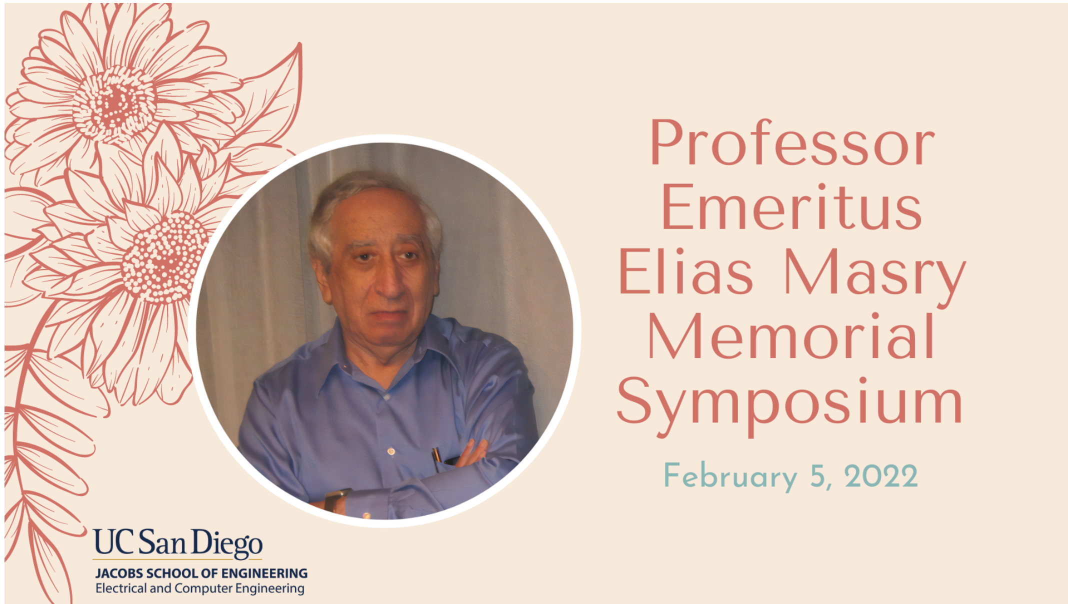
Perspectives on Contraction Theory and Neural Networks

Francesco Bullo

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

http://motion.me.ucsb.edu

Professor Emeritus Elias Masry Memorial Symposium
UC San Diego, Jacobs School of Engineering Feb 5, 2022



Acknowledgments

Alex Davydov
PhD student
UC Santa Barbara

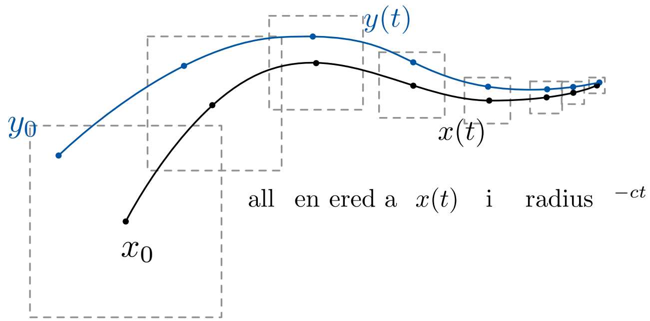
Saber Jafarpour
Postdoc
GeorgiaTech

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**. Submitted
- 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**. Submitted

Contraction theory: definition

Given $\dot{x} = \mathsf{F}(t, x)$, vector field F is contractive if its flow is a contraction map



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 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 *IEEE Conf. on Decision and Control*, pages 3835–3847, Dec. 2014.

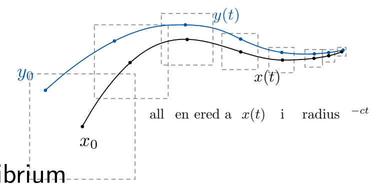
  M. Di Bernardo, D. Fiore, G. Russo, and F. Scafuti. Convergence, consensus and synchronization of complex networks via contraction theory. In J. Lü, X. Yu, G. Chen, and W. Yu, editors, *Complex Systems and Networks*, pages 313–339. Springer, 2016. ISBN 978-3-662-47824-0.

  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 on vector field do not necessarily involve Jacobians
- 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.

Contraction theory: properties of contracting systems



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

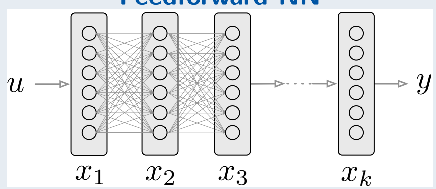
1. time-invariant F: unique globally exponential stable equilibrium
   two natural Lyapunov functions
2. periodic F: contracting system entrain to periodic inputs
3. contractivity rate is natural measure/indicator of robust stability
4. accurate numerical integration, and
5. there exist efficient methods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

Why fixed point computations?

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

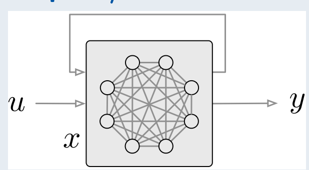
P. L. Combettes and J.-C. Pesquet. Fixed point strategies in data science. *IEEE Transactions on Signal Processing*, 2021.

Feedforward NN



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

Implicit/Recurrent NN



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

Advantages of implicit/equilibrium/fixed point formulation: simplicity, analogy with neural circuits, accuracy, memory efficiency, input-output robustness, etc

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. Y. Tsai. Implicit deep learning. 2019. URL **https://arxiv.org/abs/1908.06315**
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

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. doi
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. doi
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. doi
4. Y. Fang and T. G. Kincaid. Stability analysis of dynamical neural networks. *IEEE Transactions on Neural Networks*, 7(4):996–1006, 1996. doi
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. doi
6. W. He and J. Cao. Exponential synchronization of chaotic neural networks: a matrix measure approach. *Nonlinear Dynamics*, 55:55–65, 2009. doi
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. doi

Primer on monotone operator theory and contractions

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

Banach Contraction Theorem

If $\mathsf{Lip}(\mathsf{G}) < 1$, 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 $\mathsf{osL}(\mathsf{G}) < 1$
3. the dynamics $\dot{x} = -x + \mathsf{G}(x)$ is infinitesimally contracting

Primer on monotone operator theory and contractions: Addendum

Lim's Lemma

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

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

One-sided Lim's Lemma

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

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

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 $\mathsf{osL}(\mathsf{G}) < 1$
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 $\iff$ $\mathsf{osL}(\mathsf{F}) < -c \iff \mathsf{osL}(\mathsf{G}) < 1 - c$, for $c > 0$
  - in (ℓ_2, P), $\mathsf{osL}(\mathsf{F}) < -c$ 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 osL 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

Outline

Scalar maps and vector field

$F : \mathbb{R} \to \mathbb{R}$ is **one-sided Lipschitz** with $\mathsf{osL}(F) = b$ 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 $|x(t) - y(t)| \leq \mathrm{e}^{bt}|x(0) - y(0)|$

Contraction theory on inner product space $(\mathbb{R}^n, \ell_2)$

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

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

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

Main equivalences: For $c > 0$, map **F is c-strongly contracting** if

1. **osL** : $(\mathsf{F}(x) - \mathsf{F}(y))^\top P(x-y) \leq -c\|x-y\|^2_{2,P^{1/2}}$ for all x, y
2. **d-osL** : $PD\mathsf{F}(x) + D\mathsf{F}(x)^\top P \preceq -2cP$ 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. $-\operatorname{grad} V$ is m-**strongly contracting**, that is

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

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**

E. K. Ryu and W. Yin. *Large-Scale Convex Optimization via Monotone Operators*. Cambridge, 2022

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

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

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

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

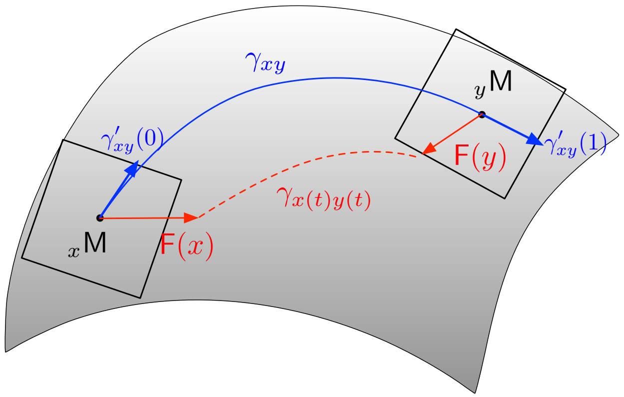
2. the optimal step size minimizing and minimum contraction factor:

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

Outline

Contraction theory on Riemannian manifold $(\mathsf{M}, \mathbb{G})$

F **contracting** if geodesic distances from x to y diminishes along the flow of F



integral test: the inner product between F and the geodesic velocity vector γ'_{xy} at x and y
differential test: condition on covariant differential of F

$$\mathbb{G}(x)\frac{\partial \mathsf{F}}{\partial x}(x) + \frac{\partial \mathsf{F}}{\partial x}(x)^\top \mathbb{G}(x) + \dot{\mathbb{G}}(x) \preceq -2c\mathbb{G}(x)$$

Outline

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

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

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

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

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

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

convexity: $\mu(\theta A + (1-\theta)B) \leq \theta\mu(A) + (1-\theta)\mu(B), \quad \forall \theta \in [0,1]$

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

Contraction theory on the normed vector spaces $(\mathbb{R}^n, \ell_1/\ell_\infty)$ 3/4

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 $x(\cdot), y(\cdot)$

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

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

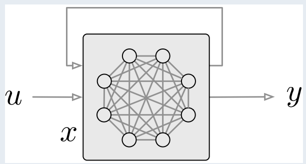
1. the map $\mathsf{Id} + \alpha\mathsf{F}$ is a contraction map with respect to $\|\cdot\|$ for

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

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

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

Example: ℓ_∞-contracting neural networks



Recurrent neural network dynamics

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

Average iteration

$$x_{k+1} = (1-\alpha)x_k + \alpha\Phi(Ax_k + Bu)$$

If

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

Then, with norm $\|\cdot\|_\infty$,

- dynamics is contracting with rate $1 - \mu_\infty(A)_+$
- average iteration is contracting with factor $1 - \dfrac{1-\mu_\infty(A)_+}{1-\min_i(a_{ii})_-}$ at $\alpha = \dfrac{1}{1-\min_i(a_{ii})_-}$

Conclusions

From Contracting Dynamics to Contracting Algorithms:

1. contraction theory and monotone operator theory are deeply connected
2. well established methodologies to tackle control, optimization and learning problems via fixed point strategies
3. same methods on Euclidean, Riemannian and non-Euclidean spaces
4. example application to recurrent neural networks

Spectacular Teacher
Thoughtful Researcher and Generous Collaborator
Marvelous Mentor

Thank you, Dr. Masry!