Contraction Theory for Optimization, Control, and Neural Networks

Francesco Bullo

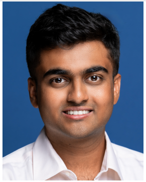
Center for Control,
Dynamical Systems & Computation
University of California at Santa Barbara
https://fbullo.github.io

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology, April 22, 2024

Acknowledgments

Veronica Centorrino
Scuola Sup Meridionale

Alexander Davydov
UC Santa Barbara



Anand Gokhale
UC Santa Barbara

Saber Jafarpour
Univ Colorado

Emiliano Dall'Anese
UC Boulder

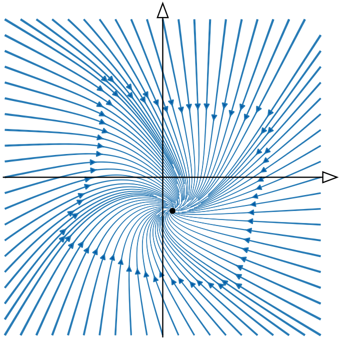
Anton Proskurnikov
Politecnico Torino

Giovanni Russo
Univ Salerno

AFOSR ARO ONR DTRA/ERDC

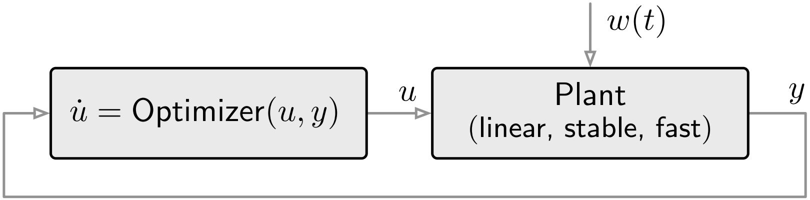
Frederick Leve @AFOSR FA9550-22-1-0059
Edward Palazzolo @ARO W911NF-22-1-0233
Marc Steinberg @ONR N00014-22-1-2813
Paul Tandy @DTRA W912HZ-22-2-0010
Donald Wagner @AFOSR FA9550-21-1-0203

Outline



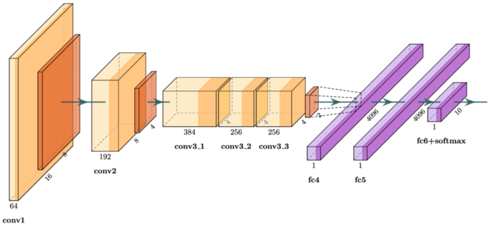
contractivity = robust computationally-friendly stability

fixed point theory + Lyapunov stability theory + geometry of metric spaces

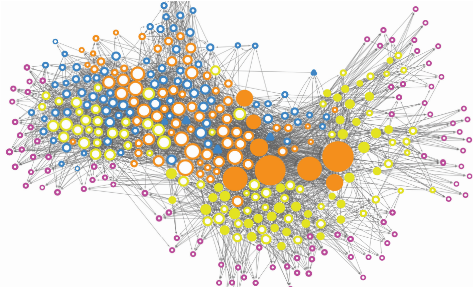


optimization via dynamical systems

online time-varying optimization, optimization-based feedback control, ...



artificial neural network AlexNet '12



C. elegans connectome '17

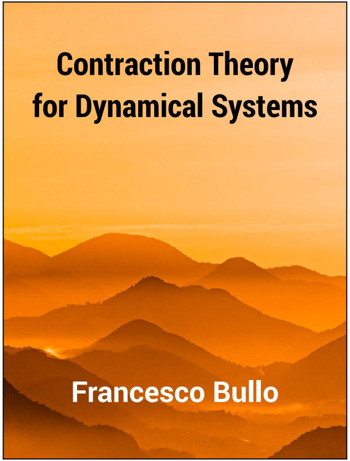
recurrent neural networks

well-posedness, stability, computation and input/output robustness

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.

Ongoing education and research on Contraction Theory



"Continuous improvement is better than delayed perfection"
Mark Twain

- Textbook: Contraction Theory for Dynamical Systems, Francesco Bullo, rev 1.1, Mar 2023. (Book and slides freely available) **https://fbullo.github.io/ctds**
- 2023 Comprehensive tutorial slides: **https://fbullo.github.io/ctds**
- 2023 Sep: Youtube lectures: "Minicourse on Contraction Theory" **https://youtu.be/FQV5PrRHks8** 12h in 6 lectures
- 2024 CDC Workshop "Contraction Theory for Systems, Control, Optimization, and Learning" (under review)

Outline

§1. A story in three chapters

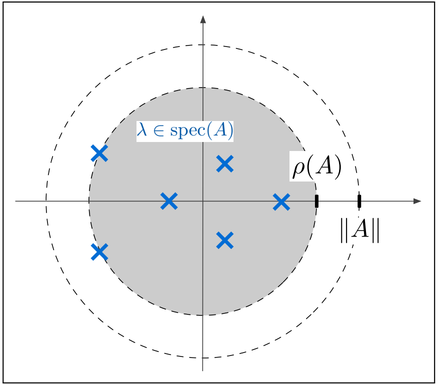
§2. Chapter #1: Contraction theory
- Basic notions on finite-dimensional vector spaces
- Examples: gradient systems and relationship with convexity
- Selected properties

§3. Chapter #2: Optimization-based control
- Equilibrium tracking
- Gradient controller

§4. Chapter #3: Artificial and biological neural networks
- Implicit and reservoir computing models in ML
- Functionality and analysis of biological networks

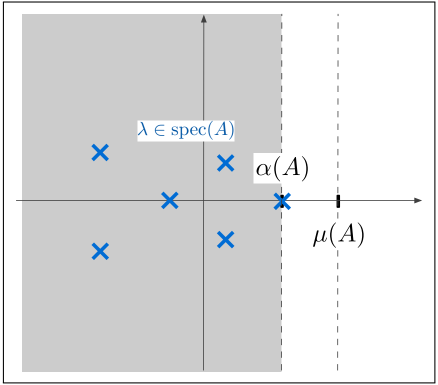
§5. Conclusions

given $n \times n$ matrix A with spectrum $\mathrm{spec}(A)$



$$\rho(A) \leq \|A\|$$

discrete-time dynamics



$$\alpha(A) \leq \mu(A) \leq \|A\|$$

continuous-time dynamics

Continuous-time dynamics and one-sided Lipschitz constants

$$\dot{x} = \mathsf{F}(x) \qquad \text{on } \mathbb{R}^n \text{ with norm } \|\cdot\| \text{ and induced log norm } \mu(\cdot)$$

One-sided Lipschitz constant **($\approx$ maximum expansion rate)**

$$\mathsf{osLip}(\mathsf{F}) = \sup_x \mu(D\mathsf{F}(x))$$

For **scalar map** f, $\mathsf{osLip}(f) = \sup_x f'(x)$

For **affine map** $\mathsf{F}_A(x) = Ax + a$

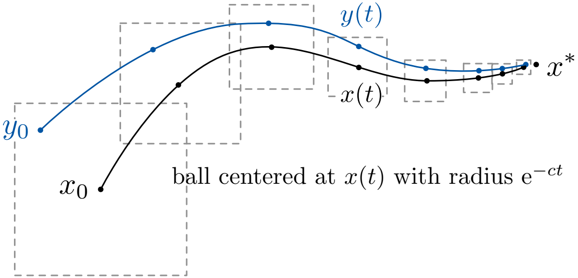
$$\mathsf{osLip}_{2,P}(\mathsf{F}_A) = \mu_{2,P}(A) \leq \ell \qquad \iff \qquad A^\top P + AP \preceq 2\ell P$$

$$\mathsf{osLip}_{\infty,\eta}(\mathsf{F}_A) = \mu_{\infty,\eta}(A) \leq \ell \qquad \iff \qquad a_{ii} + \sum_{j\neq i} |a_{ij}|\eta_i/\eta_j \leq \ell$$

Banach contraction theorem for continuous-time dynamics:

If $-c := \mathrm{osLip}(\mathsf{F}) < 0$, then

1. F is **infinitesimally contracting:** $\|x(t) - y(t)\| \leq \mathrm{e}^{-ct}\|x_0 - y_0\|$
2. F has a unique, glob exp stable equilibrium x^*
3. global Lyapunov functions $V_1(x) = \|x - x^*\|^2$ and $V_2(x) = \|\mathsf{F}(x)\|^2$



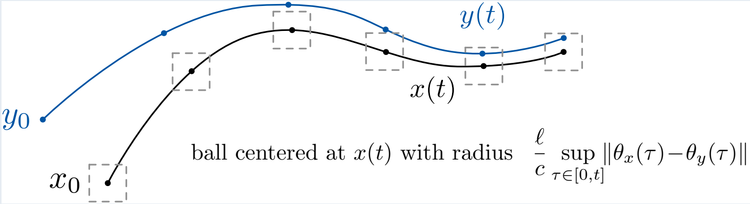
Property #1: Incremental ISS Theorem. Consider

$$\dot{x} = \mathsf{F}(x, \theta(t))$$

- **contractivity wrt x**: $\mathsf{osLip}_x(\mathsf{F}) \leq -c < 0$, uniformly in θ
- **Lipschitz wrt θ**: $\mathsf{Lip}_\theta(\mathsf{F}) \leq \ell$, uniformly in x

Then **incrementally ISS property**:

$$\|x(t) - y(t)\| \leq \mathrm{e}^{-ct}\|x_0 - y_0\| + \frac{\ell}{c}(1 - \mathrm{e}^{-ct}) \sup_\tau \|\theta_x(\tau) - \theta_y(\tau)\|$$



Example contracting systems

1. *gradient descent flows* under strong convexity assumptions
   (primal-dual, distributed, saddle, pseudo, proximal, etc)
2. *neural network dynamics* under assumptions on synaptic matrix
   (recurrent, implicit, reservoir computing, etc)
3. incremental ISS systems
4. Lur'e-type systems under LMI conditions
5. feedback linearizable systems with stabilizing controllers
6. data-driven learned models
7. nonlinear systems with a locally exponentially stable equilibrium
   are contracting with respect to appropriate Riemannian metric

Example #1: Gradient dynamics for strongly convex function

Given differentiable, strongly convex $f : \mathbb{R}^n \to \mathbb{R}$ with parameter $\nu > 0$, **gradient dynamics**

$$\dot{x} = \mathsf{F}_{\mathsf{G}}(x) := -\nabla f(x)$$

F_{G} is infinitesimally contracting wrt $\|\cdot\|_2$ with rate ν

unique globally exp stable point is global minimum

Property #2: Kachurovskii's Theorem: For differentiable $f : \mathbb{R}^n \to \mathbb{R}$, equivalent statements:

1. f is **strongly convex** with parameter ν (and minimum x^*)
2. $-\nabla f$ is **ν-strongly infinitesimally contracting** (with equilibrium x^*)

Property #3: Euler Discretization Theorem for Contracting Dynamics

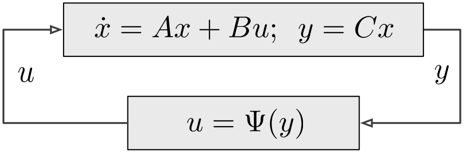
Given norm $\|\cdot\|$ and differentiable and Lipschitz $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$, equivalent statements

1. $\dot{x} = \mathsf{F}(x)$ is infinitesimally contracting
2. there exists $\alpha > 0$ such that $x_{k+1} = x_k + \alpha\mathsf{F}(x_k)$ is contracting

R. I. Kachurovskii. Monotone operators and convex functionals. *Uspekhi Matematicheskikh Nauk*, 15(4):213–215, 1960

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.

Example #2: Systems in Lur'e form



For $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{m \times n}$ and $C \in \mathbb{R}^{n \times m}$, **nonlinear system in Lur'e form**

$$\dot{x} = Ax + B\Psi(Cx) \quad =: \mathsf{F}_{\text{Lur'e}}(x)$$

where $\Psi : \mathbb{R}^m \to \mathbb{R}^m$ is ρ-*cocoercive*, that is, for all $y_1, y_2 \in \mathbb{R}^m$

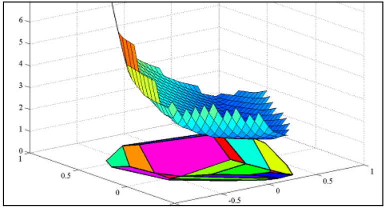
$$(\Psi(y_1) - \Psi(y_2))^\top (y_1 - y_2) \geq \rho \|\Psi(y_1) - \Psi(y_2)\|_2^2$$

For $P = P^\top \succ 0$, following statements are equivalent:

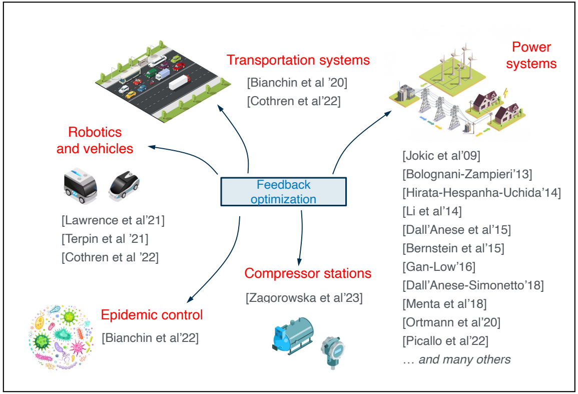
1. $\mathsf{F}_{\text{Lur'e}}$ infinitesimally contracting wrt $\|\cdot\|_{2,P^{1/2}}$ with rate $\eta > 0$ for each ρ-cocoercive Ψ
2. there exists $\lambda \geq 0$ such that $\begin{bmatrix} A^\top P + PA + 2\eta P & PB + \lambda C^\top \\ B^\top P + \lambda C & -2\lambda\rho I_m \end{bmatrix} \preceq 0$

Outline

1. parametric optimization
2. **online feedback optimization**
3. model predictive control
4. control barrier functions
5. …



parametric QP. YALMIP + Multi-Parametric Toolbox



Online feedback optimization. Courtesy of Emiliano Dall'Anese.

$$\min \mathcal{E}(x) \quad \Longleftrightarrow \quad \dot{x} = \mathsf{F}(x) \qquad \rightsquigarrow \quad x^*$$

Parametric and time-varying convex optimization

1. **parametric contracting dynamics for parametric convex optimization**

$$\min \mathcal{E}(x,\theta) \quad \Longleftrightarrow \quad \dot{x} = \mathsf{F}(x,\theta) \qquad \rightsquigarrow \quad x^*(\theta)$$

2. **contracting dynamics for time-varying strongly-convex optimization**

$$\min \mathcal{E}\big(x,\theta(t)\big) \quad \Longleftrightarrow \quad \dot{x} = \mathsf{F}\big(x,\theta(t)\big) \qquad \rightsquigarrow \quad x^*\big(\theta(t)\big)$$

A. Davydov, V. Centorrino, A. Gokhale, G. Russo, and F. Bullo. Time-varying convex optimization: A contraction and equilibrium tracking approach. *IEEE Transactions on Automatic Control*, June 2023. doi. Conditionally accepted

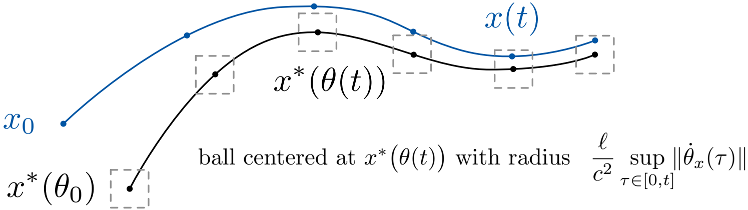
Property #4: Equilibrium Tracking Theorem. Consider

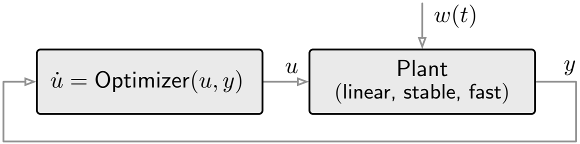
$$\dot{x} = \mathsf{F}(x, \theta(t))$$

- **contractivity wrt x**: $\mathrm{osLip}_x(\mathsf{F}) \leq -c < 0$, uniformly in θ
- **Lipschitz wrt θ**: $\mathrm{Lip}_\theta(\mathsf{F}) \leq \ell$, uniformly in x

Then **equilibrium tracking property**:

$$\|x(t) - x^\star(\theta(t))\| \;\leq\; \mathrm{e}^{-ct}\|x_0 - x^\star(\theta_0)\| \;+\; \frac{\ell}{c^2}(1 - \mathrm{e}^{-ct}) \sup_{\tau\in[0,t]} \|\dot{\theta}(\tau)\|$$





$$\begin{cases} \min & \text{cost}_1(u) + \text{cost}_2(y) \\ \text{subj. to} & y = \text{Plant}\big(u, w(t)\big) \end{cases} \implies \begin{cases} \dot{u} = \text{Optimizer}(u, y) \\ y = \text{Plant}\big(u, w(t)\big) \end{cases}$$

Example #3: Gradient controller

Online feedback optimization

$$\begin{aligned} u^*\big(w(t)\big) \;:=\; & \underset{u}{\operatorname{argmin}} && \phi(u) + \psi(y(t)) \qquad \text{(ν-strongly convex ϕ, convex ψ)} \\ & \text{subj to} && y(t) = Y_u u + Y_w w(t) \end{aligned}$$

gradient controller

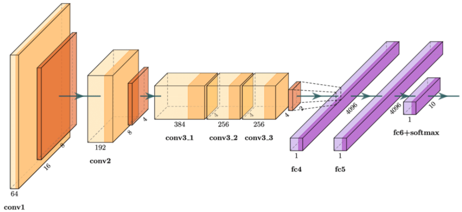
$$\dot{u} \;=\; \mathsf{F}_{\mathsf{GradCtrl}}(u, w) := -\nabla_u\big(\phi(u) + \psi(y(t))\big) \;=\; -\nabla\phi(u) - Y_u^\top \nabla\psi(Y_u u + Y_w w)$$

Equilibrium tracking for the gradient controller

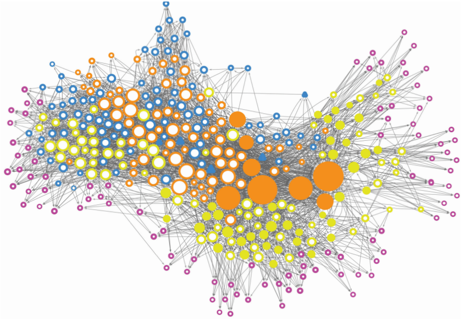
$$\limsup_{t\to\infty} \|u(t) - u^*\big(w(t)\big)\| \;\leq\; \frac{\ell_w}{\nu^2} \limsup_{t\to\infty} \|\dot{w}(t)\|$$

Outline

Artificial and biological neural networks



artificial neural network AlexNet '12



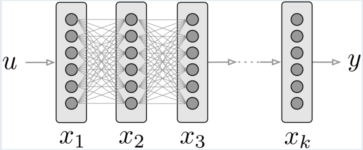
C. elegans connectome '17

Aim: dynamics of neural networks:

- reproducible and robust behavior in face of uncertain stimuli and dynamics
- functionality: regression, clustering, prediction, dimensionality reduction
- learning models, efficient computational tools, periodic behaviors ...

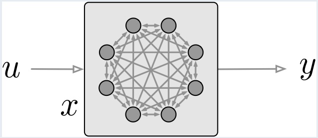
A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 2012
G. Yan, P. E. Vértes, E. K. Towlson, Y. L. Chew, D. S. Walker, W. R. Schafer, and A.-L. Barabási. Network control principles predict neuron function in the Caenorhabditis elegans connectome. *Nature*, 550(7677):519–523, 2017.

Feedforward NN



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

Implicit/Recurrent NN



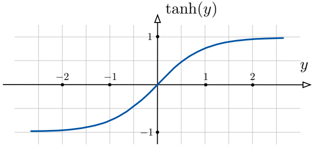
$$x = \Phi(Ax + Bu + b),$$
$$y = Cx + d$$

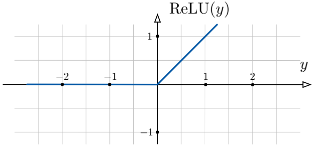
$$\dot{x} = \mathsf{F}_{\mathsf{FR}}(x) := -x + \Phi(Ax + Bu)$$

hyperbolic tangent

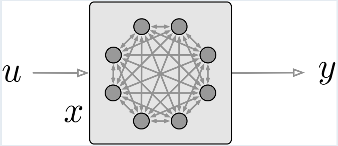
ReLU $= (x)_+$

$0 \leq \Phi_i'(y) \leq 1$





Example #4: Firing-rate networks for implicit ML



$$\dot{x} = -x + \Phi(Ax + Bu + b) \qquad \text{(recurrent NN)}$$

$$x = \Phi(Ax + Bu + b) \qquad \text{(implicit NN)}$$

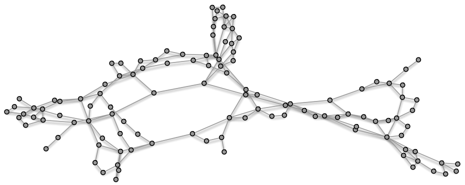
$$x_{k+1} = (1-\alpha)x_k + \alpha\Phi(Ax_k + Bu + b) \qquad \text{(Euler discrt.)}$$

If

$$\mu_\infty(A) < 1 \qquad \left(\text{i.e., } a_{ii} + \sum_{j\neq i} |a_{ij}| < 1 \text{ for all } i\right)$$

- **recurrent NN is infinitesimally contracting** with rate $1 - \mu_\infty(A)_+$
- **implicit NN is well posed**
- **Euler discretization is contracting** at $\alpha^* = (1 - \min_i(a_{ii})_-)^{-1}$

- **input-state Lipschitz constant** $\|B\|_\infty/(1 - \mu_\infty(A)_+)$
- **sensitivity to unmodeled dynamics** $\frac{\|\Delta x^*\|_\infty}{\|x^*\|_\infty} \leq \frac{\|\Delta A\|_\infty}{1-\mu_\infty(A)_+}$
- **robustness to signal delays** and more



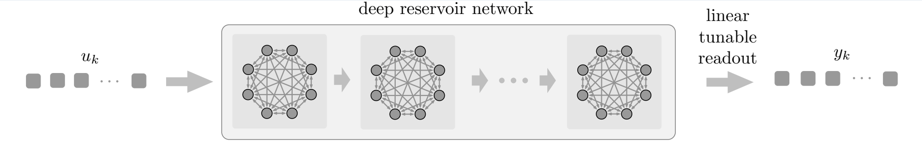
Property #5: Network Contraction Theorem. Consider interconnected subsystems

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

- **contractivity wrt x_i:** $\quad \text{osLip}_{x_i}(\mathsf{F}_i) \leq -c_i < 0, \quad$ uniformly in x_{-i}
- **Lipschitz wrt x_j, $j \neq i$:** $\quad \text{Lip}_{x_j}(\mathsf{F}_i) \leq \ell_{ij}, \quad$ uniformly in x_{-j}
- gain matrix $\begin{bmatrix} -c_1 & \dots & \ell_{1n} \\ \vdots & \ddots & \vdots \\ \ell_{n1} & \dots & -c_n \end{bmatrix}$ is **Hurwitz**

$\Longrightarrow$ **interconnected system** is contracting with rate $|\alpha(\text{gain matrix})|$

Example #5: Firing-rate networks for ML reservoir computing



$$x_{k+1}^{(1)} = (1-\alpha)x_k^{(1)} + \alpha\Phi\big(A^{(1)}x_k^{(1)} + B^{(1)}u_k + b^{(1)}\big)$$
$$x_{k+1}^{(i)} = (1-\alpha)x_k^{(i)} + \alpha\Phi\big(A^{(i)}x_k^{(i)} + B^{(i)}x_k^{(i-1)} + b^{(i)}\big)$$

(leaky integrator reservoirs)

Deep reservoir network is contracting (and "echo state property") if

$\mu_\infty(A^{(i)}) < 1$ for each i and for $\alpha \leq \alpha^{**}$

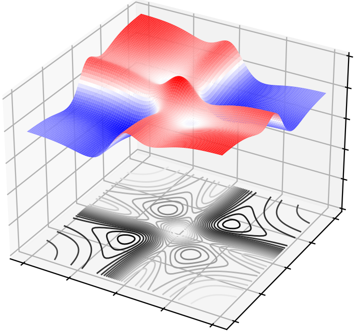
H. Jaeger. The "echo state" approach to analysing and training recurrent neural networks. Technical report, German National Research Center for Information Technology, 2001

Outline

$$\dot{x} = \mathsf{F}_{\mathsf{FR}}(x) := -x + \Phi(Ax + Bu)$$

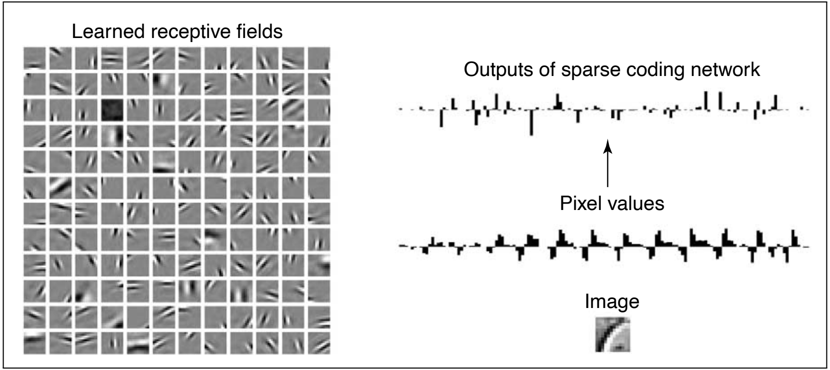
1. What is F_{FR} optimizing?
2. What is its functionality?
3. Is a normative framework for neural circuits?

4. Case study: dimensionality reduction



Energy landscape for associative memory in Hopfield models

Sparse signal reconstruction in biological neuronal circuits



- primary visual area (V1) sparsifies signals
- receptive fields (≈ dictionary) are learned empirically

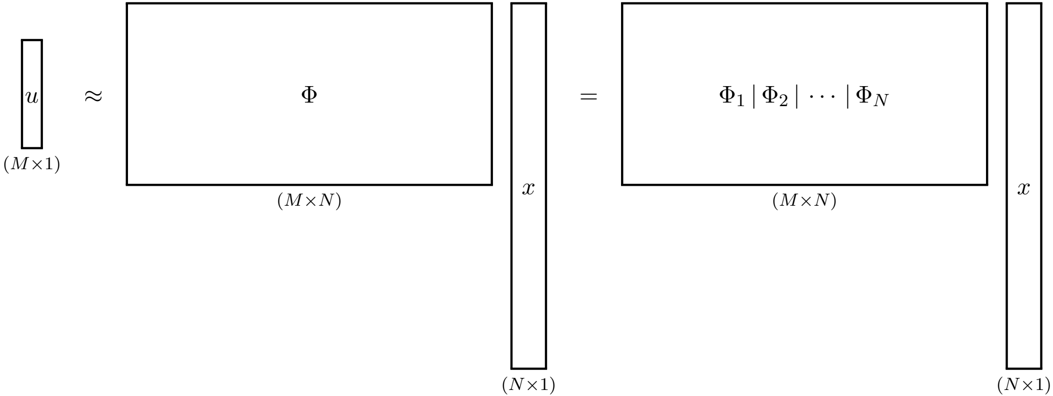
B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609, 1996.

B. A. Olshausen and D. J. Field. Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, 14(4):481–487, 2004.

Sparse reconstruction by minimizing the lasso energy

$$\min_{x \in \mathbb{R}^N} \quad \mathcal{E}_{\text{lasso}}(x) := \frac{1}{2}\|u - \Phi x\|_2^2 + \lambda \|x\|_1$$

where Φ dictionary matrix, with $\|\Phi_i\| = 1$ and $\Phi_i \cdot \Phi_j$ = similarity between elements



where x is k-sparse and $k \ll M \ll N$

Proximal gradient descent

Minimization of composite cost:

$$\min \quad \underbrace{f(x,u)}_{\text{convex in } x} \quad + \quad \underbrace{g(x)}_{\text{regularizer}}$$

proximal gradient descent:

$$\dot{x} \; = \; -x \; + \; \mathrm{prox}_{\gamma g}\big(x - \gamma \nabla_x f(x,u)\big) \quad =: \quad \mathsf{F}_{\mathsf{ProxG}}(x,u)$$

where **proximal operator** (generalized projection) of convex, closed, proper g is

$$\mathrm{prox}_{\gamma g}(z) \; := \; \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \; g(x) \; + \; \frac{1}{2\gamma}\|x - z\|_2^2$$

Properties of proximal gradient descent

1. **well-posed Lipschitz**

2. **equivalence:** x^* minimizes $f + g \quad \iff \quad \mathsf{F}_{\mathsf{ProxG}}(x^*) = 0$

3. **decreasing energy:**

   (when bounded) composite cost $f + g$ non-increasing along flow

4. **a recurrent neural network:**

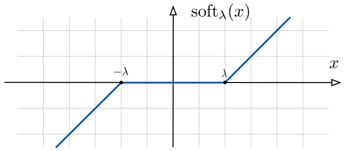
   f quadratic and $g(x) = \sum_{i=1}^{n} g_i(x_i) \quad \implies \quad \mathsf{F}_{\mathsf{ProxG}} = \mathsf{F}_{\mathsf{FR}}$

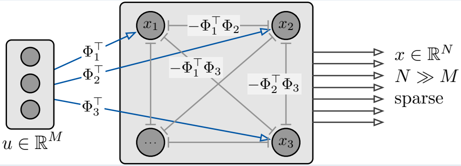
5. **contractivity:**

   $W \prec I_n \quad \implies \quad \mathsf{F}_{\mathsf{FR}}$ infinitesimally contracting
   $W \preceq I_n \quad \implies \quad \mathsf{F}_{\mathsf{FR}}$ infinitesimally non-expansive

Example #6: Biologically-plausible circuits for sparse reconstruction

$$\dot{x}(t) = \mathsf{F}_{\text{competitive}}(x,u) := -x + \text{soft}_\lambda\big((I_N - \Phi^\top\Phi)x + \Phi^\top u\big)$$





1. $x^\star$ is equilibrium $\iff$ $x^\star$ minimizes $\mathcal{E}_{\text{lasso}}(x)$
2. $\mathcal{E}_{\text{lasso}}$ is convex $\implies$ $\mathsf{F}_{\text{competitive}}$ is weakly contracting
3. Φ satisfies isometry property $\implies$ $x^\star$ is locally exp stable

$\implies$ $x^\star$ is globally linearly-exponentially stable

V. Centorrino, A. Gokhale, A. Davydov, G. Russo, and F. Bullo. Positive competitive networks for sparse reconstruction. *Neural Computation*, 36(6):1163–1197, 2024.

Outline

Selected references from my group

Contraction theory:

- A. Davydov, S. Jafarpour, and F. Bullo. Non-Euclidean contraction theory for robust nonlinear stability. *IEEE Transactions on Automatic Control*, 67(12): 6667–6681, 2022a. doi
- S. Jafarpour, A. Davydov, and F. Bullo. Non-Euclidean contraction theory for monotone and positive systems. *IEEE Transactions on Automatic Control*, 68(9):5653–5660, 2023. doi
- L. Cothren, F. Bullo, and E. Dall'Anese. Online feedback optimization and singular perturbation via contraction theory. *SIAM Journal on Control and Optimization*, Aug. 2024. doi. Submitted

Contracting neural networks:

- 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. doi
- A. Davydov, A. V. Proskurnikov, and F. Bullo. Non-Euclidean contractivity of recurrent neural networks. In *American Control Conference*, pages 1527–1534, Atlanta, USA, May 2022b. doi
- V. Centorrino, A. Gokhale, A. Davydov, G. Russo, and F. Bullo. Euclidean contractivity of neural networks with symmetric weights. *IEEE Control Systems Letters*, 7:1724–1729, 2023. doi
- V. Centorrino, A. Gokhale, A. Davydov, G. Russo, and F. Bullo. Positive competitive networks for sparse reconstruction. *Neural Computation*, 36(6): 1163–1197, 2024. doi

Weak and semicontraction theory:

- 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
- G. De Pasquale, K. D. Smith, F. Bullo, and M. E. Valcher. Dual seminorms, ergodic coefficients, and semicontraction theory. *IEEE Transactions on Automatic Control*, 69(5):3040–3053, 2024. doi
- R. Delabays and F. Bullo. Semicontraction and synchronization of Kuramoto-Sakaguchi oscillator networks. *IEEE Control Systems Letters*, 7:1566–1571, 2023. doi

Optimization:

- A. Davydov, S. Jafarpour, A. V. Proskurnikov, and F. Bullo. Non-Euclidean monotone operator theory and applications. *Journal of Machine Learning Research*, 25(307):1–33, 2024. doi. URL **http://jmlr.org/papers/v25/23-0805.html**
- A. Davydov, V. Centorrino, A. Gokhale, G. Russo, and F. Bullo. Time-varying convex optimization: A contraction and equilibrium tracking approach. *IEEE Transactions on Automatic Control*, June 2023. doi. Conditionally accepted
- A. Gokhale, A. Davydov, and F. Bullo. Contractivity of distributed optimization and Nash seeking dynamics. *IEEE Control Systems Letters*, 7:3896–3901, 2023. doi

Key references

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

Parametric optimization:
F. Borrelli, A. Bemporad, and M. Morari. *Predictive Control for Linear and Hybrid Systems*. Cambridge University Press, 2017. ISBN 1107016886

Gradient controllers:
M. Colombino, E. Dall'Anese, and A. Bernstein. Online optimization as a feedback controller: Stability and tracking. *IEEE Transactions on Control of Network Systems*, 7(1):422–432, 2020. doi

Implicit neural networks:
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

Deep reservoir computing:
C. Gallicchio, A. Micheli, and L. Pedrelli. Deep reservoir computing: A critical experimental analysis. *Neurocomputing*, 268:87–99, 2017. doi

Proximal gradient descent:
S. Hassan-Moghaddam and M. R. Jovanović. Proximal gradient flow and Douglas-Rachford splitting dynamics: Global exponential stability via integral quadratic constraints. *Automatica*, 123:109311, 2021. doi

Competitive neural networks for sparse reconstruction:
C. J. Rozell, D. H. Johnson, R. G. Baraniuk, and B. A. Olshausen. Sparse coding via thresholding and local competition in neural circuits. *Neural Computation*, 20(10):2526–2563, 2008. doi

Conclusions

contractivity = robust computationally-friendly stability
fixed point theory + Lyapunov stability theory + geometry of metric spaces

- theory (basic defs + 5 properties)
- examples (6 examples)
- applications to control, ML and neuroscience

Ongoing work

1. optimization-based control designs:
   model predictive control, control barrier functions, low-gain integral control
2. ML and biologically-inspired neural networks

search for contraction properties
design engineering systems to be contracting
verify correct/safe behavior via known Lipschitz constants

Supplementary Slides

Example #7: Primal-dual gradient dynamics

strongly convex function f s.t. $0 \prec \nu_{\text{min}} I_n \preceq \text{Hess} f \preceq \nu_{\text{max}} I_n$

constraint matrix A s.t. $0 \prec a_{\text{min}} I_m \preceq AA^\top \preceq a_{\text{max}} I_m$ (independent rows)

linearly constrained optimization:

$$\min_{x \in \mathbb{R}^n} \quad f(x)$$
$$\text{subj. to} \quad Ax = b$$

primal-dual gradient dynamics:

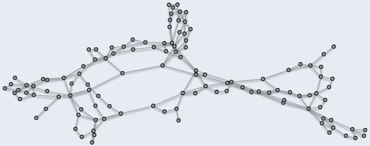
$$\begin{bmatrix} \dot{x} \\ \dot{\lambda} \end{bmatrix} = \mathsf{F}_{\text{PDG}}(x, \lambda) := \begin{bmatrix} -\nabla f(x) - A^\top \lambda \\ Ax - b \end{bmatrix}$$

F_{PDG} is infinitesimally contracting wrt $\|\cdot\|_{2,P^{1/2}}$ with rate c

$$P = \begin{bmatrix} I_n & \alpha A^\top \\ \alpha A & I_m \end{bmatrix} \text{ with } \alpha = \frac{1}{2}\min\left\{\frac{1}{\nu_{\text{max}}}, \frac{\nu_{\text{min}}}{a_{\text{max}}}\right\} \quad \text{and} \quad c = \frac{1}{4}\min\left\{\frac{a_{\text{min}}}{\nu_{\text{max}}}, \frac{a_{\text{min}}}{a_{\text{max}}}\nu_{\text{min}}\right\}$$

A. Gokhale, A. Davydov, and F. Bullo. Contractivity of distributed optimization and Nash seeking dynamics. *IEEE Control Systems Letters*, 7:3896–3901, 2023.

Example #8: Distributed gradient dynamics



decomposable cost: $\min_{x\in\mathbb{R}} \sum_{i=1}^n f_i(x)$ where each f_i is ν_i-strongly convex

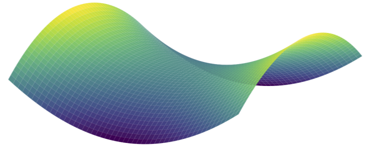
$$\begin{cases} \min_{x_{[i]}\in\mathbb{R}} & \sum_{i=1}^n f_i(x_{[i]}) \\ \text{subj. to} & \sum_{j=1}^n a_{ij}(x_i - x_j) = 0 \end{cases}$$

Laplacian-based distributed gradient (primal-dual gradient, $2n$ vars):

$$\begin{cases} \dot{x}_{[i]} = -\nabla f_i(x_{[i]}) - \sum_{j=1}^n a_{ij}(\lambda_i - \lambda_j) & \text{for each node } i \\ \dot{\lambda}_i = \sum_{j=1}^n a_{ij}(x_i - x_j) & \text{for each node } i \end{cases}$$

$\mathsf{F}_{\text{Laplacian-DistributedG}}$ **is infinitesimally contracting**† **with** $c = \frac{1}{4}\left(\frac{\lambda_2}{\lambda_n}\right)^2 \min_i \nu_i$

Example #9: Saddle dynamics



Assume $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$

- $x \mapsto f(x,y)$ is ν_x-strongly convex, uniformly in y
- $y \mapsto f(x,y)$ is ν_y-strongly concave, uniformly in x

saddle dynamics (primal-descent / dual-ascent):

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \mathsf{F_S}(x,y) := \begin{bmatrix} -\nabla_x f(x,y) \\ \nabla_y f(x,y) \end{bmatrix}$$

$\mathsf{F_S}$ is infinitesimally contracting wrt $\|\cdot\|_2$ with rate $\min\{\nu_x, \nu_y\}$

unique globally exp stable point is saddle point (min in x, max in y)

Example #10: Pseudogradient and best response play

Each player i aims to minimize its own cost function $J_i(x_i, x_{-i})$ (not a potential game)

pseudogradient dynamics (aka gradient play in game theory) $\mathsf{F}_{\mathbf{PseudoG}}$**:**

$$\dot{x}_i = -\nabla_i J_i(x_i, x_{-i})$$

- **strong convexity wrt x_i:** J_i is μ_i strongly convex wrt x_i, uniformly in x_{-i}
- **Lipschitz wrt x_{-i}:** $\mathrm{Lip}_{x_j}(\nabla_i J_i) \leq \ell_{ij}$, uniformly in x_{-j}
- $\mathsf{F}_{\mathsf{PseudoG}}$ gain matrix is Hurwitz

$\Longrightarrow$ $\mathsf{F}_{\mathbf{PseudoG}}$ **is infinitesimally contracting wrt appropriate diag-weighted** $\|\cdot\|_2$

Example #11: Best response play

Each player i aims to minimize its own cost function $J_i(x_i, x_{-i})$
$\mathsf{BR}_i : x_{-i} \to \operatorname{argmin}_{x_i} J_i(x_i, x_{-i})$ best response of player i wrt other decisions x_{-i}

best response dynamics:

$$\dot{x} = \mathsf{F}_{\mathsf{BR}}(x) := \mathsf{BR}(x) - x$$
$$\iff \quad \dot{x}_i = \mathsf{BR}_i(x_{-i}) - x_i$$

- **strong convexity wrt x_i:** J_i is μ_i strongly convex wrt x_i, uniformly in x_{-i}
- **Lipschitz wrt x_{-i}:** $\mathsf{Lip}_{x_j}(\nabla_i J_i) \leq \ell_{ij}$, uniformly in x_{-j}
  $\implies$ BR_i **is Lipschitz wrt x_j with constant ℓ_{ij}/μ_i**
- F_{BR} gain matrix is Hurwitz $\iff$ BR is a discrete-time contraction

  $\implies$ $\mathsf{BR} - \mathsf{Id}$ **is infinitesimally contracting wrt appropriate diag-weighted $\|\cdot\|_2$**

Equivalent statements:

1. $\mathsf{F}_{\mathsf{PseudoG}}$ gain matrix: $\begin{bmatrix} -\mu_1 & \dots & \ell_{1n} \\ \vdots & & \vdots \\ \ell_{n1} & \dots & -\mu_n \end{bmatrix}$ is Hurwitz

2. F_{BR} gain matrix: $\begin{bmatrix} -1 & \dots & \ell_{1n}/\mu_1 \\ \vdots & & \vdots \\ \ell_{n1}/\mu_n & \dots & -1 \end{bmatrix}$ is Hurwitz

3. discrete-time F_{BR} gain matrix: $\begin{bmatrix} 0 & \dots & \ell_{1n}/\mu_1 \\ \vdots & & \vdots \\ \ell_{n1}/\mu_n & \dots & 0 \end{bmatrix}$ is Schur

Aggregative games: $J_i(x_i, x_{-i}) = f_i\big(x_i, \frac{1}{n}\sum_{j=1}^n x_j\big)$
assume f_i is μ_i-strongly convex wrt x_i and $\ell_i = \mathrm{Lip}_y(\nabla_{x_i} f_i(x_i, y))$

$\mu_i > \ell_i$ for each agent i $\implies$ gain matrix is Hurwitz

A. Gokhale, A. Davydov, and F. Bullo. Contractivity of distributed optimization and Nash seeking dynamics. *IEEE Control Systems Letters*, 7:3896–3901, 2023. doi

Example #12: Projected gradient controller

Constrained feedback optimization:

$$\min_u \quad \mathcal{E}(u,w) = \phi(u) + \psi(Y_u u + Y_w w) \qquad (\nu \text{ strongly convex, } \ell_u \text{ strongly smooth, } \ell_w)$$

$$\text{subj. to} \quad u \in \mathcal{U} \qquad (\text{nonempty, closed, convex. } P_{\mathcal{U}} = \text{orthogonal projection})$$

Projected gradient controller

$$\dot{u} = \mathsf{F}_{\mathsf{PGC}}(u,w) := -u + P_{\mathcal{U}}\big(u - \gamma \nabla_u \mathcal{E}(u,w)\big)$$

Equilibrium tracking for projected gradient controller At $\gamma = \frac{2}{\nu+\ell_u}$,

$$\limsup_{t\to\infty} \|u(t) - u^*(t)\| \quad \leq \quad \frac{\ell_{\mathsf{PGC}}}{c_{\mathsf{PGC}}^2} \limsup_{t\to\infty} \|\dot{w}(t)\| \qquad \text{(eq tracking)}$$

1. $\text{osLip}_u(\mathsf{F}_{\mathsf{PGC}}) \leq -c_{\mathsf{PGC}} := -\dfrac{2\nu}{\nu+\ell_u}$ (contractivity prox gradient)
2. $\text{Lip}_w(\mathsf{F}_{\mathsf{PGC}}) = \ell_{\mathsf{PGC}} := \dfrac{2}{\nu+\ell_u}\ell_w$

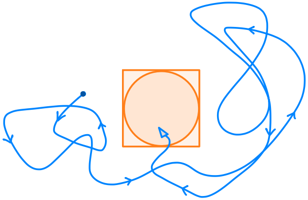
Advantages of non-Euclidean approaches

1. *well suited for certain class of systems*
   ℓ_1 for monotone flow systems
2. *computational advantages*
   ℓ_1/ℓ_∞ constraints lead to LPs, whereas ℓ_2 constraints leads to LMIs
3. *robustness to structural perturbations*
   ℓ_1/ℓ_∞ contractions are connectively robust (i.e., edge removal)
4. *adversarial input-output analysis*
   ℓ_∞ better suited for the analysis of adversarial examples than ℓ_2
5. *asynchronous distributed computation*
   ℓ_∞ contractions converge under fully asynchronous distributed execution

NonEuclidean contractions: biological transcriptional systems (Russo, Di Bernardo, and Sontag, 2010), Hopfield neural networks (Fang and Kincaid, 1996; Qiao, Peng, and Xu, 2001), chemical reaction networks (Al-Radhawi, Angeli, and Sontag, 2020), traffic networks (Coogan and Arcak, 2015; Como, Lovisari, and Savla, 2015; Coogan, 2019), multi-vehicle systems (Monteil, Russo, and Shorten, 2019), and coupled oscillators (Russo, Di Bernardo, and Sontag, 2013; Aminzare and Sontag, 2014)

Practical stability problem and the counter-intuitive nature of $\mathbb{R}^n$

Boris Polyak (1935-2023) used to say "$\mathbb{R}^n$ countradicts our intuition"



Aim: **compute settling time inside a desired set**

- since norms on $\mathbb{R}^n$ are equivalent, no formal difference in the choice of norm
- assume: can tolerate ± 1 error in each coordinate
  $\implies$ desired set is hypercube $= \ell_\infty$-ball
- assume: Lyapunov function is $V(x) = \|x\|_2^2$
  $\implies$ need to wait until solution enters unit ℓ_2-ball $\subset$ unit ℓ_∞-ball

- but n-sphere inscribed in n-hypercube is very small fraction!
  as $n \to \infty$, the ratio of volumes decreases faster than any exponential function

for large n, quadratic Lyap fnctns may provide exponentially conservative estimates

Courtesy of Anton Proskurnikov, Politecnico di Torino

$f(\mathbf{x})$	$\text{dom}(f)$	$\text{prox}_f(\mathbf{x})$	Assumptions	Reference
$\frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c$	$\mathbb{R}^n$	$(\mathbf{A}+\mathbf{I})^{-1}(\mathbf{x}-\mathbf{b})$	$\mathbf{A} \in \mathbb{S}^n_+$, $\mathbf{b} \in \mathbb{R}^n$, $c \in \mathbb{R}$	Section 6.2.3
λx^3	$\mathbb{R}_+$	$\frac{-1+\sqrt{1+12\lambda[x]_+}}{6\lambda}$	$\lambda > 0$	Lemma 6.5
μx	$[0,\alpha] \cap \mathbb{R}$	$\min\{\max\{x-\mu,0\},\alpha\}$	$\mu \in \mathbb{R}$, $\alpha \in [0,\infty]$	Example 6.14
$\lambda\\|\mathbf{x}\\|$	$\mathbb{E}$	$\left(1 - \frac{\lambda}{\max\{\\|\mathbf{x}\\|,\lambda\}}\right)\mathbf{x}$	$\\|\cdot\\|$—Euclidean norm, $\lambda > 0$	Example 6.19
$-\lambda\\|\mathbf{x}\\|$	$\mathbb{E}$	$\left(1 + \frac{\lambda}{\\|\mathbf{x}\\|}\right)\mathbf{x}$, $\mathbf{x} \neq \mathbf{0}$, $\{\mathbf{u} : \\|\mathbf{u}\\| = \lambda\}$, $\mathbf{x} = \mathbf{0}$.	$\\|\cdot\\|$—Euclidean norm, $\lambda > 0$	Example 6.21
$\lambda\\|\mathbf{x}\\|_1$	$\mathbb{R}^n$	$\mathcal{T}_\lambda(\mathbf{x}) = [\lvert\mathbf{x}\rvert - \lambda\mathbf{e}]_+ \odot \text{sgn}(\mathbf{x})$	$\lambda > 0$	Example 6.8
$\\|\boldsymbol{\omega} \odot \mathbf{x}\\|_1$	$\text{Box}[-\boldsymbol{\alpha},\boldsymbol{\alpha}]$	$\mathcal{S}_{\boldsymbol{\omega},\boldsymbol{\alpha}}(\mathbf{x})$	$\boldsymbol{\alpha} \in [0,\infty]^n$, $\boldsymbol{\omega} \in \mathbb{R}^n_+$	Example 6.23
$\lambda\\|\mathbf{x}\\|_\infty$	$\mathbb{R}^n$	$\mathbf{x} - \lambda P_{B_{\\|\cdot\\|_1}[\mathbf{0},1]}(\mathbf{x}/\lambda)$	$\lambda > 0$	Example 6.48
$\lambda\\|\mathbf{x}\\|_a$	$\mathbb{E}$	$\mathbf{x} - \lambda P_{B_{\\|\cdot\\|_{a,*}}[\mathbf{0},1]}(\mathbf{x}/\lambda)$	$\\|\mathbf{x}\\|_a$—arbitrary norm, $\lambda > 0$	Example 6.47
$\lambda\\|\mathbf{x}\\|_0$	$\mathbb{R}^n$	$\mathcal{H}_{\sqrt{2\lambda}}(x_1) \times \cdots \times \mathcal{H}_{\sqrt{2\lambda}}(x_n)$	$\lambda > 0$	Example 6.10
$\lambda\\|\mathbf{x}\\|^3$	$\mathbb{E}$	$\frac{2}{1+\sqrt{1+12\lambda\\|\mathbf{x}\\|}}\mathbf{x}$	$\\|\cdot\\|$—Euclidean norm, $\lambda > 0$,	Example 6.20
$-\lambda\sum_{j=1}^n \log x_j$	$\mathbb{R}^n_{++}$	$\left(\frac{x_j+\sqrt{x_j^2+4\lambda}}{2}\right)_{j=1}^n$	$\lambda > 0$	Example 6.9
$\delta_C(\mathbf{x})$	$\mathbb{E}$	$P_C(\mathbf{x})$	$\emptyset \neq C \subseteq \mathbb{E}$	Theorem 6.24
$\lambda\sigma_C(\mathbf{x})$	$\mathbb{E}$	$\mathbf{x} - \lambda P_C(\mathbf{x}/\lambda)$	$\lambda > 0$, $C \neq \emptyset$ closed convex	Theorem 6.46
$\lambda\max\{x_i\}$	$\mathbb{R}^n$	$\mathbf{x} - \lambda P_{\Delta_n}(\mathbf{x}/\lambda)$	$\lambda > 0$	Example 6.49
$\lambda\sum_{i=1}^k x_{[i]}$	$\mathbb{R}^n$	$\mathbf{x} - \lambda P_C(\mathbf{x}/\lambda)$, $C = H_{\mathbf{e},k} \cap \text{Box}[\mathbf{0},\mathbf{e}]$	$\lambda > 0$	Example 6.50
$\lambda\sum_{i=1}^k \lvert x_{\langle i\rangle}\rvert$	$\mathbb{R}^n$	$\mathbf{x} - \lambda P_C(\mathbf{x}/\lambda)$, $C = B_{\\|\cdot\\|_1}[\mathbf{0},k] \cap \text{Box}[-\mathbf{e},\mathbf{e}]$	$\lambda > 0$	Example 6.51
$\lambda M_f^\mu(\mathbf{x})$	$\mathbb{E}$	$\mathbf{x} + \frac{\lambda}{\mu+\lambda}\left(\text{prox}_{(\mu+\lambda)f}(\mathbf{x}) - \mathbf{x}\right)$	$\lambda,\mu > 0$, f proper closed convex	Corollary 6.64
$\lambda d_C(\mathbf{x})$	$\mathbb{E}$	$\mathbf{x} + \min\left\{\frac{\lambda}{d_C(\mathbf{x})},1\right\}(P_C(\mathbf{x}) - \mathbf{x})$	$\emptyset \neq C$ closed convex, $\lambda > 0$	Lemma 6.43
$\frac{\lambda}{2}d_C^2(\mathbf{x})$	$\mathbb{E}$	$\frac{\lambda}{\lambda+1}P_C(\mathbf{x}) + \frac{1}{\lambda+1}\mathbf{x}$	$\emptyset \neq C$ closed convex, $\lambda > 0$	Example 6.65
$\lambda H_\mu(\mathbf{x})$	$\mathbb{E}$	$\left(1 - \frac{\lambda}{\max\{\\|\mathbf{x}\\|,\mu+\lambda\}}\right)\mathbf{x}$	$\lambda,\mu > 0$	Example 6.66
$\rho\\|\mathbf{x}\\|_1^2$	$\mathbb{R}^n$	$\left(\frac{v_i x_i}{v_i+2\rho}\right)_{i=1}^n$, $\mathbf{v} = \left[\sqrt{\frac{\rho}{\mu}}\lvert\mathbf{x}\rvert - 2\rho\right]_+$, $\mathbf{e}^T\mathbf{v} = 1$ ($\mathbf{0}$ when $\mathbf{x} = \mathbf{0}$)	$\rho > 0$	Lemma 6.70
$\lambda\\|\mathbf{A}\mathbf{x}\\|_2$	$\mathbb{R}^n$	$\mathbf{x} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T + \alpha^*\mathbf{I})^{-1}\mathbf{A}\mathbf{x}$, $\alpha^* = 0$ if $\\|\mathbf{v}_0\\|_2 \leq \lambda$; otherwise, $\\|\mathbf{v}_{\alpha^*}\\|_2 = \lambda$; $\mathbf{v}_\alpha \equiv (\mathbf{A}\mathbf{A}^T + \alpha\mathbf{I})^{-1}\mathbf{A}\mathbf{x}$	$\mathbf{A} \in \mathbb{R}^{m\times n}$ with full row rank, $\lambda > 0$	Lemma 6.68

proximal operator

well-defined for all ccp functions,
generalized form of projection,
non-expansive

helps generalize gradient algorithms/dynamics to proximal algorithms/dynamics, useful for nonsmooth, constrained, large-scale, and distributed optimization

evaluation of proximal operator requires small convex optimization,
see Summary of prox computations, Beck 2017

A. Beck. *First-Order Methods in Optimization*. SIAM, 2017. ISBN 978-1-61197-498-0

N. Parikh and S. Boyd. Proximal algorithms. *Foundations and Trends in Optimization*, 1(3):127–239, 2014.

Theoretical frontiers

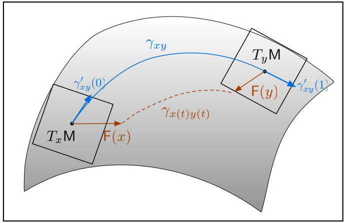
- higher order contraction and pseudocontration (dominance)
- relationship with monotone operator theory
- metric spaces
- computational methods

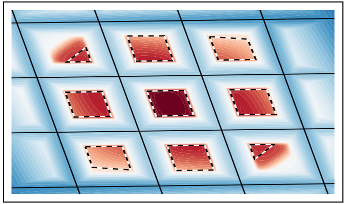
Limitations: not all stable systems are contractive:

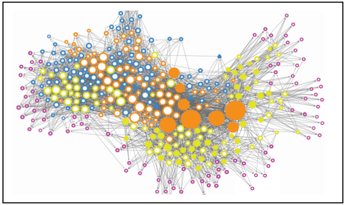
- Lyapunov-diagonally-stable networks
- multistable and locally contracting systems
- control contraction design

Application to networks, control and learning

1. reaction networks
2. control: optimization-based control design
3. ML: implicit models and energy-based learning
4. neuroscience: robust dynamical modeling, normative frameworks, biologically plausible learning







contractivity = robust computationally-friendly stability

fixed point theory + Lyapunov stability theory + geometry of metric spaces

	Lyapunov Theory	**Contraction Theory for Dynamical Systems**
	F admits global Lyapunov function	F is strongly contracting
existence of equilibrium	assumed	implied + computational methods
Lyapunov function	arbitrary	$\|x - x^*\|$ and $\|\mathsf{F}(x)\|$
inputs	ISS via $\mathcal{KL}$ and $\mathcal{L}$ functions	exponential iISS with explicit constants

	Krasovskiĭ-LaSalle Inv Principle	**Weakly Contracting Systems**
	generic V s.t. $\mathcal{L}_{\mathsf{F}}V \leq 0$	F is weakly contracting, that is, $\mathrm{osLip}(\mathsf{F}) \leq 0$
(no other assumptions)		Dichotomy Theorem
assuming bounded traj.	convergence to Krasovski-LaSalle set	each equilibrium is stable
assuming Krasovski-LaSalle set $= \{x^*\}$ is LAS	$\{x^*\}$ is GAS	$\{x^*\}$ is GAS, linear-exponential convergence, local ISS + explicit constants

Gradient dynamics for convex functions

Given differentiable convex $f : \mathbb{R}^n \to \mathbb{R}$, **gradient dynamics**

$$\dot{x} = \mathsf{F}_{\mathsf{G}}(x) := -\nabla f(x)$$

Dichotomy and Convergence

1. $-\nabla f$ has no equilibrium, f has no minimum, and every trajectory is unbounded, or
2. $-\nabla f$ has at least one equilibrium $x^* \in \mathbb{R}^n$ and the following properties hold:
   1. f is constant on convex set of equilibria, each local is a global minimum,
   2. every trajectory is bounded and converges to a minimum, each equilibrium is stable
   3. if x^* is locally asymptotically stable, then x^* is globally asymptotically stable
   4. if $\mu_2(-\mathrm{Hess}(f)(x^*)) < 0$, then linear exponential decay and $x \mapsto \|x - x^*\|_2$ is a global Lyap

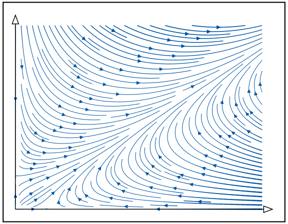
From strongly to weakly contracting systems

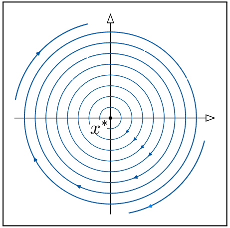
Given a norm $\|\cdot\|$, consider

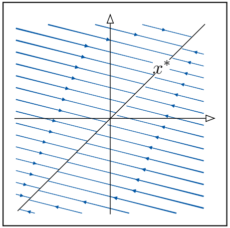
$$\dot{x} = \mathsf{F}(x) \qquad \text{satisfying} \qquad \mathsf{osLip}(\mathsf{F}) = 0$$

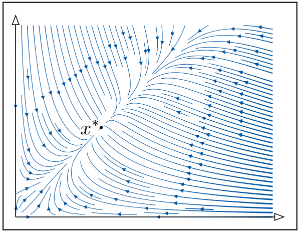
Dichotomy for weakly-contracting systems

1. no equilibrium and every trajectory is unbounded, or
2. at least one equilibrium, every trajectory is bounded, and local asy stability $\implies$ global









$$\dot{x} = \mathsf{F}(t,x) \qquad \text{on } \mathbb{R}^n \text{ with norm } \|\cdot\|_{\mathsf{glo}}$$

1. F is weakly contracting wrt $\|\cdot\|_{\mathsf{glo}}$
2. $x^\star$ is locally-exponentially-stable equilibrium

$\implies$ F is locally c-strongly contracting wrt $\|\cdot\|_{\mathsf{loc}}$ over forward-invariant $\mathcal{S}$