Robustness of Neural Networks via Non-Euclidean Contraction Theory

Saber Jafarpour

Decision and Control Laboratory
Georgia Institute of Technology

June 7, 2022

Acknowledgment

Alexander Davydov
UCSB

Anton Proskurnikov
Politecnico di Torino, Italy.

Francesco Bullo
UCSB

SJ and A. Davydov and A. Proskurnikov and F. Bullo. Robust Implicit Networks via Non-Euclidean Contractions. In *NeurIPS*, Dec 2021.

A. Davydov and SJ and F. Bullo. Non-Euclidean Contraction Theory for Robust Nonlinear Stability. In *IEEE Transactions on Automatic Control*, May 2022.

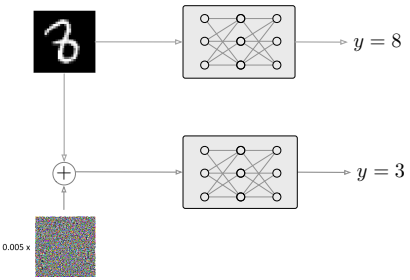
- Increase in computational power of neural networks
- **However**, neural networks can be fragile wrt to input perturbations

Adversarial examples

Small changes in the input

↓

Large changes in the output



C. Szegedy and et. al. Intriguing properties of neural networks. In *ICLR*, 2014

Critical task: ensuring safe and reliable operation of learning-based systems.

1. **Verification**: how robust is a learning algorithm?
2. **Training**: how to design robust learning models?

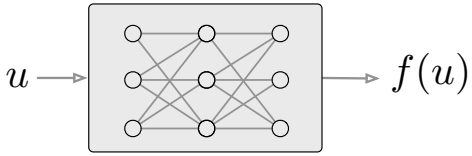
Tesla self-driving accident

This task is challenging for neural networks:

- large size of the networks
- nonlinear interactions
- unknown components
- dynamic and stochastic environment

Reachable set: Given an input set $\mathcal{X}$
safe output domain $\mathcal{S}$

$$\mathcal{R}(\mathcal{X}) = \{f(u) \mid u \in \mathcal{X}\}$$



Goal: approximate $\mathcal{R}(\mathcal{X})$ and check if $\mathcal{R}(\mathcal{X}) \subset \mathcal{S}$.

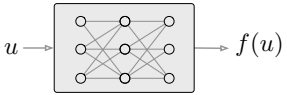
- **Lipschitz bounds:**

  A. Virmaux and K. Scaman. Lipschitz regularity of deep neural networks: analysis and efficient estimation. In *NeurIPS*, 2018
- **Interval bound propagation:**
- **Semi-definite programing:**

  M. Fazlyab, M. Morari, and G. J. Pappas. Safety verification and robustness analysis of neural networks via quadratic constraints and semidefinite programming. *IEEE Transactions on Automatic Control*, 2020.

Input-output Lipschitz constant

$$\|f(u) - f(v)\| \leq L\|u - v\|, \qquad \text{for all} \quad u, v \in \mathbb{R}^n$$



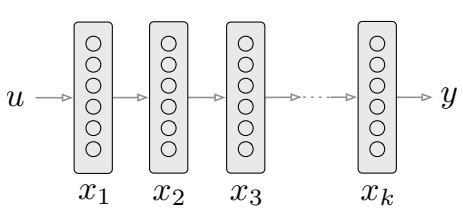
1. smaller L $\implies$ more robust neural networks but less expressive.
2. most common norms: ℓ_2 and ℓ_∞
   - ℓ_2-norm Lipschitz constant: change in energy-level
   - ℓ_∞-norm Lipschitz constant: component-wise change
3. computing the input-output Lipschitz constant is NP-hard

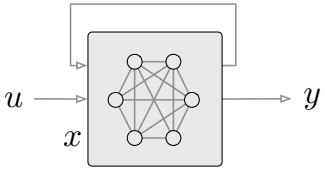
   A. Virmaux and K. Scaman. Lipschitz regularity of deep neural networks: analysis and efficient estimation. In *NeurIPS*, 2018
4. extensive research on estimating Lipschitz constant of neural networks

Presentation outline

- **implicit neural networks: motivations and challenges**
- non-Euclidean contraction theory
- well-posedness of implicit neural networks
- robustness of implicit neural networks
- Numerical experiments

- explicit hidden layers are replaced by a single implicit layer





- traditional neural networks:

$$x^{i+1} = \Phi(A_i x^i + b_i), \quad x^0 = u$$
$$y = A_k x^k + c$$

- implicit neural networks:

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

- $\Phi(y_1, \ldots, y_n) = (\phi_1(y_1), \ldots, \phi_n(y_n))^\top$ is a diagonal activation function
- activation functions are slope-restricted in $[0, 1]$, i.e., $0 \leq \frac{\phi_i(x) - \phi_i(y)}{x - y} \leq 1$ for all $x, y \in \mathbb{R}$

- Origins:


S. Bai, J. Z. Kolter, and V. Koltun. Deep equilibrium models. In *NeurIPS*, 2019


L. El Ghaoui, F. Gu, B. Travacca, A. Askari, and A. Y. Tsai. Implicit deep learning. *SIMODS*, 2019


A. Kag, Z. Zhang, and V. Saligrama. RNNs incrementally evolving on an equilibrium manifold: A panacea for vanishing and exploding gradients? In *ICLR*, 2020

- A general class of learning models
  - includes feedforward neural networks and residual networks
  - allows for arbitrary architecture
- Bio-inspired and reduce vanishing and exploding gradients
- Comparable accuracy to traditional neural networks with significant memory reduction
- More suitable for learning constrained optimization problems, stiff problems, and problems with discontinuity

- **Challenge 1**: well-posedness, i.e., existence and uniqueness of

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

- **Challenge 2**: computing robustness margin, i.e., estimate Lipschitz bound for INNs

Key insight

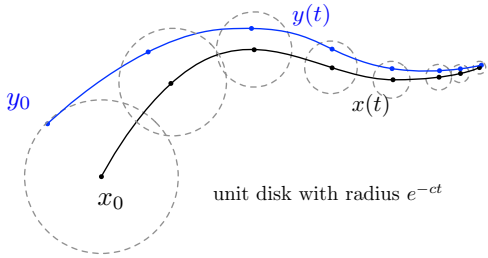
Fixed-point equation	$\Longleftrightarrow$	Dynamical system
$x = \Phi(Ax + Bu + b)$		$\dot{x} = -x + \Phi(Ax + Bu + b)$
well-posedness	$\Longleftrightarrow$	**equilibrium points**
robustness	$\Longleftrightarrow$	**reachability analysis**

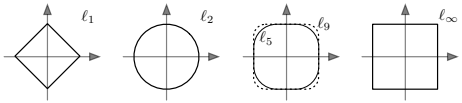
- Tools and techniques from Contraction Theory

Presentation outline

- implicit neural networks: motivations and challenges
- **non-Euclidean contraction theory**
- well-posedness of implicit neural networks
- robustness of implicit neural networks
- Numerical experiments

$\dot{x} = \mathsf{G}(x)$ is contractive if its flow is a contraction map





- **Origins**


D. C. Lewis. Metric properties of differential equations. *American Journal of Mathematics*, 71(2):294–312, 1949


B. P. Demidovich. Dissipativity of a nonlinear system of differential equations. *Uspekhi Matematicheskikh Nauk*, 16(3(99)):216, 1961


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 control theory**:

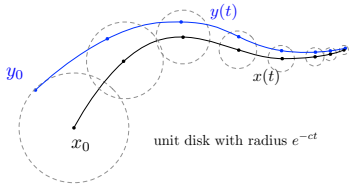

W. Lohmiller and J.-J. E. Slotine. On contraction analysis for non-linear systems. *Automatica*, 34(6):683–696, 1998

- **Reviews**:


Z. Aminzare and E. D. Sontag. Contraction methods for nonlinear systems: A brief introduction and some open problems. In *Proc CDC*, 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 *Complex Systems and Networks: Dynamics, Controls and Applications*, pages 313–339. Springer, 2016


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



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

1. time-invariant G: unique globally exponential stable equilibrium
   two natural Lyapunov functions
2. periodic G: contracting system entrain to periodic inputs
3. strong robustness properties: contractivity rate is natural measure of robust stability
   input-to-state stability in presence of un-modeled dynamics
4. accurate numerical integration and efficient methods for their equilibrium computation

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

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

- Directional derivative of norm $\|\cdot\|$ in direction of A,

$$\mu_2(A) = \tfrac{1}{2}\lambda_{\max}(A + A^\top)$$

$$\mu_1(A) = \max_j \left(a_{jj} + \sum\nolimits_{i\neq j} |a_{ij}|\right) \qquad \mu_\infty(A) = \max_i \left(a_{ii} + \sum\nolimits_{j\neq i} |a_{ij}|\right)$$

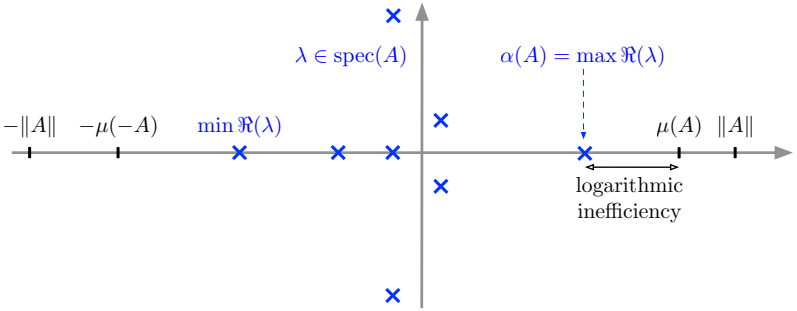
- Numerical analysis

E. Hairer, S. P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I. Nonstiff Problems*. 1993

T. Ström. On logarithmic norms. *SIAM Journal on Numerical Analysis*, 1975

Basic properties:

subadditivity:	$\mu(A+B) \leq \mu(A) + \mu(B),$	
convexity:	$\mu(\theta A + (1-\theta)B) \leq \theta\mu(A) + (1-\theta)\mu(B),$	$\forall \theta \in [0,1]$
norm/spectrum:	$\mathrm{Re}(\lambda) \leq \mu(A) \leq \|A\|,$	$\forall \lambda \in \mathrm{spec}(A)$



Lipschitz constant $\ell \in \mathbb{R}$

$$|\mathsf{G}(x) - \mathsf{G}(y)| \leq \ell |x - y| \quad \iff \quad -\ell \leq \mathsf{G}'(x) \leq \ell$$

One-sided Lipschitz constant $b \in \mathbb{R}$

$$(x - y)(\mathsf{G}(x) - \mathsf{G}(y)) \leq b(x - y)^2 \quad \iff \quad \mathsf{G}'(x) \leq b$$

Lipschitz constant $\ell \in \mathbb{R}$

$$\|\mathsf{G}(x) - \mathsf{G}(y)\|_2 \leq \ell \|x - y\|_2 \quad \iff \quad \|D\mathsf{G}(x)\|_2 \leq \ell$$

One-sided Lipschitz constant $b \in \mathbb{R}$

$$(x - y)^\top (\mathsf{G}(x) - \mathsf{G}(y)) \leq b \|x - y\|^2 \quad \iff \quad \mu_2(D\mathsf{G}(x)) \leq b$$

The second $\iff$ is non-trivial and is proven in


P. DeLellis, M. Di Bernardo, and G. Russo. On QUAD, Lipschitz, and contracting vector fields for consensus and synchronization of networks. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2011

Lipschitz constant $\ell \in \mathbb{R}$

$$\|\mathsf{G}(x) - \mathsf{G}(y)\| \leq \ell \|x - y\| \quad \iff \quad \|D\mathsf{G}(x)\| \leq \ell$$

One-sided Lipschitz constant $b \in \mathbb{R}$

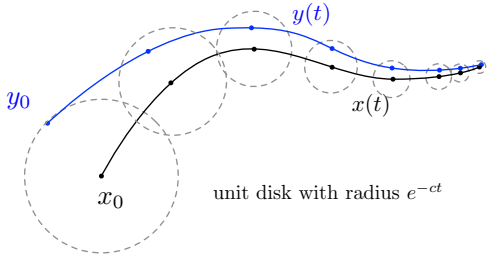
$$? \quad \iff \quad \mu(D\mathsf{G}(x)) \leq b$$

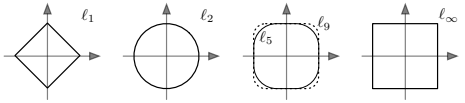
The notion of weak pairing is used instead of ?


A. Davydov, S. Jafarpour, and F. Bullo. Non-Euclidean contraction theory for robust nonlinear stability. *IEEE Trans. Autom. Control*, 2021

$\sup_x \|D\mathsf{G}(x)\| :=$ Lip G $\qquad \sup_x \mu(D\mathsf{G}(x)) :=$ osL G

$\dot{x} = \mathsf{G}(x)$ is contractive if its flow is a contraction map





Dynamical system $\dot{x} = \mathsf{G}(x)$ is contracting with respect to the norm $\|\cdot\|$ iff

$$\mu(D\mathsf{G}(x)) \leq -c, \qquad \text{for all } x$$

$$\text{or} \qquad \mathsf{osL}(\mathsf{G}) \leq -c$$

ℓ_2 – **contraction**		**LMI**
$\mu_2(D\mathsf{G}(x)) \leq -c$	$\iff$	$D\mathsf{G}(x) + D\mathsf{G}(x)^\top \preceq -cI$

- Monotone Operator Theory

E. K. Ryu and S. Boyd. Primer on monotone operator methods. *Applied Computational Mathematics*, 2016

ℓ_1/ℓ_∞ – **contraction**		**Diagonal Dominance**
$\mu_\infty(D\mathsf{G}(x)) \leq -c,$	$\iff$	$(D\mathsf{G}(x))_{ii} + \sum_{j\neq i} \lvert(D\mathsf{G}(x))_{ij}\rvert \leq -c, \quad \forall i$
$\mu_1(D\mathsf{G}(x)) \leq -c,$	$\iff$	$(D\mathsf{G}(x))_{ii} + \sum_{j\neq i} \lvert(D\mathsf{G}(x))_{ji}\rvert \leq -c, \quad \forall i$

- Non-Euclidean Monotone Operator Theory

Presentation outline

- implicit neural networks: motivations and challenges
- non-Euclidean contraction theory
- **well-posedness of implicit neural networks**
- robustness of implicit neural networks
- Numerical experiments

Challenge 1: well-posedness

Problem statement

For a fixed-point equation

$$x = \mathsf{F}(x, u) \qquad \text{(for implicit neural networks } \mathsf{F}(x, u) = \Phi(Ax + Bu + b))$$

1. when do we have a unique solution?
2. how to efficiently compute it?

Banach Fixed-point Theorem: if $\operatorname{Lip} \mathsf{F}(\cdot, u) < 1$, then $x = \mathsf{F}(x, u)$ has a unique solution by the Picard iterations

$$x^{k+1} = \mathsf{F}(x^k, u).$$

Key insight

Fixed-point of	$\Longleftrightarrow$	Equilibrium point of
$x = \mathsf{F}(x, u)$		$\dot{x} = -x + \mathsf{F}(x, u)$

- **Contraction theory**: existence and uniqueness of equilibrium point

$$\mathsf{osL}\, \mathsf{F}(\cdot, u) < 1.$$

Theorem: Fixed-point via matrix measures

If $\mathsf{osL}\, \mathsf{F}(\cdot, u) < 1$ then

1. F has a unique fixed-point x_u^*.
2. $x^{k+1} = (1-\alpha)x^k + \alpha \mathsf{F}(x^k, u)$ converges to x_u^*, for small enough α.

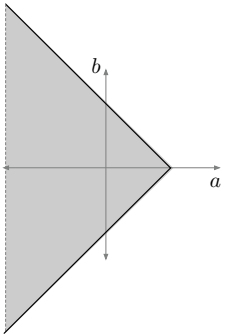
Theorem: Fixed-point via norm

If $\mathsf{Lip}\, \mathsf{F}(\cdot, u) < 1$ then

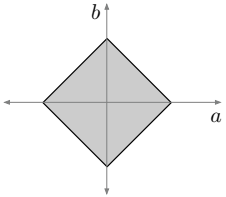
1. F has a unique fixed-point x_u^*.
2. $x^{k+1} = \mathsf{F}(x^k, u)$ converges to x_u^*.

osL $\mathsf{F}(\cdot, u) < 1$ is less conservative than Lip $\mathsf{F}(\cdot, u) < 1$.

- $\mathsf{F}(x, u) = Ax + Bu$ with $A = \begin{bmatrix} a & b \\ b & a \end{bmatrix}$



$\mu_\infty(A) \leq 1$
unbounded



$\|A\|_\infty \leq 1$
bounded

- upper bound on osL and Lip for non-Euclidean norm ℓ_∞-norm

Lemma:

$$\mathsf{osL}_\infty(\Phi(Ax+Bu+b)) = \sup_x \mu_\infty(D\phi(x)A) \leq \mu_\infty(A)$$
$$\mathsf{Lip}_\infty(\Phi(Ax+Bu+b)) = \sup_x \|D\phi(x)A\|_\infty \leq \|A\|_\infty$$

Proof:

- $D\Phi(x) = \mathrm{diag}(\phi_1'(x_1), \ldots, \phi_n'(x_n))$ and $0 \leq \phi_i'(x_i) \leq 1$
- use definition of μ_∞.

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

Theorem: Fixed-points of INNs

If $\mu_\infty(A) < 1$, then

1. there exists a unique fixed-point,
2. for small $\alpha > 0$, the average iterations:

$$x^{k+1} := (1 - \alpha)x^k + \alpha\Phi(Ax^k + Bu + b)$$

is a contraction map

Theorem: Fixed-points of INNs

If $\|A\|_\infty < 1$, then

1. there exists a unique fixed-point,
2. the Picard iterations

$$x^{k+1} := \Phi(Ax^k + Bu + b)$$

is a contraction map.

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

Theorem: Fixed-points of INNs

If $\mu_\infty(A) < 1$, then

1. there exists a unique fixed-point,
2. for small $\alpha > 0$, the average iterations:

$$x^{k+1} := (1 - \alpha)x^k + \alpha\Phi(Ax^k + Bu + b)$$

is a contraction map

Theorem: Fixed-points of INNs

If $\|A\|_\infty < 1$, then

1. there exists a unique fixed-point,
2. the Picard iterations

$$x^{k+1} := \Phi(Ax^k + Bu + b)$$

is a contraction map.

The average iteration is the Euler discretization of the dynamical system

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

Presentation outline

- implicit neural networks: motivations and challenges
- non-Euclidean contraction theory
- well-posedness of implicit neural networks
- **robustness of implicit neural networks**
- Numerical experiments

Challenge 2: Robustness margins

Problem statement

How does the fixed-point of $x = \mathsf{F}(x, u)$ change with u?

Theorem: Input-to-state Lipschitz bounds

x_u^* is a fixed-point of $x = \mathsf{F}(x, u)$ and $\operatorname{osL} \mathsf{F}(\cdot, u) < 1$,

$$\|x_u^* - x_v^*\| \leq \frac{\operatorname{Lip} \mathsf{F}(x, \cdot)}{1 - \operatorname{osL} \mathsf{F}(\cdot, u)} \|u - v\|$$

Theorem: Input-to-state Lipschitz bounds

x_u^* is a fixed-point of $x = \mathsf{F}(x, u)$ and $\operatorname{Lip} \mathsf{F}(\cdot, u) < 1$,

$$\|x_u^* - x_v^*\| \leq \frac{\operatorname{Lip} \mathsf{F}(x, \cdot)}{1 - \operatorname{Lip} \mathsf{F}(\cdot, u)} \|u - v\|$$

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

- How to compute input-output Lipschitz bounds of INNs?

$$u \underbrace{\mapsto}_{\mathrm{Lip}_{u \to x_u^*}} x_u^* \underbrace{\mapsto}_{\mathrm{Lip}_{x_u^* \to y}} y \quad \Longrightarrow \quad \mathrm{Lip}_{u \to y} = \mathrm{Lip}_{u \to x_u^*} \mathrm{Lip}_{x_u^* \to y}$$

Theorem: Input-to-output Lipschitz

if $\mu_\infty(A) < 1$ then

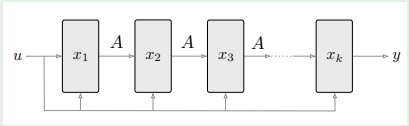
$$\mathrm{Lip}_{u \to y} \leq \frac{\|B\|_\infty \|C\|_\infty}{1 - \mu_\infty(A)_+}.$$

Theorem: Input-to-output Lipschitz

if $\|A\|_\infty < 1$ then

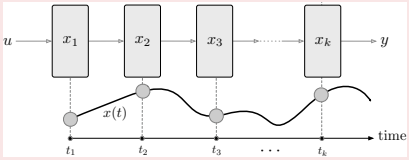
$$\mathrm{Lip}_{u \to y} \leq \frac{\|B\|_\infty \|C\|_\infty}{1 - \|A\|_\infty}.$$

Intuition #1: Weight-tied infinite-depth NN $\rightarrow$ fixed-point of INN



contraction of $x^{i+1} = \Phi(Ax^i + B_i u + b_i) \implies \lim_{i\to\infty} x^i = x^*$ solution to the INN

Intuition #2: Neural ODE model (infinite time) $\rightarrow$ fixed-point of INN



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

R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud. Neural ordinary differential equations. In *NeurIPS*, 2018

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

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

- $\gamma < 1$ is a hyperparameter and $\lambda \ge 0$ is a regularization parameter
- training optimization problem is solved via SGD
- at each step of SGD, $x_i = \Phi(Ax_i + B\widehat{u}_i + b)$ is solved using the average-iterations

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

Presentation outline

- implicit neural networks: motivations and challenges
- non-Euclidean contraction theory
- well-posedness of implicit neural networks
- robustness of implicit neural networks
- **Numerical experiments**

State-of-the-art architectures:

Implicit Deep Learning (IDL)

- ℓ_∞-norm well-posedness and robustness analysis
- results in the green boxes

L. El Ghaoui, F. Gu, B. Travacca, A. Askari, and A. Y. Tsai. Implicit deep learning. *SIMODS*, 2019

Monotone operator equilibrium networks (MON)

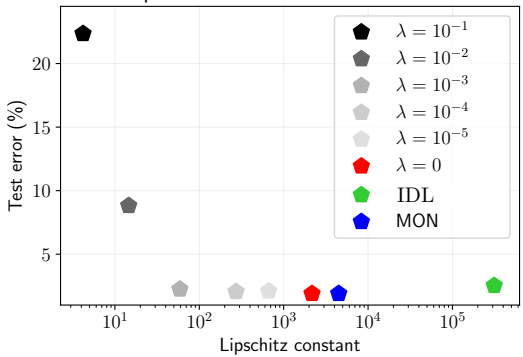
- ℓ_2-norm well-posedness and robustness analysis
- $\text{Lip}_\infty \leq \sqrt{r}\,\text{Lip}_2$ with r size of the input

E. Winston and J. Z. Kolter. Monotone operator equilibrium networks. In *NeurIPS*, 2020

C. Pabbaraju, E. Winston, and J. Z. Kolter. Estimating Lipschitz constants of monotone deep equilibrium models. In *ICLR*, 2021

- MNIST dataset: 28×28 pixel handwritten digits between $0 - 9$, $60,000$ training images and $10,000$ test images.
- implicit neural network order: $n = 100$ and $\gamma = 0.95$
- loss function: cross entropy

Test error vs Lipschitz constant on MNIST handwritten digits

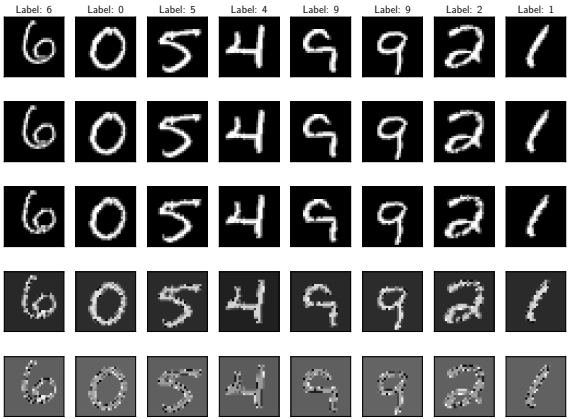


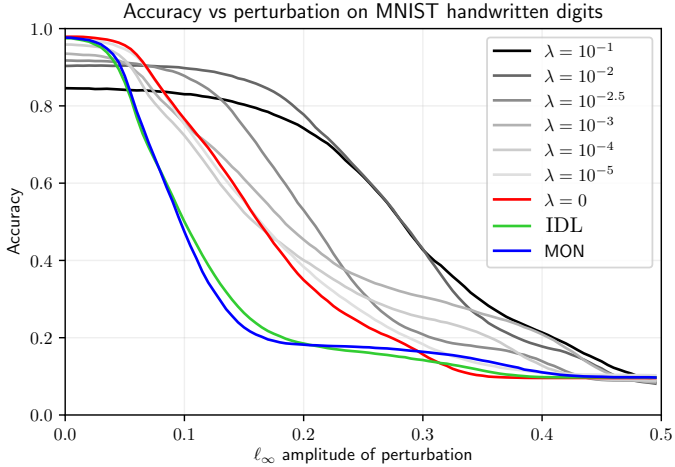
- Pareto-optimal curve

Improvements:

- $(\lambda = 0)$: two orders of magnitude wrt. IDL and wrt. MON
- $(\lambda = 10^{-3})$: three orders of magnitude wrt. IDL and one order of magnitude wrt. MON
- $(\lambda = 10^{-2})$: four orders of magnitude wrt. IDL and two orders of magnitude wrt. MON

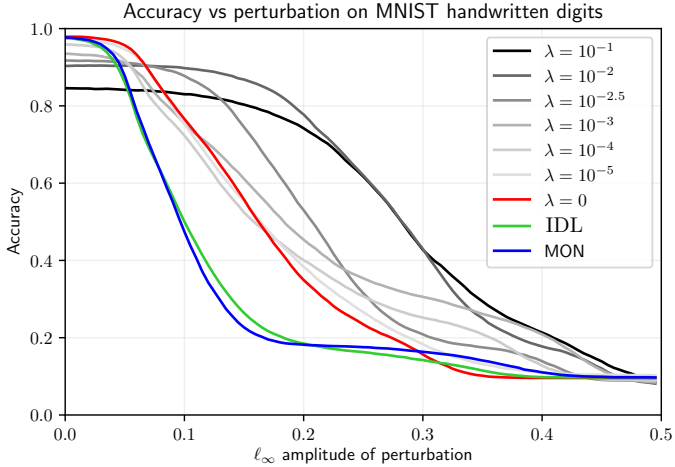
- perturbation: inversion attack $u_{\text{adv}} = u + \epsilon \, \text{sign}(\frac{1}{2}\mathbb{1}_{784} - u)$





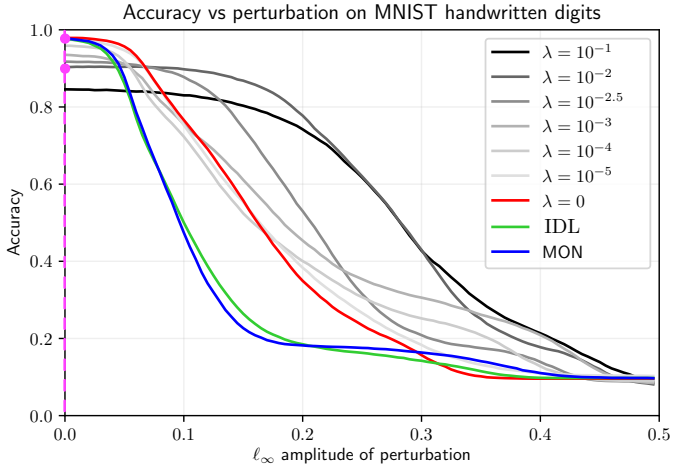
- $(\lambda = 0)$: improved robustness than IDL and MON
- $(\lambda > 0)$: improved robustness at sizable perturbations but losing some percentage accuracy in clean performance

Tradeoff between **clean performance** and **robustness**



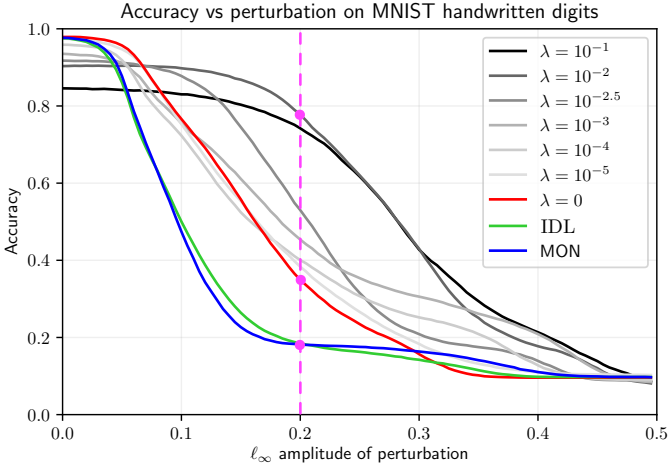
- $(\lambda = 0)$: improved robustness than IDL and MON
- $(\lambda > 0)$: improved robustness at sizable perturbations but losing some percentage accuracy in clean performance

Tradeoff between **clean performance** and **robustness**



- $(\lambda = 0)$: improved robustness than IDL and MON
- $(\lambda > 0)$: improved robustness at sizable perturbations but losing some percentage accuracy in clean performance

Tradeoff between **clean performance** and **robustness**



- $(\lambda = 0)$: improved robustness than IDL and MON
- $(\lambda > 0)$: improved robustness at sizable perturbations but losing some percentage accuracy in clean performance

Tradeoff between **clean performance** and **robustness**

Conclusions

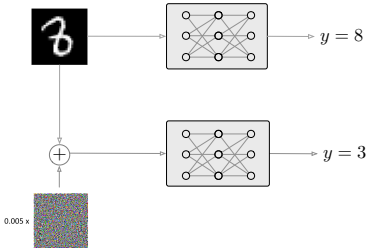
- non-Euclidean contraction theory
- well-posedness of implicit neural networks
- estimates of Lipschitz bound of implicit neural networks
- train robust implicit neural networks

Thank you for your attention!

Backup slides

Feature of adversarial perturbations:

- exist for a large class of learning algorithms
- transfer across models (not always!)
- *not* caused by overfitting (empirical evidence)



How to mitigate the effect of adversarial perturbations?

Adversarial training

- improve training using an attack
- easy to implement
- no provable guarantee

Robust optimization

- use over-approximation of the output
- hard to implement in training
- provide guarantees

- A large and flexible class of neural networks:
  includes feedforward neural networks

$$\begin{bmatrix} x^k \\ x^{k-1} \\ x^{\ell-2} \\ \vdots \\ x^2 \\ x^1 \end{bmatrix} = \sigma\left(\begin{bmatrix} 0 & A_{k-1} & 0 & \dots & 0 \\ 0 & 0 & A_{k-2} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & A_1 \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} x^k \\ x^{k-1} \\ x^{k-2} \\ \vdots \\ x^2 \\ x^1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ A_0 \end{bmatrix} u \right),$$

$$y = \begin{bmatrix} A_k & 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} x^k \\ x^{k-1} \\ x^{\ell-2} \\ \vdots \\ x^2 \\ x^1 \end{bmatrix}$$

1. **Origins**

   S. Bai, J. Z. Kolter, and V. Koltun. Deep equilibrium models. In *NeurIPS*, 2019

   L. El Ghaoui, F. Gu, B. Travacca, A. Askari, and A. Y. Tsai. Implicit deep learning. *SIMODS*, 2019

   A. Kag, Z. Zhang, and V. Saligrama. RNNs incrementally evolving on an equilibrium manifold: A panacea for vanishing and exploding gradients? In *ICLR*, 2020

2. **Monotone operator theory**

   E. Winston and J. Z. Kolter. Monotone operator equilibrium networks. In *NeurIPS*, 2020 M. Revay,

   R. Wang, and I. R. Manchester. Lipschitz bounded equilibrium networks. 2020. URL https://arxiv.org/abs/2010.01732

3. **Convergence**

   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 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

- Origins:

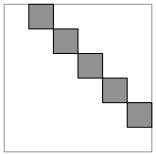
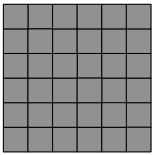
S. Bai, J. Z. Kolter, and V. Koltun. Deep equilibrium models. In *NeurIPS*, 2019

L. El Ghaoui, F. Gu, B. Travacca, A. Askari, and A. Y. Tsai. Implicit deep learning. *SIMODS*, 2019

A. Kag, Z. Zhang, and V. Saligrama. RNNs incrementally evolving on an equilibrium manifold: A panacea for vanishing and exploding gradients? In *ICLR*, 2020

- Generalizing feedforward neural networks to fully-connected synaptic matrices

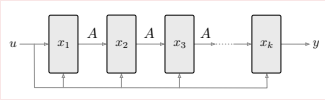
Intuition: $x^{i+1} = \phi_i(A_i x^i + b_i) \quad \Longleftrightarrow \quad x = \Phi(Ax + Bu + b)$, where A has upper diagonal structure.

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

- comparable accuracy to traditional neural networks with significant memory reduction

  S. Bai, J. Z. Kolter, and V. Koltun. Deep equilibrium models. In *NeurIPS*, 2019

Intuition: implicit neural network = weight-tied infinite-layer network



$$x^{i+1} = \phi_i(Ax^i + B_i u + b_i) \implies \lim_{i\to\infty} x^i = x^* \text{ solution to the INN}$$

- suitable for learning constrained optimization problems

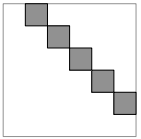
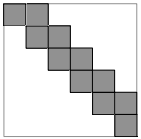
  A. Agrawal, B. Amos, S. Barratt, S. Boyd, S. Diamond, and J. Z. Kolter. Differentiable convex optimization layers. In *NeurIPS*, 2019

Intuition: casting KKT condition as an implicit layer

- vanishing and exploding gradient

A. Kag, Z. Zhang, and V. Saligrama. RNNs incrementally evolving on an equilibrium manifold: A panacea for vanishing and exploding gradients? In *ICLR*, 2020

Intuition: the notion of "autapse" (time-delayed self-feedback) from neuroscience

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

- suitable for learning stiff problems or problems with discontinuity

S. Pfrommer, M. Halm, and M. Posa. ContactNets: Learning discontinuous contact dynamics with smooth, implicit representations. *arXiv preprint*, 2020