

Biologically Plausible Optimization: Competitive Neural Circuits & Contraction Theory

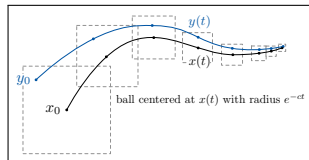
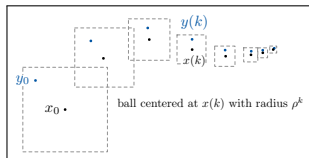
Francesco Bullo

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

<https://fbullo.github.io/ctds>

AFOSR Grant FA9550-22-1-0059

Dynamical Systems and Control Theory Program Review, Dr. Fred Leve, Aug 27, 2024

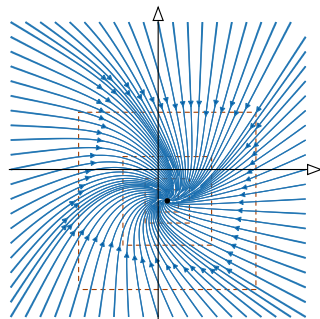


contractivity = robust computationally-friendly stability

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

highly-ordered transient and asymptotic behavior, no anonymous constants/functions:

- ① unique globally exponential stable equilibrium
& two natural Lyapunov functions
- ② robustness properties
 - bounded input, bounded output (iss)
 - finite input-state gain
 - robustness margin wrt unmodeled dynamics
 - robustness margin wrt delayed dynamics
- ③ periodic input, periodic output
- ④ modularity and interconnection properties
- ⑤ accurate numerical integration and equilibrium point computation



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


2021/22 review presentation (1st year)

① theory: **contractivity for ℓ_1/ℓ_∞ norms, network contraction theorem**

A. Davydov, S. Jafarpour, and F. Bullo. Non-Euclidean contraction theory for robust nonlinear stability. *IEEE Transactions on Automatic Control*, 67(12):6667–6681, 2022. 

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. 


② examples: **non-Euclidean contractivity & fixed point theory for 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. 

A. Davydov, A. V. Proskurnikov, and F. Bullo. Non-Euclidean contraction analysis of continuous-time neural networks. *IEEE Transactions on Automatic Control*, 2024. . To appear


2022/23 review highlights (2nd year)

① theory: **equilibrium propagation**

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. . Submitted


② examples: **symmetric neural networks, gradient dynamics**

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. 

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

③ extensions: **semicontraction and ergodicity, local contractivity, phase-coupled oscillators**


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. 

R. Delabays, S. Jafarpour, and F. Bullo. Multistabilities and anomalies in oscillator models of lossy power grids. *Nature Communications*, 13:5238, 2022. 


2023/24 review highlights (3rd year)

1 theory: **Lur'e models, singular perturbation**


control: **optimization-based controllers & online feedback optimization**


A. Davydov and F. Bullo. Exponential stability of parametric optimization-based controllers via Lur'e contractivity. *IEEE Control Systems Letters*, 8:1277–1282, 2024b. 

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. . Submitted

Z. Marvi, F. Bullo, and A. G. Alleyne. Control barrier proximal dynamics: A contraction theoretic approach for safety verification. *IEEE Control Systems Letters*, 8:880–885, 2024. 


2 applications: **biologically plausible optimization, learning contracting dynamics**

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

S. Jaffe, A. Davydov, D. Lapsekili, A. K. Singh, and F. Bullo. Learning neural contracting dynamics: Extended linearization and global guarantees. In *Advances in Neural Information Processing Systems*, 2024. . Submitted

3 **Dissemination:**

Minicourse “Contracting and Gradient Dynamics”, Linköping Sweden, 12h, Sep 2023 [youtube link](#)

A. Davydov and F. Bullo. Perspectives on contractivity in control, optimization and learning. *IEEE Control Systems Letters*, 2024a. . To appear (Invited Opinion Paper)

Contraction Theory for Dynamical Systems

Francesco Bullo

Contraction Theory for Dynamical Systems, Francesco Bullo,
KDP, 1.2 edition, 2024, ISBN 979-8836646806
(252 pages and 94 exercises)

- ① Textbook with exercises and answers.
 - ② Content:
 - Fixed point theory
 - Theory of contracting dynamics on vector spaces
 - Applications to nonlinear and interconnected systems
 - Weakly contracting, monotone and semicontracting systems
 - ③ Self-Published and Print-on-Demand at:
<https://www.amazon.com/dp/B0B4K1BTF4>
 - ④ PDF Freely available at <https://fbullo.github.io/ctds>
 - ⑤ Other formats: slides (PDF), paperback & hardcover (amazon)
 - ⑥ 12h youtube minicourse: [youtube link](#)
- "Continuous improvement is better than delayed perfection"
Mark Twain

Acknowledgments, ongoing and recently-completed collaborations



Veronica Centorrino
Scuola Sup Meridionale



Alexander Davydov
UC Santa Barbara



Anand Gokhale
UC Santa Barbara

- Ambuj Singh, UC Santa Barbara: ML and contraction
- Andrew Alleyne, Univ Minnesota: control barrier functions
- Anton Proskurnikov, Politecnico Turin, Italy: nonEuclidean S-lemma
- Elena Valcher, Univ Padova, Italy: semicontraction and ergodic coefficients
- Emiliano Dall'Anese, Boston Univ: singular perturbation and MPC
- Giovanni Russo, Univ Salerno, Italy: neuroscience
- Michael Margaliot, Tel Aviv Univ, Israel: k -contractivity
- Robin Delabays, Univ Sion, Switzerland: Kuramoto oscillators

§1. Problem setup


- Functionality and analysis of biological networks
- Sparse signal reconstruction in biological neuronal circuits
- Sparse signal reconstruction in engineering

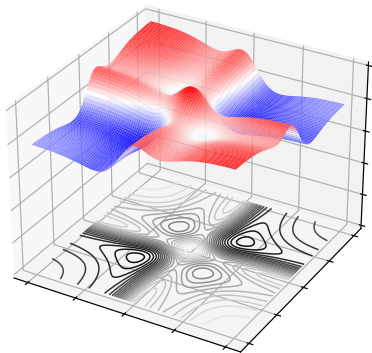
§2. Two-step analysis: proximal gradients and contractivity

- Step 1: Transcribing optimization problems into interpretable recurrent neural networks
- Step 2: Convergence analysis for non-expansive systems

§3. Conclusion and ongoing research

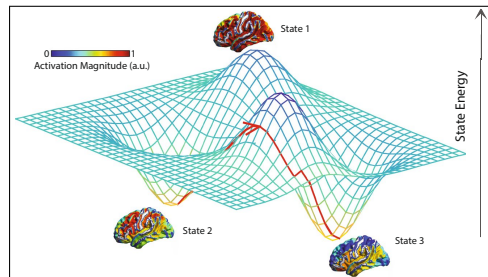
“The idea that the brain functions so as to minimize certain costs pervades theoretical neuroscience.”

S. C. Surace, J.-P. Pfister, W. Gerstner, and J. Brea. On the choice of metric in gradient-based theories of brain function. *PLOS Computational Biology*, 16(4):e1007640, 2020. 




Energy landscape for associative memory in Hopfield models

S. Betteti, G. Baggio, F. Bullo, and S. Zampieri. Stimulus-driven dynamics for robust memory retrieval in Hopfield networks. 2024. Under review



Energy of activity within a system, and energy of activity shared between systems


S. Gu, M. Cieslak, B. Baird, S. F. Muldoon, S. T. Grafton, F. Pasqualetti, and D. S. Bassett. The energy landscape of neurophysiological activity implicit in brain network structure. *Scientific Reports*, 8(1), 2018. 


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

where A is synaptic matrix, Ψ is activation function, and u is stimulus

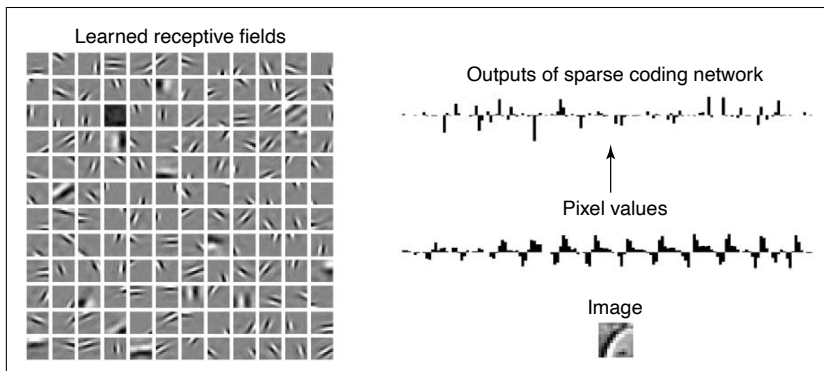
- 1 What is the functionality of F_{FR} ?
- 2 Specifically, what is F_{FR} optimizing?
- 3 Is there a normative top-down framework for neural circuits?
- 4 Can biologically-plausible circuits help design improved artificial NNs?

Case study: dimensionality reduction


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. 


C. Pehlevan and D. B. Chklovskii. Neuroscience-inspired online unsupervised learning algorithms: Artificial neural networks. *IEEE Signal Processing Magazine*, 36(6):88–96, 2019. 

Sparse signal reconstruction in biological neuronal circuits

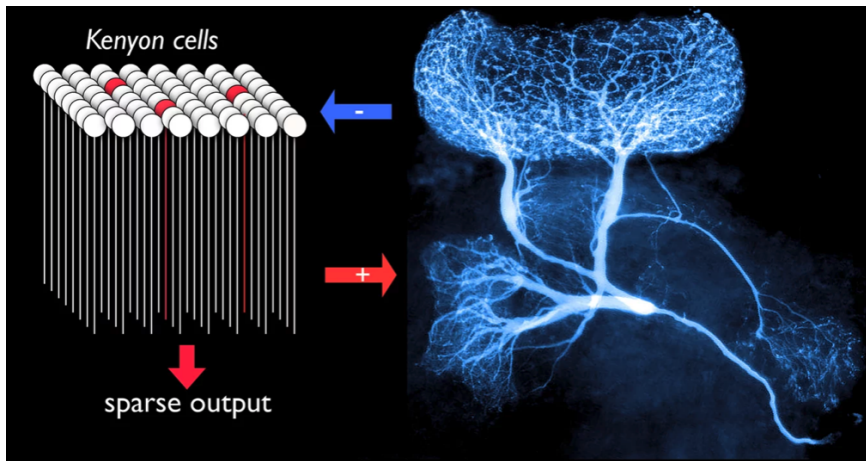


- primary visual area (V1) sparsifies signals
- receptive fields (\approx dictionary) are learned empirically

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. 

From mammals to insects



- mushroom body with giant interneuron of a locust
- all excitatory → interneuron → all excitatory enables sparse coding

M. Papadopoulou, S. Cassenaer, T. Nowotny, and G. Laurent. Normalization for sparse encoding of odors by a wide-field interneuron. *Science*, 332(6030):721–725, 2011. [doi](#)

Sparse reconstruction by minimizing the lasso energy

$$\min_{x \in \mathbb{R}^N, x \geq 0} \mathcal{E}_{\text{lasso}}(x) := \frac{1}{2} \|u - \Phi x\|_2^2 + \lambda \|x\|_1$$

where Φ overcomplete dictionary matrix, with $\|\Phi_i\| = 1$ and $\Phi_i \cdot \Phi_j = \text{similarity between } (i, j)$

The diagram illustrates the sparse reconstruction equation $u \approx \Phi x = [\Phi_1 | \Phi_2 | \dots | \Phi_N] x$. On the left, a vertical rectangle labeled u with dimensions $(M \times 1)$ is followed by an approximation symbol \approx . This is followed by a large horizontal rectangle labeled Φ with dimensions $(M \times N)$, and a vertical rectangle labeled x with dimensions $(N \times 1)$. To the right of this is an equals sign $=$, followed by a large horizontal rectangle containing the expression $\Phi_1 | \Phi_2 | \dots | \Phi_N$ with dimensions $(M \times N)$, and another vertical rectangle labeled x with dimensions $(N \times 1)$.

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

$$\begin{array}{ccc} \boxed{u} & \approx & \boxed{\Phi_1 \mid \cdots \mid \Phi_N} \\ (M \times 1) & & (M \times N) \end{array} \quad \begin{array}{c} \boxed{x} \\ (N \times 1) \end{array} \quad k \ll M \ll N, \quad \|\Phi_i\|_2 = 1$$

Sparse stimulus recovery from lasso:

If Φ is a *restricted isometry* with $(k \ll M, \delta_k < 1)$ if, for all k -sparse x ,

$$(1 - \delta_k) \|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta_k) \|x\|_2^2$$

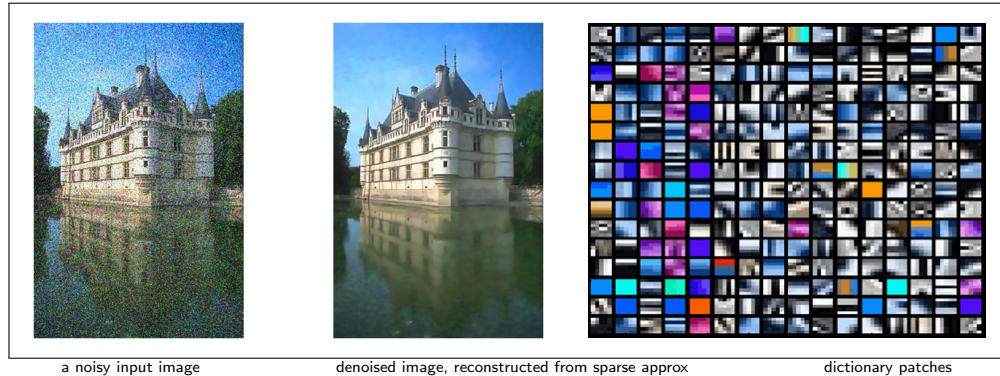
Then exact* recovery from unique solution to *lasso problem*

$$\min_{y \in \mathbb{R}^N} E_{\text{lasso}}(y) := \underbrace{\|u - \Phi y\|_2^2}_{\text{quadratic reconstruction cost}} + \lambda \underbrace{\|y\|_1}_{\text{sparsity-promoting}}$$

Gramian matrix $\Phi^\top \Phi \succeq 0$ implies E_{lasso} only convex

Restricted isometry implies $\Phi_k^\top \Phi_k \succ 0$ and restricted E_{lasso} strongly convex

Sparse signal reconstruction in engineering



- identify and exploit sparsity in signals
- dimensionality reduction in machine learning

E. J. Candes and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, 2005

J. Wright and Y. Ma. *High-Dimensional Data Analysis with Low-Dimensional Models: Principles, Computation, and Applications*. Cambridge University Press, 2022

§1. Problem setup

- Functionality and analysis of biological networks
- Sparse signal reconstruction in biological neuronal circuits
- Sparse signal reconstruction in engineering

§2. Two-step analysis: proximal gradients and contractivity

- Step 1: Transcribing optimization problems into interpretable recurrent neural networks
- Step 2: Convergence analysis for non-expansive systems

§3. Conclusion and ongoing research

Minimization of composite cost:


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

proximal gradient descent:

$$\dot{x} = -x + \text{prox}_{\gamma g}(x - \gamma \nabla_x f(x, u)) =: F_{\text{ProxG}}(x, u)$$

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

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

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. 

Prox gradient descent = firing rate network

$$\begin{aligned}\dot{x} &= F_{\text{FR}}(x) &:= -x + \Phi(Ax + Bu) \\ \dot{x} &= F_{\text{ProxG}}(x, u) &:= -x + \text{prox}_{\gamma g}(x - \gamma \nabla_x f(x, u))\end{aligned}$$

① **well-posed Lipschitz**

② **equivalence:** x^* minimizes $f + g \iff F_{\text{ProxG}}(x^*) = 0$

③ **decreasing energy:**

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

④ **a recurrent neural network:**

$$f \text{ quadratic and } g(x) = \sum_{i=1}^n g_i(x_i) \implies F_{\text{ProxG}} = F_{\text{FR}}$$

⑤ **contractivity:**

$$\begin{aligned}W \prec I_n &\implies F_{\text{FR}} \text{ infinitesimally contracting} \\ W \preceq I_n &\implies F_{\text{FR}} \text{ infinitesimally non-expansive}\end{aligned}$$

Firing rate network for sparse reconstruction

For **positive** E_{lasso}

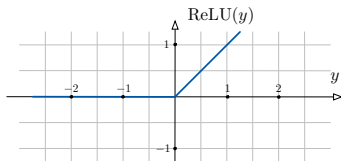
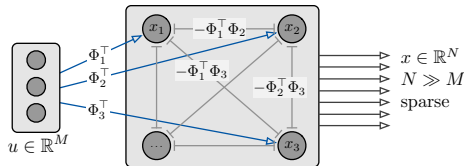
$$\min_{y \in \mathbb{R}^N, y \geq 0} E_{\text{lasso}}(y) := \|u - \Phi y\|_2^2 + \lambda \|y\|_1$$

proximal gradient dynamics is **firing rate with shifted relu**:

$$\dot{x} = -x + \text{relu}\left((I_n - \Phi^\top \Phi)x + \Phi^\top u - \lambda \mathbb{1}_n\right)$$

or, in components

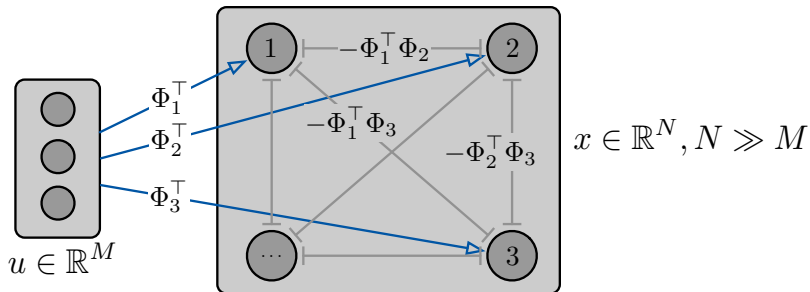
$$\dot{x}_i = -x_i + \text{relu}\left(-\sum_{j \neq i} \Phi^\top \Phi_j x_j + \Phi^\top u + \lambda\right)$$



Biological interpretation = lateral inhibition & competition

Nonnegative firing rates and non-negative dictionary elements Φ_i :

$$\dot{x}_i = -x_i + \text{relu} \left(\underbrace{\sum_{j \neq i} (-\Phi_i^\top \Phi_j)}_{\text{lateral inhibition}} x_j + \Phi_i^\top u + \lambda \right)$$



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.

A. Balavoine, J. Romberg, and C. J. Rozell. Convergence and rate analysis of neural networks for sparse approximation. *IEEE Transactions on Neural Networks and Learning Systems*, 23(9):1377–1389, 2012.

§1. Problem setup

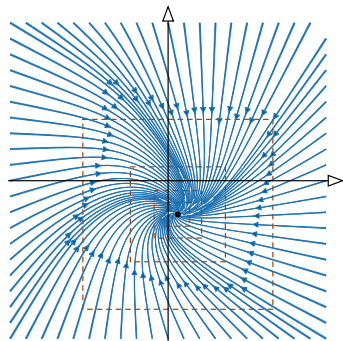
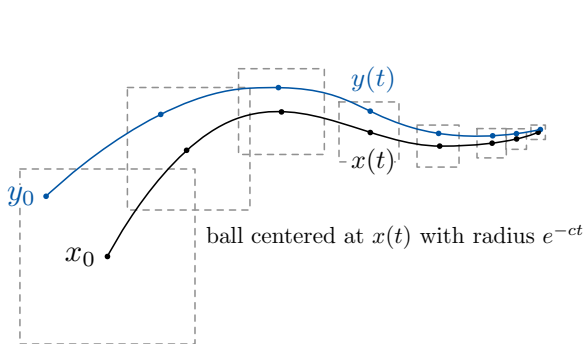
- Functionality and analysis of biological networks
- Sparse signal reconstruction in biological neuronal circuits
- Sparse signal reconstruction in engineering

§2. Two-step analysis: proximal gradients and contractivity

- Step 1: Transcribing optimization problems into interpretable recurrent neural networks
- Step 2: Convergence analysis for non-expansive systems

§3. Conclusion and ongoing research

Contracting and non-expansive dynamics



contraction rate r akin to rate of convergence

f strongly convex $\iff -\nabla f$ strongly contracting ($r > 0$)

f **convex** $\iff -\nabla f$ **non-expansive** ($r = 0$)

ℓ_2 **contractivity analysis** for:

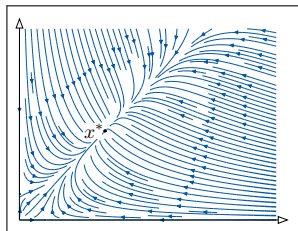
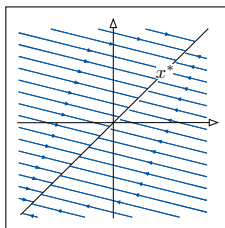
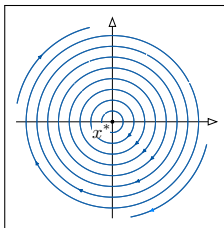
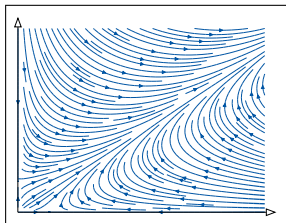
$$\begin{aligned}\dot{x} &= F_{\text{FR}}(x) = -x + \Psi(Wx + Bu) \\ 0 &\leq \Phi'_i(y) \leq 1 \quad \text{and} \quad W = W^\top\end{aligned}$$


- ① E_{lasso} is convex in x
 $\implies F_{\text{FR}}(x)$ is globally non-expansive
- ② for restricted isometric dictionaries Φ , E_{lasso} has strict minimum
 $\implies F_{\text{FR}}(x)$ is strongly contracting locally near equilibrium x^*

Do non-expansive systems converge?

Dichotomy for non-expansive 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

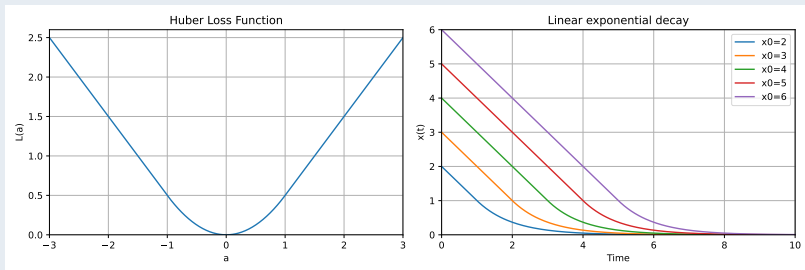


S. Jafarpour, P. Cisneros-Velarde, and F. Bullo. Weak and semi-contraction for network systems and diffusively-coupled oscillators. *IEEE Transactions on Automatic Control*, 67(3):1285–1300, 2022. 

Lin-exp convergence for non-expansive + locally contracting

Convex quadratic-linear function (Huber loss) leads to linear-exponential decay

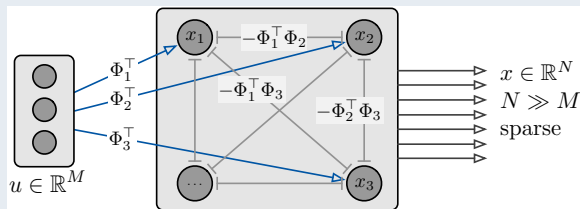
$$f_{\text{Huber}}(x) = \begin{cases} \frac{1}{2}x^2 & \text{if } |x| \leq 1 \\ |x| - \frac{1}{2} & \text{if } |x| > 1 \end{cases} \quad \Rightarrow \quad \dot{x} = -\nabla f_{\text{Huber}}(x) = -\text{sat}(x)$$



- ① as Huber loss, E_{lasso} is convex with strict minimum point x^*
- ② saturation dynamics and competitive firing rate are both:
 - globally non-expansive & locally strongly contracting near x^*
 - linear-exponentially converging

Main result: biologically-plausible & converging neural circuits

$$\dot{x}(t) = -x + \text{relu}\left((I_N - \Phi^\top \Phi)x + \Phi^\top u - \lambda \mathbb{1}_n\right)$$



- | | | |
|--|------------|---|
| 1 x^* is equilibrium | \iff | x^* minimizes $\mathcal{E}_{\text{lasso}}(x)$ |
| 2 $\mathcal{E}_{\text{lasso}}$ is convex | \implies | $F_{\text{competitive}}$ is non-expansive |
| 3 Φ satisfies isometry property | \implies | x^* is locally exp stable |

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

§1. Problem setup

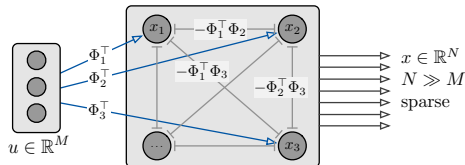
- Functionality and analysis of biological networks
- Sparse signal reconstruction in biological neuronal circuits
- Sparse signal reconstruction in engineering

§2. Two-step analysis: proximal gradients and contractivity

- Step 1: Transcribing optimization problems into interpretable recurrent neural networks
- Step 2: Convergence analysis for non-expansive systems

§3. Conclusion and ongoing research

Summary



❶ competitive positive firing rate dynamics for sparse reconstruction

❷ normative top-down explanation


from lasso \implies interpretable competitive neural networks
from gradient descent to proximal gradient descent


❸ sharp contractivity estimates for symmetric neural networks

❹ linear-exponential convergence and local input-to-state stability

1 neural networks for sparse reconstruction

- Hebbian learning of the dictionary
- ℓ_0 , entropy, and nonconvex sparsity-promoting regularizers
- similarity matching and low-rank matrix factorization


C. Pehlevan, A. M. Sengupta, and D. B. Chklovskii. Why do similarity matching objectives lead to Hebbian/anti-Hebbian networks? *Neural Computation*, 30(1):84–124, 2017. 

I. Rentzeperis, L. Calatroni, L. U. Perrinet, and D. Prandi. Beyond ℓ_1 sparse coding in V1. *PLOS Computational Biology*, 19(9):e1011459, 2023. 

2 systems-theoretic problems in cognitive neuroscience

- stimulus-driven associative memory
- astrocytes modeling
- biologically-plausible learning and control

S. Betteti, G. Baggio, F. Bullo, and S. Zampieri. Stimulus-driven dynamics for robust memory retrieval in Hopfield networks. 2024. Under review

J. J. Moore, A. Genkin, M. Tournoy, J. L. Pughe-Sanford, R. R. de Ruyter van Steveninck, and D. B. Chklovskii. The neuron as a direct data-driven controller. *Proceedings of the National Academy of Sciences*, 121(27), 2024. 

3 connections with machine learning models

- unsupervised representation learning
- self-attention dynamics and transformers
- structured state space sequence models

B. Geshkovski, C. Letrouit, Y. Polyanskiy, and P. Rigollet. A mathematical perspective on transformers. *arXiv preprint arXiv:2312.10794*, 2023. 