Contration Theory for Network Systems

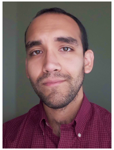
Francesco Bullo

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

Elliit Focus Period Symposium on Network Dynamics and Control
Linköping, Sweden, Sep 20, 2023

Acknowledgments

Veronica Centorrino
Scuola Sup Meridionale



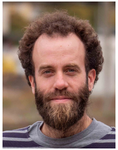
Pedro Cisneros-Velarde
UIUC

Lily Cothren
UC Boulder

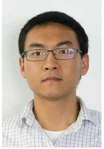
Emiliano Dall'Anese
UC Boulder

Alex Davydov
UC Santa Barbara

Giulia De Pasquale
ETH



Robin Delabays
HES-SO Sion



Xiaoming Duan
Shanghai Jiao Tong

Anand Gokhale
UC Santa Barbara

Saber Jafarpour
GeorgiaTech

Anton Proskurnikov
Politecnico Torino

Giovanni Russo
Univ Salerno

John W. Simpson-Porco
University of Toronto

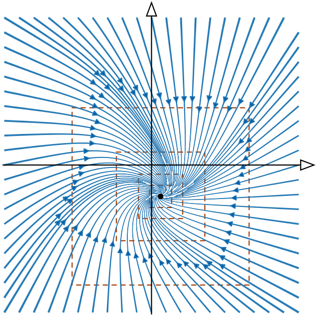
Kevin D. Smith
Utilidata

Elena Valcher
Universita di Padova

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

Contraction theory: historical notes

- **Origins**

  S. Banach. Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales. *Fundamenta Mathematicae*, 3(1):133–181, 1922. doi

- **Dynamics**:

  G. Dahlquist. *Stability and error bounds in the numerical integration of ordinary differential equations*. PhD thesis, (Reprinted in Trans. Royal Inst. of Technology, No. 130, Stockholm, Sweden, 1959), 1958

  S. M. Lozinskii. Error estimate for numerical integration of ordinary differential equations. I. *Izvestiya Vysshikh Uchebnykh Zavedenii. Matematika*, 5:52–90, 1958. URL **http://mi.mathnet.ru/eng/ivm2980**. (in Russian)

- **Computation**:

  C. A. Desoer and H. Haneda. The measure of a matrix as a tool to analyze computer algorithms for circuit analysis. *IEEE Transactions on Circuit Theory*, 19(5):480–486, 1972. doi

- **Systems and control**:

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

- **Incomplete list of scientists who influenced me**
  Aminzare, Arcak, Chung, Coogan, Corless, Di Bernardo, Manchester, Margaliot, Martins, Pavel, Pavlov, Pham, Proskurnikov, Russo, Sepulchre, Slotine, Sontag, ...

- **Surveys**:

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. 2014b.

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*. Springer, 2016.

H. Tsukamoto, S.-J. Chung, and J.-J. E. Slotine. Contraction theory for nonlinear stability analysis and learning-based control: A tutorial overview. *Annual Reviews in Control*, 52:135–169, 2021.

P. Giesl, S. Hafstein, and C. Kawan. Review on contraction analysis and computation of contraction metrics. *Journal of Computational Dynamics*, 10(1):1–47, 2023.

Our work up to 2022

1. **contraction theory on non-Euclidean norms ℓ_1/ℓ_∞**
   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, 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

2. **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. 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

Recent and ongoing work

1. theory: **equilibrium propagation**

   A. Davydov, V. Centorrino, A. Gokhale, G. Russo, and F. Bullo. Contracting dynamics for time-varying convex optimization. IEEE Transactions on Automatic Control, June 2023. doi. Submitted

2. **examples**

   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, 2023b. doi

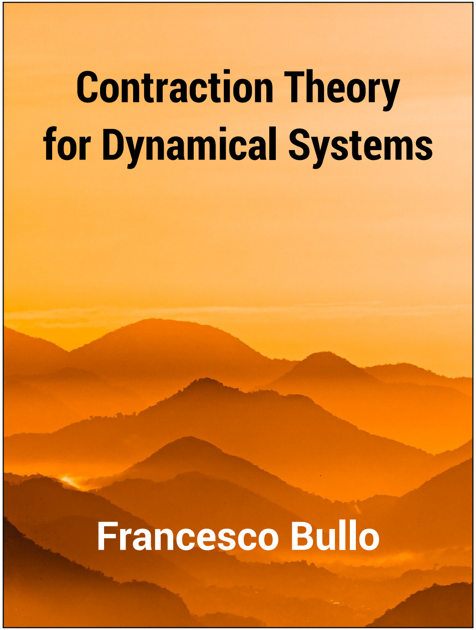
   A. Gokhale, A. Davydov, and F. Bullo. Contractivity of distributed optimization and Nash seeking dynamics. IEEE Control Systems Letters, Sept. 2023. doi. Submitted

   V. Centorrino, A. Gokhale, A. Davydov, G. Russo, and F. Bullo. Contractivity of competitive neural networks for sparse reconstruction. Technical Report, Sept. 2023a

3. **extensions**

   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), 2024. doi. To appear

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



Contraction Theory for Dynamical Systems, Francesco Bullo, KDP, 1.1 edition, 2023, ISBN 979-8836646806

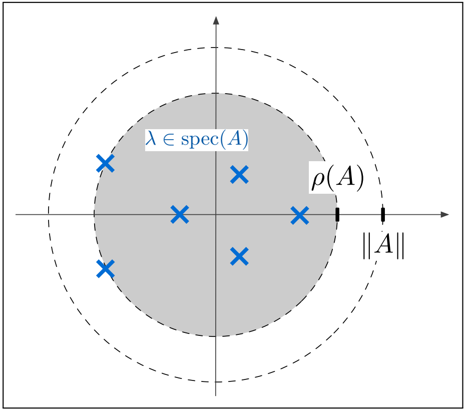
1. Textbook with exercises and answers. Format: textbook, slides, and paperbook
2. Content:
   Fixed point theory
   Theory of contracting dynamics on vector spaces
   Applications to nonlinear and interconnected systems
3. Self-Published and Print-on-Demand at:
   https://www.amazon.com/dp/B0B4K1BTF4
4. PDF Freely available at
   https://fbullo.github.io/ctds
5. 10h minicourse on youtube:
   https://youtu.be/RvR47ZbqJjc
6. Future version to include: systems on Riemannian manifolds, homogeneous spaces, and solid cones

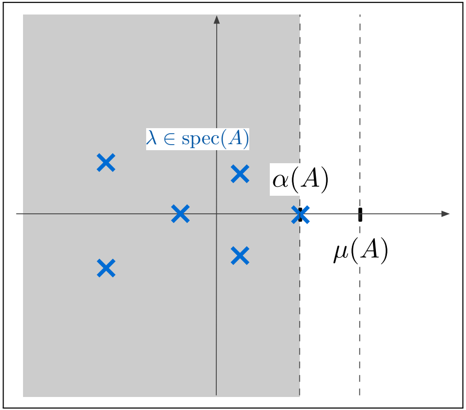
   "Continuous improvement is better than delayed perfection"
   Mark Twain

Outline

Induced matrix norms

Vector norm	Induced matrix norm	Induced matrix log norm
$\Vert x\Vert_1 = \sum_{i=1}^n \vert x_i\vert$	$\Vert A\Vert_1 = \max_{j\in\{1,\dots,n\}} \sum_{i=1}^n \vert a_{ij}\vert$ $=$ max column "absolute sum" of A	$\mu_1(A) = \max_{j\in\{1,\dots,n\}} \left(a_{jj} + \sum_{i=1,i\neq j}^n \vert a_{ij}\vert\right)$ absolute value only off-diagonal
$\Vert x\Vert_2 = \sqrt{\sum_{i=1}^n x_i^2}$	$\Vert A\Vert_2 = \sqrt{\lambda_{\text{max}}(A^\top A)}$	$\mu_2(A) = \lambda_{\text{max}}\left(\frac{A+A^\top}{2}\right)$
$\Vert x\Vert_\infty = \max_{i\in\{1,\dots,n\}} \vert x_i\vert$	$\Vert A\Vert_\infty = \max_{i\in\{1,\dots,n\}} \sum_{j=1}^n \vert a_{ij}\vert$ $=$ max row "absolute sum" of A	$\mu_\infty(A) = \max_{i\in\{1,\dots,n\}} \left(a_{ii} + \sum_{j=1,j\neq i}^n \vert a_{ij}\vert\right)$ absolute value only off-diagonal


$\lambda \in \mathrm{spec}(A)$
$\rho(A)$
$\|A\|$


$\lambda \in \mathrm{spec}(A)$
$\alpha(A)$
$\mu(A)$

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

$$\begin{aligned}\mathsf{osLip}(\mathsf{F}) &= \inf\{b \in \mathbb{R} \text{ such that } [\![\mathsf{F}(x) - \mathsf{F}(y), x - y]\!] \le b\|x - y\|^2 \quad \text{for all } x, y\} \\ &= \sup_x \mu(D\mathsf{F}(x))\end{aligned}$$

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

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

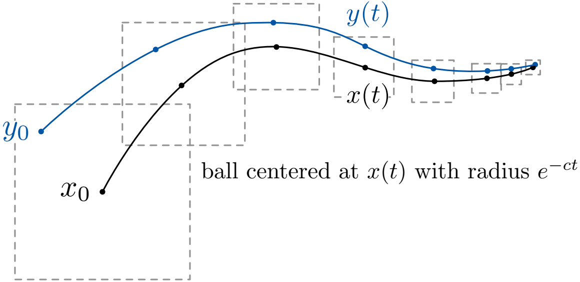
$$\mathsf{osLip}_{2,P}(\mathsf{F}_A) = \mu_{2,P}(A) \le \ell \qquad \iff \qquad A^\top P + AP \preceq 2\ell P$$

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

Banach contraction theorem for continuous-time dynamics:

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

1. F is **infinitesimally contracting** = distance between trajectories decreases exp fast (e^{-ct})
2. F has a unique, glob exp stable equilibrium x^*



For all matrices $A, B \in \mathbb{R}^{n\times n}$, Lipschitz maps $\mathsf{F}, \mathsf{G} : \mathbb{R}^n \to \mathbb{R}^n$ and $a \in \mathbb{R}$

"the modulus properties"

	matrix norms	Lipschitz constants
(positive definiteness)	$\\|A\\| \geq 0$ and $\\|A\\| = 0 \iff A = \mathbb{0}_{n\times n}$	$\mathrm{Lip}(\mathsf{F}) \geq 0$ and $\mathrm{Lip}(\mathsf{F}) = 0 \iff \mathsf{F}$ is constant
(homogeneity)	$\\|aA\\| = \|a\| \\, \\|A\\|$	$\mathrm{Lip}(a\mathsf{F}) = \|a\| \\, \mathrm{Lip}(\mathsf{F})$
(subadditivity)	$\\|A + B\\| \leq \\|A\\| + \\|B\\|$	$\mathrm{Lip}(\mathsf{F} + \mathsf{G}) \leq \mathrm{Lip}(\mathsf{F}) + \mathrm{Lip}(\mathsf{G})$
(sub-multiplicativity)	$\\|AB\\| \leq \\|A\\| \\|B\\|$	$\mathrm{Lip}(\mathsf{F} \circ \mathsf{G}) \leq \mathrm{Lip}(\mathsf{F}) \\, \mathrm{Lip}(\mathsf{G})$

"the real part properties"

	matrix log norms	one-sided Lipschitz constants
(positive homogeneity)	$\mu(aA) = \|a\| \\, \mu(\mathrm{sign}(a)A)$	$\mathrm{osLip}(a\mathsf{F}) = \|a\| \\, \mathrm{osLip}(\mathrm{sign}(a)\mathsf{F})$
(subadditivity)	$\mu(A + B) \leq \mu(A) + \mu(B)$	$\mathrm{osLip}(\mathsf{F} + \mathsf{G}) \leq \mathrm{osLip}(\mathsf{F}) + \mathrm{osLip}(\mathsf{G})$
(translation property)	$\mu(A + aI_n) = \mu(A) + a$	$\mathrm{osLip}(\mathsf{F} + a \\, \mathsf{Id}) = \mathrm{osLip}(\mathsf{F}) + a$
(uniform monotonicity)	$\mu(A) < 0$ $\implies A$ invertible, $\\|A^{-1}\\| \leq -1/\mu(A)$	$\mathrm{osLip}(\mathsf{F}) < 0$ $\implies \mathsf{F}$ injective, $\mathrm{Lip}(\mathsf{F}^{-1}) \leq -1/\mathrm{osLip}(\mathsf{F})$

F. Bullo. *Contraction Theory for Dynamical Systems*. Kindle Direct Publishing, 1.1 edition, 2023. ISBN 979-8836646806. URL **https://fbullo.github.io/ctds**

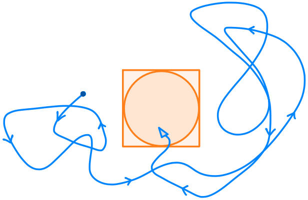
Advantages of non-Euclidean approaches

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, 2014a)

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

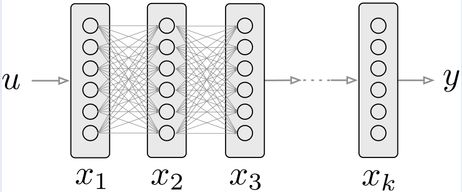
for large n, quadratic Lyap fnctns may provide exponentially conservative estimates

Courtesy of Anton Proskurnikov, Politecnico di Torino

Outline

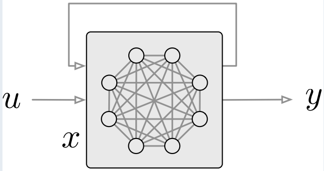
Outline

Feedforward NN



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

Recurrent NN



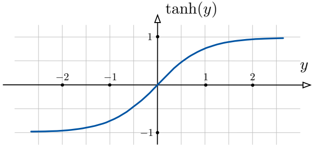
$$\dot{x} = -x + \Phi(Ax + Bu + b),$$
$$y = Cx + d$$

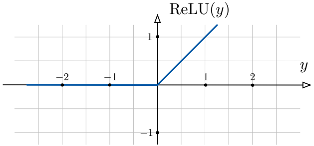
A. Davydov, A. V. Proskurnikov, and F. Bullo. Non-Euclidean contractivity of recurrent neural networks. In *American Control Conference*, pages 1527–1534, Atlanta, USA, May 2022b.

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, 2023b.

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

sigmoid, hyperbolic tangent
ReLU $= \max\{x, 0\} = (x)_+$
$0 \leq \Phi_i'(y) \leq 1$





F_{FR} is infinitesimally contracting wrt $\|\cdot\|_\infty$ with rate $1 - \mu_\infty(W)_+$ if

$\mu_\infty(W) < 1$ $\quad$ (i.e., $w_{ii} + \sum_j |w_{ij}| < 1$ for all i)

Note: clear **graphical interpretation** + **generalization to interconnection theorem**

Example #2: Firing-rate network with symmetric synapses

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

$$0 \leq \Phi_i'(y) \leq 1 \qquad \text{and} \qquad W = W^\top \text{ with } \lambda_W = \lambda_{\mathsf{max}}(W)$$

F_{FR} is infinitesimally contracting:

(for $\lambda_W < 0$)	**with rate** 1 **wrt** $\|\cdot\|_{2,(-W)^{1/2}}$
(for $\lambda_W = 0$)	**with rate** $1-\epsilon$ **wrt** $\|\cdot\|_{2,Q_{\mathsf{FR},\epsilon}}$, **for each** $\epsilon > 0$
(for $0 < \lambda_W < 1$)	**with rate** $1-\lambda_W$ **wrt** $\|\cdot\|_{2,Q_{\mathsf{FR},\lambda_W}}$

For $\lambda_W = 1$, F_{FR} **is weakly infinitesimally contracting wrt** $\|\cdot\|_{2,Q_{\mathsf{FR},\lambda_W}}$

Note: when $W = W^\top$, **sharper result**, but no graph interpretation and hard to generalize

Outline

Example #3: 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

Euler discretization theorem

Euler discretization theorem for contracting dynamics

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

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 #4: Primal-dual gradient dynamics

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

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

linearly constrained optimization:

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

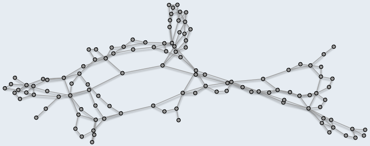
primal-dual gradient dynamics:

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

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

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

Distributed optimization



undirected, weighted and connected graph with n nodes and m edges
Laplacian $L \in \mathbb{R}^{n\times n}$, λ_2 = algebraic connectivity, λ_2/λ_n = synchronizability
oriented incidence matrix $B \in \mathbb{R}^{n\times m}$

Distributed optimization setup

cost function f is decomposable into sum of private cost function

$$f(x) = \sum_{i=1}^{n} f_i(x) \qquad \text{where each } f_i \text{ is private to node } i$$

each node i has a local estimate $x_{[i]}$ of global variable x and $\mathbf{x} = [x_{[1]}, \dots, x_{[n]}]$

Example #5: Incidence-based distributed gradient

Assume graph is a tree

$$0 \prec \lambda_2 I_{n-1} \preceq B^\top B \preceq \lambda_n I_{n-1}$$

decomposable cost: $\min_{x \in \mathbb{R}} \sum_i f_i(x)$ where each f_i is ν_i-strongly convex

$$\begin{cases} \min_{x_{[i]} \in \mathbb{R}} & \sum_{i=1}^n f_i(x_{[i]}) \\ \text{subj. to} & x_{[i]} - x_{[j]} = 0 \quad \text{for each edge } e = (i,j) \end{cases}$$

incidence-based distributed gradient (primal-dual gradient, $n+m$ vars):

$$\begin{cases} \dot{x}_{[i]} = -\nabla f_i(x_{[i]}) - \sum_{e=(i,j)} \lambda_e + \sum_{e=(j,i)} \lambda_e & \text{for each node } i \\ \dot{\lambda}_e = x_{[i]} - x_{[j]} & \text{for each edge } e = (i,j) \end{cases}$$

$\mathsf{F}_{\textbf{Incidence-DistributedG}}$ **is infinitesimally contracting with** $c = \dfrac{1}{4} \dfrac{\lambda_2}{\lambda_n} \min_i \nu_i$

Example #6: Laplacian-based distributed gradient

Given $\Pi_n = I_n - \mathbb{1}_n \mathbb{1}_n^\top / n =$ orthogonal projection onto $\mathrm{span}\{\mathbb{1}_n\}^\perp$,

$$0 \prec \lambda_2 \Pi_n \preceq L \preceq \lambda_n I_n$$

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

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

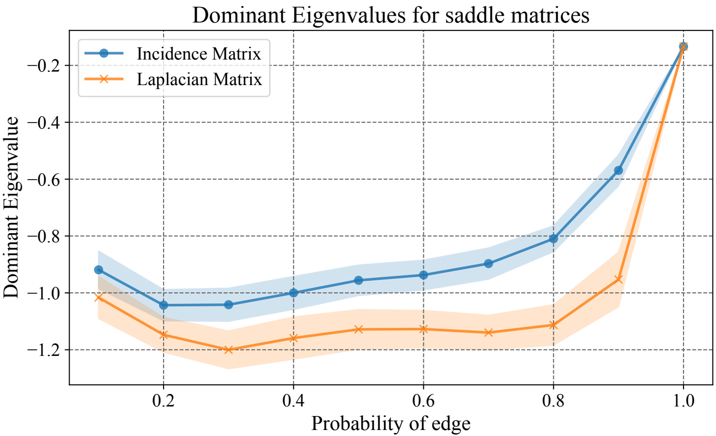
Laplacian-based distributed gradient (primal-dual gradient, $2n$ vars):

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

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

$\lambda_2/\lambda_n =$ **synchronizability** parameter from study of oscillator networks via the MSF approach

private functions $q_i(x_i - v_i)^2$, for $x_i \in \mathbb{R}$, v_i and q_i uniformly sampled from $[0, 10]$
symmetric connected Erdős-Rényi graph with 40 nodes, 50 graphs for each probability value



L. M. Pecora and T. L. Carroll. Synchronization in chaotic systems. *Physical Review Letters*, 64(8):821–824, 1990

G. Chen. Searching for best network topologies with optimal synchronizability: A brief review. *IEEE/CAA Journal of Automatica Sinica*, 9 (4):573–577, 2022.

Composite optimization

composite minimization (cost = sum of terms with structurally different properties):

$$x^\star = \operatorname*{argmin}_{x \in \mathbb{R}^n} f(x) + g(x)$$

$f : \mathbb{R}^n \to \mathbb{R}$ is strongly convex and strongly smooth
$g : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex, closed, and proper (ccp)

proximal operator: for $\gamma > 0$, define $\mathrm{prox}_{\gamma g} : \mathbb{R}^n \to \mathbb{R}^n$ by

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

Equivalence:

1. x^* is minimizer for: $\min_{x \in \mathbb{R}^n} f(x) + g(x)$
2. x^* is fixed point for: $x = \mathrm{prox}_{\gamma g}(x - \gamma \nabla f(x))$ for all γ

proximal gradient dynamics: $\dot{x} = \mathsf{F}_{\mathsf{ProxG}}(x) := -x + \mathrm{prox}_{\gamma g}(x - \gamma \nabla f(x))$

Examples

constraint and projections:

$$g(x) = \begin{cases} 0, & \text{if } x \in C \\ +\infty, & \text{if } x \notin C \end{cases}$$

$$\text{prox}_g(x) = \Pi_C(x)$$

e.g.: saturation for box constraints

separable cost and diagonal functions:

$$g(x) = \sum_i g_i(x_i)$$

$$\big(\text{prox}_g(x)\big)_i = \text{prox}_{g_i}(x_i)$$

proximal operator

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

gradient algorithms/dynamics for proximal algorithms/dynamics – nonsmooth, constrained, large-scale, and distributed optimization

evaluation of proximal operator requires small convex optimization,

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

Example #7: Proximal gradient dynamics

proximal gradient dynamics:

$$\dot{x} = \mathsf{F}_{\mathsf{ProxG}}(x) := -x + \mathrm{prox}_{\gamma g}(x - \gamma \nabla f(x))$$

1. **$\mathsf{F}_{\mathsf{ProxG}}$ is infinitesimally contracting wrt $\|\cdot\|_2$**

   for $0 < \gamma < \frac{2}{\ell}$, with rate $c = 1 - \max\{|1-\gamma\nu|, |1-\gamma\ell|\}$,

   for $\gamma^\star = \frac{2}{\nu+\ell}$, with maximal rate $c^\star = \frac{2\nu}{\nu+\ell}$

2. **$\mathsf{F}_{\mathsf{ProxG}}$ is infinitesimally contracting wrt $\|\cdot\|_{2,(\gamma A - I_n)^{1/2}}$ with rate $c = 1$**

   if $f(x) = \frac{1}{2}x^\top A x + b^\top x$ with $A \succ 0$ and $\gamma > 1/\lambda_{\min}(A)$

Outline

Gradient dynamics and Nash equilibria in games

- Nash equilibria: existence, uniqueness, computation, convergence for gradient-like dynamics, robustness
- games with partial information
- aggregative games: demand-side management in the smart grid, charging control for plug-in electric vehicles, spectrum sharing in wireless networks, and network congestion control

S. Li and T. Başar. Distributed algorithms for the computation of noncooperative equilibria. *Automatica*, 23(4):523–533, 1987. doi

M. Arcak and N. C. Martins. Dissipativity tools for convergence to Nash equilibria in population games. *IEEE Transactions on Control of Network Systems*, 8(1):39–50, 2021. doi

L. Pavel. Dissipativity theory in game theory: On the role of dissipativity and passivity in Nash equilibrium seeking. *IEEE Control Systems*, 42(3):150–164, June 2022. doi

G. Belgioioso, P. Yi, S. Grammatico, and L. Pavel. Distributed generalized Nash equilibrium seeking: An operator-theoretic perspective. *IEEE Control Systems*, 42(4):87–102, 2022. doi

A. Gokhale, A. Davydov, and F. Bullo. Contractivity of distributed optimization and Nash seeking dynamics. *IEEE Control Systems Letters*, Sept. 2023. doi. Submitted

Example #8: 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):

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

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 #9: Pseudogradient 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):

$$\dot{x} = \mathsf{F}_{\mathsf{PseudoG}}(x) = -\big(\nabla_1 J_1(x_1, x_{-1}), \dots, \nabla_n J_n(x_n, x_{-n})\big) \qquad \text{(stacked vector)}$$

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

- **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

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

if $\mathsf{F}_{\mathsf{PseudoG}}$ is infinitesimally contracting (wrt any norm)

then **unique globally exp stable Nash equilibrium** $J_i(x_i^*, x_{-i}^*) \leq J_i(y_i, x_{-i}^*)$ for all y_i

Example #10: 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** $\|\cdot\|_2$

if F_{BR} is infinitesimally contracting (wrt any norm)

then **unique globally exp stable Nash equilibrium** (fixed point of BR)

Equivalent statements:

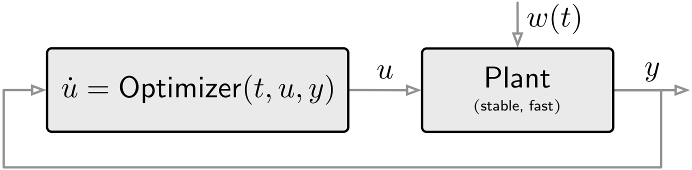
1. $\mathsf{F}_{\mathsf{PseudoG}}$ gain matrix: $\begin{bmatrix} -\mu_1 & \dots & \ell_{1n} \\ \vdots & & \vdots \\ \ell_{n1} & \dots & -\mu_n \end{bmatrix}$ is Hurwitz
2. F_{BR} gain matrix: $\begin{bmatrix} -1 & \dots & \ell_{1n}/\mu_1 \\ \vdots & & \vdots \\ \ell_{n1}/\mu_n & \dots & -1 \end{bmatrix}$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

Outline

Solving optimization problems via dynamical systems



- studies in linear and nonlinear programming (Arrow, Hurwicz, and Uzawa 1958)
- neural networks (Hopfield and Tank 1985) and analog circuits (Kennedy and Chua 1988)
- optimization on manifolds (Brockett 1991)
- ...
- online and dynamic feedback optimization (Dall'Anese, Dörfler, Simonetto, ...)

A. Davydov, V. Centorrino, A. Gokhale, G. Russo, and F. Bullo. Contracting dynamics for time-varying convex optimization. *IEEE Transactions on Automatic Control*, June 2023. doi. Submitted

L. Cothren, F. Bullo, and E. Dall'Anese. Singular perturbation via contraction theory. *Technical Report*, Sept. 2023

From convex optimization to contracting dynamics – time-varying

Many convex optimization problems can be solved with contracting dynamics

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

	Convex Optimization	**Contracting Dynamics**
Unconstrained	$\min_{x \in \mathbb{R}^n} \quad f(x, \theta)$	$\dot{x} = -\nabla_x f(x, \theta)$
Constrained	$\min_{x \in \mathbb{R}^n} \quad f(x, \theta)$ s.t. $\quad x \in \mathcal{X}(\theta)$	$\dot{x} = -x + \mathrm{Proj}_{\mathcal{X}(\theta)}(x - \gamma \nabla_x f(x, \theta))$
Composite	$\min_{x \in \mathbb{R}^n} \quad f(x, \theta) + g(x, \theta)$	$\dot{x} = -x + \mathrm{prox}_{\gamma g_\theta}(x - \gamma \nabla_x f(x, \theta))$
Equality	$\min_{x \in \mathbb{R}^n} \quad f(x, \theta)$ s.t. $\quad Ax = b(\theta)$	$\dot{x} = -\nabla_x f(x, \theta) - A^\top \lambda,$ $\dot{\lambda} = Ax - b(\theta)$
Inequality	$\min_{x \in \mathbb{R}^n} \quad f(x, \theta)$ s.t. $\quad Ax \leq b(\theta)$	$\dot{x} = -\nabla f(x, \theta) - A^\top \nabla M_{\gamma, b(\theta)}(Ax + \gamma \lambda),$ $\dot{\lambda} = \gamma(-\lambda + \nabla M_{\gamma, b(\theta)}(Ax + \gamma \lambda))$

For parameter-dependent vector field $\mathsf{F} : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^n$ and differentiable $\theta : \mathbb{R}_{\geq 0} \to \Theta \subset \mathbb{R}^d$

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

Assume there exist norms $\|\cdot\|_{\mathcal{X}}$ and $\|\cdot\|_{\Theta}$ s.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

Theorem: Incremental ISS any two soltns: $x(t)$ with input θ_x and $y(t)$ with input θ_y

$$D^+\|x(t) - y(t)\|_{\mathcal{X}} \;\leq\; -c\|x(t) - y(t)\|_{\mathcal{X}} \;+\; \ell\|\theta_x(t) - \theta_y(t)\|_{\Theta}$$

Equilibrium tracking

For parameter-dependent vector field $\mathsf{F} : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^n$ and differentiable $\theta : \mathbb{R}_{\geq 0} \to \Theta \subset \mathbb{R}^d$

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

Assume there exist norms $\|\cdot\|_{\mathcal{X}}$ and $\|\cdot\|_{\Theta}$ s.t.

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

Theorem: Equilibrium tracking for contracting dynamics

1. for each fixed θ, there exists a unique equilbrium $x^\star(\theta)$
2. the equilibrium map $x^\star(\cdot)$ is Lipschitz with constant $\dfrac{\ell}{c}$
3. $D^+\|x(t) - x^\star(\theta(t))\|_{\mathcal{X}} \;\leq\; -c\|x(t) - x^\star(\theta(t))\|_{\mathcal{X}} \;+\; \dfrac{\ell}{c}\,\|\dot{\theta}(t)\|_{\Theta}$

Consequences for tracking error

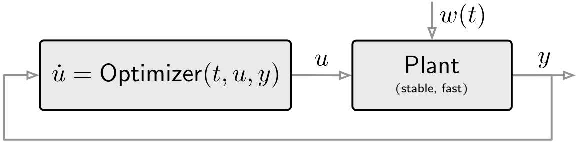
$$D^+\|x(t)-x^\star(\theta(t))\|_\mathcal{X} \;\leq\; -c\|x(t)-x^\star(\theta(t))\|_\mathcal{X} \;+\; \frac{\ell}{c}\,\|\dot\theta(t)\|_\Theta$$

- bounded input, bounded error
  with asymptotic bound:

$$\limsup_{t\to\infty}\|x(t)-x^\star(\theta(t))\|_\mathcal{X} \quad\leq\quad \frac{\ell}{c^2}\,\limsup_{t\to\infty}\|\dot\theta(t)\|_\Theta$$

- bounded energy input, bounded energy error
- vanishing input, vanishing error
- exponentially vanishing input $\sim \mathrm{e}^{-ht}$, exponentially vanishing error $\sim \mathrm{e}^{-\min\{c,h\}t}$
- periodic input, periodic error

Application: Dynamic feedback optimization



dynamic feedback optimization

online optimization, optimization-based feedback, input/output regulation ...

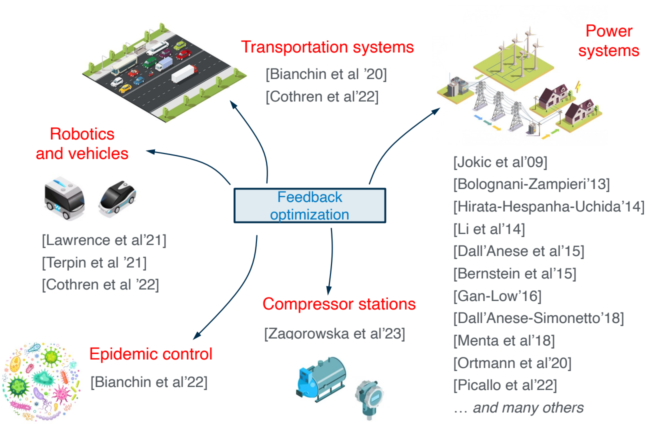
$$\begin{cases} \min & \text{cost}_1(u) + \text{cost}_2(y) \\ \text{subj. to} & y = \text{Plant}\big(u, w(t)\big) \end{cases} \implies \begin{cases} \dot{u} = \text{Optimizer}(t,u,y) \\ y = \text{Plant}\big(u, w(t)\big) \end{cases}$$

A. Jokic, M. Lazar, and P. van den Bosch. On constrained steady-state regulation: Dynamic KKT controllers. *IEEE Transactions on Automatic Control*, 54(9):2250–2254, 2009.

A. Hauswirth, S. Bolognani, G. Hug, and F. Dorfler. Timescale separation in autonomous optimization. *IEEE Transactions on Automatic Control*, 66(2):611–624, 2021.

G. Bianchin, J. Cortés, J. I. Poveda, and E. Dall'Anese. Time-varying optimization of LTI systems via projected primal-dual gradient flows. *IEEE Transactions on Control of Network Systems*, 9(1):474–486, 2022.

Some works on feedback optimization



Slide courtesy of Emiliano Dall'Anese, University of Colorado Boulder

Example #11: Gradient controller

Fast/stable LTI plant with control input u and state/measurement disturbance $w(t)$:

$$\begin{aligned} \epsilon \dot{x} &= Ax + Bu + Ew(t) \qquad\qquad A \text{ Hurwitz} \\ y &= Cx + Dw(t) \end{aligned}$$

In singular perturbation limit as $\epsilon \to 0^+$, steady state map (Y_u and Y_w)

$$y = \underbrace{-CA^{-1}B}_{=:\ Y_u} u + \underbrace{(D - CA^{-1}E)}_{=:\ Y_w} w$$

Feedback optimization

equilibrium trajectory $u^*(t)$ is solution to

$$\begin{aligned} \min_u \quad & \phi(u) + \psi(y(t)) \\ \text{subj to} \quad & y(t) = Y_u u + Y_w w(t) \end{aligned}$$

(ν-strongly convex ϕ, convex ψ)

Example #11: Gradient controller

In singular perturbation limit as $\epsilon \to 0^+$,

$$\mathcal{E}(u,w) = \phi(u) + \psi(Y_u u + Y_w w), \qquad (\nu\text{-strongly convex in } u)$$

$$\begin{aligned}\nabla_u \mathcal{E}(u,w) &= \nabla\phi(u) + Y_u^\top \nabla\psi(Y_u u + Y_w w) \\ &= \nabla\phi(u) + Y_u^\top \nabla\psi(y) \qquad (\text{no need to measure } w(t) \text{ to compute } \dot{u}(t))\end{aligned}$$

Hence, **gradient controller** is equivalently defined by

$$\dot{u} = \mathsf{F}_{\mathsf{GradCtrl}}(u,w) \;:=\; -\nabla\mathcal{E}_u(u,w) = -\nabla\phi(u) - Y_u^\top \nabla\psi(Y_u u + Y_w w)$$

Equilibrium tracking for the gradient controller

1. $\mathsf{osLip}_u(\mathsf{F}_{\mathsf{GradCtrl}}) \leq -\nu$ (gradient of ν-strongly convex function)
2. $\mathsf{Lip}_w(\mathsf{F}_{\mathsf{GradCtrl}}) = \ell_w := \|Y_u^\top\| \,\mathsf{Lip}(\nabla\psi) \|Y_w\|$

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

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 (example of proximal gradient dynamics):

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

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

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

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

Outline

Conclusions

contractivity = robust computationally-friendly stability

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

Ongoing work

1. equilibrium tracking with noise
   applications to optimization-based control
2. non-expansive dynamics for weakly convex optimization
   sparse reconstruction in biologically plausible neural networks
   coupled neural-synaptic dynamics for representation learning
3. polyhedral norms
4. singular perturbation for feedback optimization, bilevel optimization, Stackelberg games
5. primal-dual dynamics for inequality constraints
6. semicontractivity for population games