equivalence (2) $\iff$ (3) is just a transcription:
  - $\mathsf{F} = -\mathsf{Id} + \mathsf{G}$ contracting with rate $c \iff \mathsf{osLip}(\mathsf{F}) < -c \iff \mathsf{osLip}(\mathsf{G}) < 1 - c$, for $c > 0$
  - in (ℓ_2, P), $\mathsf{osLip}(\mathsf{F}) < -c$ 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. 2021. 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.