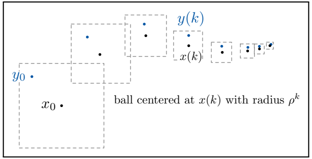
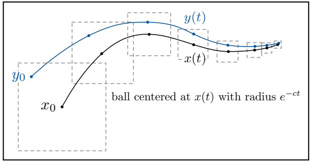
Contracting Dynamical Systems: A Tutorial on Theory and Applications

Francesco Bullo

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

Minicourse, Focus Period "Network Dynamics and Control," University of Linköping, Sweden, 2023/9/13-15
Tutorial (based on lectures @ SSM in Napoli Nov '22, ACC @ San Diego Jun '23). This version: 2026/01/28

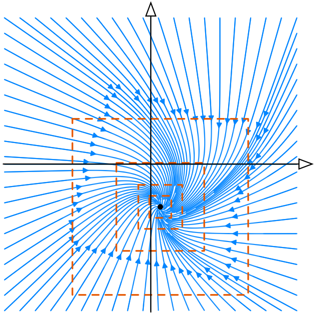




contractivity = robust computationally-friendly stability

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

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

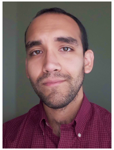


1. unique globally exponential stable equilibrium
   & two natural Lyapunov functions
2. robustness properties
   - bounded input, bounded output (iss)
   - finite input-state gain
   - robustness margin wrt unmodeled dynamics
   - robustness margin wrt delayed dynamics
3. periodic input, periodic output
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

Acknowledgments

Veronica Centorrino
Scuola Sup Meridionale



Pedro Cisneros-Velarde
VMware

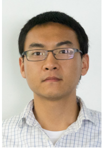
Lily Cothren
UC Boulder

Emiliano Dall'Anese
Boston University

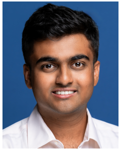
Alex Davydov
UC Santa Barbara

Giulia De Pasquale
ETH

Robin Delabays
HES-SO Sion



Xiaoming Duan
Shanghai Jiao Tong



Anand Gokhale
UC Santa Barbara

Saber Jafarpour
UC Boulder

Anton Proskurnikov
Politecnico Torino

Giovanni Russo
Univ Salerno

John W. Simpson-Porco
University of Toronto

Kevin D. Smith
Utilidata

Elena Valcher
Universita di Padova

With support from AFOSR project FA9550-22-1-0059.

Outline

§1. History and resources

§2. Basic definitions: discrete and continuous-time dynamics on vector spaces
- The linear algebra of matrix norms; see CTDS Chapter 2
- Properties of induced matrix norms and Lipschitz constants

§3. Example systems
- Constrained, distributed and proximal gradient dynamics
- Continuous-time recurrent neural networks
- Nonlinear dynamics in Lur'e form

§4. Properties of contracting dynamics
- Equilibria, Lyapunov functions, and Euler discretization
- Incremental input-to-state stability
- Contractivity of interconnected systems
- Additional properties: entrainment, robustness wrt unmodeled dynamics and delays

§5. Example applications
- Gradient dynamics and Nash equilibria in games
- Time-varying gradient dynamics and feedback optimization
- Recurrent and implicit neural networks

§6. Generalizations with examples
- G1: Local contractivity: Small-residual theorem and the Kuramoto coupled oscillators
- G2: Weak contractivity: Biologically-plausible circuits for sparse reconstruction
- G3: Contractivity on Riemannian manifolds and the Karcher mean
- G4: Semicontractivity: Primal-dual gradient with redundant constraints

§7. Conclusions and future research

§8. Advanced Topics
- More on semicontractivity: ergodic coefficients and duality
- Network small-gain theorem for Metzler matrices
- Proof of semicontractivity of saddle matrices
- Proof of Euler discretization theorem
- Non-Euclidean Monotone Operator Theory

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, Andrieu, Arcak, Astolfi, Chung, Coogan, Corless, Dall'Anese, Di Bernardo, Giesl, Kawano, Manchester, Margaliot, Martins, Ngoc, Pavel, Pavlov, Praly, Pham, Proskurnikov, Russo, Sepulchre, Slotine, Sontag, Tarbouriech, ...

- **Surveys and Perspectives**:

  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. doi

  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. doi

  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. doi

  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. doi

  A. Davydov and F. Bullo. Perspectives on contractivity in control, optimization and learning. *IEEE Control Systems Letters*, 8:2087–2098, 2024. doi

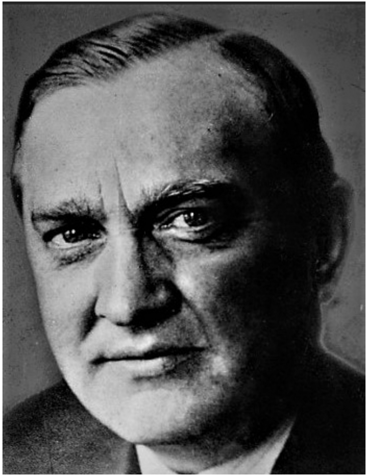


Figure: Stefan Banach (Krakow, 30 Mar 1892 – Lviv, 31 Aug 1945) was a self-taught Polish mathematician

1920: doctoral thesis on Banach spaces @ University of Lviv
1920-1922: Assistant Professor @ Lwow Polytechnic
1922: Full Professor @ Lwow Polytechnic
1924: Member of the Polish Academy of Arts and Sciences
1929: Founder, Lvov School of Mathematics
1931: first functional analysis: "Theory of Linear Operations"
1939-45: dark years

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

The Banach Contraction Theorem is also referred to as the *Picard-Banach-Caccioppoli*, because of the earlier work by Picard (1890) on the "method of successive approximations" and the later independent work by Renato Caccioppoli (1930).

Figure: Renato Caccioppoli (Napoli, 20 Jan 1904 – Napoli, 8 May 1959) was an Italian mathematician

1921-1932 student and researcher @ Napoli
1931-1934 professor @ Padova
1934-1959 professor @ Napoli

R. Caccioppoli. Un teorema generale sull'esistenza di elementi uniti in una trasformazione funzionale. *Rendiconti dell'Accademia Nazionale dei Lincei*, 11:794–799, 1930

Contraction conditions without Jacobians

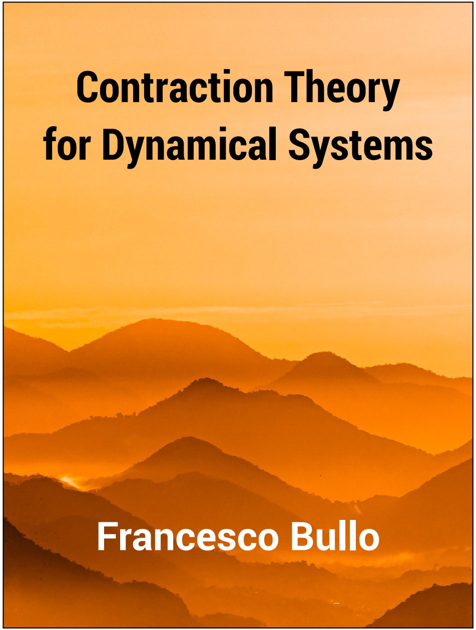
1. **one-sided Lipschitz maps** in: G. Dahlquist. Error analysis for a class of methods for stiff non-linear initial value problems. In G. A. Watson, editor, *Numerical Analysis*, pages 60–72. Springer, 1976. and E. Hairer, S. P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I. Nonstiff Problems*. Springer, 1993. (Section 1.10, Exercise 6)
2. **uniformly decreasing maps** in: L. Chua and D. Green. A qualitative analysis of the behavior of dynamic nonlinear networks: Stability of autonomous networks. *IEEE Transactions on Circuits and Systems*, 23(6): 355–379, 1976.
3. no-name in: A. F. Filippov. *Differential Equations with Discontinuous Righthand Sides*. Kluwer, 1988. ISBN 902772699X (Chapter 1, page 5)
4. **maps with negative nonlinear measure** in: H. Qiao, J. Peng, and Z.-B. Xu. Nonlinear measures: A new approach to exponential stability analysis for Hopfield-type neural networks. *IEEE Transactions on Neural Networks*, 12(2):360–370, 2001.
5. **dissipative Lipschitz maps** in: T. Caraballo and P. E. Kloeden. The persistence of synchronization under environmental noise. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 461(2059):2257–2267, 2005.
6. **maps with negative lub log Lipschitz constant** in: G. Söderlind. The logarithmic norm. History and modern theory. *BIT Numerical Mathematics*, 46(3):631–652, 2006.
7. **QUAD maps** in: W. Lu and T. Chen. New approach to synchronization analysis of linearly coupled ordinary differential systems. *Physica D: Nonlinear Phenomena*, 213(2):214–230, 2006.
8. **incremental quadratically stable maps** in: L. D'Alto and M. Corless. Incremental quadratic stability. *Numerical Algebra, Control and Optimization*, 3:175–201, 2013.

Contraction conditions with Jacobians

1. **Demidovich LMI condition** in: B. P. Demidovič. On the dissipativity of a certain non-linear system of differential equations. I. *Vestnik Moskovskogo Universiteta. Serija I. Matematika, Mehanika*, 6:19–27, 1961
2. Krasovskiĭ's method for Lyapunov functions
3. common Lyapunov function approach
4. Pointwise quadratic constraints
5. Incremental multiplier matrices
6. Lyapunov functions for the variational system

Links to recent related educational and research events

- 2023 ACC Workshop on "Contraction Theory for Systems, Control, and Learning" **http://motion.me.ucsb.edu/contraction-workshop-2023**
- 2024 CDC Workshop on "Contraction Theory for Systems, Control, Optimization, and Learning" **http://motion.me.ucsb.edu/contraction-workshop-2024**
- Tutorial session: **https://sites.google.com/view/contractiontheory** "Contraction Theory for Machine Learning" (PDFs and youtube videos) at the 2021 IEEE CDC conference, by Soon-Jo Chung, Jean-Jacques Slotine, and Hiroyasu Tsukamoto
- Tutorial paper at CDC2021 "Contraction-Based Methods for Stable Identification and Robust Machine Learning: a Tutorial" by Ian Manchester and coauthors: **https://arxiv.org/abs/2110.00207**, **https://ieeexplore.ieee.org/abstract/document/9683128**
- Plenary presentation: (Slides **https://fbullo.github.io/talks/2022-12-FBullo-ContractionSystemsControl-CDC.pdf**) "Contraction Theory in Systems and Control" by Francesco Bullo at the 2022 IEEE CDC
- Youtube lectures: "Lectures on Nonlinear Systems" by Jean-Jacques Slotine, Fall 2013: **https://web.mit.edu/nsl/www/videos/lectures.html**, Lectures 14-20 (approximately 1h20min each)
- Youtube lectures: "Minicourse on Contraction Theory" by Francesco Bullo, Fall 2022. Youtube lectures **https://youtu.be/RvR47ZbqJjc**: 10h in 4 lectures, with chapters
- Textbook: Contraction Theory for Dynamical Systems, Francesco Bullo, rev 1.1, Mar 2023. (Book and slides freely available) **https://fbullo.github.io/ctds**



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

- **Table of Contents:**

  Examples: neural networks, gradient dynamics, Lur’e systems, traffic networks, diffusively-coupled dynamical systems, and more

- PDF text and slides freely available at **https://fbullo.github.io/ctds**
- paperback and hardcover at: **(link to amazon)**
- 12h recorded minicourse at: **(link to youtube)**

"Continuous improvement is better than delayed perfection"
Mark Twain

Selected references from my group

Contraction theory on normed spaces and Riemannian manifolds:

- 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.
- 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.
- J. W. Simpson-Porco and F. Bullo. Contraction theory on Riemannian manifolds. *Systems & Control Letters*, 65:74–80, 2014.

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.
- 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 2022c.
- 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.

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, 2022a.
- 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 and F. Bullo. Semicontraction and synchronization of Kuramoto-Sakaguchi oscillator networks. *IEEE Control Systems Letters*, 7:1566–1571, 2023.

Optimization:

- F. Bullo, P. Cisneros-Velarde, A. Davydov, and S. Jafarpour. From contraction theory to fixed point algorithms on Riemannian and non-Euclidean spaces. In *IEEE Conf. on Decision and Control*, Dec. 2021.
- 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. 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*, 70(11):7446–7460, 2025.

Outline

§1. History and resources

§2. Basic definitions: discrete and continuous-time dynamics on vector spaces
- The linear algebra of matrix norms; see CTDS Chapter 2
- Properties of induced matrix norms and Lipschitz constants

§3. Example systems
- Constrained, distributed and proximal gradient dynamics
- Continuous-time recurrent neural networks
- Nonlinear dynamics in Lur'e form

§4. Properties of contracting dynamics
- Equilibria, Lyapunov functions, and Euler discretization
- Incremental input-to-state stability
- Contractivity of interconnected systems
- Additional properties: entrainment, robustness wrt unmodeled dynamics and delays

§5. Example applications
- Gradient dynamics and Nash equilibria in games
- Time-varying gradient dynamics and feedback optimization
- Recurrent and implicit neural networks

§6. Generalizations with examples
- G1: Local contractivity: Small-residual theorem and the Kuramoto coupled oscillators
- G2: Weak contractivity: Biologically-plausible circuits for sparse reconstruction
- G3: Contractivity on Riemannian manifolds and the Karcher mean
- G4: Semicontractivity: Primal-dual gradient with redundant constraints

§7. Conclusions and future research

§8. Advanced Topics
- More on semicontractivity: ergodic coefficients and duality
- Network small-gain theorem for Metzler matrices
- Proof of semicontractivity of saddle matrices
- Proof of Euler discretization theorem
- Non-Euclidean Monotone Operator Theory

For a non-empty set $\mathcal{X}$, a map $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a *metric* (or a *distance*) on $\mathcal{X}$ if

(separation):	$d(x, y) = 0$ if and only if $x = y$
(symmetry):	$d(x, y) = d(y, x)$ for all $x, y \in \mathcal{X}$
(triangle inequality):	$d(x, y) \leq d(x, z) + d(z, y)$ for all $x, y, z \in \mathcal{X}$

A map $T : \mathcal{X} \to \mathcal{X}$ is

1. *Lipschitz* if there exists $\ell \geq 0$, called *a Lipschitz constant* of T, such that

$$d(T(x), T(y)) \leq \ell \, d(x, y) \quad \text{for all } x, y \in \mathcal{X},$$

2. a *contraction* if it is Lipschitz with constant $\ell < 1$.

Banach Contraction Theorem Let $(\mathcal{X}, d)$ be a *complete metric space*

If $T : \mathcal{X} \to \mathcal{X}$ is Lipschitz with constant $\ell < 1$ (called the *contraction factor*), then

1. T has a unique fixed point x^* in $\mathcal{X}$
2. the sequence $\{x_k\}_{k\in\mathbb{N}}$ generated by the *Picard iteration* $x_{k+1} = T(x_k)$ converges to x^* for all initial conditions $x_0 \in \mathcal{X}$
3. the following error estimates hold for all $k \in \mathbb{N}$:

(geometric convergence): $$d(x_k, x^*) \le \ell^k d(x_0, x^*)$$

(a-priori upper bound): $$d(x_k, x^*) \le \frac{\ell^k}{1-\ell} d(x_0, x_1)$$

(a-posteriori upper bound): $$d(x_k, x^*) \le \frac{\ell}{1-\ell} d(x_{k-1}, x_k)$$

Proof of Banach Contraction Theorem

For $x_{k+1} = T(x_k)$, note $d(x_{k+1}, x_k) \le \ell d(x_k, x_{k-1})$.

- we show the sequence $\{x_k\}_{k\in\mathbb{N}}$ is Cauchy. For all k and h,

$$\begin{aligned} d(x_{k+h}, x_k) &\le d(x_{k+h}, x_{k+h-1}) + \cdots + d(x_{k+1}, x_k) && \text{(triangle inequality)} \\ &\le (\ell^{h-1} + \cdots + 1) d(x_{k+1}, x_k) && \text{(Lipschitzness)} \\ &\le \frac{1}{1-\ell} d(x_{k+1}, x_k) && \text{(geometric series, } \ell < 1\text{)} \\ &\le \frac{\ell^k}{1-\ell} d(x_1, x_0) && \text{(Lipschitzness)} \end{aligned}$$

- hence $\{x_k\}$ is Cauchy sequence, i.e., elements become arbitrarily close to each other as the sequence progresses
- since $\mathcal{X}$ is complete, there exists x^* to which the sequence converges
- uniqueness can be proved from $\ell < 1$
- $T(x^*) = x^*$ can be proved through triangle inequality
- geometric convergence

$$d(x_k, x^*) = d(T(x_{k-1}), x^*) \le \ell d(x_{k-1}, x^*) \le \ell^k d(x_0, x^*)$$

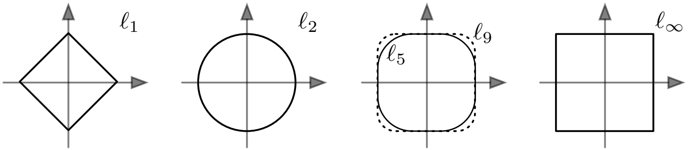
Examples of metric spaces

1. finite dimensional vector spaces with a norm ($\mathbb{R}^n$ and $d(x,y) = \|x-y\|$)
2. Riemannian manifolds (e.g., matrix Lie groups, Grassmanian/Stiefel ...)
3. infinite-dimensional Hilbert and Banach spaces
4. cones with the Thomson metric (e.g., positive definite matrices)
5. ...

Note: in this slides, contractivity = *contractivity on* $(\mathbb{R}^n, \|\cdot\|)$. Available for this case: all discrete/continuous-time theorems, numerous examples, amenable to analysis.

Linear algebra: induced 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
$\Vert x\Vert_2 = \sqrt{\sum_{i=1}^n x_i^2}$	$\Vert A\Vert_2 = \sqrt{\lambda_{\mathsf{max}}(A^\top A)}$	$\mu_2(A) = \lambda_{\mathsf{max}}\left(\frac{A + A^\top}{2}\right)$
$\Vert x\Vert_\infty = \max_{i\in\{1,\dots,n\}} \vert x_i\vert$	$\Vert A\Vert_\infty = \max_{i\in\{1,\dots,n\}} \sum_{j=1}^n \vert a_{ij}\vert$ $=$ 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



Discrete-time dynamics and Lipschitz constants

$$x_{k+1} = \mathsf{F}(x_k) \qquad \text{on } \mathbb{R}^n \text{ with norm } \|\cdot\| \text{ and induced norm } \|\cdot\|$$

Lipschitz constant

$$\begin{aligned}\mathsf{Lip}(\mathsf{F}) &= \inf\{\ell > 0 \text{ such that } \|\mathsf{F}(x) - \mathsf{F}(y)\| \leq \ell\|x-y\| \quad \text{for all } x, y\} \\ &= \sup_x \|D\mathsf{F}(x)\|\end{aligned}$$

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

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

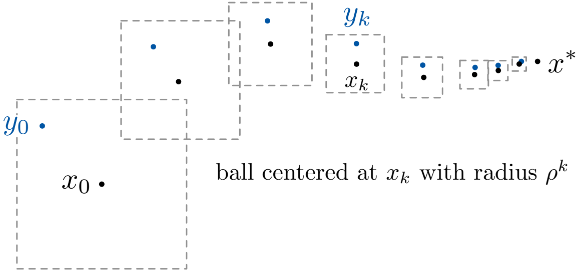
$$\|x\|_{2,P^{1/2}} = (x^\top P x)^{1/2} \qquad \mathsf{Lip}_{2,P^{1/2}}(\mathsf{F}_A) = \|A\|_{2,P^{1/2}} \leq \ell \qquad \iff \qquad A^\top P A \preceq \ell^2 P$$

$$\|x\|_\infty = \max_i |x_i| \qquad \mathsf{Lip}_\infty(\mathsf{F}_A) = \|A\|_\infty \leq \ell \qquad \iff \qquad |A|\mathbb{1}_n \leq \ell \mathbb{1}_n$$

Banach contraction theorem for discrete-time dynamics:

If $\rho := \mathsf{Lip}(\mathsf{F}) < 1$, then

1. F is **contracting** $=$ distance between trajectories decreases exp fast (ρ^k)
2. F has a unique, glob exp stable equilibrium x^*



From induced norms to induced log norms

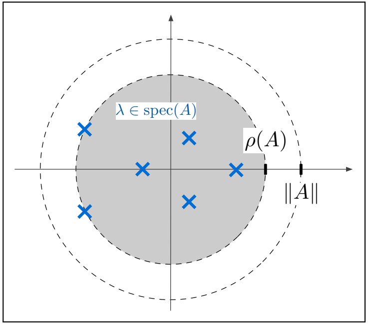
The **induced log norm** of $A \in \mathbb{R}^{n\times n}$ wrt to $\|\cdot\|$:

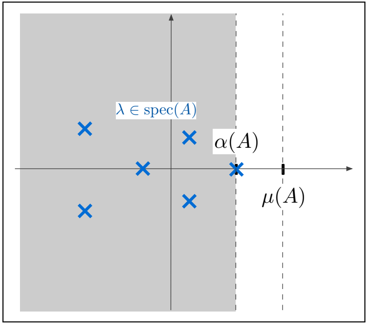
$$\mu(A) := \lim_{h\to 0^+} \frac{\|I_n + hA\| - 1}{h}$$

subadditivity: $\mu(A+B) \leq \mu(A) + \mu(B)$

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

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





Example induced log 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
$\Vert x\Vert_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$	$\Vert A\Vert_2 = \sqrt{\lambda_{\mathsf{max}}(A^\top A)}$	$\mu_2(A) = \lambda_{\mathsf{max}}\left(\frac{A+A^\top}{2}\right)$
$\Vert x\Vert_\infty = \max_{i\in\{1,\dots,n\}} \vert x_i\vert$	$\Vert A\Vert_\infty = \max_{i\in\{1,\dots,n\}} \sum_{j=1}^{n} \vert a_{ij}\vert$ $=$ 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

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

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

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

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

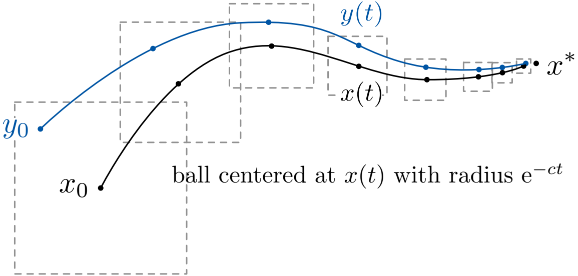
$$\mathsf{osLip}_{2,P^{1/2}}(\mathsf{F}_A) = \mu_{2,P^{1/2}}(A) \leq \ell \qquad \iff \qquad A^\top P + PA \preceq 2\ell P$$

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

Banach contraction theorem for continuous-time dynamics:

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

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



Key properties of inner products

Curve norm derivative formula: $\frac{1}{2}D^+\|x(t)\|^2 = \langle\!\langle \dot{x}(t), x(t) \rangle\!\rangle = \dot{x}^\top x$

Sup of Euclidean numerical range: $\mu_2(A) = \lambda_{\mathsf{max}}(\frac{A+A^\top}{2}) = \sup_{\|x\|=1} \langle\!\langle Ax, x \rangle\!\rangle = \sup_{x^\top x=1} x^\top Ax$

An **inner product** is $\langle\!\langle \cdot, \cdot \rangle\!\rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ satisfying

1. $\langle\!\langle x_1 + x_2, y \rangle\!\rangle = \langle\!\langle x_1, y \rangle\!\rangle + \langle\!\langle x_2, y \rangle\!\rangle$ (additivity)
2. $\langle\!\langle bx, y \rangle\!\rangle = \langle\!\langle x, by \rangle\!\rangle = b\,\langle\!\langle x, y \rangle\!\rangle$ for $b \in \mathbb{R}$ (homogeneity)
3. $\langle\!\langle x, x \rangle\!\rangle > 0$, for all $x \neq \mathbb{0}_n$ (definiteness)
4. $|\,\langle\!\langle x, y \rangle\!\rangle\,| \leq \langle\!\langle x, x \rangle\!\rangle^{1/2}\,\langle\!\langle y, y \rangle\!\rangle^{1/2}$ (Cauchy-Schwarz)

Given norm $\|\cdot\|$, compatibility: $\langle\!\langle x, x \rangle\!\rangle = \|x\|^2$ for all x

Definition and key properties of pairings

A **weak pairing** is $[\![\cdot,\cdot]\!] : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ satisfying

1. $[\![x_1 + x_2, y]\!] \leq [\![x_1, y]\!] + [\![x_2, y]\!]$, (sub-additivity)
2. $[\![bx, y]\!] = [\![x, by]\!] = b[\![x, y]\!]$ for $b \geq 0$ and $[\![-x, -y]\!] = [\![x, y]\!]$, (positive homogeneity)
3. $[\![x, x]\!] > 0$, for all $x \neq \mathbb{0}_n$, (definiteness)
4. $|[\![x, y]\!]| \leq [\![x, x]\!]^{1/2}[\![y, y]\!]^{1/2}$, (Cauchy-Schwarz)

Given norm $\|\cdot\|$, compatibility: $[\![x, x]\!] = \|x\|^2$ for all x

If weak pairing additionally satisfies $[\![-x, x]\!] = -\|x\|^2$, (straight angle property)
then it is called **regular** and satisfies:

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

Sup of non-Euclidean numerical range (Lumer): $\mu(A) = \sup_{\|x\|=1}[\![Ax, x]\!]$

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.

A. V. Proskurnikov and F. Bullo. Regular pairings for non-quadratic Lyapunov functions and contraction analysis. *SIAM Journal on Control and Optimization*, 2026. . To appear

Example regular pairings

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

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

Table of continuous-time contractivity conditions

Log Norm bound	Demidovich condition	One-sided Lipschitz condition
$\mu_{2,P^{1/2}}(D\mathsf{F}(x)) \leq b$	$PD\mathsf{F}(x) + D\mathsf{F}(x)^\top P \preceq 2bP$	$(x-y)^\top P\big(\mathsf{F}(x)-\mathsf{F}(y)\big) \leq b\\|x-y\\|^2_{P^{1/2}}$
$\mu_1(D\mathsf{F}(x)) \leq b$	$\mathrm{sign}(v)^\top D\mathsf{F}(x)v \leq b\\|v\\|_1$	$\mathrm{sign}(x-y)^\top(\mathsf{F}(x)-\mathsf{F}(y)) \leq b\\|x-y\\|_1$
$\mu_\infty(D\mathsf{F}(x)) \leq b$	$\max_{i\in I_\infty(v)} v_i\,(D\mathsf{F}(x)v)_i \leq b\\|v\\|^2_\infty$	$\max_{i\in I_\infty(x-y)}(x_i-y_i)(\mathsf{F}_i(x)-\mathsf{F}_i(y)) \leq b\\|x-y\\|^2_\infty$

Equivalent contractivity conditions

J. A. Jacquez and C. P. Simon. Qualitative theory of compartmental systems. *SIAM Review*, 35(1):43–79, 1993.

H. Qiao, J. Peng, and Z.-B. Xu. Nonlinear measures: A new approach to exponential stability analysis for Hopfield-type neural networks. *IEEE Transactions on Neural Networks*, 12(2):360–370, 2001.

G. Como, E. Lovisari, and K. Savla. Throughput optimality and overload behavior of dynamical flow networks under monotone distributed routing. *IEEE Transactions on Control of Network Systems*, 2(1):57–67, 2015.

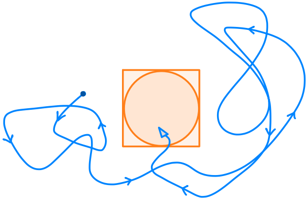
Advantages of non-Euclidean approaches

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 "$\mathbb{R}^n$ contradicts 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

  $\Longrightarrow$ desired set is hypercube $= \ell_\infty$-ball
- assume: Lyapunov function is $V(x) = \|x\|_2^2$

  $\Longrightarrow$ 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 (see also **https://youtu.be/sZqGWy0hxe8**)

Proof of Banach contraction theorem for continuous-time dynamics

For $\dot{x} = \mathsf{F}(x)$ with $\mathrm{osLip}(\mathsf{F}) = -c < 0$ and unit-time flow map ϕ:

- using the properties of the regular pairing, we compute

$$\begin{aligned} \|x-y\| D^+ \|x-y\| &= [\![\dot{x}-\dot{y}, x-y]\!] && (\textit{curve norm derivative}) \\ &= [\![\mathsf{F}(x)-\mathsf{F}(y), x-y]\!] && (\dot{x} = \mathsf{F}(x)) \\ &\leq -c\|x-y\|^2 && (\mathrm{osLip}(\mathsf{F}) = -c) \end{aligned}$$

- By the Grönwall Comparison,

$$D^+\|x-y\| \leq -c\|x-y\| \quad \implies \quad \|x(t)-y(t)\| \leq \mathrm{e}^{-ct}\|x(0)-y(0)\|$$

  and ϕ is a contraction with factor $\mathrm{e}^{-c} < 1$
- recall $(\mathbb{R}^n, \|\cdot\|)$ is complete metric space,
- the Banach Contraction Theorem implies **existence** of a unique x^* fixed point of ϕ
- $\phi(x^*) = x^*$ implies that
  - either x^* is an equilibrium
  - or x^* is a point in a periodic orbit with period 1,
- by contradiction, assume a periodic orbit of period 1 exists. Then each point in the orbit is a fixed point of ϕ, which violates the uniqueness of x^* as a fixed point,
- hence, x^* is the **unique** equilibrium of F.

The **upper right Dini derivative** of a continuous function $f :]a,b[\to \mathbb{R}$ at a point $t \in]a,b[$ is

$$D^+ f(t) = \limsup_{\Delta t > 0, \Delta t \to 0} \frac{f(t+\Delta t) - f(t)}{\Delta t}$$

where the limit superior of a sequence $\{a_n\}_{n\in\mathbb{N}} \subset \mathbb{R}$ is $\limsup_{n\to\infty} a_n = \lim_{n\to\infty} \sup_{m\geq n} a_m$.

Properties of the upper right Dini derivative

Given a continuous function $f :]a,b[\to \mathbb{R}$,

1. if f is differentiable at $t \in]a,b[$, then $D^+ f(t) = \frac{d}{dt} f(t)$ is the usual derivative of f at t,
2. if $D^+ f(t) \leq 0$ for all $t \in]a,b[$, then f is non-increasing on $]a,b[$.

Grönwall Comparison Lemma for absolutely continuous functions

Given $a \in \mathbb{R}$ and a continuous function $t \mapsto \gamma(t) \in \mathbb{R}$, assume the absolutely continuous function $t \mapsto z(t)$ satisfies the differential inequality

$$D^+ z(t) \leq a z(t) + \gamma(t).$$

Then, for $t \in [t_0, \infty)$,

$$z(t) \leq \mathrm{e}^{a(t-t_0)} z(t_0) + \int_{t_0}^{t} \mathrm{e}^{a(t-\tau)} \gamma(\tau) d\tau.$$

In other words, $z(t)$ is upper bounded by the solution to the corresponding differential equality.

Equivalence between integral and differential osLip For continuously-diff $\mathsf{F}:\mathbb{R}^n\to\mathbb{R}^n$

$$\mathsf{Lip}(\mathsf{F})=\sup_x\|D\mathsf{F}(x)\| \qquad \text{and} \qquad \mathsf{osLip}(\mathsf{F})=\sup_x\mu(D\mathsf{F}(x))$$

Proof Mean Value Theorem for vector-valued C^1 function $\mathsf{F}(x)-\mathsf{F}(y)=\big(\int_0^1 D\mathsf{F}\big(y+s(x-y)\big)ds\big)(x-y)$ for any x,y:

$$\begin{aligned}
\mathsf{osLip}(\mathsf{F}) &= \sup_{x\neq y}\frac{[\![\big(\int_0^1 D\mathsf{F}\big(y+s(x-y)\big)ds\big)(x-y),x-y]\!]}{\|x-y\|^2}\\
&\leq \sup_{x\neq y}\int_0^1\frac{[\![D\mathsf{F}\big(y+s(x-y)\big)(x-y),x-y]\!]}{\|x-y\|^2}ds && \text{(subadditivity of } [\![\cdot,\cdot]\!])\\
&\leq \int_0^1\sup_{x\neq y}\frac{[\![D\mathsf{F}\big(y+s(x-y)\big)(x-y),x-y]\!]}{\|x-y\|^2}ds=\int_0^1\sup_{y,z\neq\mathbb{0}_n}\frac{[\![D\mathsf{F}\big(y+sz\big)z,z]\!]}{\|z\|^2}ds\\
&= \int_0^1\sup_{y,z\neq\mathbb{0}_n}\mu\big(D\mathsf{F}\big(y+sz\big)\big)ds\leq\sup\nolimits_{x\in\mathbb{R}^n}\mu(D\mathsf{F}(x)) && \text{(Lumer's equality)}
\end{aligned}$$

Vice versa, recall $D\mathsf{F}(y)v=\lim_{h\to0^+}(\mathsf{F}(y+hv)-\mathsf{F}(y))/h$. Pick $x=y+hv$ for arbitrary $v\in\mathbb{R}^n$, $\|v\|=1$, and $h>0$,

$$\begin{aligned}
\mathsf{osLip}(\mathsf{F}) &= \sup_{y\in\mathbb{R}^n,v\in\mathbb{R}^n,\|v\|=1,h>0}\frac{[\![\mathsf{F}(x)-\mathsf{F}(y),x-y]\!]}{\|x-y\|^2}\Big|_{x=y+hv}\\
&\geq \sup_{y\in\mathbb{R}^n,v\in\mathbb{R}^n,\|v\|=1}\lim_{h\to0^+}\frac{[\![\mathsf{F}(y+hv)-\mathsf{F}(y),v]\!]}{h} && \text{(weak homogeneity)}\\
&= \sup_{y\in\mathbb{R}^n,v\in\mathbb{R}^n,\|v\|=1}[\![D\mathsf{F}(y)v,v]\!] && \text{(continuity of } w\mapsto[\![w,v]\!])\\
&= \sup_{y\in\mathbb{R}^n}\mu(D\mathsf{F}(y)). && \text{(Lumer's equality)}
\end{aligned}$$

Outline

§1. History and resources

§2. Basic definitions: discrete and continuous-time dynamics on vector spaces
- The linear algebra of matrix norms; see CTDS Chapter 2
- Properties of induced matrix norms and Lipschitz constants

§3. Example systems
- Constrained, distributed and proximal gradient dynamics
- Continuous-time recurrent neural networks
- Nonlinear dynamics in Lur'e form

§4. Properties of contracting dynamics
- Equilibria, Lyapunov functions, and Euler discretization
- Incremental input-to-state stability
- Contractivity of interconnected systems
- Additional properties: entrainment, robustness wrt unmodeled dynamics and delays

§5. Example applications
- Gradient dynamics and Nash equilibria in games
- Time-varying gradient dynamics and feedback optimization
- Recurrent and implicit neural networks

§6. Generalizations with examples
- G1: Local contractivity: Small-residual theorem and the Kuramoto coupled oscillators
- G2: Weak contractivity: Biologically-plausible circuits for sparse reconstruction
- G3: Contractivity on Riemannian manifolds and the Karcher mean
- G4: Semicontractivity: Primal-dual gradient with redundant constraints

§7. Conclusions and future research

§8. Advanced Topics
- More on semicontractivity: ergodic coefficients and duality
- Network small-gain theorem for Metzler matrices
- Proof of semicontractivity of saddle matrices
- Proof of Euler discretization theorem
- Non-Euclidean Monotone Operator Theory

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

"the real part properties"

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

F. Bullo. *Contraction Theory for Dynamical Systems*. Kindle Direct Publishing, 1.3 edition, 2026. ISBN 979-8836646806. URL **https://fbullo.github.io/ctds**

The linear algebra of matrix norms and log norms

Now review Chapter 2 in CTDS

Lemma 2.12 (Weighted matrix and log norms). *Given an invertible matrix R and a norm $\|\cdot\|$,*

$$\|A\|_R = \|RAR^{-1}\| \quad \text{and} \quad \mu_R(A) = \mu(RAR^{-1}). \tag{2.34}$$

Theorem 2.13 (Composite induced norms and log norms). *For any set of local norms $\|\cdot\|_i$ and an aggregating norm $\|\cdot\|_{\text{agg}}$ over a decomposition of $\mathbb{R}^n$, consider a matrix $A \in \mathbb{R}^{n\times n}$,*

(i) the composite norm $\|\cdot\|_{\text{cmpst}}$ is a well-defined, i.e., it satisfies the norm properties;

(ii) if the aggregating norm $\|\cdot\|_{\text{agg}}$ is monotonic, then

$$\max_{i\in\{1,\dots,r\}} \|A_{ii}\|_i \leq \|A\|_{\text{cmpst}} \leq \|\lceil A\rceil\|_{\text{agg}}, \tag{2.49}$$

$$\max_{i\in\{1,\dots,r\}} \mu_i(A_{ii}) \leq \mu_{\text{cmpst}}(A) \leq \mu_{\text{agg}}(\lceil A\rceil_{\text{M}}). \tag{2.50}$$

Log norms	Quadratic forms (for all $x \in \mathbb{R}^n$)	Ref
$\mu_{2,P^{1/2}}(A) = \lambda_{\max}\Big(\frac{PAP^{-1}+A^\top}{2}\Big)$	$\mu_{2,P^{1/2}}(A) = \max\{x^\top PAx : x^\top Px = 1\}$ $= \min\{b\in\mathbb{R} : A^\top P + PA \preceq 2bP\}$	Lemma 2.16
$\mu_{1,[\eta]}(A) = \max\big(\eta^\top\|A\|_{\text{M}}[\eta]^{-1}\big)$	$\mu_{1,[\eta]}(A) = \sup\{\text{sign}(x)^\top[\eta]Ax : \|x\|_{1,[\eta]} = 1\}$ $= \min\{b\in\mathbb{R} : \eta^\top\|A\|_{\text{M}} \leq b\eta^\top\}$	Lemma 2.21
$\mu_{\infty,[\eta]^{-1}}(A) = \max\big([\eta]^{-1}\|A\|_{\text{M}}\eta\big)$	$\mu_{\infty,[\eta]^{-1}}(A) = \max\{\max_{i\in I_\infty(x)} x_i(Ax)_i : \|x\|_{\infty,[\eta]^{-1}} = 1\}$ $= \min\{b\in\mathbb{R} : \|A\|_{\text{M}}\eta \leq b\eta\}$	Lemma 2.21

Table 2.2: Table of quadratic forms for weighted ℓ_1, ℓ_2, and ℓ_∞ log norms; $P \in \mathbb{S}^n_{\succ 0}$ and $\eta \in \mathbb{R}^n_{>0}$. We adopt the shorthand $I_\infty(x) = \{i \in \{1,\dots,n\} : |x_i| = \|x\|_\infty\}$ as the set of indices where x takes its maximal absolute value.

Theorem 2.23 (Spectrum-norm properties). *Given a matrix $A \in \mathbb{R}^{n\times n}$ and a norm $\|\cdot\|$,*

(i) for any eigenvalue λ of A, the spectral-radius norm property is

$$\text{(spectral-radius norm property)} \qquad 0 \leq |\lambda| \leq \rho(A) \leq \|A\|, \tag{2.71}$$

and, if A is invertible,

$$0 \leq 1/\|A^{-1}\| \leq |\lambda| \leq \rho(A) \leq \|A\|; \tag{2.72}$$

(ii) for any eigenvalue λ of A, the spectral-abscissa log-norm property is

$$\text{(spectral-abscissa log-norm property)} \qquad -\|A\| \leq -\mu(-A) \leq \Re(\lambda) \leq \alpha(A) \leq \mu(A) \leq \|A\|; \tag{2.73}$$

(iii) if the norm $\|\cdot\|$ is monotonic and A is diagonal, then

$$\|A\| = \max_{i\in\{1,\dots,n\}} |A_{ii}| = \rho(A), \tag{2.74}$$

$$\mu(A) = \max_{i\in\{1,\dots,n\}} A_{ii} = \alpha(A). \tag{2.75}$$

Theorem 2.24 (Monotonicity properties). *Consider a monotonic norm $\|\cdot\|$, a matrix $A \in \mathbb{R}^{n\times n}$, and a non-negative matrix $B \in \mathbb{R}^{n\times n}_{\geq 0}$. Then*

$$\text{(monotonicity property of spectral radius)} \qquad \rho(A) \leq \rho(|A|) \leq \rho(|A| + B), \tag{2.77a}$$

$$\text{(monotonicity property of induced norm)} \qquad \|A\| \leq \||A|\| \leq \||A| + B\| \tag{2.77b}$$

and

$$\text{(monotonicity property of spectral abscissa)} \qquad \alpha(A) \leq \alpha(|A|_{\text{M}}) \leq \alpha(|A|_{\text{M}} + B), \tag{2.78a}$$

$$\text{(monotonicity property of log norms)} \qquad \mu(A) \leq \mu(|A|_{\text{M}}) \leq \mu(|A|_{\text{M}} + B). \tag{2.78b}$$

Lemma 2.27 (Optimal weighted norms via the Jordan normal form). *Given a matrix $A \in \mathbb{R}^{n\times n}$, a monotonic norm $\|\cdot\|$, and $\varepsilon > 0$, define*

$$T \in \mathbb{C}^{n\times n} \text{ as an invertible matrix such that } TAT^{-1} \text{ is in Jordan normal form,} \tag{2.89}$$

$$Q \in \mathbb{C}^{n\times n} \text{ as an unitary matrix such that } QAQ^{-1} \text{ is in Schur normal form, and} \tag{2.90}$$

$$\delta = \varepsilon/\|J_{0,n}\| > 0, \text{ where } J_{0,n} \text{ is a Jordan block with eigenvalue } 0 \text{ and dimension } n. \tag{2.91}$$

Then

(i) the norm $\|\cdot\|_{\text{sdg}(\delta^{-1})T}$ is ε-optimal and ε-logarithmically optimal;

(ii) if A is diagonalizable, the norm $\|\cdot\|_T$ is optimal and logarithmically optimal; and

(iii) the norm $\|\cdot\|_{\text{sdg}(\delta^{-1})Q}$ is ε-optimal and ε-logarithmically optimal for sufficiently small δ.

Lemma 2.29 (Quasiconvex dependence upon matrix weights). *Given any $A \in \mathbb{R}^{n\times n}$,*

(i) the function $P \in \mathbb{S}^n_{\succ 0} \mapsto \mu_{2,P^{1/2}}(A)$ is quasiconvex with sublevel sets

$$\{P \in \mathbb{S}^n_{\succ 0} : \mu_{2,P^{1/2}}(A) \leq b\} = \{P \in \mathbb{S}^n_{\succ 0} : A^\top P + PA \preceq 2bP\}, \tag{2.93}$$

(ii) the functions $\eta \in \mathbb{R}^n_{>0} \mapsto \mu_{1,[\eta]}(A)$ and $\eta \in \mathbb{R}^n_{>0} \mapsto \mu_{\infty,[\eta]^{-1}}(A)$ are quasiconvex with sublevel sets

$$\{\eta \in \mathbb{R}^n_{>0} : \mu_{1,[\eta]}(A) \leq b\} = \{\eta \in \mathbb{R}^n_{>0} : \eta^\top|A|_{\text{M}} \leq b\eta^\top\}, \tag{2.94}$$

$$\{\eta \in \mathbb{R}^n_{>0} : \mu_{\infty,[\eta]^{-1}}(A) \leq b\} = \{\eta \in \mathbb{R}^n_{>0} : |A|_{\text{M}}\eta \leq b\eta\}. \tag{2.95}$$

Lemma 2.30 (Weighted ℓ_2 log norms and Lyapunov inequalities). *Given a matrix $A \in \mathbb{R}^{n\times n}$ with spectral abscissa $\alpha(A)$, define for any nonnegative tolerance $\varepsilon \geq 0$*

$$P_\varepsilon := \text{any element of } \{P \in \mathbb{S}^n_{\succ 0} : AP^\top + PA \preceq 2(\alpha(A)+\varepsilon)P\}. \tag{2.97}$$

Then

(i) for any $\varepsilon > 0$, P_ε is well defined and $\|\cdot\|_{2,P_\varepsilon^{1/2}}$ is logarithmically ε-optimal for A,

(ii) if each eigenvalue $\lambda_i(A)$ with $\Re(\lambda_i(A)) = \alpha(A)$ is semisimple, then P_0 is well defined and $\|\cdot\|_{2,P_0^{1/2}}$ is logarithmically optimal for A.

Lemma 2.31 (Optimal diagonally-weighted norms for non-negative and Metzler matrices). *Consider a nonnegative matrix $A \in \mathbb{R}^{n\times n}_{\geq 0}$ and a Metzler matrix $M \in \mathbb{R}^{n\times n}$. For any $p \in [1,\infty]$ and $\delta > 0$, define*

v *and* $w \in \mathbb{R}^n_{>0}$ *to be the right and left dominant eigenvectors of* $A + \delta\mathbb{1}_n\mathbb{1}_n^\top$ *(respectively,* $M + \delta\mathbb{1}_n\mathbb{1}_n^\top$*),*

$q \in [1,\infty]$ *to satisfy* $1/p + 1/q = 1$ *(with the convention* $1/\infty = 0$*), and*

$$\eta_\delta = \left(\frac{w_1^{1/p}}{v_1^{1/q}}, \dots, \frac{w_n^{1/p}}{v_n^{1/q}}\right) \in \mathbb{R}^n_{>0}.$$

Then

(i) for sufficiently small δ, the norm $\|\cdot\|_{p,[\eta_\delta]}$ is ε-optimal for A (respectively, ε-logarithmically optimal for M), and

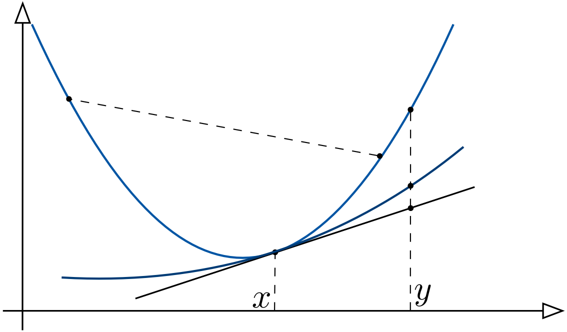
(ii) if A (respectively, M) is irreducible, then the norm $\|\cdot\|_{p,[\eta_0]}$ is optimal for A (respectively, logarithmically optimal for M).

Specifically, for $p \in \{1,2,\infty\}$ and for an irreducible A with spectral radius $\rho(A)$ and an irreducible M with spectral abscissa $\alpha(A)$,

$$\rho(A) = \|A\|_{1,[w]} = \|A\|_{\infty,[v]^{-1}} = \|A\|_{2,[w\oslash v]^{1/2}}, \tag{2.98}$$

$$\alpha(M) = \mu_{1,[w]}(M) = \mu_{\infty,[v]^{-1}}(M) = \mu_{2,[w\oslash v]^{1/2}}(M). \tag{2.99}$$

Outline



$f : \mathbb{R}^n \to \mathbb{R}$ is **ν-strongly convex** if, for all x, y,

1. $f(\chi x + (1-\chi)y) \leq \chi f(x) + (1-\chi)f(y) - \frac{1}{2}\nu\chi(1-\chi)\|x-y\|_2^2$ for each $0 \leq \chi \leq 1$
2. (if f is differentiable) $f(y) \geq f(x) + \nabla f(x)^\top (y-x) + \dfrac{\nu}{2}\|y-x\|_2^2$
3. (if f is differentiable) $\big(\nabla f(x) - \nabla f(y)\big)^\top (x-y) \geq \nu\|x-y\|_2^2$
4. (if f is twice differentiable) $\mathrm{Hess}\, f(x) \succeq \nu I_n$

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

If f is twice-differentiable, then $\mathrm{Hess}\, f(x) \succeq \nu I_n$ for all x

$$\begin{aligned} & D(-\nabla f)(x) = -\,\mathrm{Hess}\, f(x) \preceq -\nu I_n \\ \iff \qquad & I_n\, D(-\nabla f)(x) + D(-\nabla f)(x)^\top I_n \preceq -2\nu I_n \end{aligned}$$

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^*), that is

$$\big(-\nabla f(x) + \nabla f(y) \big)^\top (x-y) \leq -\nu \|x-y\|_2^2$$

For map $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$, equivalent statements:

1. F is a **monotone operator**[a] (or a **coercive operator**) with parameter ν,
2. $-\mathsf{F}$ is **ν-strongly contracting** wrt $\|\cdot\|_2$

[a] $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ is a *ν-strongly monotone operator* if $\langle\!\langle \mathsf{F}(x) - \mathsf{F}(y), x-y \rangle\!\rangle \geq \nu \|x-y\|_2^2$

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

Example #2: Primal-dual gradient dynamics

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

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:

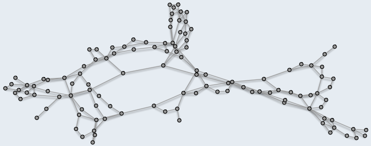
$$\begin{bmatrix} \dot{x} \\ \dot{\lambda} \end{bmatrix} = \mathsf{F}_{\mathsf{PDG}}(x, \lambda) := \begin{bmatrix} -\nabla f(x) - A^\top \lambda \\ Ax - b \end{bmatrix}$$

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

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

For each $\nu_{\mathsf{min}} I_n \preceq Q \preceq \nu_{\mathsf{max}} I_n$,

$$\begin{bmatrix} -Q & -A^\top \\ A & 0 \end{bmatrix}^\top P + P \begin{bmatrix} -Q & -A^\top \\ A & 0 \end{bmatrix} \preceq -2cP$$



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 $\mathbf{x} = [x_{[1]}, \ldots, x_{[n]}]$
consensus constraints among estimates are imposed in two ways:

1. incidence constraint: $B^\top \mathbf{x} = \mathbb{0}_m$
2. Laplacian constraint: $L\mathbf{x} = \mathbb{0}_n$

F. Dörfler. Distributed consensus-based optimization. Advanced Topics in Control 2018: Distributed Systems & Control, 2018

G. Qu and N. Li. On the exponential stability of primal-dual gradient dynamics. *IEEE Control Systems Letters*, 3(1):43–48, 2019.

Example #3: Incidence-based distributed gradient

Assume graph is a tree, so that (see LNS.Exercise9.2)

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

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

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

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

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

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

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

Example #4: Laplacian-based distributed gradient

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

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

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

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

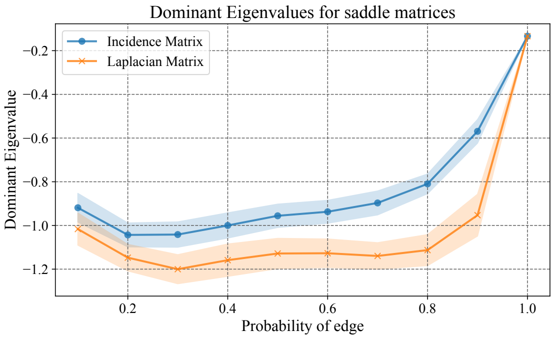
Laplacian-based distributed gradient (primal-dual gradient, $2n$ vars):

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

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

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

Empirically, define 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 $N = 40$ nodes, with varying edge probability parameters, 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

$$g(x) = \frac{1}{2}\|x\|_2^2 \qquad \mathrm{prox}_{\lambda g}(v) = \frac{v}{1+\lambda}$$

$$g(x) = \iota_C(x) = \begin{cases} 0 & \text{if } x \in C \\ \infty & \text{otherwise} \end{cases} \qquad \mathrm{prox}_{\lambda g}(v) = \Pi_C(v)$$

$$g(x) = \|x\|_1 \qquad \mathrm{prox}_{\lambda g}(v) = \mathrm{sign}(v) \cdot \max(|v| - \lambda, 0)$$

$$g(x) = \begin{cases} \frac{1}{2}x^2 & \text{if } |x| \le \delta \\ \delta|x| - \frac{1}{2}\delta^2 & \text{if } |x| > \delta \end{cases} \qquad \mathrm{prox}_{\lambda g}(v) = \begin{cases} \frac{v}{1+\lambda} & \text{if } |v| \le \delta + \lambda \\ v - \lambda \mathsf{sign}(v) & \text{if } |v| > \delta + \lambda \end{cases}$$

$f(\mathbf{x})$	$\mathrm{dom}(f)$	$\mathrm{prox}_f(\mathbf{x})$	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
$\lambda\|\mathbf{x}\|_1$	$\mathbb{R}^n$	$\mathcal{T}_\lambda(\mathbf{x})=[\lvert\mathbf{x}\rvert-\lambda\mathbf{e}]_+\odot\mathrm{sgn}(\mathbf{x})$	$\lambda>0$	Example 6.8
$\|\boldsymbol{\omega}\odot\mathbf{x}\|_1$	$\mathrm{Box}[-\boldsymbol{\alpha},\boldsymbol{\alpha}]$	$\mathcal{S}_{\boldsymbol{\omega},\boldsymbol{\alpha}}(\mathbf{x})$	$\boldsymbol{\alpha}\in[0,\infty]^n$, $\boldsymbol{\omega}\in\mathbb{R}^n_+$	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
$\lambda\sum_{i=1}^k x_{[i]}$	$\mathbb{R}^n$	$\mathbf{x}-\lambda P_C(\mathbf{x}/\lambda)$, $C=H_{\mathbf{e},k}\cap\mathrm{Box}[\mathbf{0},\mathbf{e}]$	$\lambda>0$	Example 6.50
$\lambda\sum_{i=1}^k\lvert x_{\langle i\rangle}\rvert$	$\mathbb{R}^n$	$\mathbf{x}-\lambda P_C(\mathbf{x}/\lambda)$, $C=B_{\|\cdot\|_1}[\mathbf{0},k]\cap\mathrm{Box}[-\mathbf{e},\mathbf{e}]$	$\lambda>0$	Example 6.51
$\lambda M_f^\mu(\mathbf{x})$	$\mathbb{E}$	$\mathbf{x}+\frac{\lambda}{\mu+\lambda}\left(\mathrm{prox}_{(\mu+\lambda)f}(\mathbf{x})-\mathbf{x}\right)$	$\lambda,\mu>0$, 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
$\rho\|\mathbf{x}\|_1^2$	$\mathbb{R}^n$	$\left(\frac{v_ix_i}{v_i+2\rho}\right)_{i=1}^n$, $\mathbf{v}=\left[\sqrt{\frac{\rho}{\mu}}\lvert\mathbf{x}\rvert-2\rho\right]_+$, $\mathbf{e}^T\mathbf{v}=1$ ($\mathbf{0}$ 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.

Example #5: Proximal gradient dynamics

proximal gradient dynamics:

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

$f : \mathbb{R}^n \to \mathbb{R}$ is ν-strongly convex and ℓ-strongly smooth
$g : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex, closed, proper

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 $c^\star = \frac{2\nu}{\nu+\ell}$

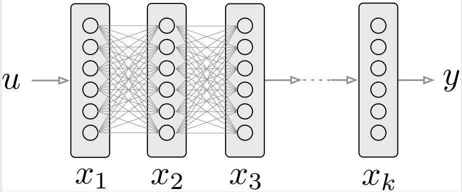
2. **$\mathsf{F}_{\mathsf{ProxG}}$ is infinitesimally contracting wrt $\|\cdot\|_{2,(\gamma A - I_n)^{1/2}}$ with rate $c = 1$**

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

Outline

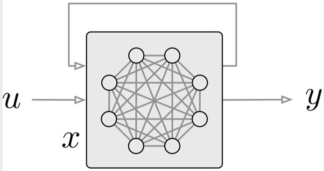
Neural network models

Feedforward NN



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

Recurrent NN



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

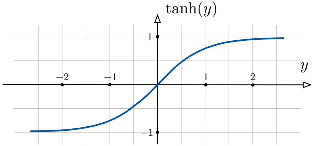
square matrix W = *synaptic matrix* — diagonal nonlinear Φ = *activation function*

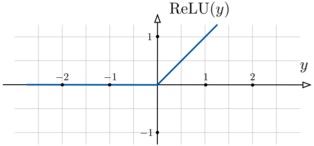
A. Davydov, A. V. Proskurnikov, and F. Bullo. Non-Euclidean contractivity of recurrent neural networks. In *American Control Conference*, pages 1527–1534, Atlanta, USA, May 2022c. doi

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

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

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





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

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

$$\text{osLip}_\infty(\mathsf{F}_{\mathsf{FR}}) = \sup_{x,u} \mu_\infty\big(-I_n + (D\Phi(Wx + Bu))W\big) = -1 + \sup_{x,u} \mu_\infty\big(D\Phi(Wx + Bu)W\big)$$

$$\le -1 + \max_{d \in [0,1]^n} \mu_\infty(\text{diag}(d)W) \qquad \text{(max convex polytope, } 2^n \text{ vertices)}$$

$$= -1 + \max\big\{\mu_\infty(0), \mu_\infty(W)\big\} = -1 + \mu_\infty(W)_+$$

For each row i, define the ith absolute row-sum of A by

$$\mathbb{r}_i(A) = a_{ii} + \sum_{j=1,j\neq i}^{n} |a_{ij}|$$

and note $\mu_\infty(A) = \max_i \mathbb{r}_i$.

Since $d_i \geq 0$ and $([d]A)_{ij} = d_i a_{ij}$, we note

$$\mathbb{r}_i([d]A) = d_i \mathbb{r}_i(A)$$

and compute

$$\begin{aligned}
\max_{d\in[0,1]^n} \mu_\infty([d]A) &\overset{\text{(by def)}}{=} \max_{d\in[0,1]^n} \max_i d_i \mathbb{r}_i(A) \\
&\overset{\text{(the } n \text{ functions are decoupled)}}{=} \max_i \max_{d_i\in[0,1]} d_i \mathbb{r}_i(A) \\
&\overset{(d_i \in [0,1])}{=} \max_i \begin{cases} \mathbb{r}_i, & \text{if } \mathbb{r}_i(A) \geq 0 \\ 0, & \text{if } \mathbb{r}_i(A) < 0 \end{cases} \\
&\overset{\text{(dropping the if clause)}}{\leq} \max\{\max_i \mathbb{r}_i(A), 0\} = \max\{\mu_\infty(A), 0\}.
\end{aligned}$$

Example #7: Firing-rate network with symmetric synapses

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

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

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** $\|\cdot\|_{2,Q_{\mathsf{FR},\lambda_W}}$

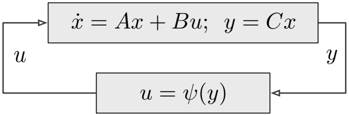
For $\lambda_W = 1$, F_{FR} **is weakly infinitesimally contracting wrt** $\|\cdot\|_{2,Q_{\mathsf{FR},\lambda_W}}$

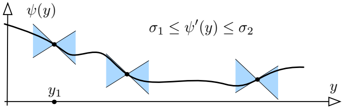
- $Q_{\mathsf{FR},a} := U h_a(\Lambda) U^\top \succ 0$, where $W = U \Lambda U^\top$ and $h_a(z) := 2a\big(1 + \sqrt{1 - z/a}\big)$
- optimal rates
- proof based upon LMI calculations and Sylvester's law of inertia

Outline

nonlinear system in Lur'e form $x \in \mathbb{R}^n, u \in \mathbb{R}, y \in \mathbb{R}$:

$$\dot{x} = Ax + Bu \qquad y = Cx$$
$$u = \psi(y) \qquad \psi : \mathbb{R} \to \mathbb{R}$$





$M = M^\top \in \mathbb{R}^{2\times 2}$ is an *incremental multiplier matrix* for ψ if

$$\begin{bmatrix} y_1 - y_2 \\ \psi(y_1) - \psi(y_2) \end{bmatrix}^\top M \begin{bmatrix} y_1 - y_2 \\ \psi(y_1) - \psi(y_2) \end{bmatrix} \geq 0 \qquad \text{for all } y_1, y_2 \in \mathbb{R}$$

Eg, *slope constraint* $\sigma_1 \leq \psi'(y) \leq \sigma_2$ is described by $M_{\sigma_1,\sigma_2} = \begin{bmatrix} -\sigma_1\sigma_2 & (\sigma_1+\sigma_2)/2 \\ (\sigma_1+\sigma_2)/2 & -1 \end{bmatrix}$

Example #8: Systems in Lur'e form

$$\mathsf{F}_{\mathsf{Lur'e}}(x) = Ax + B\psi(Cx)$$

assume

1. nonlinearity $\psi : \mathbb{R} \to \mathbb{R}$ described by incremental multiplier M
2. there exist an $n \times n$ matrix $P = P^\top \succ 0$ and a scalar $c > 0$ satisfying LMI

$$\begin{bmatrix} PA + A^\top P + 2cP & PB \\ B^\top P & 0 \end{bmatrix} + \begin{bmatrix} C & 0 \\ \mathbb{0}_{1\times n} & 1 \end{bmatrix}^\top M \begin{bmatrix} C & 0 \\ \mathbb{0}_{1\times n} & 1 \end{bmatrix} \preceq 0$$

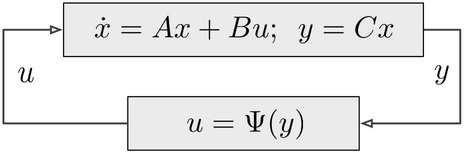
$\mathsf{F}_{\mathsf{Lur'e}}(x)$ is infinitesimally contracting wrt $\|\cdot\|_{2,P^{1/2}}$ with rate c

- proof based upon S-lemma
- LMIs defining P and M together imply contractivity LMI
- typical vector valued constraints: monotonic or sector bound

L. D'Alto and M. Corless. Incremental quadratic stability. *Numerical Algebra, Control and Optimization*, 3:175–201, 2013.

M. Giaccagli, V. Andrieu, S. Tarbouriech, and D. Astolfi. Infinite gain margin, contraction and optimality: An LMI-based design. *European Journal of Control*, 68:100685, 2022.

Example #8: Systems in Lur'e form: multivariable characterization



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

Equilibrium and Lyapunov functions for a contracting vector field

For a time-invariant F, c-strongly contracting wrt $\|\cdot\|$

1. for each $t > 0$, flow at time t of F is a contraction with factor e^{-ct}, i.e., distance between solutions exponentially decreases with rate c
2. there exists an equilibrium x^*, that is unique, globally exponentially stable with global Lyapunov functions

$$V_1(x) = \|x - x^*\|^2 \quad \text{and} \quad V_2(x) = \|\mathsf{F}(x)\|^2$$

3. if additionally $D\mathsf{F}(x) = D\mathsf{F}(x)^\top$ for all x, then another global Lyapunov function is

$$V_3(x) = -\int_0^1 x^\top \mathsf{F}(tx)dt + w \qquad \text{for each scalar } w$$

Also, V_3 is c-strongly convex and $\mathsf{F} = -\nabla V_3$

Proof of global Lyapunov functions

Regarding $V_1(x) = \|x - x^*\|^2$, from $D^+\|x - y\| \leq -c\|x - y\|$, we immediately have

$$\|x(t) - x^*\| \leq \mathrm{e}^{-ct}\|x(0) - x^*\|$$

Regarding $V_2(x) = \|\mathsf{F}(x)\|^2$, note $\dfrac{d}{dt}\mathsf{F}(x(t)) = D\mathsf{F}(x(t))\dot{x}(t) = D\mathsf{F}(x(t))\mathsf{F}(x(t))$ and

$$\begin{aligned}
\|\mathsf{F}(x(t))\| D^+\|\mathsf{F}(x(t))\| &= [\![\frac{d}{dt}\mathsf{F}(x(t)), \mathsf{F}(x(t))]\!] && \text{(curve norm derivative)}\\
&= [\![D\mathsf{F}(x(t))\mathsf{F}(x(t)), \mathsf{F}(x(t))]\!] \\
&\leq \mu(D\mathsf{F}(x(t)))[\![\mathsf{F}(x(t)), \mathsf{F}(x(t))]\!] && \text{(Lumer inequality)}\\
&\leq \sup_{z\in\mathbb{R}^n} \mu(D\mathsf{F}(z))\|\mathsf{F}(x(t))\|^2 = -c\|\mathsf{F}(x(t))\|^2
\end{aligned}$$

Regarding V_3, see M. Fitzsimmons and J. Liu. A note on the equivalence of a strongly convex function and its induced contractive differential equation. *Automatica*, page 110349, 2022.

Euler discretization theorem

Euler discretization theorem for contracting dynamics

Given arbitrary norm $\|\cdot\|$ 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

Optimal* contractivity of Euler discretization $\mathsf{Id} + \alpha\mathsf{F}$

Given $c := -\mathsf{osLip}(\mathsf{F}) > 0$ and $\ell := \mathsf{Lip}(\mathsf{F})$, define *condition number* $\kappa = \ell/c \geq 1$:

3. $$0 < \alpha < \frac{1}{c\kappa(1+\kappa)} \implies \mathsf{Lip}\big(\mathsf{Id} + \alpha\mathsf{F}\big) \leq \left(1 + \alpha c - \frac{\alpha^2\ell^2}{1-\alpha\ell}\right)^{-1} < 1$$

4. the optimal* step size and contraction factor are

$$\alpha^* = \frac{1}{c}\left(\frac{1}{2\kappa^2} - \frac{3}{8\kappa^3} + \mathcal{O}\left(\frac{1}{\kappa^4}\right)\right), \qquad \mathsf{Lip}(\mathsf{Id} + \alpha^*\mathsf{F}) = 1 - \frac{1}{4\kappa^2} + \frac{1}{8\kappa^3} + \mathcal{O}\left(\frac{1}{\kappa^4}\right)$$

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.

Optimal* contractivity of Euler discretization $\mathsf{Id} + \alpha\mathsf{F}$: inner-product norms $\|\cdot\|_{2,P^{1/2}}$

Given $c := -\,\mathsf{osLip}(\mathsf{F}) > 0$ and $\ell := \mathsf{Lip}(\mathsf{F})$, define *condition number* $\kappa = \ell/c \geq 1$:

1. $0 < \alpha < \dfrac{2}{c\kappa^2} \quad \Longrightarrow \quad \mathsf{Lip}\big(\mathsf{Id} + \alpha\mathsf{F}\big) \leq \sqrt{1 - 2\alpha c + \alpha^2\ell^2} < 1$
2. the optimal* step size and contraction factor are

$$\alpha^* = \frac{1}{c\kappa^2}, \qquad \mathsf{Lip}(\mathsf{Id} + \alpha^*\mathsf{F}) = 1 - \frac{1}{2\kappa^2} + \mathcal{O}\Big(\frac{1}{\kappa^4}\Big)$$

Standard proof from monotone operator theory. For $\alpha > 0$, compute

$$\begin{aligned}
\|(\mathsf{Id} + \alpha\mathsf{F})x - (\mathsf{Id} + \alpha\mathsf{F})y\|^2 &= \|x - y + \alpha(\mathsf{F}(x) - \mathsf{F}(y))\|^2 \\
&= \|x - y\|^2 + 2\alpha\,\langle\!\langle \mathsf{F}(x) - \mathsf{F}(y), x - y\rangle\!\rangle + \alpha^2\|\mathsf{F}(x) - \mathsf{F}(y)\|^2 \\
&\leq \big(1 - 2\alpha c + \alpha^2\ell^2\big)\|x - y\|^2
\end{aligned}$$

Next, study convex parabola $\alpha \mapsto 1 - 2\alpha c + \alpha^2\ell^2$. Eg, $1 - 2\alpha c + \alpha^2\ell^2 < 1$ iff $0 < \alpha < 2c/\ell^2$

E. K. Ryu and S. Boyd. Primer on monotone operator methods. *Applied Computational Mathematics*, 15(1):3–43, 2016

Outline

Fot time and input-dependent vector F,

$$\dot{x} = \mathsf{F}(t, x, u(t)), \qquad x(0) = x_0 \in \mathcal{X}, \qquad u(t) \in \mathcal{U}$$

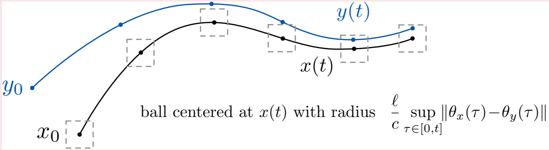
Given norms $\|\cdot\|_{\mathcal{X}}$ and $\|\cdot\|_{\mathcal{U}}$, assume

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

Then

1. any soltns: $x(t)$ with input u_x and $y(t)$ with input u_y

$$D^+\|x(t)-y(t)\|_{\mathcal{X}} \;\leq\; -c\|x(t)-y(t)\|_{\mathcal{X}} \;+\; \ell\|u_x(t)-u_y(t)\|_{\mathcal{U}}$$



2. F is **incrementally ISS**, that is, for all x_0, y_0

$$\|x(t)-y(t)\|_{\mathcal{X}} \;\leq\; \mathrm{e}^{-ct}\|x_0-y_0\|_{\mathcal{X}} \;+\; \frac{\ell(1-\mathrm{e}^{-ct})}{c} \sup_{\tau\in[0,t]} \|u_x(\tau)-u_y(\tau)\|_{\mathcal{U}}$$

Proof of iISS property

Using the properties of regular pairings, we compute

$$
\begin{aligned}
\|x(t)-y(t)\|\,D^+\|x(t)-y(t)\| &= [\![\dot{x}(t)-\dot{y}(t), x-y]\!] && (\textit{curve norm derivative})\\
&= [\![\mathsf{F}(t,x,u_x)-\mathsf{F}(t,y,u_y), x-y]\!] \\
&\leq [\![\mathsf{F}(t,x,u_x)-\mathsf{F}(t,y,u_x), x-y]\!] \\
&\qquad + [\![\mathsf{F}(t,y,u_x)-\mathsf{F}(t,y,u_y), x-y]\!] && (\textit{subadditivity})\\
&\leq -c\|x-y\|^2 + [\![\mathsf{F}(t,y,u_x)-\mathsf{F}(t,y,u_y), x-y]\!] && (\text{contractivity})\\
&\leq -c\|x-y\|^2 + \|\mathsf{F}(t,y,u_x)-\mathsf{F}(t,y,u_y)\|\,\|x-y\| && (\textit{Cauchy-Schwarz})\\
&\leq -c\|x-y\|^2 + \ell\|u_x-u_y\|\,\|x-y\|. && (\text{Lipschitzness})
\end{aligned}
$$

Given norm $\|\cdot\|_{\mathcal{X}}$ on $\mathbb{R}^n$ (or $\|\cdot\|_{\mathcal{U}}$ on $\mathbb{R}^k$),

- $\mathcal{L}^q_{\mathcal{X}}$, $q \in [1, \infty]$, is vector space of continuous signals, $x : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$, with well-defined bounded norm

$$\|x(\cdot)\|_{\mathcal{X},q} = \begin{cases} \left(\int_0^\infty \|x(t)\|^q_{\mathcal{X}} dt \right)^{1/q} & \text{if } q \in [1, \infty[\\ \sup_{t \geq 0} \|x(t)\|_{\mathcal{X}} & \text{if } q = \infty \end{cases}$$

- Input-state system has **$\mathcal{L}^q_{\mathcal{X},\mathcal{U}}$-induced gain** upper bounded by $\gamma > 0$ if, for all $u \in \mathcal{L}^q_{\mathcal{U}}$, the state x from zero initial state satisfies

$$\|x(\cdot)\|_{\mathcal{X},q} \leq \gamma \|u(\cdot)\|_{\mathcal{U},q}$$

3. F has **incremental $\mathcal{L}^q_{\mathcal{X},\mathcal{U}}$ gain equal to ℓ/c, for $q \in [1, \infty]$**,

$$\|x(\cdot) - y(\cdot)\|_{\mathcal{X},q} \leq \frac{\ell}{c} \|u_x(\cdot) - u_y(\cdot)\|_{\mathcal{U},q} \qquad (\text{for } x_0 = y_0)$$

$$\mathbb{0}_n = \mathsf{F}(x,u), \quad x \in \mathcal{X}, u \in \mathcal{U}$$

Given norms $\|\cdot\|_{\mathcal{X}}$ and $\|\cdot\|_{\mathcal{U}}$, assume

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

1. for each fixed u, there exists a unique equilibrium $x^\star(u)$
2. the equilibrium map $x^* : \mathcal{U} \to \mathcal{X}$ is Lipschitz with constant $\dfrac{\ell}{c}$

Sensitivity analysis in convex optimization

If $f(x,u)$ is ν-strongly convex and differentiable wrt x,
$\nabla_x f$ is ℓ-Lipschitz wrt u,

then $\quad$ global minimum $u \mapsto x^*(u)$ is Lipschitz with constant $\dfrac{\ell}{\nu}$

Proof of Parametrized continuous-time Banach Contraction Theorem

Recall iISS: any soltns $x_1(t)$ with input $u_1(t)$ and $x_2(t)$ with input $u_2(t)$

$$D^+\|x_1(t) - x_2(t)\|_{\mathcal{X}} \;\leq\; -c\|x_1(t) - x_2(t)\|_{\mathcal{X}} \;+\; \ell\|u_1(t) - u_2(t)\|_{\mathcal{U}}$$

For constant $u_1(t) = u_1$ and $u_2(t) = u_2$, as $t \to +\infty$,

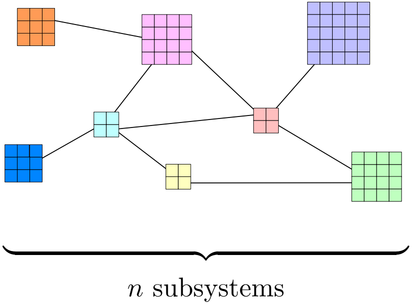
$$x_1(t) \to x^*(u_1) \quad \text{and} \quad x_2(t) \to x^*(u_2)$$

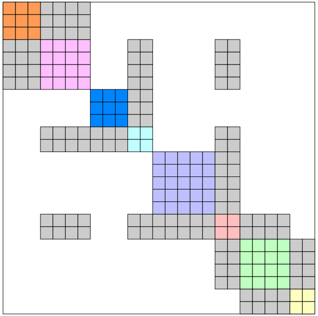
Taking the limit, we obtain

$$0 \leq -c\|x^*(u_1) - x^*(u_2)\|_{\mathcal{X}} \;+\; \ell\|u_1 - u_2\|_{\mathcal{U}}$$

that is, $\|x^*(u_1) - x^*(u_2)\|_{\mathcal{X}} \leq \dfrac{\ell}{c}\|u_1 - u_2\|_{\mathcal{U}}$

Outline





1. n *local norms* $\|\cdot\|_i$ on $\mathbb{R}^{N_i}$, $i \in \{1,\dots,n\}$
2. a *aggregating norm* $\|\cdot\|_{\text{agg}}$ on $\mathbb{R}^n$
3. $\implies$ a *composite norm* on $\mathbb{R}^N$, $N = N_1 + \dots + N_n$

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

O. Pastravanu and M. Voicu. Generalized matrix diagonal stability and linear dynamical systems. *Linear Algebra and its Applications*, 419 (2):299–310, 2006.

G. Russo, M. Di Bernardo, and E. D. Sontag. A contraction approach to the hierarchical analysis and design of networked systems. *IEEE Transactions on Automatic Control*, 58(5):1328–1331, 2013.

Interconnected subsystems: $x_i \in \mathbb{R}^{N_i}$ and $x_{-i} \in \mathbb{R}^{N-N_i}$:

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

Network contraction theorem. Given local norms, assume

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

$\Longrightarrow$ the **interconnected system** is infinitesimally contracting
wrt composite $\| \cdot \|$ generated by log optimal norm for Γ and $c = |\alpha(\Gamma)|$

$$\begin{bmatrix} -c_1 & \dots & \ell_{1n} \\ \vdots & & \vdots \\ \ell_{n1} & \dots & -c_n \end{bmatrix}$$ is **Metzler** (Perron-Frobenius Theorem applies)

(see LNS.Section10.4)

Hurwitzness depends upon both topology and edge weights

- M Hurwitz iff there exists a positive ξ such that $M\xi < \mathbb{0}_n$ (power method)
- For $n = 2$, Hurwitz if and only if **small gain condition**

$$\text{cycle gain} := \frac{\ell_{12}}{c_1}\frac{\ell_{21}}{c_2} < 1$$

- For $n \geq 3$, Hurwitz if **network small gain condition**

see **network small-gain theorem for Metzler matrices**

Proof of Network Contraction Theorem (via Jacobians and aggregate majorants)

For $\dot{x}_i = \mathsf{F}_i(x_i, x_{-i})$ we compute $D\mathsf{F}(x) = \begin{bmatrix} D_{x_1}\mathsf{F}_1 & \dots & D_{x_n}\mathsf{F}_1 \\ \vdots & & \vdots \\ D_{x_1}\mathsf{F}_n & \dots & D_{x_n}\mathsf{F}_n \end{bmatrix}$

Assuming local norms, aggregate norm and composite norm (Section 2.4.4), recalling the definition of aggregate Metzler majorant:

$$\begin{aligned} \mathsf{osLip}_{\mathsf{cmpst}}\big(\mathsf{F}\big) &= \sup_x \mu_{\mathsf{cmpst}}\big(D\mathsf{F}(x)\big) \\ &\leq \sup_x \mu_{\mathsf{agg}}(\lceil D\mathsf{F}(x)\rceil_{\mathsf{M}}) && \text{(composite norm Theorem 2.13)} \\ &\overset{\text{entry-wise}}{\leq} \mu_{\mathsf{agg}}(\sup_x \lceil D\mathsf{F}(x)\rceil_{\mathsf{M}}) && \text{(monotonicity properties Theorem 2.24)} \\ &= \mu_{\mathsf{agg}}(\Gamma) && \text{(definition of } \Gamma) \end{aligned}$$

and, when aggregate norm is ϵ-logarithmically optimal for Metzler matrix Γ,

$$\mathsf{osLip}_{\mathsf{cmpst}}\big(\mathsf{F}\big) \leq \mu_{\mathsf{agg}}(\Gamma) \leq \alpha(\Gamma) + \epsilon \qquad \text{(for arbitrarily small } \epsilon)$$

Note: The same proof method works for discrete time systems.
Note: Sharper-but-harder-to-check sufficient condition: there exists an aggregate norm (say row/colun sum or Demidovich) such that $\mu_{\mathsf{agg}}(\lceil D\mathsf{F}(x)\rceil_{\mathsf{M}})(x) \leq -c < 0$

Proof of Network Contraction Theorem (via pairings)

First, design a log optimal norm for $\Gamma = \begin{bmatrix} -c_1 & \dots & \ell_{1r} \\ \vdots & & \vdots \\ \ell_{r1} & \dots & -c_r \end{bmatrix} \in \mathbb{R}^{r \times r}$

From Lemma 3.21 on Metzler matrices in CTDS, for arbitrarily small ϵ, one can compute $\eta \in \mathbb{R}^n_{>0}$ such that $\|\cdot\|_{2,\mathrm{diag}(\eta)^{1/2}}$ is ϵ-log optimal for Γ:

$$\mu_{2,\mathrm{diag}(\eta)^{1/2}}(\Gamma) \leq \alpha(\Gamma) + \epsilon \quad \iff \quad \mathrm{diag}(\eta)\Gamma + \Gamma^\top \mathrm{diag}(\eta) \preceq 2(\alpha(\Gamma) + \epsilon)\,\mathrm{diag}(\eta)$$

Next, define the composite norm $\|\cdot\|_\eta$ on $\mathbb{R}^N$ by

$$\|(x_1, \dots, x_r)\|_\eta^2 = \sum\nolimits_{i=1}^r \eta_i \|x_i\|_i^2$$

with pairing

$$[\![(x_1, \dots, x_r), (y_1, \dots, y_r)]\!]_\eta = \sum\nolimits_{i=1}^r \eta_i [\![x_i, y_i]\!]_i$$

For each i, compute

$$
\begin{aligned}
&[\![\mathsf{F}_i(t,x_i,x_{-i}) - \mathsf{F}_i(t,y_i,y_{-i}), x_i - y_i]\!]_i \\
&\qquad \le [\![\mathsf{F}_i(t,x_i,x_{-i}) - \mathsf{F}_i(t,y_i,x_{-i}), x_i - y_i]\!]_i + [\![\mathsf{F}_i(t,y_i,x_{-i}) - \mathsf{F}_i(t,y_i,y_{-i}), x_i - y_i]\!]_i \\
&\qquad \le -c_i \|x_i - y_i\|_i^2 + \sum\nolimits_{j=1, j\neq i}^{r} \ell_{ij} \|x_j - y_j\|_j \|x_i - y_i\|_i
\end{aligned}
$$

Next, we check the one-sided Lipschitz condition for the vector field on $\mathbb{R}^N$:

$$
\begin{aligned}
&\sum\nolimits_{i=1}^{r} \eta_i [\![\mathsf{F}_i(t,x_i,x_{-i}) - \mathsf{F}_i(t,y_i,y_{-i}), x_i - y_i]\!]_i \\
&\qquad \le -\sum\nolimits_{i=1}^{r} \eta_i c_i \|x_i - y_i\|_i^2 + \sum\nolimits_{i,j=1, j\neq i}^{r} \eta_i \ell_{ij} \|x_j - y_j\|_j \|x_i - y_i\|_i \\
&\qquad = \begin{bmatrix} \|x_1 - y_1\|_1 \\ \vdots \\ \|x_r - y_r\|_r \end{bmatrix}^\top \operatorname{diag}(\eta)\Gamma \begin{bmatrix} \|x_1 - y_1\|_1 \\ \vdots \\ \|x_r - y_r\|_r \end{bmatrix} \\
&\qquad = \begin{bmatrix} \|x_1 - y_1\|_1 \\ \vdots \\ \|x_r - y_r\|_r \end{bmatrix}^\top \frac{\operatorname{diag}(\eta)\Gamma + \Gamma^\top \operatorname{diag}(\eta)}{2} \begin{bmatrix} \|x_1 - y_1\|_1 \\ \vdots \\ \|x_r - y_r\|_r \end{bmatrix}
\end{aligned}
$$

so that the interconnected system is contracting if the gain matrix Γ is diagonally stable.

Application: Singularly perturbed matrices

Given a constant $\epsilon > 0$, consider block matrix

$$\mathcal{A}_\epsilon = \begin{bmatrix} \epsilon A & \epsilon B \\ C & D \end{bmatrix} \in \mathbb{R}^{(n+m)\times(n+m)}.$$

$$\begin{array}{ccc}
\begin{array}{c}\mu(A) < 0,\ \mu(D) < 0, \text{ and} \\ \mu(A)\mu(D) > \|B\|\|C\|\end{array} & \Longrightarrow & \mathcal{A}_\epsilon \text{ is Hurwitz for all } \epsilon > 0 \\
\Downarrow & & \Downarrow \\
D \text{ and } A - BD^{-1}C \text{ are Hurwitz} & \Longrightarrow & \exists \epsilon^* \text{ s.t. } \mathcal{A}_\epsilon \text{ is Hurwitz for each } \epsilon < \epsilon^*
\end{array}$$

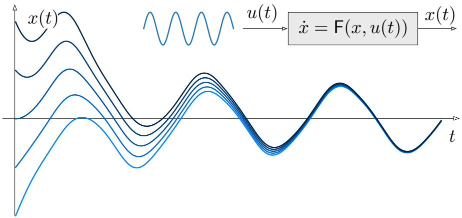
Additionally, a valid choice of ϵ^* is:

$$\epsilon^* := \frac{|\mu(A - BD^{-1}C)| \cdot |\mu(D)|}{\|B\|\|D^{-1}C(A - BD^{-1}C)\| + |\mu(A - BD^{-1}C)|\|D^{-1}CB\|}$$

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. To appear

Outline

Entrainment in systems with periodic time-dependence



For time-varying vector field $\mathsf{F}(t, x)$ and norm $\|\cdot\|$

1. $\mathrm{osLip}_x(\mathsf{F}) \leq -c < 0$, uniformly in t
2. F is T-periodic in t

Then

1. there exists a unique periodic solution $x^* : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$ with period T
2. for every initial condition x_0,

$$\|x(t, x_0) - x^*(t)\| \leq \mathrm{e}^{-ct}\|x_0 - x^*(0)\|$$

G. Russo, M. Di Bernardo, and E. D. Sontag. Global entrainment of transcriptional systems to periodic inputs. *PLoS Computational Biology*, 6(4):e1000739, 2010.

Robustness of systems with unmodeled dynamics

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

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

Assume:

- **contractivity**: $\mathrm{osLip}(\mathsf{F}) \leq -c < 0$
- **bounded disturbance**: $\mathrm{osLip}(\Delta) \leq d < c$

Then

1. $\mathsf{F} + \Delta$ is strongly contracting with rate $c - d$
2. the unique equilibria x^*_{F} of F and $x^*_{\mathsf{F}+\Delta}$ of $\mathsf{F} + \Delta$ satisfy

$$\|x^*_{\mathsf{F}} - x^*_{\mathsf{F}+\Delta}\| \leq \frac{\|\Delta(x^*_{\mathsf{F}})\|}{c - d}$$

$$\dot{x}(t) = \mathsf{F}\big(x(t), x(t-s), u(t)\big), 0 \le s \le S, \qquad \|\cdot\|_{\mathcal{X}}, \|\cdot\|_{\mathcal{U}} \tag{2}$$

assume there exist positive constants $c, \ell_{\mathcal{U}}, \ell_{\mathcal{X}}$ such that, for all variables,

$$\textbf{osL } x: \qquad [\![\mathsf{F}(x,d,u) - \mathsf{F}(y,d,u), x-y]\!]_{\mathcal{X}} \le -c\|x-y\|_{\mathcal{X}}^2 \tag{3}$$

$$\textbf{Lip } x(t-s): \qquad \|\mathsf{F}(x,x_1,u) - \mathsf{F}(x,x_2,u)\|_{\mathcal{X}} \le \ell_{\mathcal{X}}\|x_1 - x_2\|_{\mathcal{X}} \tag{4}$$

$$\textbf{Lip } u: \qquad \|\mathsf{F}(x,d,u) - \mathsf{F}(x,d,v)\|_{\mathcal{X}} \le \ell_{\mathcal{U}}\|u-v\|_{\mathcal{U}} \tag{5}$$

By the curve norm derivative formula, subadditivity, and Cauchy-Schwarz inequality,

$$\begin{aligned}
\|x(t)-y(t)\|_{\mathcal{X}} D^+\|x(t)-y(t)\|_{\mathcal{X}} &= [\![\mathsf{F}(x(t), x(t-s), u_x(t)) - \mathsf{F}(y(t), y(t-s), u_y(t)), x(t)-y(t)]\!]_{\mathcal{X}} \\
&\le [\![\mathsf{F}(x(t), x(t-s), u_x(t)) - \mathsf{F}(y(t), x(t-s), u_x(t)), x(t)-y(t)]\!]_{\mathcal{X}} \\
&\quad + [\![\mathsf{F}(y(t), x(t-s), u_x(t)) - \mathsf{F}(y(t), y(t-s), u_x(t)), x(t)-y(t)]\!]_{\mathcal{X}} \\
&\quad + [\![\mathsf{F}(y(t), y(t-s), u_x(t)) - \mathsf{F}(y(t), y(t-s), u_y(t)), x(t)-y(t)]\!]_{\mathcal{X}} \\
&\le -c\|x(t)-y(t)\|_{\mathcal{X}}^2 + \ell_{\mathcal{X}}\|x(t-s)-y(t-s)\|_{\mathcal{U}}\|x(t)-y(t)\|_{\mathcal{X}}, \\
&\quad + \ell_{\mathcal{U}}\|u_x(t)-u_y(t)\|_{\mathcal{U}}\|x(t)-y(t)\|_{\mathcal{X}}.
\end{aligned}$$

Thus, with $z(t) = \|x(t)-y(t)\|_{\mathcal{X}}$, delay differential inequality:

$$D^+z(t) \le -cz(t) + \ell_{\mathcal{X}} \sup\nolimits_{0\le s\le S} z(t-s) + \ell_{\mathcal{U}}\|u_x(t)-u_y(t)\|_{\mathcal{U}}, \tag{6}$$

Halanay inequality is applicable (see Chapter 3). If $c > \ell_{\mathcal{X}}$, then

$$z(t) \le z_0 \mathrm{e}^{-\rho(t-t_0)} + \ell_{\mathcal{U}} \int_{t_0}^{t} \mathrm{e}^{-\rho(t-\tau)}\|u_x(\tau)-u_y(\tau)\|_{\mathcal{U}} d\tau, \tag{7}$$

where $\rho > 0$ is the unique positive root of $\rho = c - \ell_{\mathcal{X}}\mathrm{e}^{\rho S}$ and $z_0 = \sup_{0\le s\le S} z(t_0 - s)$.

Interconnected subsystems $i \in \{1, \dots, n\}$

$$\dot{x}_i = \mathsf{F}_i(x_i, x_{-i}, x_{-i}(t-s), u_i), \qquad 0 \le s \le S, \qquad \|\cdot\|_i, \|\cdot\|_{i,\mathcal{U}} \tag{8}$$

Assume there exist positive constants st

osL x_i : $\quad [\![\mathsf{F}_i(x_i, \dots) - \mathsf{F}_i(y_i, \dots), x_i - y_i]\!]_i \le -c_i \|x_i - y_i\|_i^2$

Lip x_{-i} : $\quad \|\mathsf{F}_i(\dots, x_{-i}, \dots) - \mathsf{F}_i(\dots, y_{-i}, \dots)\|_i \le \sum_{j=1, j\neq i}^{n} \gamma_{ij} \|x_j - y_j\|_j$

Lip x_{-1}^{-s} : $\quad \|\mathsf{F}_i(\dots, x_{-i}^{-s}, \dots) - \mathsf{F}_i(\dots, y_{-i}^{-s}, \dots)\|_i \le \sum_{j=1, j\neq i}^{n} \widehat{\gamma}_{ij} \|x_j^{-s} - y_j^{-s}\|_j$

Lip u_i : $\quad \|\mathsf{F}_i(\dots, u_i) - \mathsf{F}_i(\dots, v_i)\|_i \le \ell_{i,\mathcal{U}} \|u_i - v_i\|_{i,\mathcal{U}}$

With $z_i(t) = \|x_i(t) - y_i(t)\|_i$, delay differential inequality on $\mathbb{R}^n_{\ge 0}$:

$$D^+ z(t) \le -C z(t) + \Gamma z(t) + \widehat{\Gamma} \sup_{0 \le s \le S} z(t-s) + \ell_{\mathcal{U}} \|u_x(t) - u_y(t)\|_{\mathcal{U}}$$

and, if the Metzler matrix $-C + \Gamma + \widehat{\Gamma}$ is Hurwitz, then (8) is incremental ISS

F. Mazenc, M. Malisoff, and M. Krstic. Vector extensions of Halanay's inequality. *IEEE Transactions on Automatic Control*, 67(3):1453–1459, 2022.

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.

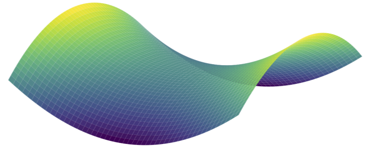
D. Gadjov and L. Pavel. A passivity-based approach to Nash equilibrium seeking over networks. *IEEE Transactions on Automatic Control*, 64(3):1077–1092, 2019.

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.

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.

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

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

Aim: $\min_x \max_y f(x, y)$

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

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

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)

If f is twice-differentiable, then

$$\sup_x \mu_2(D\mathsf{F}_{\mathsf{S}}(x,y)) = \sup_x \mu_2\left(\begin{bmatrix}-\operatorname{Hess}_x f(x,y) & -D_y\nabla_x f(x,y)\\ D_x\nabla_y f(x,y) & \operatorname{Hess}_y f(x,y)\end{bmatrix}\right)$$

$$\overset{\mu_2(A)=\mu_2(\frac{A+A^\top}{2})}{=} \sup_x \mu_2\left(\begin{bmatrix}-\operatorname{Hess}_x f(x,y) & 0\\ 0 & \operatorname{Hess}_y f(x,y)\end{bmatrix}\right) = -\min\{\nu_x,\nu_y\}$$

Example #10: 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):

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

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

- **strong convexity wrt x_i:** J_i is μ_i strongly convex wrt x_i, uniformly in x_{-i}
- **Lipschitz wrt x_{-i}:** $\mathsf{Lip}_{x_j}(\nabla_i J_i) \leq \ell_{ij}$, 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 #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 $\|\cdot\|_2$**

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

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

Equivalent statements:

1. F_{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 $\ell_i = \text{Lip}_y(\nabla_{x_i} f_i(x_i, y))$

$$\mu_i > \ell_i \text{ for each agent } i \qquad \Longrightarrow \qquad \text{Hurwitz}$$

A game-theoretic antagonistic opinion dynamics example

Agents with variable opinions $x_i \in \mathbb{R}$ and fixed private opinion p_i

Given unsigned interpersonal weights a_{ij} and attachment parameter $b_i > 0$, cost function:

$$J_i(x) = \underbrace{\tfrac{1}{2}\sum_j a_{ij}(x_i - x_j)^2}_{\text{tendency to consensus/dissensus}} + \underbrace{b_i(x_i - p_i)^2}_{\text{attachment to private opinion}} + \underbrace{\psi(x_i)}_{\text{convex penalty}}$$

- $\mathrm{Hess}_i\, J_i = 2b_i + \sum_j a_{ij} + \mathrm{Hess}\,\psi(x_i)$
- $\nabla_i J_i = \sum_j a_{ij}(x_j - x_i) + 2b_i(x_i - p_i) + \partial_i \psi(x_i)$ and $\mathrm{Lip}_{x_j}(\nabla_i J_i) = |a_{ij}|$

If **weak antagonistic relations**

$$b_i > \sum_{j \text{ s.t. } a_{ij}<0} |a_{ij}|$$

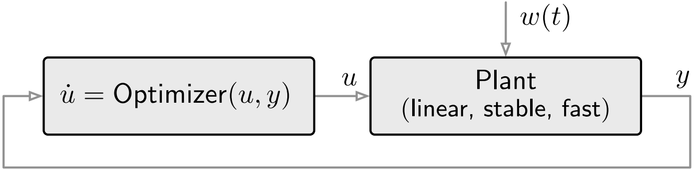
then

1. gain matrix of $\mathsf{F}_{\text{PseudoG}}$ has negative row sums
2. pseudogradient and best response play are strongly contracting wrt $\|\cdot\|_\infty$

P. Wankhede, N. Mandal, S. Martínez, and P. Tallapragada. Opinion dynamics for utility maximizing agents: Exploring the impact of the resource penalty. *IEEE Transactions on Control of Network Systems*, 12(1):5–17, 2025.

Outline

Solving optimization problems via dynamical systems



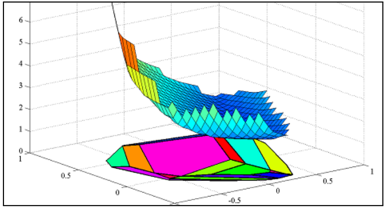
- studies in linear and nonlinear programming (Arrow, Hurwicz, and Uzawa 1958)
- neural networks (Hopfield and Tank 1985) and analog circuits (Kennedy and Chua 1988)
- optimization on manifolds (Brockett 1991)
- ...
- online and dynamic feedback optimization (Dall'Anese, Dörfler, Simonetto, ...)

A. Davydov, V. Centorrino, A. Gokhale, G. Russo, and F. Bullo. Time-varying convex optimization: A contraction and equilibrium tracking approach. *IEEE Transactions on Automatic Control*, 70(11):7446–7460, 2025.

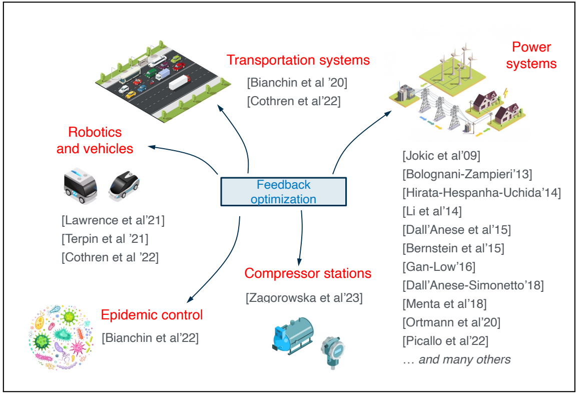
L. Cothren, F. Bullo, and E. Dall'Anese. Online feedback optimization and singular perturbation via contraction theory. *SIAM Journal on Control and Optimization*, Aug. 2024. . To appear

Motivation: Optimization-based control

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.

Parametric and time-varying convex optimization

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

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*, 70(11):7446–7460, 2025.

Parametric convex optimization and contracting dynamics

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. $\quad Ax \leq b(\theta)$	$\dot{x} = -\nabla f(x, \theta) - A^\top \nabla M_{\gamma, b(\theta)}(Ax + \gamma\lambda),$ $\dot{\lambda} = \gamma(-\lambda + \nabla M_{\gamma, b(\theta)}(Ax + \gamma\lambda))$

Tracking equilibrium curves

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

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

Tracking equilibrium curves

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

Theorem: Equilibrium tracking for contracting dynamics.

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

Proof of equilibrium tracking

Given: $\dot{x} = \mathsf{F}(x, \theta(t))$ with $\mathsf{osLip}_x(\mathsf{F}) \leq -c$ and $\mathsf{Lip}_u(\mathsf{F}) \leq \ell$

Task: compare **traj** $x(t)$ with **equilibrium curve** $x^*(\theta(t))$ implicitly defined by $\mathsf{F}(x, \theta(t)) = 0$

Consider **auxiliary dynamics** with two trajectories:

$$\dot{x} = \mathsf{F}(x, \theta(t)) + v(t) \quad =: \quad \mathsf{F}_{\mathsf{aux}}(x, \theta, v)$$

1. $v(t) = 0 \qquad \Longrightarrow \qquad$ trajectory $x(t)$
2. $v(t) = \dot{x}^*(\theta(t)) \qquad \Longrightarrow \qquad$ equilibrium trajectory $x^*(\theta(t))$

$\mathsf{F}_{\mathsf{aux}}$ is contracting with $\mathsf{osLip}_x(\mathsf{F}_{\mathsf{aux}}) \leq -c$ and $\mathsf{Lip}_v(\mathsf{F}_{\mathsf{aux}}) = 1.$ Hence, iISS:

$$\begin{aligned} D^+\|x(t) - x^\star(\theta(t))\|_{\mathcal{X}} &\leq -c \cdot \|x(t) - x^\star(\theta(t))\|_{\mathcal{X}} + 1 \cdot \|0 - \dot{x}^*(\theta(t))\|_{\mathcal{X}} \\ &\leq -c \cdot \|x(t) - x^\star(\theta(t))\|_{\mathcal{X}} + \frac{\ell}{c} \cdot \|\dot{\theta}(t)\|_{\Theta} \qquad \Big(\text{since } \mathsf{Lip}(x^*) = \frac{\ell}{c}\Big) \end{aligned}$$

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

Implications for set invariance

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

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

if $\|\theta_x(t) - \theta_y(t)\|_\Theta \leq \delta$ for all t,

then each trajectory $x(t) - y(t)$ approaches or remains inside $\left\{z \in \mathcal{X} \;\middle|\; \|z\|_{\mathcal{X}} \leq \frac{\ell\delta}{c}\right\}$

Theorem: Equilibrium tracking. Given trajectory $\theta(t)$, any solution $x(t)$ satisfies

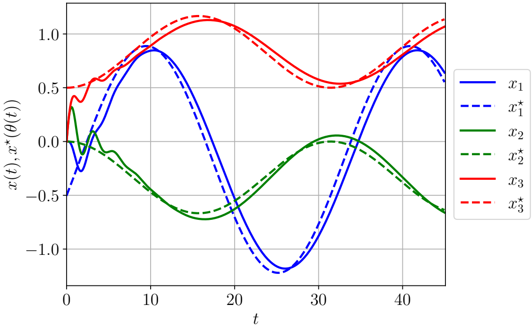
$$D^+\|x(t) - x^\star(\theta(t))\|_{\mathcal{X}} \;\leq\; -c\|x(t) - x^\star(\theta(t))\|_{\mathcal{X}} \;+\; \frac{\ell}{c}\|\dot{\theta}(t)\|_\Theta$$

if $\|\dot{\theta}(t)\|_\Theta \leq \delta$ for all t,

then each trajectory $x(t) - x^\star(\theta(t))$ approaches or remains inside $\left\{z \in \mathcal{X} \;\middle|\; \|z\|_{\mathcal{X}} \leq \frac{\ell\delta}{c^2}\right\}$

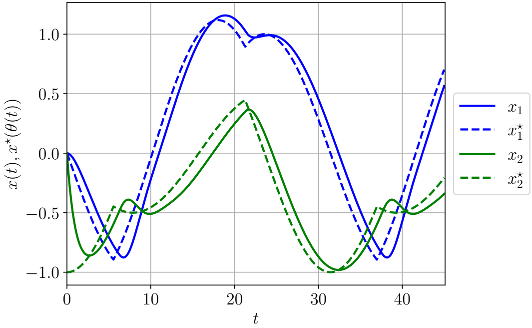
$$\min_{x\in\mathbb{R}^3} \quad \frac{1}{2}\|x - r(t)\|_2^2$$
$$\text{subj. to} \quad x_1 + 2x_2 + x_3 = \sin(\omega t),$$

$$r(t) = (\sin(\omega t), \cos(\omega t), 1), \omega = 0.2$$

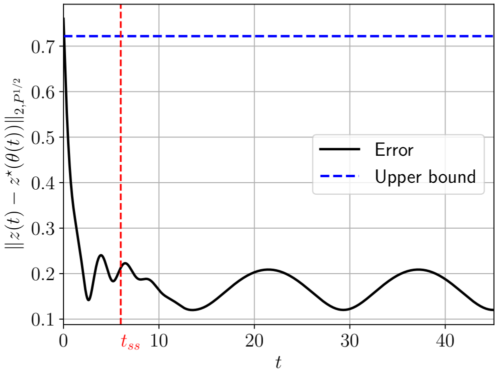


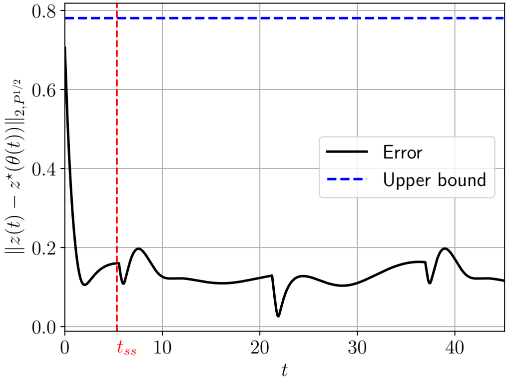
$$\min_{x\in\mathbb{R}^2} \quad \frac{1}{2}\|x + r(t)\|_2^2$$
$$\text{subj. to} \quad -x_1 + x_2 \leq \cos(\omega t),$$

$$r(t) = (\sin(\omega t), \cos(\omega t)), \omega = 0.2$$

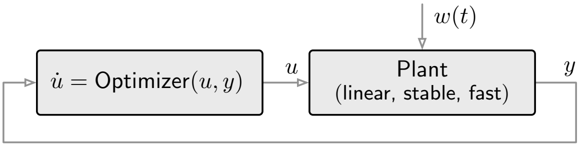


Empirical error versus theoretical upper bound





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. doi

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

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. doi

Some works on feedback optimization

Feedback optimization

Transportation systems

[Bianchin et al '20]

[Cothren et al'22]

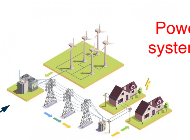
Robotics and vehicles

[Lawrence et al'21]

[Terpin et al '21]

[Cothren et al '22]

Power systems



[Jokic et al'09]

[Bolognani-Zampieri'13]

[Hirata-Hespanha-Uchida'14]

[Li et al'14]

[Dall'Anese et al'15]

[Bernstein et al'15]

[Gan-Low'16]

[Dall'Anese-Simonetto'18]

[Menta et al'18]

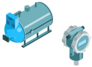
[Ortmann et al'20]

[Picallo et al'22]

… *and many others*

Compressor stations

[Zagorowska et al'23]



Epidemic control

[Bianchin et al'22]

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

Example #12: Gradient controller

Setup 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 problem

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

$$\begin{aligned} \min_{u} \quad & \phi(u) + \psi(y(t)) \qquad\qquad (\nu\text{-strongly convex } \phi, \text{ convex } \psi) \\ \text{subj to} \quad & y(t) = Y_u u + Y_w w(t) \end{aligned}$$

Gradient controller (as function of measured output):

$$\begin{cases} \dot{u}(t) = -\nabla\phi(u(t)) - Y_u^\top \nabla\psi(y(t)), \qquad u(0) = u_0 \\ \text{fast/stable LTI} \end{cases}$$

Example #12: Gradient controller

Equivalent rewriting 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

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

Equilibrium tracking for the gradient controller

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

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

Example #13: Projected gradient controller

Constrained feedback optimization:

$$\begin{aligned} \min_u \quad & \mathcal{E}(u,w) = \phi(u) + \psi(Y_u u + Y_w w) && (\nu \text{ strongly convex, } \ell_u \text{ strongly smooth, } \ell_w) \\ \text{subj. to} \quad & u \in \mathcal{U} && (\text{nonempty, closed, convex. } P_{\mathcal{U}} = \text{orthogonal projection}) \end{aligned}$$

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 $\gamma = \frac{2}{\nu+\ell_u}$,

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

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

Exact tracking with knowledge of external signal

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

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

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

Additionally, assume F differentiable in both arguments. Inverse function theorem implies

$$D_\theta x^\star(\theta) = -\big(D_x\mathsf{F}(x^\star(\theta), \theta)\big)^{-1} D_\theta\mathsf{F}(x^\star(\theta), \theta).$$

(To verify this equality, differentiate wrt θ the equilibrium equation $0 = \mathsf{F}(x^\star(\theta), \theta)$.)

time-varying contracting dynamics with feedforward prediction

$$\dot{x}(t) = \mathsf{F}(x(t), \theta(t)) - \big(D_x\mathsf{F}(x(t), \theta(t))\big)^{-1} D_\theta\mathsf{F}(x(t), \theta(t))\, \dot{\theta}(t)$$

For example, if $F = -\nabla_x f$:

$$\dot{x} = -\nabla_x f(x, \theta) + \big(\mathrm{Hess}\, f(x, \theta)\big)^{-1} D_\theta \nabla_x f(x, \theta)\, \dot{\theta}$$

Exact tracking with knowledge of external signal

Time-varying contracting dynamics with feedforward prediction

$$\dot{x}(t) = \mathsf{F}(x(t), \theta(t)) - \big(D_x\mathsf{F}(x(t), \theta(t))\big)^{-1} D_\theta \mathsf{F}(x(t), \theta(t))\, \dot{\theta}(t)$$

Asymptotically exact equilibrium tracking

1. $\|\mathsf{F}\big(x(t), \theta(t)\big)\| \leq \mathrm{e}^{-ct} \|\mathsf{F}\big(x(0), \theta(0)\big)\|$
2. $\|x(t) - x^\star(\theta(t))\| \leq \dfrac{\ell}{c} \mathrm{e}^{-ct} \|x(0) - x^\star(\theta(0))\|$

Proof sketch

First compute

$$\frac{d}{dt}\mathsf{F}(x(t), \theta(t)) = D_x\mathsf{F}(x, \theta)\dot{x} + D_\theta\mathsf{F}(x, \theta)\dot{\theta} = D_x\mathsf{F}(x, \theta)\mathsf{F}(x, \theta)$$

and so

$$\|\mathsf{F}(x(t), \theta(t))\| D^+ \|\mathsf{F}(x(t), \theta(t))\| = \left[\!\left[\frac{d}{dt}\mathsf{F}(x(t), \theta(t)), \mathsf{F}(x(t), \theta(t)) \right]\!\right] \leq \cdots \leq -c\|\mathsf{F}(x(t), \theta(t))\|^2$$

Separately,

$$c\|x - x^*\| \leq \|F(x)\| \leq \ell\|x - x^*\|$$

Summary and future work on time-varying and feedback optimization

Summary:

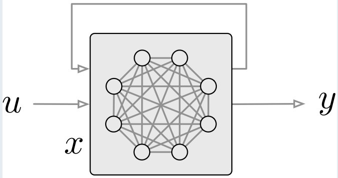
1. from convex optimization to contracting dynamics
2. tracking-bounds for time-varying contracting systems
3. applications to convex optimization and feedback optimization

Ongoing work and open problems:

1. contracting predictor-corrector methods
2. tracking bounds in time-varying norms
3. convex but not strongly convex problems

Outline

ℓ_∞-contracting neural networks



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

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

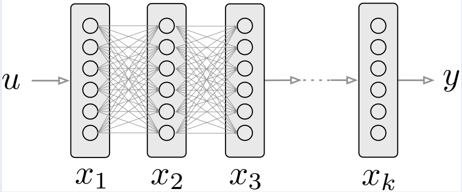
$$x_{k+1} = (1-\alpha)x_k + \alpha\Phi(Ax_k + Bu + b) \qquad \text{(\textit{Euler discretization})}$$

If

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

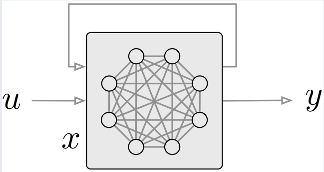
- recurrent NN is contracting with rate $1 - \mu_\infty(A)_+$
- implicit NN is well posed
- Euler discretization is contracting with factor $1 - \dfrac{1-\mu_\infty(A)_+}{1-\min_i(a_{ii})_-}$ at $\alpha^* = \dfrac{1}{1-\min_i(a_{ii})_-}$
- input-state Lipschitz constant $\mathrm{Lip}_{u\to x} = \dfrac{\|B\|_\infty}{1-\mu_\infty(A)_+}$

Feedforward NN



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

Implicit/Recurrent NN



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

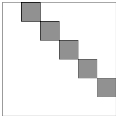
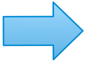
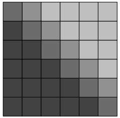
ML advantages of implicit/equilibrium/fixed point formulation:

1. bio-inspired
2. expressivity and ability to model I/O behavior, instead of modalities
3. simplicity and memory efficiency
4. accuracy
5. input-output robustness

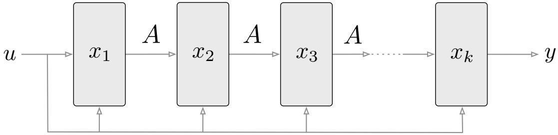
S. Jafarpour, A. Davydov, A. V. Proskurnikov, and F. Bullo. Robust implicit networks via non-Euclidean contractions. In *Advances in Neural Information Processing Systems*, Dec. 2021.

Motivation #1: Generalizing FF to fully-connected synaptic matrices

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

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

Motivation #2: Weight-tied infinite-depth NN $\rightarrow$ fixed-point of INN



$$x^{i+1} = \Phi(Ax^i + Bu + b) \quad \Longrightarrow \quad \lim_{i\to\infty} x^i = x^* \text{ solution to the INN}$$

Recent literature on implicit NNs

1. S. Bai, J. Z. Kolter, and V. Koltun. Deep equilibrium models. In *Advances in Neural Information Processing Systems*, 2019. URL **https://arxiv.org/abs/1909.01377**
2. L. El Ghaoui, F. Gu, B. Travacca, A. Askari, and A. Tsai. Implicit deep learning. *SIAM Journal on Mathematics of Data Science*, 3(3):930–958, 2021. doi
3. E. Winston and J. Z. Kolter. Monotone operator equilibrium networks. In *Advances in Neural Information Processing Systems*, 2020. URL **https://arxiv.org/abs/2006.08591**
4. M. Revay, R. Wang, and I. R. Manchester. Lipschitz bounded equilibrium networks. *arXiv preprint arXiv:2010.01732*, 2020. doi
5. A. Kag, Z. Zhang, and V. Saligrama. RNNs incrementally evolving on an equilibrium manifold: A panacea for vanishing and exploding gradients? In *International Conference on Learning Representations*, 2020. URL **https://openreview.net/forum?id=HylpqA4FwS**
6. K. Kawaguchi. On the theory of implicit deep learning: Global convergence with implicit layers. In *International Conference on Learning Representations*, 2021. URL **https://openreview.net/forum?id=p-NZIuwqhI4**
7. S. W. Fung, H. Heaton, Q. Li, D. McKenzie, S. Osher, and W. Yin. Fixed point networks: Implicit depth models with Jacobian-free backprop, 2021. URL **https://arxiv.org/abs/2103.12803**. ArXiv e-print

Implicit Neural Networks (INNs) for regression

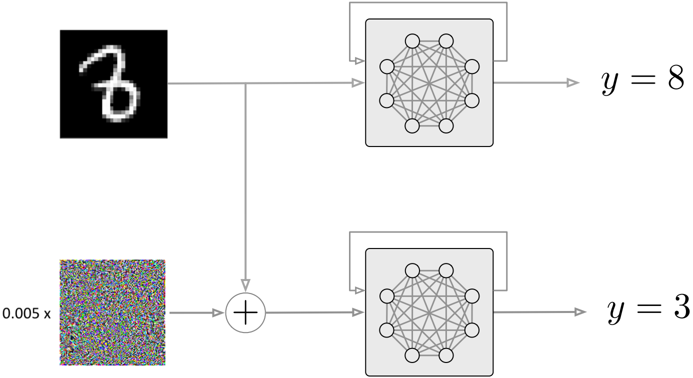
- Training INNs:
  1. loss function $\mathcal{L}$
  2. training data $(\widehat{u}_i, \widehat{y}_i)_{i=1}^N$
  3. **training optimization problem**

$$\min_{A,B,C,b,x} \quad \sum_{i=1}^{N} \mathcal{L}(\widehat{y}_i, Cx_i + c)$$
$$x_i = \Phi(Ax_i + B\widehat{u}_i + b)$$

- Efficient back-propagation through implicit differentiation
- Stochastic gradient descent: at each step solve $x = \Phi(Ax + Bu + b)$.

Robustness of INNs

Adversarial examples: small input change can cause large output change!

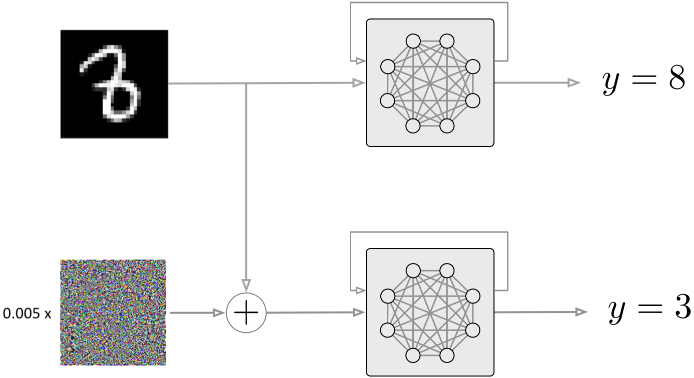


Robustness measures: **input-output Lipschitz constant**

1. ℓ_2-**norm Lipschitz constant**: not informative in many scenarios
2. ℓ_∞-**norm Lipschitz constant**: large-scale input wrt wide-spread perturbations

Robustness of INNs

Adversarial examples: small input change can cause large output change!



Robustness measures: **input-output Lipschitz constants**

1. NP-hard to compute exactly
2. Approximations provide only coarse certified robustness guarantees

Training INNs

Training optimization problem:

$$\min_{A,B,C,b} \quad \sum_{i=1}^{N} \mathcal{L}(\widehat{y}_i, Cx_i + c) + \quad \lambda \quad \mathsf{Lip}_{u\to y}$$

$$x_i = \Phi(Ax_i + B\widehat{u}_i + b)$$

$$\mu_\infty(A) \leq \gamma$$

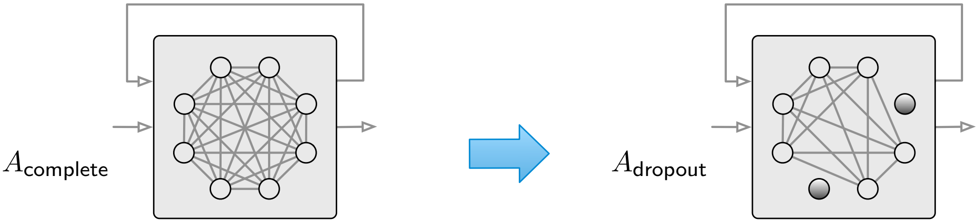
- $\lambda \geq 0$ is a regularization parameter
- $\gamma < 1$ is a hyperparameter

Parametrization of μ_∞ constraint:

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

Graph-Theoretic Regularization

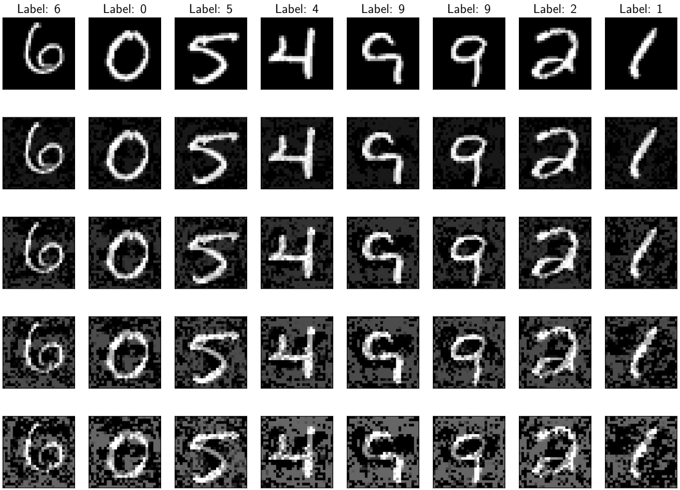
Synaptic matrix A encodes interactions between neurons



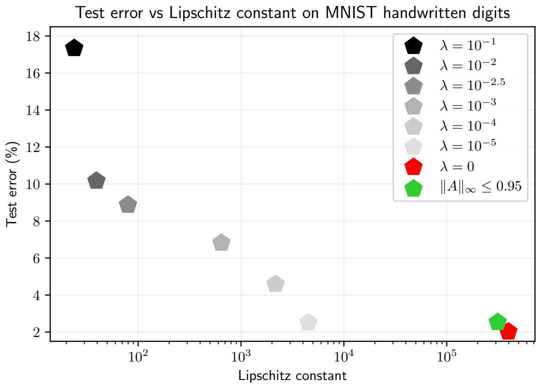
- A_{dropout} is a principal submatrix of A_{complete}
- $\mu_\infty(A_{\text{dropout}}) \leq \mu_\infty(A_{\text{complete}})$
  - Well-posedness of original INN implies well-posedness of INN with subset of neurons
  - Promotes *compression* and *sparsity* of overparametrized models

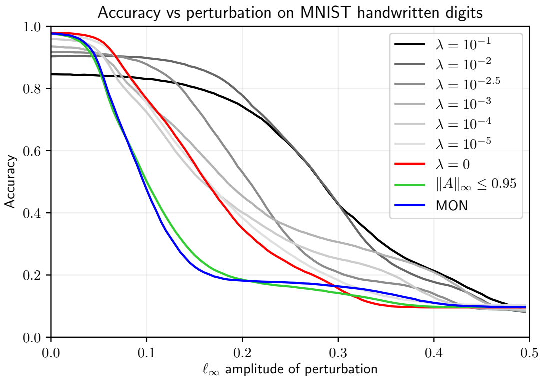
Numerical Experiments

- MNIST handwritten digit dataset (60K+10K, 28x28, grayscale)
- implicit neural network order: $n = 100$



Tradeoff between **accuracy** and **robustness**





- Pareto-optimal curve
- Clean performance vs. robustness

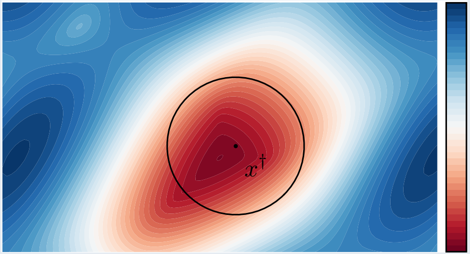
Outline

Outline

Local contractivity theory

$$\dot{x}(t) = \mathsf{F}\big(x(t)\big) \qquad \text{and} \qquad x(k+1) = \mathsf{F}\big(x(k)\big)$$

for a norm $\|\cdot\|$, recall $\mathsf{Lip}(\mathsf{F}) = \sup_x \|D\mathsf{F}(x)\|$ and $\mathsf{osLip}(\mathsf{F}) = \sup_x \mu(D\mathsf{F}(x))$



Example contour plot of $x \mapsto \mu(D\mathsf{F}(x))$

Red values are points x where $\mu(D\mathsf{F}(x)) < 0$

Blue values are points where $\mu(D\mathsf{F}(x)) > 0$

contracting set $S :=$ red region

closed ball $\overline{B}_r(x^*) = \{x \text{ such that } \|x - x^*\| \leq r\}$

Theorem: if contracting region S is invariant and convex (so that $\mathsf{Lip}(\mathsf{F}) = \sup_x \|D\mathsf{F}(x)\|$),
then one can restrict F to S and usual contractivity properties (with caveats) apply

1. invariance of contracting set S?
2. convexity of contracting set S?

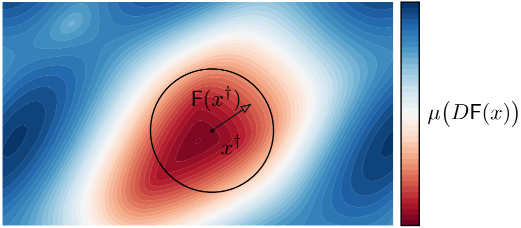
Preliminary equilibrium lemma

If $x^* \in S$ satisfies $\mathsf{F}(x^*) = \mathbb{0}_n$, then each $\overline{B}_r(x^*) \subset S$ is invariant for $\dot{x} = \mathsf{F}(x)$.
If $x^* \in S$ satisfies $\mathsf{F}(x^*) = x^*$, then each $\overline{B}_r(x^*) \subset S$ is invariant for $x(k+1) = \mathsf{F}(x(k))$.

Proof: $x \mapsto \|x - x^*\|$ is a Lyapunov function decreasing along the flow

Hence, we may look for largest equilibrium-centered ball inside S. However,

1. as the ball grows inside S, the contraction rate (i.e., $\sup_{x \in \text{ball}} \mu(D\mathsf{F}(x))$) goes to zero
2. what if an equilibrium point is not known?



The small-residual theorem (continuous time)

For $\dot{x} = \mathsf{F}(x)$ infinitesimally contracting with rate $c > 0$ in region S

$$\overline{B}_r(x^*) \subset S \quad \text{and} \quad \|\mathsf{F}(x^*)\| \leq cr \quad \Longrightarrow \quad \overline{B}_r(x^*) \text{ is invariant}$$

The small-residual theorem (discrete time)

For $x(k+1) = \mathsf{F}(x(k))$ contracting with factor $\ell < 1$ in region S

$$\overline{B}_r(x^*) \subset S \quad \text{and} \quad \|\mathsf{F}(x^*) - x^*\| \leq r(1-\ell) \quad \Longrightarrow \quad \overline{B}_r(x^*) \text{ is invariant}$$

Proof of small-residual theorem (discrete time): Pick $x \in \overline{B}_r(x^*)$ and compute

$$\|\mathsf{F}(x) - x^*\| \overset{\text{triangle ineq}}{\leq} \|\mathsf{F}(x) - \mathsf{F}(x^*)\| + \|\mathsf{F}(x^*) - x^*\|$$

where $\|\mathsf{F}(x) - \mathsf{F}(x^*)\| \;\leq\; \ell\|x - x^*\| \;\leq\; \ell r$ by contractivity on S and by $x \in \overline{B}_r(x^*)$
where $\|\mathsf{F}(x^*) - x^*\| \;\leq\; (1-\ell)r$ by small residual

$$\|\mathsf{F}(x) - x^*\| \leq \ell r + (1-\ell)r = r \qquad \Longrightarrow \qquad \mathsf{F}(x) \in \overline{B}_r(x^*)$$

Proof of small-residual theorem (continuous time): Pick $x \in \partial\overline{B}_r(x^*)$ and compute

$$\|\phi_t(x) - x^*\| \overset{\text{triangle ineq}}{\leq} \|\phi_t(x) - \phi_t(x^*)\| + \|\phi_t(x^*) - x^*\| \qquad \text{(equality when } t=0)$$

where

$$D^+\big|_{t=0}\|\phi_t(x) - \phi_t(x^*)\| \leq -c\|\phi_t(x) - \phi_t(x^*)\|\big|_{t=0} = -cr \qquad \text{(contractivity)}$$

$$D^+\big|_{t=0}\tfrac{1}{2}\|\phi_t(x^*) - x^*\|^2 = [\![\mathsf{F}(\phi_t(x^*)), \phi_t(x^*) - x^*]\!]\big|_{t=0} \leq \|\mathsf{F}(x^*)\|\,\|\phi_t(x^*) - x^*\| \qquad \text{(c.n.d)}$$

$$\Longrightarrow \qquad D^+\big|_{t=0}\|\phi_t(x) - x^*\| \leq -cr + \|\mathsf{F}(x^*)\|.$$

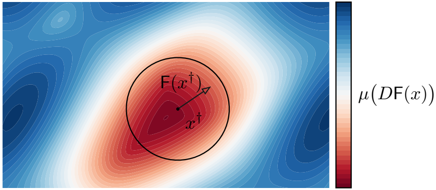
If a continuous h satisfies $D^+\big|_{t=0}h(t) \leq 0$, then $h(t) \leq h(0)$ for small t.
Invariance follows from Nagumo's Theorem.

Local contractivity

Given a norm $\|\cdot\|$ and a set S, consider

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

satisfying $\sup_{x\in S} \mu(D\mathsf{F}(x)) \leq -c < 0$



trajectories slow down and
approach each other while inside S

1. **integral and differential conditions do not coincide**
   In general $\mathrm{osLip}(\mathsf{F}|_S) \geq \sup\{\mu(D\mathsf{F}(x))$ s.t. $x \in S\}$, with equality when S is convex
2. **x^* exists if "residual is below threshold"**
   if $\exists$ a closed ball with center $\bar{x}$ and radius $r > 0$ inside S such that $\|\mathsf{F}(\bar{x})\| \leq cr$,
   then ball is F-invariant and contains a unique exponentially stable equilibrium x^*
3. **x^* exists if complete trajectory in set**
   if $\exists\ \phi_t(x_0) \in S$ for all $t \geq 0$, then $x^* := \lim_{t\to+\infty} \phi_t(x_0) \in S$ is an equilibrium
4. **there exists either 0 or 1 equilibrium x^* in each convex subset of S**
   each convex subset of S possesses 0 or 1 equilibrium

Local contractivity near each Hurwitz equilibrium

Consider a continuously-differentiable F with an equilibrium x^* such that $D\mathsf{F}(x^*)$ is Hurwitz. Pick a sufficiently small $\epsilon > 0$ and compute $P = P^\top \succ 0$ such that

$$\mu_{2,P^{1/2}}(D\mathsf{F}(x^*)) \leq \alpha(D\mathsf{F}(x^*)) + \epsilon$$

Then

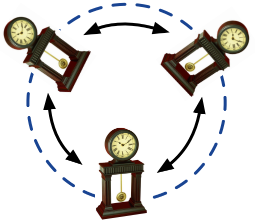
1. by the continuity of $D\mathsf{F}$, there exists a radius $r > 0$ such that

$$\mu_{2,P^{1/2}}(D\mathsf{F}(x)) < 0$$

   in a ball of radius r centered at x^* with respect to the norm $\|\cdot\|_{2,P^{1/2}}$
2. each trajectory starting inside this ball converges to x^*

Pendulum clocks & *"an odd kind of sympathy"*

[Christiaan Huygens, Horologium Oscillatorium, 1673]



Canonical coupled oscillator model

[Arthur Winfree '67, Yoshiki Kuramoto '75]

[find on youtube 2015 remarks by "Kuramoto talks about Kuramoto model"]

Kuramoto model

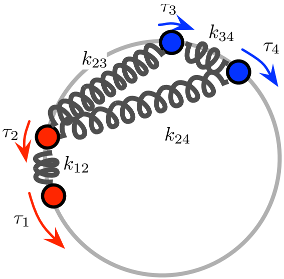
- n **oscillators** with angle $\theta_i \in \mathbb{S} = \mathbb{T}$
- **natural frequencies** $\omega_i \in \mathbb{R}$
- **coupling strengths** $a_{ij} = a_{ji}$

$$\dot{\theta}_i = \omega_i - \sum_{j=1}^{n} a_{ij} \sin(\theta_i - \theta_j)$$

$$\omega_i = \sum_{j=1}^{n} a_{ij} \sin(\theta_i - \theta_j)$$

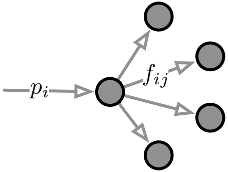
Spring network

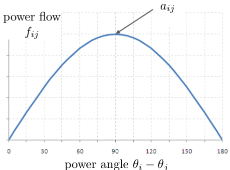
- $\omega_i = \tau_i$: torque at i
- $a_{ij} = k_{ij}$: spring stiffness i, j
- $\sin(\theta_i - \theta_j)$: modulation
- elastic energy
  $\mathcal{E} = \sum_{ij}(1 - \cos(\theta_i - \theta_j))$

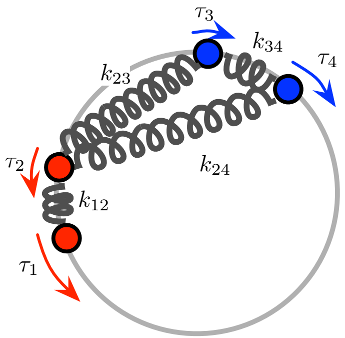


Power network

- $\omega_i = p_i$: injected power
- a_{ij} : max power flow i, j
- $\sin(\theta_i - \theta_j)$: modulation
- KCL flow conservation and Ohm's law







Coupled swing equations

Euler-Lagrange eq for spring network on ring:

$$m_i\ddot{\theta}_i + d_i\dot{\theta}_i = \tau_i - \sum\nolimits_j k_{ij}\sin(\theta_i - \theta_j)$$

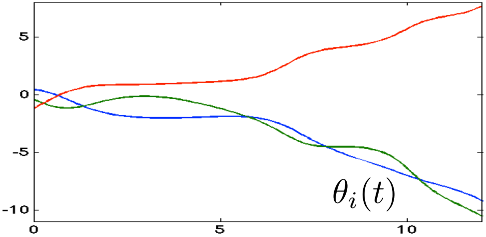
Kuramoto coupled oscillators

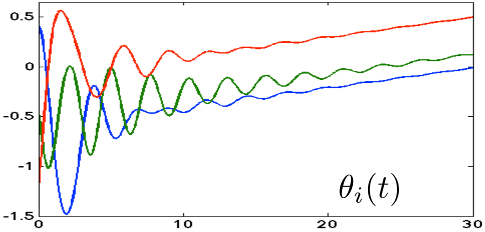
$$\dot{\theta}_i = \omega_i - \sum\nolimits_j a_{ij}\sin(\theta_i - \theta_j)$$

Kuramoto equilibrium equation

$$0 = \omega_i - \sum\nolimits_j a_{ij}\sin(\theta_i - \theta_j)$$

Incoherence or synchronization?





Frequency sync: $\dot{\theta}_i = \dot{\theta}_j$
Phase sync: $\theta_i = \theta_j$

Preliminary observations

1. For $\alpha \in [-\pi, \pi[$ and $\theta = (\theta_1, \dots, \theta_n) \in \mathbb{T}^n$, $\mathrm{rot}_\alpha(\theta)$ is counterclockwise rotation of each entry $(\theta_1, \dots, \theta_n)$ by α
   Kuramoto model *invariant under rotations*: rotated solutions are solutions
   $\Longrightarrow$ system can be written in $n-1$ angle differences (so that it is really $n-1$ dim)

2. Note $\sum_i \dot{\theta}_i = \sum_i \omega_i$. Define $\omega_{\text{sync}} := \frac{1}{n} \sum_{i=1}^{n} \omega_i = \mathrm{average}(\omega)$ and change reference frame to rotating frame with ω_{sync}.
   $\Longrightarrow$ restrict to $\omega_{\text{sync}} = 0 \iff \mathbb{1}_n^\top \omega = 0$

3. Let B denote directed incidence matrix. Jacobian of the Kuramoto model is:

$$J(\theta) = -B \operatorname{diag}(\{a_{ij} \cos(\theta_i - \theta_j)\}_{\{i,j\} \in E}) B^\top$$

   $\Longrightarrow$ $J(\theta) = -\text{Laplacian}(\theta)$, but weights $a_{ij} \cos(\theta_i - \theta_j)$ may be negative

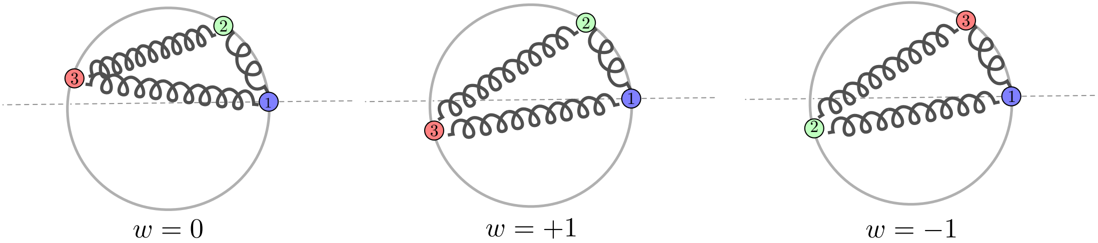
4. define **phase cohesive subset** $\{\theta \in \mathbb{T}^n$ such that $|\theta_i - \theta_j| \le \pi/2$, for all edges $\}$
   $\Longrightarrow$ $J(\theta) \preceq 0$ on phase cohesive θ

Winding numbers and partitions

Given a cycle $\sigma = (1, \ldots, n_\sigma)$ and orientation

1. **winding number of $\theta \in \mathbb{T}^n$ along σ**

   $=$ number of times the **shortest-arc path wraps around torus**



2. given basis $\sigma_1, \ldots, \sigma_r$ for cycles, **winding vector of θ** is

$$w(\theta) = (w_{\sigma_1}(\theta), \ldots, w_{\sigma_r}(\theta))$$

Theorem: Kirchhoff angle law on $\mathbb{T}^n$

$$\text{winding number is at most } \pm \lfloor n_\sigma/2 \rfloor - 1$$

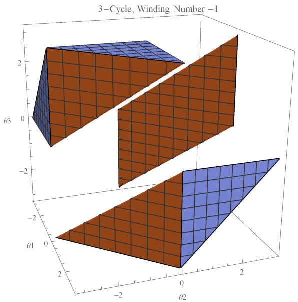
Theorem: Winding partition For each possible winding vector u, define

$$\mathsf{WindingCell}(u) := \{\theta \in \mathbb{T}^n \text{ such that } w(\theta) = u\}$$

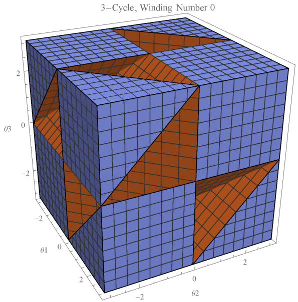
Then

$$\mathbb{T}^n = \cup_u \mathsf{WindingCell}(u)$$

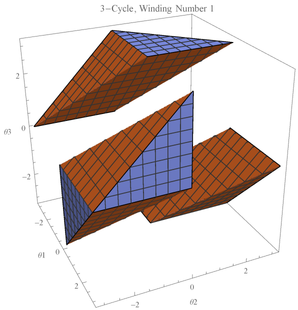
Winding partition: example and properties



$w = -1$



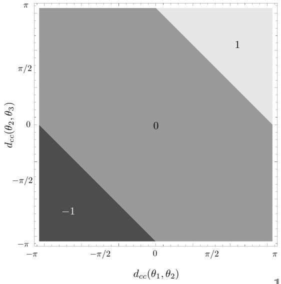
$w = 0$



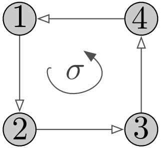
$w = +1$

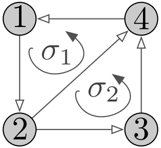
Theorem: Reduced cell is convex polytope

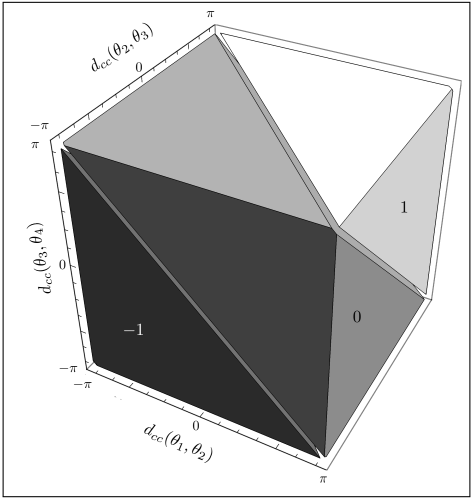
- each winding cell is connected and invariant under rotation
- **bijection:**
  reduced winding cell $\longleftrightarrow$ open convex polytope

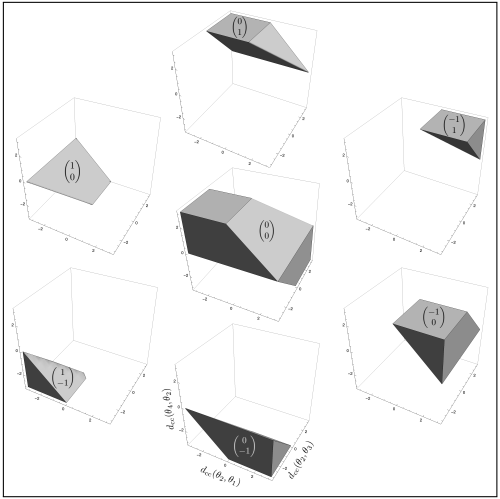


Two other examples

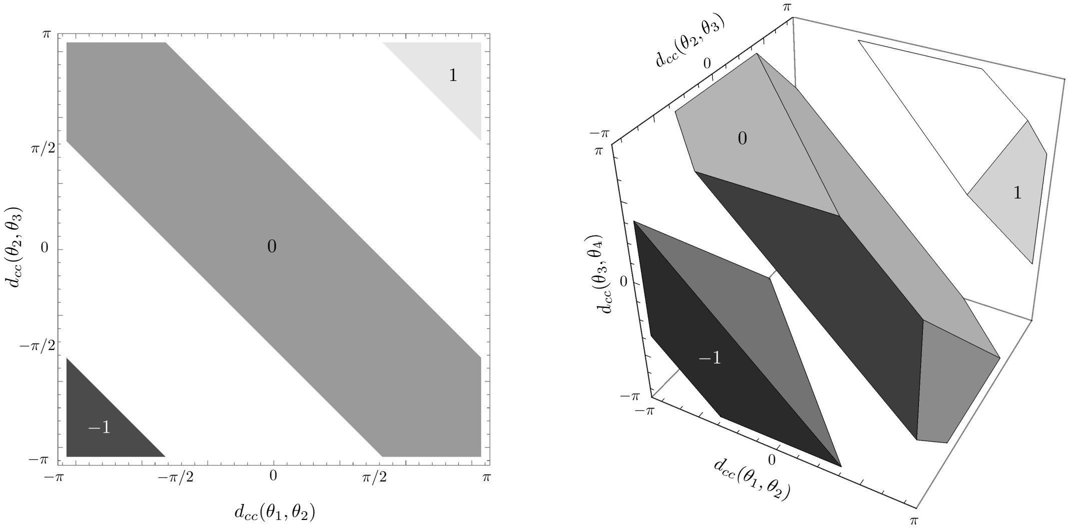








Phase cohesive winding cells



(a) (b)

cohesive subset $|\theta_i - \theta_j| \leq \pi/2$

Multistable Sync = global partition + local contraction

$$\dot{\theta}_i = \omega_i - \sum_{j=1}^{n} a_{ij} \sin(\theta_i - \theta_j)$$

1. in each winding cell the energy is well posed:

$$\mathcal{E}(\theta) = \sum_{ij} (1 - \cos(\theta_i - \theta_j)) + \omega^\top \theta$$

2. in each winding cell Kuramoto model is precisely: $\dot{\theta} = -\nabla \mathcal{E}(\theta)$
3. $\mathrm{Hess}\,\mathcal{E}(\theta) = \mathrm{Hess} \sum_{ij} (1 - \cos(\theta_i - \theta_j)) = -\text{Laplacian}(\theta)$ (possibly negative weights)
4. $\mathrm{Hess}\,\mathcal{E}(\theta) \preceq 0$ on the **cohesive subset** $|\theta_i - \theta_j| \leq \pi/2$

   hence, modulo the symmetry, $\mathcal{E}$ is strongly convex on cohesive subset
5. modulo the symmetry, local strong contractivity (on each connected cohesive subset)

At most uniqueness theorem:

1. each winding cell has at most one cohesive equilibrium
2. contraction algorithm to decide/compute in each winding cell

S. Jafarpour, E. Y. Huang, K. D. Smith, and F. Bullo. Flow and elastic networks on the n-torus: Geometry, analysis and computation. *SIAM Review*, 64(1):59–104, 2022b.

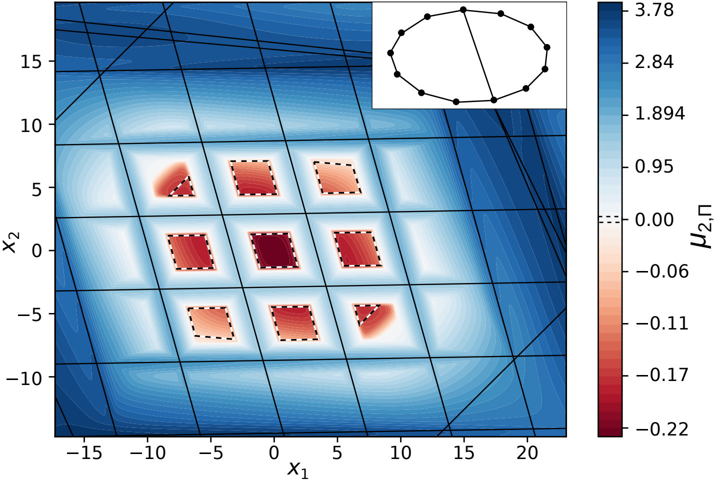
Extentions: the Kuramoto-Sakaguchi model with phase delays

$$\dot{\theta}_i = \omega_i - \sum_{j=1}^{n} a_{ij} \sin(\theta_i - \theta_j + \phi_{ij})$$

same properties, by robustness of contracting dynamics

R. Delabays and F. Bullo. Semicontraction and synchronization of Kuramoto-Sakaguchi oscillator networks. *IEEE Control Systems Letters*, 7:1566–1571, 2023.

Two-dimensional slice of $\mathbb{R}^{13}$, showing log seminorm of Jacobian of Kuramoto-Sakaguchi model with delay $\varphi = 0.01$.



plain black lines = boundaries of the winding cells.

dashed black lines = boundaries of the $\bar{\gamma}$-cohesive winding cells

red color $\implies$ system is semicontracting in phase-cohesive winding cells

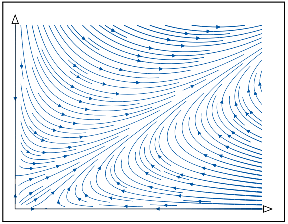
Outline

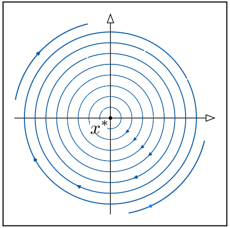
Given a norm $\|\cdot\|$, consider

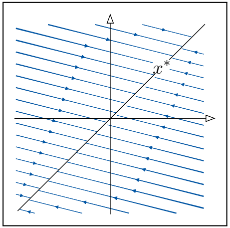
$$\dot{x} = \mathsf{F}(x) \qquad \text{satisfying} \qquad \mathrm{osLip}(\mathsf{F}) = 0$$

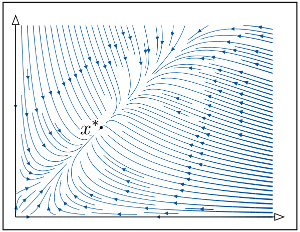
Dichotomy for weakly-contracting systems

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









Weakly contracting dynamics + locally-exp-stable equilibrium

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

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

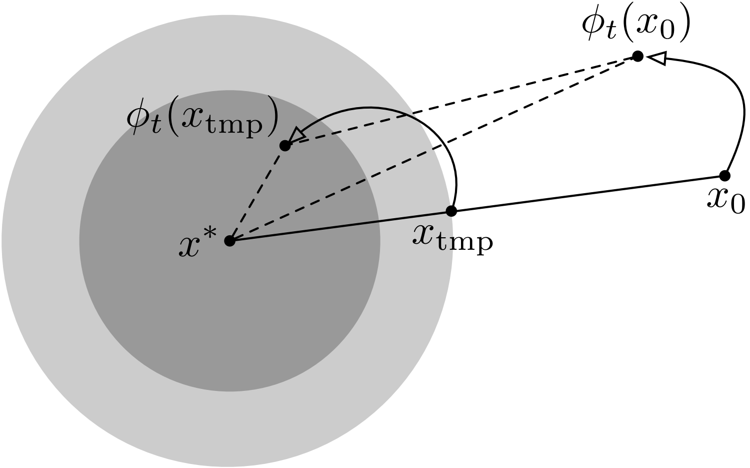
   $\implies$ F is locally c-strongly contracting wrt $\|\cdot\|_{\mathsf{loc}}$ over forward-invariant $\mathcal{S}$

   $\implies$ exists $\mathcal{B}_{\mathsf{glo}} = \{x \mid \|x - x^\star\|_{\mathsf{glo}} \leq r\} \subset \mathcal{S}$

Equivalently:

1. F is globally weakly contracting wrt $\|\cdot\|_{\mathsf{glo}}$
2. F is locally strongly contracting wrt $\|\cdot\|_{\mathsf{loc}}$ in $\mathcal{S}$
3. equilibrium point in $\mathcal{S}$

Proof of globally-weakly + locally-strongly



1. **finite decay in finite time**: For each $x(0) \notin \mathcal{S}$ and each $\rho < 1$,

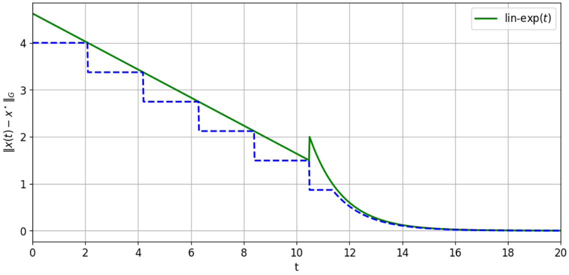
$$\|x(t_\rho) - x^\star\|_{\text{glo}} \leq \|x(0) - x^\star\|_{\text{glo}} - \rho r \qquad \text{for } t_\rho = \ln(\kappa_{\text{loc,glo}}(1-\rho)^{-1})c^{-1}$$

$\implies$ *average linear decay rate* $\qquad c_{\text{ld}} = \rho r / t_\rho$

$\implies$ $x(t) \in \mathcal{B}_{\text{glo}}$ after *linear decay time* $\qquad t_{\text{ld}} = \left\lceil \frac{\|x(0)-x^\star\|_{\text{glo}} - r}{\rho r} \right\rceil t_\rho$

2. **linear exponential decay**:

$$\|x(t) - x^\star\|_{\text{glo}} \leq \begin{cases} \left(\|x(0) - x^\star\|_{\text{glo}} + \rho r\right) - c_{\text{ld}} t & \text{if } t \leq t_{\text{ld}} \\ \kappa_{\text{loc,glo}}\, r\, \mathrm{e}^{-c(t-t_{\text{ld}})} & \text{if } t > t_{\text{ld}} \end{cases}$$



Example #13: 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

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

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

Example #14: Biologically-plausible circuits for sparse reconstruction

Φ dictionary matrix:

- full row rank, each columm Φ_i has unit norm
- $\Phi_i \cdot \Phi_j$ = similarity between dictionary elements

$$\underset{(M\times 1)}{u} \approx \underset{(M\times N)}{\Phi}\ \underset{(N\times 1)}{x} = \underset{(M\times N)}{\left[\Phi_1 \mid \Phi_2 \mid \cdots \mid \Phi_N\right]}\ \underset{(N\times 1)}{x}\,, \qquad \underbrace{\Phi^\top \Phi}_{\text{rank at most } M} = \underset{(N\times M)}{\Phi^\top}\ \underset{(M\times N)}{\Phi} = \underset{(N\times N)}{(\Phi^\top\Phi)_{ij} = \Phi_i^\top \Phi_j}$$

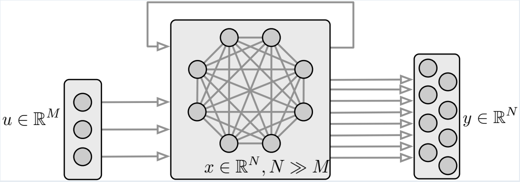
Sparse reconstruction:

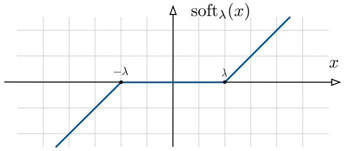
$$\min_{x\in\mathbb{R}^N} \quad \mathcal{E}_{\text{lasso}}(x) := \frac{1}{2}\|u - \Phi x\|_2^2 + \lambda\|x\|_1$$

Competitive neural network

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

or, in components $\quad \dot{x}_i(t) = -x_i + \text{soft}_\lambda\big(-\sum_{j=1,j\neq i}^{n} \Phi_i^\top \Phi_j x_j + \Phi_i^\top u\big)$





Equilibria, weak contractivity and convergence of $\mathsf{F}_{\text{competitive}}$

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

$\implies$ each trajectory linearly-exponentially-decays to $x^\star$

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

Contraction dynamics on Riemannian manifolds

Contraction theory on Riemannian manifolds originates in

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

A formal coordinate-free analysis (with connection to monotone operators) is given in

J. W. Simpson-Porco and F. Bullo. Contraction theory on Riemannian manifolds. *Systems & Control Letters*, 65:74–80, 2014.

In the differential geometry literature, geodesically monotonic vector fields are studied by

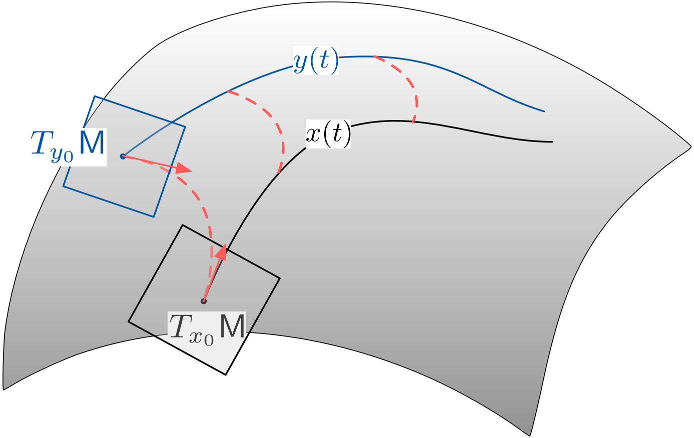
S. Z. Németh. Geodesic monotone vector fields. *Lobachevskii Journal of Mathematics*, 5:13–28, 1999. URL **http://mi.mathnet.ru/eng/ljm145**

J. X. Da Cruz Neto, O. P. Ferreira, and L. R. Lucambio Pérez. Contributions to the study of monotone vector fields. *Acta Mathematica Hungarica*, 94(4):307–320, 2002.

J. H. Wang, G. López, V. Martín-Márquez, and C. Li. Monotone and accretive vector fields on Riemannian manifolds. *Journal of Optimization Theory and Applications*, 146(3):691–708, 2010.

Assume: existence and uniqueness of geodesic curve $\gamma(t) = x \#_t y$ between each (x, y)

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



integral test: the inner product between F and the geodesic velocity vector γ' at x and y

differential test: condition on covariant differential of F

Given vector field F on a Riemannian manifold $(\mathsf{M}, \mathbb{G})$ and $c > 0$, equivalent statements:

1. **integral condition:** for each $x, y \in \mathsf{M}$ and geodesic $\gamma : [0,1] \to \mathsf{M}$ with $\gamma(0) = x$, $\gamma(1) = y$,

$$\langle\!\langle \mathsf{F}(y), \gamma'(1) \rangle\!\rangle_{\mathbb{G}} - \langle\!\langle \mathsf{F}(x), \gamma'(0) \rangle\!\rangle_{\mathbb{G}} \leq -c\, \mathrm{d}_{\mathbb{G}}(x,y)^2$$

   or, equivalently, using the parallel transport map $P_{y \to x} : T_y\mathsf{M} \to T_x\mathsf{M}$,

$$\langle\!\langle P_{y\to x}\mathsf{F}(y) - \mathsf{F}(x), \gamma'(0) \rangle\!\rangle_{\mathbb{G}} \leq -c\, \mathrm{d}_{\mathbb{G}}(x,y)^2$$

2. **differential condition:** for all $v_x \in T_x\mathsf{M}$

$$\langle\!\langle \nabla_{v_x}\mathsf{F}(x), v_x \rangle\!\rangle_{\mathbb{G}} \leq -c \|v_x\|_{\mathbb{G}}^2,$$

   where $\nabla\mathsf{F}$ is covariant derivative. In components, generalized Demidovich condition:

$$\mathbb{G}(x) D\mathsf{F}(x) + D\mathsf{F}(x)^\top \mathbb{G}(x) + \mathcal{L}_{\mathsf{F}}\mathbb{G}(x) \preceq -2c\mathbb{G}(x)$$

3. **trajectory condition:** for all solutions $x(\cdot), y(\cdot)$

$$D^+ \mathrm{d}_{\mathbb{G}}(x(t), y(t)) \leq -c\, \mathrm{d}_{\mathbb{G}}(x(t), y(t))$$

Given Riemannian manifold $(\mathsf{M}, \mathbb{G})$,
a function $f : \mathsf{M} \to \mathbb{R}$ is **ν-strongly geodesically convex** if, for each x, y,

1. $f\big(x \#_t y\big) \leq (1-\chi) f(x) + \chi f(y) - \frac{1}{2}\nu\chi(1-\chi)\mathrm{d}_{\mathbb{G}}(x,y)^2$
2. (if f is twice differentiable) $\mathrm{Hess}\, f(x) \succeq \nu\mathbb{G}(x)$

natural gradient dynamics

$$\dot{x} = \mathsf{F}_{\mathsf{G}}(x) := -\mathbb{G}(x)^{-1}\nabla f(x)$$

F_{G} is infinitesimally contracting wrt $\mathbb{G}$ with rate ν
unique globally exp stable point is global minimum

Let $f : \mathbb{R}^n \to \mathbb{R}$ smooth, strongly convex
natural gradient on $(\mathbb{R}^n, \mathrm{Hess}(f))$ = Newton's continuous-time method
infinitesimally contracting with rate 1

J. W. Simpson-Porco and F. Bullo. Contraction theory on Riemannian manifolds. *Systems & Control Letters*, 65:74–80, 2014.

Example #16: Rosenbrock function

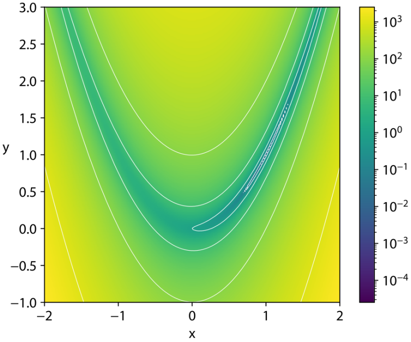
$$f_{\mathsf{Rsnbrck}}(x_1, x_2) = 100(x_1^2 - x_2)^2 + (x_1 - 1)^2$$

is 2-strongly geodesically convex wrt

$$\mathbb{G}(x) = \begin{bmatrix} 400x_1^2 + 1 & -200x_1 \\ -200x_1 & 100 \end{bmatrix}$$

and natural gradient is 2-strongly contracting

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = -\mathbb{G}(x)^{-1} \nabla f_{\mathsf{Rsnbrck}} = -2 \begin{bmatrix} x_1 - 1 \\ x_1^2 - 2x_1 + x_2 \end{bmatrix}$$



contour plot for f_{Rsnbrck}
long, shallow parabolic valley
global minimum $(1, 1)$

P. M. Wensing and J.-J. E. Slotine. Beyond convexity — Contraction and global convergence of gradient descent. *PLoS One*, 15(8):1–29, 2020.

Example #17: Karcher mean on manifold of positive-definite matrices

$\mathbb{S}^n_{>0}$ = manifold of positive-definite matrices with *affine-invariant Riemannian metric*:

$$\mathbb{G}(X)(\xi,\eta) = \operatorname{trace}(X^{-1}\xi X^{-1}\eta) \qquad \text{(Riemannian metric)}$$

$$X\#_t Y = X^{1/2}(X^{-1/2}YX^{-1/2})^t X^{1/2} \qquad \text{(geodesic)}$$

$$d_{\mathbb{G}}(X,Y) = \|\log(X^{-1/2}YX^{-1/2})\|_{\mathsf{F}} \qquad \text{(geodesic distance)}$$

Given dataset $\{A_i \in \mathbb{S}^n_{>0}\}_{i\in\{1,\dots,N\}}$, define **Karcher loss function**

$$f_{\text{Karcher}}(X) = \sum_{i=1}^{N} d_{\mathbb{G}}(X, A_i)^2$$

f_{Karcher} is $2N$-strongly geodesically convex on $\mathbb{S}^n_{>0}$

Karcher mean = global minimizer = globally exp stable point of natural gradient

H. Zhang and S. Sra. First-order methods for geodesically convex optimization. In *29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pages 1617–1638, 2016. URL **https://proceedings.mlr.press/v49/zhang16b.html**

Consider a vector field $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$, and let $\xi, \eta \in \mathbb{R}^n$

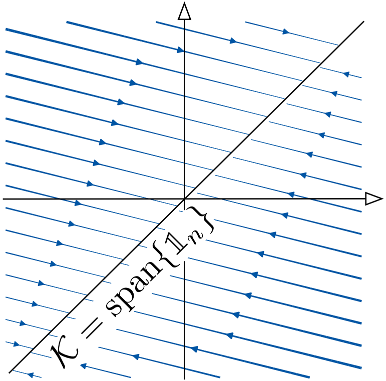
- Invariance property: for all $x, y \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$,

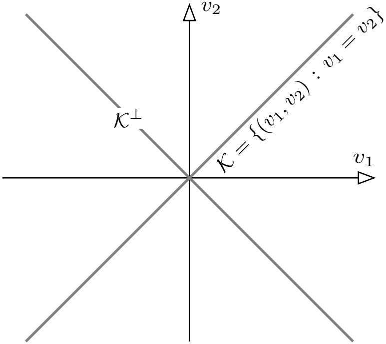
$$\mathsf{F}(x + \alpha\xi) = \mathsf{F}(x) \qquad \text{or equivalently} \qquad D\mathsf{F}(x)\xi = \mathbb{0}_n$$

- Conservation property: for all $x, y \in \mathbb{R}^n$,

$$\eta^\top \mathsf{F}(x) = \eta^\top \mathsf{F}(y) \qquad \text{or equivalently} \qquad \eta^\top D\mathsf{F}(x) = \mathbb{0}_n^\top$$

systems with invariance or conservation properties are not strongly contracting





For $\dot{x} = -Lx$

1. $\mathcal{K} = \mathrm{span}\{\mathbb{1}_n\}$
2. $x_{\mathsf{avg}} = \frac{1}{n}\mathbb{1}_n^\top x$ along $\mathcal{K}$
3. $x_\perp = x - x_{\mathsf{avg}}\mathbb{1}_n \in \mathcal{K}^\perp$

decomposition: perpendicular dynamics + reconstruction equation:

$$\dot{x}_\perp := -\Pi_n L x_\perp \qquad \in \mathbb{1}_m^\perp$$

$$\dot{x}_{\mathsf{avg}} = -\frac{1}{n}\mathbb{1}_n^\top L x_\perp \qquad \in \mathbb{R}$$

Systems with symmetry and their reduced dynamics

Model	**Symmetry**	**Reduced space**
Laplacian	$\dot{x} = \mathsf{F}_{\mathsf{Laplacian}}(x) := -Lx$ $\mathsf{F}_{\mathsf{Laplacian}}(x + \alpha \mathbb{1}_n) = \mathsf{F}_{\mathsf{Laplacian}}(x)$	$\mathbb{R}^n/\mathbb{R}$
Kuramoto-Sakaguchi	$\dot{\theta} = \mathsf{F}_{\mathsf{KS}}(\theta) := \omega + B\mathcal{A}\big(\sin(B^\top\theta - \varphi) + \sin(\varphi)\big)$ $\mathsf{F}_{\mathsf{KS}}(\theta + \alpha \mathbb{1}_n) = \mathsf{F}_{\mathsf{KS}}(\theta)$	$\mathbb{T}^n/\mathbb{S} \;\; \to \;\; \mathbb{R}^n/\mathbb{R}$
Primal-dual gradient with k redundant constraints	$\begin{bmatrix} \dot{x} \\ \dot{\lambda} \end{bmatrix} = \mathsf{F}_{\mathsf{PDG}}\left(\begin{bmatrix} x \\ \lambda \end{bmatrix}\right) := \begin{bmatrix} -\nabla f(x) - A^\top\lambda \\ Ax - b \end{bmatrix}$ $\mathsf{F}_{\mathsf{PDG}}\left(\begin{bmatrix} x \\ \lambda \end{bmatrix} + \begin{bmatrix} 0 \\ \alpha\nu \end{bmatrix}\right) = \mathsf{F}_{\mathsf{PDG}}\left(\begin{bmatrix} x \\ \lambda \end{bmatrix}\right)$ for all $\nu \in \ker(A^\top)$	$\mathbb{R}^{n+m}/\mathbb{R}^k$

If $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ is invariant under $\mathbb{R}^k$ translations, then
perpendicular dynamics $\mathsf{F}_\perp : \mathbb{R}^n/\mathbb{R}^k \to \mathbb{R}^n/\mathbb{R}^k$ is well defined
full solution obtained via **reconstruction equation**

Seminorms and semicontraction

A **seminorm** is a function $|||\cdot||| : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ such that $\forall a \in \mathbb{R}$ and $\forall x, y \in \mathbb{R}^n$

1. (*homogeneity*): $|||ax||| = |a| |||x|||$
2. (*subadditivity*): $|||x+y||| \leq |||x||| + |||y|||$

kernel is a subspace $\quad \mathcal{K} = \{x \in \mathbb{R}^n$ such that $|||x||| = 0\}$
seminorm is invariant $\quad |||x+\kappa||| = |||x|||$ for all $\kappa \in \mathcal{K}$

seminorm on $\mathbb{R}^n$ with kernel $\mathcal{K} \sim \mathbb{R}^k$ $\iff$ **norm on $\mathcal{K}^\perp \sim \mathbb{R}^n/\mathbb{R}^k$**

- *matrix seminorm* is $|||A||| = \max_{\substack{|||v|||=1 \\ v \perp \mathcal{K}}} |||Av|||$
- *matrix log seminorm* $\mu_{|||\cdot|||}(A) = \lim_{h \to 0^+} \dfrac{|||I_n + hA||| - 1}{h}$
- F is *infinitesimally semicontracting* if $\quad \sup_x \mu_{|||\cdot|||}(D\mathsf{F}(x)) \leq -c$

F is inf semicontracting on $\mathbb{R}^n$ $\iff$ **$\mathsf{F}_\perp$ is inf contracting on $\mathbb{R}^n/\mathbb{R}^k$**

ℓ_2 seminorm with kernel $\mathcal{K}$ $\iff$ $P = P^\top \succeq 0$ and $\ker(P) = \mathcal{K}$

$$\|x\|^2_{2,P^{1/2}} := x^\top P x$$

consensus ℓ_2 seminorm with $\mathcal{K} = \operatorname{span}\{\mathbb{1}_n\}$

$$\|x\|^2_{2,\Pi_n} := \sum\nolimits_{i,j} (x_i - x_j)^2,$$

$$\Pi_n = I_n - \mathbb{1}_n \mathbb{1}_n^\top / n = \text{ orthogonal projection onto } \mathcal{K}^\top = \operatorname{span}\{\mathbb{1}_n\}^\perp$$

Given ℓ_2 seminorm defined by $P = P^\top \succeq 0$ and $\ker(P) = \mathcal{K}$,
semicontractivity LMIs for $A\mathcal{K} \subset \mathcal{K}$

$$\|A\|_{2,P^{1/2}} \leq \ell \quad \iff \quad A^\top P A \preceq \ell^2 P$$

$$\mu_{2,P^{1/2}}(A) \leq \ell \quad \iff \quad A^\top P + AP \preceq 2\ell P$$

Laplacian flow

$$\dot{x} = \mathsf{F}_{\text{Laplacian}}(x) := -Lx$$

where L is the Laplacian of a weighted undirected graph

$\mathsf{F}_{\text{Laplacian}}$ is semicontracting wrt $\|\cdot\|_{2,\Pi_n}$ with rate λ_2

- $L \succeq \lambda_2 \Pi_n$
- $\Pi_n L = L \Pi_n = L$
- $\Pi_n(-L) + (-L)\Pi_n \preceq -2\lambda_2 \Pi_n$
- $\mathrm{osLip}_{2,\Pi_n}(\mathsf{F}_{\text{Laplacian}}) := \mu_{2,\Pi_n}(-L) = -\lambda_2$

Example #19: Kuramoto-Sakaguchi model and synchronization

graph: incidence matrix B, weight matrix $\mathcal{A}$, max degree d_{max} and algebraic connectivity λ_2
natural frequency ω, frustration parameter φ

$$\dot{\theta}_i = \omega_i + \sum_j a_{ij} \sin(\theta_i - \theta_j + \varphi_{ij})$$

F_{KS} is locally infinitesimally semicontracting wrt $\|\cdot\|_{2,\Pi_n}$

Proof sketch

$$\dot{\theta} = \omega + \cos(\varphi) \underbrace{B\mathcal{A}\sin(B^\top\theta)}_{\mathsf{F}_{\mathsf{odd}}(\theta)} - \sin(\varphi) \underbrace{B\mathcal{A}\big(\mathbb{1} - \cos(B^\top\theta)\big)}_{\mathsf{F}_{\mathsf{even}}(\theta)}$$

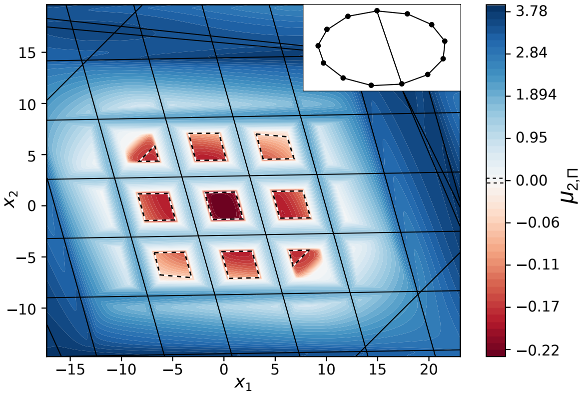
For $\theta \in \mathbb{T}^n$, define $\gamma(\theta) = \max_{(i,j)} |\theta_i - \theta_j|$

$$\mu_{2,\Pi_n}(D\mathsf{F}_{\mathsf{odd}}(\theta)) = \mu_{2,\Pi_n}(-L(\theta)) \leq -\lambda_2 \cos(\gamma(\theta)) \qquad (\text{Jacobian} = -\text{Laplacian}, \ L = B\mathcal{A}B^\top)$$

$$\mu_{2,\Pi_n}(D\mathsf{F}_{\mathsf{even}}(\theta)) \leq d_{\mathsf{max}} \sin(\gamma(\theta))$$

$$\implies \qquad \mu_{2,\Pi_n}(D\mathsf{F}_{\mathsf{KS}}(\theta)) < 0 \text{ locally in } \left\{\theta \in \mathbb{T}^n \;\middle|\; \gamma(\theta) < \arctan \frac{\lambda_2}{d_{\mathsf{max}} \tan(\varphi)}\right\}$$

Local semicontractivity of KS system, inside cells



$\mu_{2,\Pi_n}(D\mathsf{F}_{\mathsf{KS}}(\theta))$ for θ in two-dimensional slice of $\mathbb{R}^{13}$
model parameters: frustration $\varphi = 0.01$, graph in inset

R. Delabays and F. Bullo. Semicontraction and synchronization of Kuramoto-Sakaguchi oscillator networks. *IEEE Control Systems Letters*, 7:1566–1571, 2023.

Example #20: Primal-dual gradient dynamics with redundant constraints

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

constraint matrix A s.t. $0 \preceq a_{\mathsf{min}} \Pi_A \preceq AA^\top \preceq a_{\mathsf{max}} I_m$

where Π_A is the orthogonal projection onto $\operatorname{Im}(A)$

i.e., redundant constraints are allowed

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 semicontracting wrt $\|\cdot\|_{2,P^{1/2}}$ with rate c

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

For each $\nu_{\mathsf{min}} I_n \preceq Q \preceq \nu_{\mathsf{max}} I_n$,

$$\begin{bmatrix} -Q & -A^\top \\ A & 0 \end{bmatrix}^\top P + P \begin{bmatrix} -Q & -A^\top \\ A & 0 \end{bmatrix} \preceq -2cP$$

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

Outline

Contractivity (and its generalizations) in dynamical network systems

1. **Lotka-Volterra population dynamics** (Lotka, 1920; Volterra, 1928):
   ℓ_1- semiglobally strongly contracting (after a rescaling change of coordinates)

2. **Matrosov-Bellman interconnected stable systems** (Bellman, 1962; Matrosov, 1962):
   strongly contracting wrt composite norm

3. **Kuramoto coupled oscillators** (Kuramoto, 1975):
   strongly semicontracting wrt (ℓ_2, Π_n) norm, in neighb'd of each phase-cohesive equilibrium

4. **Yorke multigroup SIS epidemic model** (Lajmanovich and Yorke, 1976):
   equilibrium contracting wrt weighted ℓ_1/ℓ_∞ norms (at disease-free and endemic eq.)

5. **Hopfield and cellular neural networks** (Hopfield, 1982):
   ℓ_1/ℓ_∞-strongly contracting

6. **Daganzo cell transmission model for traffic networks** (Daganzo, 1994):
   ℓ_1-weakly contracting, when the dynamics is monotone

7. **Chua's diffusively-coupled dynamical systems** (Wu and Chua, 1995):
   strongly semi-contracting wrt $(2, p)$ tensor norm on $\mathbb{R}^n \otimes \mathbb{R}^k$

8. ...

contractivity = robust computationally-friendly stability

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

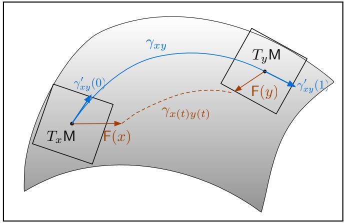
	Lyapunov Theory	**Contraction Theory for Dynamical Systems**
	F admits global Lyapunov function	F is strongly contracting
existence of equilibrium	assumed	implied + computational methods
Lyapunov function	arbitrary	$\|x - x^*\|$ and $\|\mathsf{F}(x)\|$
inputs	ISS via $\mathcal{KL}$ and $\mathcal{L}$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

search for contraction properties

design engineering systems to be contracting

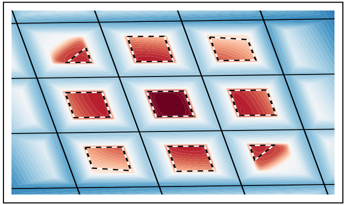
Theoretical frontiers

- higher order contraction
- relationship with monotone operator theory
- metric spaces
- computational methods



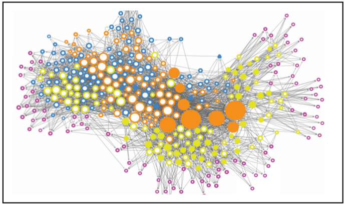
Limitations: not all stable systems are contractive:

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



Application to control and learning

1. control: optimization-based control design
2. ML: implicit models and energy-based learning
3. neuroscience: robust dynamical modeling



Outline

Consider a vector field $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$, and let $\xi, \eta \in \mathbb{R}^n$.

- Invariance property: for all $x, y \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$,

$$\mathsf{F}(x + \alpha\xi) = \mathsf{F}(x) \qquad \text{or equivalently} \qquad D\mathsf{F}(x)\xi = \mathbb{0}_n$$

- Conservation property: for all $x, y \in \mathbb{R}^n$,

$$\eta^\top \mathsf{F}(x) = \eta^\top \mathsf{F}(y) \qquad \text{or equivalently} \qquad \eta^\top D\mathsf{F}(x) = \mathbb{0}_n^\top$$

Example #21: Averaging and dynamical flow systems

Prototypical dynamics with invariance and conservation

Let $A \in \mathbb{R}^{n \times n}$ be row-stochastic: $A\mathbb{1}_n = \mathbb{1}_n$ and $A \geq 0$

Averaging Systems

$$x_{k+1} = Ax_k$$

Invariance: dynamics unaffected by translations in $\operatorname{span}\{\mathbb{1}_n\}$

Examples: distributed optimization, robotic coordination, frequency synchronization, ...

Dynamical Flow Systems

$$x_{k+1} = A^\top x_k$$

Conservation: quantity $\mathbb{1}_n^\top x$ is constant

Examples: compartmental models, Markov chains

Given row-stochastic $A \in \mathbb{R}^{n\times n}$,

Markov-Dobrushin ergodic coefficient

$$\tau_1(A) = \max_{\|z\|_1=1, \mathbb{1}_n^\top z=0} \|A^\top z\|_1$$

$\tau_1(A) < 1$ under mild connectivity conditions

$\tau_p(A)$ also defined for general $p \in [1,\infty]$

How is τ_1 an induced norm?

A. A. Markov. Extensions of the law of large numbers to dependent quantities. *Izvestiya Fiziko-matematicheskogo obschestva pri Kazanskom universitete*, 15, 1906. (in Russian)

R. L. Dobrushin. Central limit theorem for nonstationary Markov chains. I. *Theory of Probability & Its Applications*, 1(1):65–80, 1956. doi

$A \in \mathbb{R}^{n\times n}$ row-stochastic

Classical Property of Averaging Systems $\quad x_{k+1} = Ax_k$

Given $x \in \mathbb{R}^n$, max-min disagreement:

$$d_{\text{max-min}}(Ax) \;\leq\; \tau_1(A)\; d_{\text{max-min}}(x), \qquad \text{where} \quad d_{\text{max-min}}(x) = \max_i\{x_i\} - \min_j\{x_j\}$$

Classical Property of Markov Chains $\quad x_{k+1} = A^\top x_k$

Given π, σ in the simplex Δ_n, total variation distance:

$$d_{\text{TV}}(A^\top\pi, A^\top\sigma) \;\leq\; \tau_1(A)\; d_{\text{TV}}(\pi,\sigma), \qquad \text{where} \quad d_{\text{TV}}(\pi,\sigma) = \frac{1}{2}\sum_i |\pi_i - \sigma_i|$$

Why is the same τ_1 relevant in both cases?

A **seminorm** is a function $|||\cdot||| : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ s.t., $\forall a \in \mathbb{R}$ and $\forall x, y \in \mathbb{R}^n$:

1. (*homogeneity*): $|||ax||| = |a|\,|||x|||$
2. (*subadditivity*): $|||x + y||| \leq |||x||| + |||y|||$

The *kernel* is the vector space:

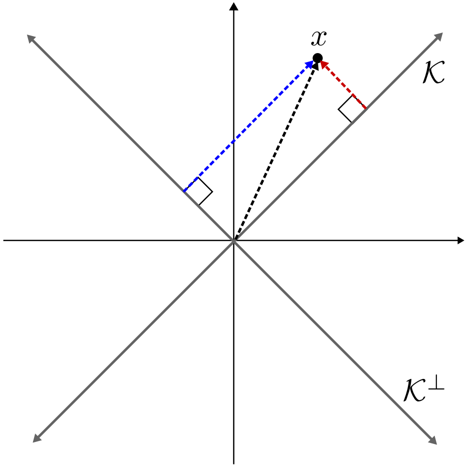
$$\mathcal{K} = \{x \in \mathbb{R}^n : |||x||| = 0\}$$

We focus on *consensus seminorms*, where $\mathcal{K} = \operatorname{span}\{\mathbb{1}_n\}$.

Note: $|||\cdot|||$ is invariant under translations in $\mathcal{K}$

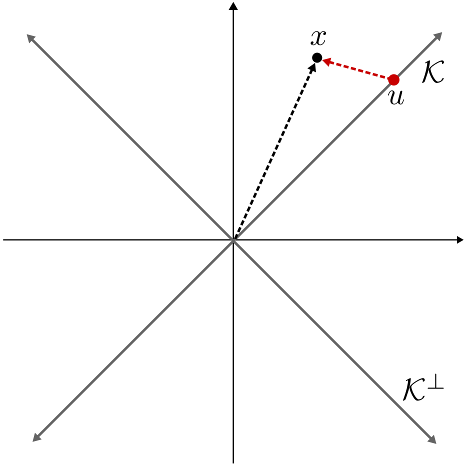
Projection and distance-based seminorms: graphical definitions

Projection seminorms



$$|||x|||_{\mathsf{proj},p} \triangleq \|\Pi_\perp x\|_p, \quad \Pi_\perp = \Pi_\perp^\top$$

Distance seminorms



$$|||x|||_{\mathsf{dist},p} \triangleq \min_{u\in\mathcal{K}} \|x-u\|_p$$

Projection and distance-based consensus seminorms ($\mathcal{K} = \operatorname{span}\{\mathbb{1}_n\}$)

	$\lvert\lvert\lvert x \rvert\rvert\rvert_{\mathsf{proj},p}$	$\lvert\lvert\lvert x \rvert\rvert\rvert_{\mathsf{dist},p}$
ℓ_1	$\displaystyle\sum_{i=1}^{n} \lvert x_i - x_{\mathrm{avg}} \rvert$	$\displaystyle\sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} x_{(i)} - \sum_{j=\lceil \frac{n}{2} \rceil+1}^{n} x_{(j)}$
ℓ_2	$\sqrt{\frac{1}{n}\sum_{i,j}(x_i - x_j)^2}$	$\sqrt{\frac{1}{n}\sum_{i,j}(x_i - x_j)^2}$
ℓ_∞	$\max_i \lvert x_i - x_{\mathrm{avg}} \rvert$	$\frac{1}{2}\big(x_{(1)} - x_{(n)}\big)$

where we have sorted $x_{(1)} \geq x_{(2)} \geq \cdots \geq x_{(n)}$

Therefore

$$d_{\mathsf{max\text{-}min}}(x) = 2\lvert\lvert\lvert x \rvert\rvert\rvert_{\mathsf{dist},\infty} \qquad \text{and} \qquad d_{\mathsf{TV}}(\pi,\sigma) = \lvert\lvert\lvert \pi - \sigma \rvert\rvert\rvert_{\mathsf{proj},1}$$

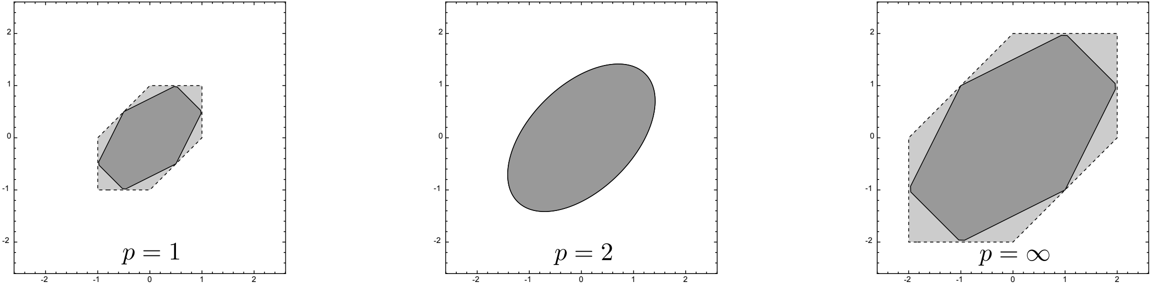


Figure: Two-dimensional sections of three-dimensional unit disks of projection (solid contours) and distance (dashed contours) consensus seminorms. We plot the sections corresponding to $(x_1, x_2, x_3 = 0)$.

Induced matrix seminorms

Consider a seminorm $|||\cdot|||$ on $\mathbb{R}^n$ with kernel $\mathcal{K}$.

Induced matrix seminorm: function $|||\cdot||| : \mathbb{R}^{n\times n} \to \mathbb{R}_{\geq 0}$ where

$$|||A||| = \max_{\substack{|||x|||\leq 1 \\ x\perp\mathcal{K}}} |||Ax|||, \qquad \forall A \in \mathbb{R}^{n\times n}$$

⚠ In general, $|||Ax||| \nleq |||A||| |||x|||$

Inequality is true if $x \in \mathcal{K}^\perp$ or $A\mathcal{K} \subseteq \mathcal{K}$

Key facts about dual and induced norms

Properties of dual and induced norms

1. **ℓ_p and ℓ_q norms are dual**, for $1/p + 1/q = 1$

$$\|\cdot\|_p \;=\; (\|\cdot\|_q)_\star \qquad\qquad \|\cdot\|_q \;=\; (\|\cdot\|_p)_\star$$

2. dual norm satisfies (sharp) *Hölder inequality*: $\quad x^\top y \;\leq\; \|x\|_p\,\|y\|_q$
3. equality between dual induced norms: $\quad \|A\|_p \;=\; \|A^\top\|_q$
4. induced norm is submultiplicative: $\quad \|AB\| \;\leq\; \|A\|\,\|B\|$

Key facts about dual and induced seminorms

Properties of dual and induced seminorms

1. **ℓ_p-distance and ℓ_q-projection seminorms are dual**, for $1/p + 1/q = 1$

$$|||\cdot|||_{\mathrm{dist},p} = \left(|||\cdot|||_{\mathrm{proj},q}\right)_\star \qquad\qquad |||\cdot|||_{\mathrm{proj},q} = \left(|||\cdot|||_{\mathrm{dist},p}\right)_\star$$

2. dual seminorm satisfies (sharp) *Markov inequality*: $x^\top \Pi_\perp y \leq |||x|||_{\mathrm{dist},p}\, |||y|||_{\mathrm{proj},q}$
3. equality between dual induced seminorms: $|||A|||_{\mathrm{dist},p} = |||A^\top|||_{\mathrm{proj},q}$
4. induced seminorm is submultiplicative: $|||AB||| \leq |||A|||\,|||B|||$ if $A\mathcal{K} \subseteq \mathcal{K}$ or $B\mathcal{K}^\top \subseteq \mathcal{K}^\top$

Ergodic coefficients are induced seminorms

$$|||A|||_{\mathrm{dist},p} = |||A^\top|||_{\mathrm{proj},q} = \tau_q(A) := \max_{\|z\|_q = 1,\ z \perp \mathbb{1}_n} \|A^\top z\|_q$$

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.

Classical Property of Averaging Systems

Given row-stochastic $A \in \mathbb{R}^{n\times n}$ and $x, y \in \mathbb{R}^n$:

$$\begin{aligned} \|A(x-y)\|_{\mathsf{dist},\infty} &\leq \tau_1(A)\|x-y\|_{\mathsf{dist},\infty} \\ &= \big|\big|\big|A\big|\big|\big|_{\mathsf{dist},\infty}\big|\big|\big|x-y\big|\big|\big|_{\mathsf{dist},\infty} \end{aligned}$$

Classical Property of Markov Chains

Given row-stochastic $A \in \mathbb{R}^{n\times n}$ and π, σ in the simplex Δ_n:

$$\begin{aligned} \|A^\top(\pi-\sigma)\|_{\mathsf{proj},1} &\leq \tau_1(A)\|\pi-\sigma\|_{\mathsf{proj},1} \\ &= \big|\big|\big|A^\top\big|\big|\big|_{\mathsf{proj},1}\big|\big|\big|\pi-\sigma\big|\big|\big|_{\mathsf{proj},1} \end{aligned}$$

Summary and future work

1. ergodic coefficients are contraction factors
2. duality explains their roles in both averaging and flow systems
3. nonEuclidean norms play a key role
4. **semicontraction theory**
   1. discrete/continuous-time Markov chains
   2. discrete/continuous-time nonlinear consensus algorithms
   3. primal-dual dynamics with redundant constraints
   4. local contractivity of Kuramoto-Sakaguchi models

Future work

consider the set of undirected, unweighted connected graphs + selfloops
for each adjacency A_i, define row-stochastic $\mathcal{A}_i = \mathrm{diag}(A_i \mathbb{1}_n)^{-1} A_i$ (equal neighbor)
find a consensus seminorm $|||\cdot|||$ such that, for each i,

$$|||\mathcal{A}_i||| < 1$$

or **prove** that it does not exist

The *induced log seminorm* of $A \in \mathbb{R}^{n\times n}$ is

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

Laplacian L, corresponding to weighted digraph with adj. matrix A:

$$\mu_{\mathsf{dist},1}(-L) = -\min_j \left\{ (d_{\mathrm{out}})_j - \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor - 1} a_{(i),j} + \sum_{i=\lceil \frac{n}{2} \rceil}^{n-1} a_{(i),j} \right\}, \quad d_{\mathrm{out}} = A\mathbb{1}_n$$

$$\mu_{\mathsf{dist},2}(-L) = \min \left\{ b : \Pi_\perp L + L^\top \Pi_\perp \succeq -2b\Pi_\perp \right\}, \quad \Pi_\perp = I_n - \tfrac{1}{n}\mathbb{1}_n\mathbb{1}_n^\top$$

$$\mu_{\mathsf{dist},\infty}(-L) = -\min_{i\neq j} \left\{ a_{ij} + a_{ji} + \sum_{k\neq i,j} \min\{a_{ik}, a_{jk}\} \right\}$$

Let $p, q \in [1,\infty]$ such that $p^{-1} + q^{-1} = 1$. For any matrix $M \in \mathbb{R}^{n\times n}$, and any kernel $\mathcal{K}$,

$$\mu_{\mathrm{dist},p}(M) = \mu_{\mathrm{proj},q}(M^\top)$$

Outline

- graph theoretic conditions for Metzler matrix to be Hurwitz
- combination of theory of Metzler Hurwitz matrices and graph theory
- critical role played by cycles (acyclic digraphs easy to handle)

X. Duan, S. Jafarpour, and F. Bullo. Graph-theoretic stability conditions for Metzler matrices and monotone systems. *SIAM Journal on Control and Optimization*, 59(5):3447–3471, 2021.

Hurwitz Metzler Theorem (see LNS.Section10.4)

1. M is Hurwitz,
2. there exists $\eta \in \mathbb{R}^n_{>0}$ such that $\eta^\top M < \mathbb{0}_n^\top$ or, equivalently, $\mu_{1,\mathrm{diag}(\eta)}(M) < 0$,
3. there exists $\xi \in \mathbb{R}^n_{>0}$ such that $M\xi < \mathbb{0}_n$ or, equivalently, $\mu_{\infty,\mathrm{diag}(\xi)^{-1}}(M) < 0$,
4. there exists a diagonal $P = P^\top \succ 0$ satisfying $M^\top P + PM \prec 0$ or, equivalently, $\mu_{2,P^{1/2}}(M) < 0$, and
5. all leading principal minors of $-M$ are positive.

The *leading principal minors* of a matrix are the determinants of its top-left $i \times i$ submatrices, for $i \in \{1, \dots, n\}$

Let $\mathcal{G}$ be a weighted directed graph such that

- $m_{ii} < 0$ is weight of self-loop at node i
- $m_{ij} > 0$ is weight of directed edge (i, j)

that is, the adjacency matrix M of $\mathcal{G}$ is Metzler with negative diagonal entries

1. a *simple cycle* in $\mathcal{G}$ is a directed cycle (with at least 2 nodes) in which only the first and last vertices are equal. Self-loops are not simple cycles.

2. the *gain* of a cycle $\phi = (i_1, i_2, \dots, i_k, i_1)$ is

$$\gamma_\phi(M) = \left(\frac{m_{i_1 i_2}}{-m_{i_2 i_2}}\right)\left(\frac{m_{i_2 i_3}}{-m_{i_3 i_3}}\right)\cdots\left(\frac{m_{i_k i_1}}{-m_{i_1 i_1}}\right) \qquad \text{(rational function of entries of } M\text{)}$$

3. two cycles ϕ and ψ are *disjoint*, denoted by $\phi \perp \psi$, if they have no node in common

Input: a Metzler matrix $M \in \mathbb{R}^n$ with associated digraph $\mathcal{G}$
Output: set of rational functions $\{\gamma_{\mathcal{C}_2}(M), \dots, \gamma_{\mathcal{C}_n}(M)\}$

1: **for** i from 2 to n
2: $\quad \mathcal{C}_i \;:=\;$ the set of simples cycles in $\mathcal{G}$ passing through only nodes $\{1, \dots, i\}$
3: $$\underbrace{\gamma_{\mathcal{C}_i}(M)}_{\text{set gain}} := \sum_{\phi \in \mathcal{C}_i} \underbrace{\gamma_\phi}_{\text{cycle gain}} - \sum_{\substack{\phi,\psi \in \mathcal{C}_i \\ \phi \perp \psi}} \gamma_\phi \gamma_\psi + \sum_{\substack{\phi,\psi,\rho \in \mathcal{C}_i \\ \phi\perp\psi,\ \phi\perp\rho,\ \psi\perp\rho}} \gamma_\phi \gamma_\psi \gamma_\rho - \cdots$$

Network small-gain theorem for Metzler matrices

$$\text{Metzler } M \text{ is Hurwitz} \quad \iff \quad \gamma_{\mathcal{C}_2} < 1, \dots, \gamma_{\mathcal{C}_n} < 1$$

These Hurwitzness conditions are: At most $n-1$. Polynomial after rewriting. Not unique (because nodes may be renumbered). Possibly redundant. Computational efficient (except precomputation of simple cycles)

$$M = \begin{bmatrix} -c_1 & \ell_{12} & 0 \\ \ell_{21} & -c_2 & \ell_{23} \\ 0 & \ell_{32} & -c_3 \end{bmatrix}$$

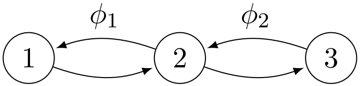


Figure: digraph associated to M and simple cycles $\phi_1 = (1,2,1)$ and $\phi_2 = (2,3,2)$

- cycle gains: $\gamma_{\phi_1} = \dfrac{\ell_{12}\ell_{21}}{c_1c_2}$ and $\gamma_{\phi_2} = \dfrac{\ell_{23}\ell_{32}}{c_2c_3}$
- cycle set $\mathcal{C}_2 = \{\phi_1\} \implies$ set gain $\gamma_{\mathcal{C}_2} = \gamma_{\phi_1}$ (note: $\gamma_{\phi_1} < 1$ is redundant)
- cycle set $\mathcal{C}_3 = \{\phi_1, \phi_2\} \implies$ set gain $\gamma_{\mathcal{C}_3} = \gamma_{\phi_1} + \gamma_{\phi_2}$ (no 2nd order terms since ϕ_1 and ϕ_2 are not disjoint)

$$\begin{bmatrix} -c_1 & \ell_{12} & 0 \\ \ell_{21} & -c_2 & \ell_{23} \\ 0 & \ell_{32} & -c_3 \end{bmatrix} \text{ Hurwitz} \iff \gamma_{\phi_1} + \gamma_{\phi_2} < 1 \quad \text{i.e.,} \quad \frac{\ell_{12}\ell_{21}}{c_1c_2} + \frac{\ell_{23}\ell_{32}}{c_2c_3} < 1$$

E.g., for $c = c_1 = c_2 = c_3$ and $\ell_{12} = \ell_{21} = \ell_{23} = \ell_{32} = 1$, $\quad M$ Hurwitz $\iff c > \sqrt{2}$.

This can be easily verified since: $\operatorname{spec}\left(\begin{bmatrix} -c & 1 & 0 \\ 1 & -c & 1 \\ 0 & 1 & -c \end{bmatrix}\right) = \{-c, -c-\sqrt{2}, -c+\sqrt{2}\}$.

$$M = \begin{bmatrix} -c_1 & 0 & 0 & \ell_{14} \\ 0 & -c_2 & \ell_{23} & \ell_{24} \\ 0 & \ell_{32} & -c_3 & \ell_{34} \\ \ell_{41} & \ell_{42} & \ell_{43} & -c_4 \end{bmatrix}$$

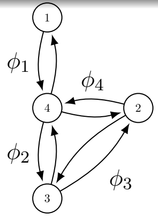


Figure: associated digraph and simple cycles

- cycle gains: $\gamma_{\phi_1} = \frac{\ell_{14}\ell_{41}}{c_1 c_4}$, $\gamma_{\phi_2} = \frac{\ell_{34}\ell_{43}}{c_3 c_4}$, $\gamma_{\phi_3} = \frac{\ell_{23}\ell_{32}}{c_2 c_3}$, and $\gamma_{\phi_4} = \frac{\ell_{24}\ell_{42}}{c_2 c_4}$
- $\mathcal{C}_2 = \emptyset$
- $\mathcal{C}_3 = \{\phi_3\}$: $\gamma_{\mathcal{C}_3} = \gamma_{\phi_3}$
- $\mathcal{C}_4 = \{\phi_1, \dots, \phi_4\}$: $\gamma_{\mathcal{C}_4} = \sum_{i=1}^4 \gamma_{\phi_i} - \gamma_{\phi_1}\gamma_{\phi_3}$

$$\begin{bmatrix} -c_1 & 0 & 0 & \ell_{14} \\ 0 & -c_2 & \ell_{23} & \ell_{24} \\ 0 & \ell_{32} & -c_3 & \ell_{34} \\ \ell_{41} & \ell_{42} & \ell_{43} & -c_4 \end{bmatrix} \text{ Hurwitz} \quad \iff \quad \gamma_{\phi_3} < 1 \text{ and } \gamma_{\phi_1} + \gamma_{\phi_2} + \gamma_{\phi_3} + \gamma_{\phi_4} - \gamma_{\phi_1}\gamma_{\phi_3} < 1$$

$$\begin{bmatrix} -c_1 & 0 & 0 & 0 & \ell_{15} & \ell_{16} \\ 0 & -c_2 & 0 & \ell_{24} & \ell_{25} & 0 \\ 0 & 0 & -c_3 & \ell_{34} & 0 & \ell_{36} \\ 0 & \ell_{42} & \ell_{43} & -c_4 & 0 & 0 \\ \ell_{51} & \ell_{52} & 0 & 0 & -c_5 & 0 \\ \ell_{61} & 0 & \ell_{63} & 0 & 0 & -c_6 \end{bmatrix}$$

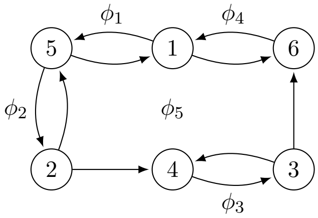


Figure: associated digraph and simple cycles

- $\mathcal{C}_2$, $\mathcal{C}_3$ empty
- $\mathcal{C}_4 = \{\phi_3\}$: $\gamma_3 < 1$ (redundant)
- $\mathcal{C}_5 = \{\phi_1, \phi_2, \phi_3\}$: $\gamma_{\mathcal{C}_5} = \gamma_1 + \gamma_2 + \gamma_3 - \gamma_1\gamma_3 - \gamma_2\gamma_3 < 1$
- $\mathcal{C}_6 = \{\phi_1, \ldots, \phi_5\}$: $\gamma_{\mathcal{C}_6} = \sum_{i=1}^{5} \gamma_i - \gamma_1\gamma_3 - \gamma_2\gamma_3 - \gamma_3\gamma_4 - \gamma_2\gamma_4 + \gamma_2\gamma_3\gamma_4 < 1$

Outline

Outline

Given $Q \in \mathbb{R}^{n\times n}$, $A \in \mathbb{R}^{m\times n}$, and a time-scale parameter $\tau > 0$, define

saddle matrix $$\mathcal{S} = \begin{bmatrix} -Q & -A^\top \\ \tau^{-1}A & 0 \end{bmatrix} \in \mathbb{R}^{(m+n)\times(m+n)}$$

$q_{\mathsf{min}} := \lambda_{\mathsf{min}}(Q+Q^\top)/2 > 0$
$q_{\mathsf{max}} := \min\{q \text{ such that } Q^\top Q \preceq q(Q+Q^\top)/2\} \leq \sigma^2_{\mathsf{max}}(Q)/q_{\mathsf{min}}$
$a_{\mathsf{min}}\Pi_A \preceq AA^\top \preceq a_{\mathsf{max}}I_m$, where $\Pi_A \in \mathbb{R}^{m\times m}$ is orthogonal projection onto image of A

Semi-contractivity LMI

$$\mathcal{S}^\top P + P\mathcal{S} \preceq -2cP$$

where

$$P = \begin{bmatrix} I_n & \alpha A^\top \\ \alpha A & \tau\Pi_A \end{bmatrix} \succeq 0 \quad \text{with} \quad \alpha = \frac{1}{2}\min\left\{\frac{1}{\nu_{\mathsf{max}}}, \tau\frac{\nu_{\mathsf{min}}}{a_{\mathsf{max}}}\right\}$$
$$c = \frac{1}{2}\tau^{-1}\alpha a_{\mathsf{min}} = \frac{1}{4}\min\left\{\frac{a_{\mathsf{min}}}{\tau q_{\mathsf{max}}}, \frac{a_{\mathsf{min}}}{a_{\mathsf{max}}}q_{\mathsf{min}}\right\}$$

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

Proof of saddle matrix semicontractivity I: $P \succeq 0$
Use Schur complement to show that $P \succeq 0$. Clearly the $(1,1)$ block is positive definite. Therefore,

$$P \succeq 0 \quad \Longleftrightarrow \quad \tau \Pi_A - \alpha^2 A A^\top \succ 0 \quad \Longleftrightarrow \quad \tau - \alpha^2 a_{\mathsf{max}} > 0 \quad \Longleftrightarrow \quad \alpha^2 < \tau / a_{\mathsf{max}}.$$

The inequality $\alpha^2 < \tau/a_{\mathsf{max}}$ follows from the stronger inequality $(2\alpha)^2 < \frac{\tau}{a_{\mathsf{max}}}$ with the following argument:

$$\min\left\{\frac{1}{q_{\mathsf{max}}}, \tau\frac{q_{\mathsf{min}}}{a_{\mathsf{max}}}\right\}^2 \leq \min\left\{\frac{1}{q_{\mathsf{max}}}, \tau\frac{q_{\mathsf{min}}}{a_{\mathsf{max}}}\right\} \cdot \max\left\{\frac{1}{q_{\mathsf{max}}}, \tau\frac{q_{\mathsf{min}}}{a_{\mathsf{max}}}\right\} = \frac{q_{\mathsf{min}}}{q_{\mathsf{max}}} \cdot \frac{\tau}{a_{\mathsf{max}}} \leq \frac{\tau}{a_{\mathsf{max}}}.$$

Proof of saddle matrix semicontractivity II: factorization of LMI

Next, we aim to show that $-\mathcal{S}^\top P - P\mathcal{S} - 2cP \succeq 0$. After some bookkeeping, we compute

$$
\begin{aligned}
-\mathcal{S}^\top P - P\mathcal{S} - 2cP &= \begin{bmatrix} Q^\top & -\tau^{-1}A^\top \\ A & 0 \end{bmatrix}\begin{bmatrix} I_n & \alpha A^\top \\ \alpha A & \tau\Pi_A \end{bmatrix} + \begin{bmatrix} I_n & \alpha A^\top \\ \alpha A & \tau\Pi_A \end{bmatrix}\begin{bmatrix} Q & A^\top \\ -\tau^{-1}A & 0 \end{bmatrix} - 2c\begin{bmatrix} I_n & \alpha A^\top \\ \alpha A & \tau\Pi_A \end{bmatrix} \\
&= \begin{bmatrix} Q + Q^\top - 2\tau^{-1}\alpha A^\top A - 2cI_n & \alpha Q^\top A^\top - A^\top + A^\top \Pi_A^\top - 2c\alpha A^\top \\ A + \alpha A Q - \Pi_A A - 2c\alpha A & 2\alpha A A^\top - 2c\tau\Pi_A \end{bmatrix}.
\end{aligned}
$$

The (2,2) block satisfies the lower bound

$$
2\alpha AA^\top - 2c\tau\Pi_A = 2\left(\tfrac{1}{2}\alpha AA^\top - c\tau\Pi_A\right) + \alpha AA^\top \succeq 2\big(\tfrac{1}{2}\alpha a_{\mathsf{min}} - c\tau\big)\Pi_A + \alpha AA^\top = \alpha AA^\top \succ 0.
$$

Given this lower bound and the equality $\Pi_A A = A$, we can factorize the resulting matrix as follows:

$$
-\mathcal{S}^\top P - P\mathcal{S} - cP \succeq \begin{bmatrix} I_n & 0 \\ 0 & A \end{bmatrix} \underbrace{\begin{bmatrix} Q + Q^\top - 2(\tau^{-1}\alpha A^\top A + cI_n) & \alpha Q^\top - 2c\alpha I_n \\ \alpha Q - 2c\alpha I_n & \alpha I_n \end{bmatrix}}_{n\times n} \begin{bmatrix} I_n & 0 \\ 0 & A^\top \end{bmatrix}.
$$

Proof of saddle matrix semicontractivity III: Schur complement and final bounds

Since $\alpha I_n \succ 0$, it suffices to show that the Schur complement of the (2,2) block is positive semidefinite:

$$
\begin{aligned}
& Q + Q^\top - 2(\tau^{-1}\alpha A^\top A + cI_n) - \alpha\big(Q^\top - 2cI_n\big)\big(Q - 2cI_n\big) \succeq 0 && (9)\\
\iff \quad & (Q + Q^\top - \alpha Q^\top Q) + 2\alpha c(Q + Q^\top) \succeq 2(\tau^{-1}\alpha A^\top A + cI_n) + 4\alpha c^2 I_n && (10)\\
\Longleftarrow \quad & Q + Q^\top - \alpha Q^\top Q \succeq 2(\tau^{-1}\alpha A^\top A + cI_n) \quad \text{and} \quad 2\alpha c(Q + Q^\top) \succeq 4\alpha c^2 I_n. && (11)
\end{aligned}
$$

To prove the first inequality in (11), we upper bound the right hand side as follows:

$$
\begin{aligned}
2(\tau^{-1}\alpha A^\top A + cI_n) \preceq 2(\tau^{-1}\alpha a_{\text{max}} + c)I_n & \overset{c=\frac{1}{2}\tau^{-1}\alpha a_{\text{min}}}{=} \tau^{-1}\alpha(2a_{\text{max}} + a_{\text{min}})I_n \\
& \overset{\alpha\leq\frac{1}{2}\tau q_{\text{min}}/a_{\text{max}}}{\preceq} \frac{1}{2}\frac{q_{\text{min}}}{a_{\text{max}}}(2a_{\text{max}} + a_{\text{min}})I_n \preceq \frac{3}{2}q_{\text{min}}I_n.
\end{aligned}
$$

Next, since $\alpha \leq \frac{1}{2q_{\text{max}}}$, we know $-\alpha q_{\text{max}} \geq -\frac{1}{2}$. We then lower bound the left hand side as follows:

$$
Q + Q^\top - \alpha Q^\top Q \overset{\text{by definition}}{\succeq} Q + Q^\top - \alpha q_{\text{max}}(Q + Q^\top)/2 \succeq (2 - \tfrac{1}{2})(Q + Q^\top) \succeq \frac{3}{2}q_{\text{min}}I_n.
$$

Finally, we prove the second inequality in (11) that is, $2\alpha c(Q + Q^\top) \succeq 4\alpha c^2 I_n$. This is equivalent to $Q + Q^\top \succeq 2cI_n$ and follows from noting $c \leq \frac{1}{2}\frac{a_{\text{min}}}{a_{\text{max}}}q_{\text{min}} < q_{\text{min}}$.

Outline

Euler discretization theorem

Euler discretization theorem for contracting dynamics

Given arbitrary norm $\|\cdot\|$ 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

Optimal* contractivity of Euler discretization $\mathsf{Id} + \alpha \mathsf{F}$

Given $c := -\mathsf{osLip}(\mathsf{F}) > 0$ and $\ell := \mathsf{Lip}(\mathsf{F})$, define *condition number* $\kappa = \ell/c \geq 1$:

3. $$0 < \alpha < \frac{1}{c\kappa(1+\kappa)} \quad \implies \quad \mathsf{Lip}\big(\mathsf{Id} + \alpha \mathsf{F}\big) \leq \left(1 + \alpha c - \frac{\alpha^2 \ell^2}{1 - \alpha \ell}\right)^{-1} < 1$$

4. the optimal* step size and contraction factor are

$$\alpha^* = \frac{1}{c}\left(\frac{1}{2\kappa^2} - \frac{3}{8\kappa^3} + \mathcal{O}\left(\frac{1}{\kappa^4}\right)\right), \qquad \mathsf{Lip}(\mathsf{Id} + \alpha^* \mathsf{F}) = 1 - \frac{1}{4\kappa^2} + \frac{1}{8\kappa^3} + \mathcal{O}\left(\frac{1}{\kappa^4}\right)$$

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, 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**

Euler discretization theorem: Additional equivalences

Given $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ and $\alpha \geq 0$, define

- **shifted map** $\mathsf{G} := \mathsf{Id} + \mathsf{F} \quad \iff \quad \mathsf{F} = -\mathsf{Id} + \mathsf{G}$
- $\mathsf{F}_\alpha := \underbrace{\mathsf{Id} + \alpha\mathsf{F}}_{\text{Euler discretization of F}} = \underbrace{(1-\alpha)\,\mathsf{Id} + \alpha\mathsf{G}}_{\text{average map of G}} =: \mathsf{G}_\alpha$

For a differentiable F and $x \in \mathbb{R}^n$

$$\underset{\text{equilibrium point}}{\mathsf{F}(x) = 0} \quad \iff \quad \underset{\text{fixed point}}{\mathsf{F}_\alpha(x) = \mathsf{G}_\alpha(x) = \mathsf{G}(x) = x}$$

$$\underset{\text{F is infinitesimally contracting}}{\mathrm{osLip}(\mathsf{F}) < 0} \quad \iff \quad \mathrm{osLip}(\mathsf{G}) = \mathrm{osLip}(\mathsf{F}) + 1 < 1$$

$$\Updownarrow \qquad\qquad\qquad\qquad \Updownarrow$$

$$\underset{\mathsf{F}_{\alpha^*}\text{ is contracting}}{\exists \alpha^* \text{ s.t. } \mathrm{Lip}(\mathsf{F}_{\alpha^*}) < 1} \quad \iff \quad \underset{\mathsf{G}_{\alpha^*}\text{ is contracting}}{\exists \alpha^* \text{ s.t. } \mathrm{Lip}(\mathsf{G}_{\alpha^*}) < 1}$$

Optimal* contractivity of Euler discretization $\mathsf{Id} + \alpha \mathsf{F}$: inner-product norms $\|\cdot\|_{2,P^{1/2}}$

Given $c := -\operatorname{osLip}(\mathsf{F}) > 0$ and $\ell := \operatorname{Lip}(\mathsf{F})$, define *condition number* $\kappa = \ell/c \geq 1$:

1. $0 < \alpha < \dfrac{2}{c\kappa^2} \quad \Longrightarrow \quad \operatorname{Lip}\big(\mathsf{Id} + \alpha \mathsf{F}\big) \leq \sqrt{1 - 2\alpha c + \alpha^2 \ell^2} < 1$
2. the optimal* step size and contraction factor are

$$\alpha^* = \frac{1}{c\kappa^2}, \qquad \operatorname{Lip}(\mathsf{Id} + \alpha^* \mathsf{F}) = 1 - \frac{1}{2\kappa^2} + \mathcal{O}\Big(\frac{1}{\kappa^4}\Big)$$

Standard proof from monotone operator theory. For $\alpha > 0$, compute

$$\begin{aligned}
\|(\mathsf{Id} + \alpha \mathsf{F})x - (\mathsf{Id} + \alpha \mathsf{F})y\|^2 &= \|x - y + \alpha(\mathsf{F}(x) - \mathsf{F}(y))\|^2 \\
&= \|x-y\|^2 + 2\alpha \,\langle\!\langle \mathsf{F}(x) - \mathsf{F}(y), x - y \rangle\!\rangle + \alpha^2 \|\mathsf{F}(x) - \mathsf{F}(y)\|^2 \\
&\leq \big(1 - 2\alpha c + \alpha^2 \ell^2\big) \|x - y\|^2
\end{aligned}$$

Next, study convex parabola $\alpha \mapsto 1 - 2\alpha c + \alpha^2\ell^2$. Eg, $1 - 2\alpha c + \alpha^2 \ell^2 < 1$ iff $0 < \alpha < 2c/\ell^2$

E. K. Ryu and S. Boyd. Primer on monotone operator methods. *Applied Computational Mathematics*, 15(1):3–43, 2016

Optimal* contractivity of Euler discretization $\mathsf{Id} + \alpha \mathsf{F}$: nonEuclidean $\|\cdot\|_{\infty,\mathrm{diag}(\eta)^{-1}}$, $\|\cdot\|_{1,\mathrm{diag}(\eta)}$

Let $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ be differentiable and Lipschitz

define *contraction rate* $c := -\,\mathsf{osLip}(\mathsf{F}) > 0$

define *diagonal Lipschitz constant* $\ell_{\mathsf{diag}} = \max_{i \in \{1,\dots,n\}} \sup_{x \in \mathbb{R}^n} |D\mathsf{F}_{ii}(x)|$; can show $\ell_{\mathsf{diag}} \geq c$

1. $$0 < \alpha \leq \frac{1}{\ell_{\mathsf{diag}}} \quad \implies \quad \mathsf{Lip}\big(\mathsf{Id} + \alpha \mathsf{F}\big) \leq 1 - \alpha c < 1$$

2. the optimal* step size and contraction factor are

$$\alpha^* = \frac{1}{\ell_{\mathsf{diag}}}, \qquad \mathsf{Lip}\big(\mathsf{Id} + \alpha^* \mathsf{F}\big) = 1 - \frac{c}{\ell_{\mathsf{diag}}}$$

Acceleration: (i) the condition number improves/diminishes $\kappa \geq \kappa_\infty := \frac{c}{\ell_{\mathsf{diag}}}$, and (ii) $\mathsf{Lip}(\mathsf{Id} + \alpha^* \mathsf{F}) = 1 - \frac{1}{4\kappa^2} + \mathcal{O}(\frac{1}{\kappa^4})$ improves/decreases to $\mathsf{Lip}(\mathsf{Id} + \alpha^* \mathsf{F}) = 1 - \frac{1}{\kappa_\infty}$.

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.

Proof of ℓ_∞/ℓ_1 Euler discretization theorem

For every $A=[a_{ij}]\in\mathbb{R}^{n\times n}$, $\eta\in\mathbb{R}^n_{>0}$, and $\alpha\in\mathbb{R}$ such that $|\alpha|\le(\max_i|a_{ii}|)^{-1}$, **norm=lognorm identity**:

$$\|I_n+\alpha A\|_{1,\mathrm{diag}(\eta)}=1+\alpha\mu_{1,\mathrm{diag}(\eta)}(A),\qquad \|I_n+\alpha A\|_{\infty,\mathrm{diag}(\eta)^{-1}}=1+\alpha\mu_{\infty,\mathrm{diag}(\eta)^{-1}}(A),\tag{12}$$

whose proof is an algebraic exercise (hint: diagonal of $I_n+\alpha A$ is nonnegative).

Next, consider $\|\cdot\|_{\infty,\mathrm{diag}(\eta)^{-1}}$; the proof for $\|\cdot\|_{1,\mathrm{diag}(\eta)}$ is omitted. Regarding part 1, for each $i\in\{1\dots,n\}$ and $x\in\mathbb{R}^n$

$$\begin{aligned}\ell_{\mathsf{diag}}&=\max_{i\in\{1,\dots,n\}}\sup_{x\in\mathbb{R}^n}|D\mathsf{F}_{ii}(x)|\overset{(\mathsf{osLip}(\mathsf{F})<0\implies D\mathsf{F}_{ii}(x)<0)}{=}\max_{i\in\{1,\dots,n\}}\sup_{x\in\mathbb{R}^n}\big(-D\mathsf{F}_{ii}(x)\big),\\ &\ge\max_{i\in\{1,\dots,n\}}\sup_{x\in\mathbb{R}^n}\Big(-D\mathsf{F}_{ii}(x)-\sum\nolimits_{j\neq i}|D\mathsf{F}_{ij}(x)|\tfrac{\eta_i}{\eta_j}\Big)\\ &=-\max_{i\in\{1,\dots,n\}}\sup_{x\in\mathbb{R}^n}\mu_{\infty,\mathrm{diag}(\eta)^{-1}}(D\mathsf{F}(x))=-\mathsf{osLip}(\mathsf{F})=c.\end{aligned}$$

Since $\ell_{\mathsf{diag}}=\sup_x\max_i|D\mathsf{F}_{ii}(x)|\ge\max_i|D\mathsf{F}_{ii}(x)|$ for all x and $\alpha\le\frac{1}{\ell_{\mathsf{diag}}}\le\frac{1}{\max_i|D\mathsf{F}_{ii}(x)|}$, equation (12) implies

$$\|I_n+\alpha D\mathsf{F}(x)\|_{\infty,\mathrm{diag}(\eta)^{-1}}=1+\alpha\mu_{\infty,\mathrm{diag}(\eta)^{-1}}(D\mathsf{F}(x))\le 1+\alpha\,\mathsf{osLip}(\mathsf{F})=1-\alpha c.$$

Finally, $\mathsf{Lip}(\mathsf{Id}+\alpha\mathsf{F})\le\sup_x\|I_n+\alpha D\mathsf{F}(x)\|_{\infty,\mathrm{diag}(\eta)^{-1}}\le 1-\alpha c$.

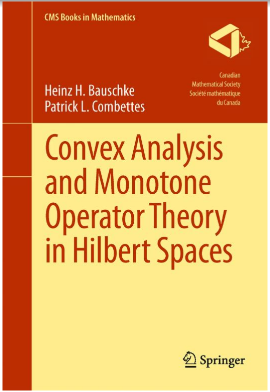
Regarding part 2, $\alpha\to\mathsf{Lip}(\mathsf{Id}+\alpha\mathsf{F})$ is decreasing and therefore minimum at the maximum of allowable value of α. Note that $\alpha^*=\ell_{\mathsf{diag}}^{-1}$ is the maximum value of α and $\mathsf{Lip}(\mathsf{Id}+\alpha^*\mathsf{F})=1-c/\ell_{\mathsf{diag}}>0$ since $c/\ell_{\mathsf{diag}}\le 1$.

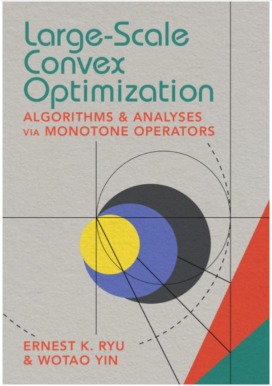
Outline

Monotone operator methods

Success in many disparate fields

1. Optimization and control
   - Subdifferentials are monotone
2. Game theory
   - Monotone games
3. Systems analysis
   - Input-output behavior
4. Machine learning





L. Pavel. Distributed GNE seeking under partial-decision information over networks via a doubly-augmented operator splitting approach. *IEEE Transactions on Automatic Control*, 65(4):1584–1597, 2020.

P. L. Combettes and J.-C. Pesquet. Fixed point strategies in data science. *IEEE Transactions on Signal Processing*, 69:3878–3905, 2021.

E. Winston and J. Z. Kolter. Monotone operator equilibrium networks. In *Advances in Neural Information Processing Systems*, 2020. URL **https://arxiv.org/abs/2006.08591**

A. Davydov, S. Jafarpour, A. V. Proskurnikov, and F. Bullo. Non-Euclidean monotone operator theory with applications to recurrent neural networks. In *IEEE Conf. on Decision and Control*, Cancún, México, Dec. 2022b.

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. . URL **http://jmlr.org/papers/v25/23-0805.html**

Background on monotone operators

operator $A : \mathbb{R}^n \to \mathbb{R}^n$ is **monotone** with parameter $m \geq 0$ if

$$\langle\langle A(x) - A(y), x - y \rangle\rangle \geq m\|x - y\|_2^2 \qquad (\mathrm{osLip}_2(-A) \leq -m)$$

A **monotone inclusion problem** is of the form

$$\text{find } x \in \mathbb{R}^n \text{ s.t. } 0 \in A(x)$$

A **monotone splitting problem** is of the form

$$\text{find } x \in \mathbb{R}^n \text{ s.t. } 0 \in (A + B)(x)$$

Existing algorithms based on Banach contractions or Krasnosel'skii–Mann iterations:

- Forward step method, proximal-point algorithm, etc.
- Forward-backward splitting, Peaceman-Rachford splitting, etc.

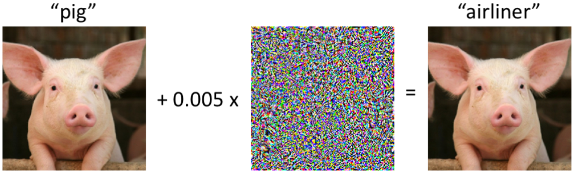
E. K. Ryu and S. Boyd. Primer on monotone operator methods. *Applied Computational Mathematics*, 15(1):3–43, 2016

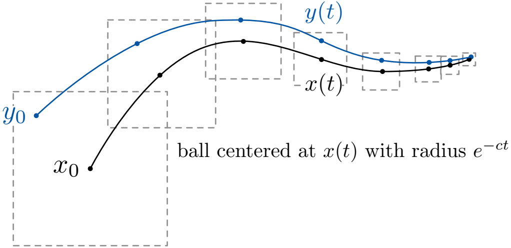
Why non-Euclidean?

Algorithms for inclusions and splittings are limited to Hilbert settings

Many problems are better stated in **Banach spaces**!

1. ℓ_∞ robustness analysis of neural networks
2. L_∞ norm systems analysis
3. Non-Euclidean contracting dynamics
4. Totally asynchronous distributed optimization





A differentiable $\mathsf{F}: \mathbb{R}^n \to \mathbb{R}^n$ is **strongly monotone w.r.t $\|\cdot\|$ with parameter m** if

$$-\mu(-D\mathsf{F}(x)) \geq m, \quad \forall x \in \mathbb{R}^n. \qquad (\mathsf{osLip}(-\mathsf{F}) \leq -m)$$

The **resolvent** and **reflected resolvent** of F with parameter $\alpha > 0$ are given by:

$$\mathsf{J}_{\alpha\mathsf{F}} := (\mathsf{Id} + \alpha\mathsf{F})^{-1}, \qquad \mathsf{R}_{\alpha\mathsf{F}} := 2\mathsf{J}_{\alpha\mathsf{F}} - \mathsf{Id}$$

Fixed points of $\mathsf{J}_{\alpha\mathsf{F}}$ and $\mathsf{R}_{\alpha\mathsf{F}}$ correspond to zeros of F

Lipschitz constants: Suppose F is monotone w.r.t. a diagonally-weighted ℓ_1/ℓ_∞ norm

$$\mathsf{Lip}(\mathsf{J}_{\alpha\mathsf{F}}) = \frac{1}{1+\alpha m}, \quad \forall \alpha > 0$$
$$\mathsf{Lip}(\mathsf{R}_{\alpha\mathsf{F}}) = \frac{1-\alpha m}{1+\alpha m}, \quad \forall \alpha \in \left]0, \mathsf{diagL}(\mathsf{F})^{-1}\right]$$

$\mathsf{diagL}(\mathsf{F}) = \sup_{x\in\mathbb{R}^n} \max_{i\in\{1,\dots,n\}} (D\mathsf{F}(x))_{ii}$

Non-Euclidean monotone inclusions

Monotone inclusion problem $\mathsf{F}(x) = 0$

The **forward step method** of F (ℓ_1/ℓ_∞ monotone) is the iteration

$$x_{k+1} = (\mathsf{Id} - \alpha\mathsf{F})(x_k)$$

1. if $m > 0$, $\|x_{k+1} - x^*\| \leq (1 - \alpha m)\|x_k - x^*\|, \qquad \forall \alpha \in]0, \mathsf{diagL}(\mathsf{F})^{-1}]$
2. if $m = 0$ and $\mathrm{zero}(\mathsf{F}) \neq \emptyset$, then convergence to an element of $\mathrm{zero}(\mathsf{F})$ with rate $\mathcal{O}(1/\sqrt{k})$

The **proximal point method** of F (ℓ_1/ℓ_∞ monotone) is the iteration

$$x_{k+1} = \mathsf{J}_{\alpha\mathsf{F}}(x_k)$$

1. if $m > 0$, $\|x_{k+1} - x^*\| \leq \dfrac{1}{1 + \alpha m}\|x_k - x^*\|, \qquad \forall \alpha > 0$
2. if $m = 0$ and $\mathrm{zero}(\mathsf{F}) \neq \emptyset$, then convergence to an element of $\mathrm{zero}(\mathsf{F})$ with rate $\mathcal{O}(1/\sqrt{k})$

Comparison to standard convergence rates

Algorithm	F strongly monotone and globally Lipschitz			
	ℓ_2		Diagonally weighted ℓ_1 or ℓ_∞	
	α range	Optimal Lip	α range	Optimal Lip
Forward step	$\left]0, \frac{2m}{\ell^2}\right[$	$1 - \frac{1}{2\kappa^2} + \mathcal{O}\left(\frac{1}{\kappa^3}\right)$	$\left]0, \frac{1}{\mathsf{diagL}(\mathsf{F})}\right]$	$1 - \frac{1}{\kappa_\infty}$
Proximal point	$]0, \infty[$	N/A	$]0, \infty[$	N/A
Cayley method	$]0, \infty[$	$1 - \frac{1}{2\kappa} + \mathcal{O}\left(\frac{1}{\kappa^2}\right)$	$\left]0, \frac{1}{\mathsf{diagL}(\mathsf{F})}\right]$	$1 - \frac{2}{\kappa_\infty} + \mathcal{O}\left(\frac{1}{\kappa_\infty^2}\right)$

Step size ranges and Lipschitz constants for algorithms for finding zeros of monotone operators. $\kappa := \ell/m \geq 1$ and $\kappa_\infty := \mathsf{diagL}(\mathsf{F})/m \in [1, \kappa]$

Non-Euclidean operator splitting

Monotone splitting problem $(\mathsf{F}+\mathsf{G})(x)=0$

The **forward-backward splitting method** of F and G (ℓ_1/ℓ_∞ monotone) is

$$x_{k+1} = (\mathsf{J}_{\alpha\mathsf{G}} \circ (\mathsf{Id} - \alpha\mathsf{F}))(x_k)$$

1. if F s.m., $m>0$, $\|x_{k+1}-x^*\| \leq (1-\alpha m)\|x_k - x^*\|, \quad \forall\alpha \in\,]0, \mathsf{diagL}(\mathsf{F})^{-1}]$
2. if $m=0$ and $\mathrm{zero}(\mathsf{F}+\mathsf{G}) \neq \emptyset$, then iteration converges to an element of $\mathrm{zero}(\mathsf{F}+\mathsf{G})$ with rate $\mathcal{O}(1/\sqrt{k})$

The **Peaceman-Rachford splitting method** of F and G (ℓ_1/ℓ_∞ monotone) is

$$\begin{aligned} x_{k+1} &= \mathsf{J}_{\alpha\mathsf{G}}(z_k), \\ z_{k+1} &= z_k + 2\mathsf{J}_{\alpha\mathsf{F}}(2x_{k+1}-z_k) - 2x_{k+1}. \end{aligned}$$

1. If F s.m., $m>0$,

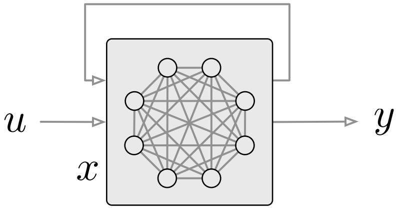
$$\|x_{k+1}-x^*\| \leq \frac{1-\alpha m}{1+\alpha m}\|x_k - x^*\|, \quad \forall\alpha \in\,]0, \min\{\mathsf{diagL}(\mathsf{F})^{-1}, \mathsf{diagL}(\mathsf{G})^{-1}\}]$$

Application to recurrent neural networks

Equilibrium computation of RNN

$$\dot{x} = -x + \Phi(Ax + Bu + b) =: F(x, u)$$

$$\Phi(x) = \mathrm{LeakyReLU}(x) = \max\{x, ax\}$$



A sufficient condition for contractivity is $\mu_\infty(A) = \gamma < 1$.

In this case, $-F(x, u)$ is strongly monotone and can apply **forward step method**

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

converges for $\alpha \in]0, \alpha^*]$ with linear convergence rate $1 - \alpha(1 - \Phi(\gamma))$

$\alpha^* = (1 - \min_{i \in \{1,\dots,n\}} \min\{a \cdot (A)_{ii}, (A)_{ii}\})^{-1}$

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 2022c.

Splitting methods for equilibrium computation

Finding an equilibrium point $x^*(u)$ is equivalent to $(\mathsf{F}+\mathsf{G})(x^*(u))=0$ where

$$\mathsf{F}(z)=(I_n-A)z-Bu-b,\quad \mathsf{G}(z)=\frac{1-a}{a}\min\{z,0\}$$

Apply **forward-backward splitting**

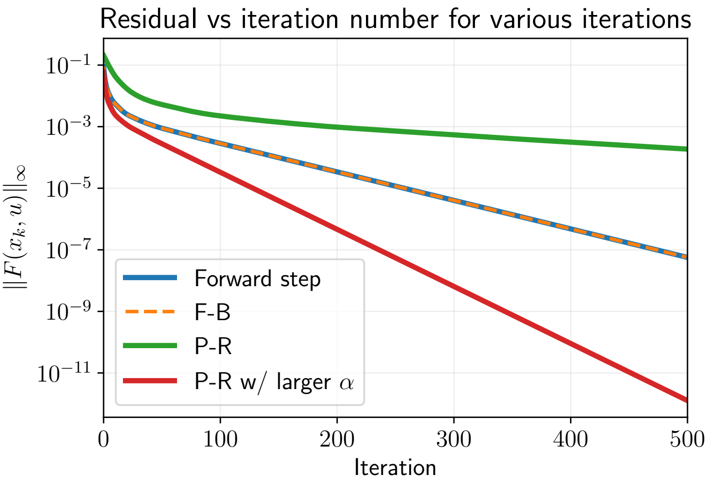
$$x_{k+1}=\mathsf{J}_{\alpha\mathsf{G}}((1-\alpha)x_k+\alpha(Ax_k+Bu+b)).$$

Converges with rate $1-\alpha(1-\gamma)$ for $\alpha\in\,]0,(1-\min_i(A)_{ii})^{-1}]$

Apply **Peaceman-Rachford splitting**

$$x_{k+1}=(I_n+\alpha(I_n-A))^{-1}(z_k+\alpha(Bu+b)),$$
$$z_{k+1}=z_k+2\mathsf{J}_{\alpha\mathsf{G}}(2x_{k+1}-z_k)-2x_{k+1}.$$

Converges with rate $\dfrac{1-\alpha(1-\gamma)}{1+\alpha(1-\gamma)}$ for $\alpha\in\,\left]0,\min\left\{(1-\min_i(A)_{ii})^{-1},\dfrac{a}{1-a}\right\}\right]$,



We generate $A \in \mathbb{R}^{200\times 200}, B \in \mathbb{R}^{200\times 50}, b \in \mathbb{R}^{200}, u \in \mathbb{R}^{50}$ with entries according to a normal distribution and then project A so that $\mu_\infty(A) \leq 0.99$

Conclusions

Summary:

1. provide a transcription of monotone operator theory for non-Euclidean norms
2. provable convergence of classical iterations for monotone inclusions and splittings
3. application to continuous-time recurrent neural network

Extensions and open problems:

1. tightening Lipschitz estimates for operator splittings
2. infinite-dimensional Banach spaces and set-valued F
3. further applications to systems analysis and neural networks

Thank you for reading!

For any questions, please do not hesitate to email me

Let F denote a ℓ_1 weakly-contracting analytic vector field on a subset C of $\mathbb{R}^n$. Assume there exists a bounded solution $x(\cdot)$ in C of $\dot{x} = \mathsf{F}(x)$ defined for all $t \in \mathbb{R}_{\geq 0}$. If the function $t \mapsto \|\mathsf{F}(x(t))\|_1$ is constant, then the solution $x(\cdot)$ is an equilibrium of F, that is, $x(t) = x^*$ for all t and $\mathsf{F}(x^*) \equiv 0$.

Proof For simplicity take $n = 2$. By analyticity, and unless $\|f(x(t))\|_1$ is identically zero (in case we are done), we can pick an interval J where both $f_i(x(t))$ have no zeroes, and hence a constant sign. (If one is identically zero, the proof is the same ignoring that variable.) Without loss of generality (take $-f_i$ if necessary), assume that both have positive sign, so $\|f(x(t))\|_1 = f_1(x(t)) + f_2(x(t)) = \frac{dx_1}{dt} + \frac{dx_2}{dt} = \frac{d(x_1+x_2)}{dt}$. Since $\|f(x)\|_1$ is constant, this means that $\frac{d(x_1+x_2)}{dt} \equiv c$ on the interval, and therefore $x_1(t) + x_2(t) = ct + b$ on the interval J. By analytic continuation, this is true for all $t \in \mathbb{R}_{\geq 0}$, contradicting boundedness of $x(\cdot)$ unless $c = 0$. So we have that $\frac{d(x_1+x_2)}{dt} \equiv 0$, that is, $\|f(x(t))\|_1 \equiv 0$, as desired. The proof for ℓ_∞ norm is even easier - just take an interval where one of the two terms is max.

How to train an implicit layer

An implicit layer is defined by a fixed point equation

$$x^\star = f_\theta(x^\star, u). \tag{13}$$

Forward evaluation means: given u find $x^\star$. In practice one runs a solver until convergence (Picard, Anderson acceleration, quasi–Newton).
We train via the implicit function theorem. Let the loss be $L(x^\star, \theta)$ and define

$$g(x, \theta) = x - f_\theta(x). \tag{14}$$

At solution $g(x^\star, \theta) = 0$ and the gradient is

$$\frac{\partial L}{\partial \theta} = \frac{\partial L}{\partial x^\star}\Big(I - \frac{\partial f_\theta}{\partial x}\Big)^{-1}\frac{\partial f_\theta}{\partial \theta}. \tag{15}$$

We never form the inverse explicitly. We only apply it as a Jacobian–solve: in reverse mode one solves a second linear fixed point problem to compute the vector–Jacobian product.
Training loop:

- forward: solve $x^\star$ from the fixed point
- backward: solve the adjoint fixed point to apply $\big(I - \partial_x f_\theta\big)^{-1}$
- update θ with SGD/Adam

Good practice: parameterize f_θ such that it is contractive or strongly monotone, so that both forward fixed point and adjoint fixed point converge fast and robustly.

We define g because the implicit function theorem is written for $g(x,\theta) = 0$ rather than $x = f_\theta(x)$. Take

$$g(x,\theta) = x - f_\theta(x). \tag{16}$$

A solution $x^\star$ satisfies $g(x^\star,\theta) = 0$. The implicit function theorem says: if $D_x g$ at $(x^\star,\theta)$ is invertible, then locally one can view $x^\star$ as a differentiable function of θ, and

$$\frac{\partial x^\star}{\partial \theta} = -\big(D_x g\big)^{-1} D_\theta g. \tag{17}$$

Now

$$D_x g = I - D_x f_\theta \quad \text{and} \quad D_\theta g = -D_\theta f_\theta. \tag{18}$$

Thus

$$\frac{\partial x^\star}{\partial \theta} = \big(I - D_x f_\theta\big)^{-1} D_\theta f_\theta. \tag{19}$$

For the loss $L(x^\star,\theta)$ one gets

$$\frac{\partial L}{\partial \theta} = \frac{\partial L}{\partial x^\star}\big(I - D_x f_\theta\big)^{-1} D_\theta f_\theta. \tag{20}$$

In practice we never form the inverse. Instead we solve the linear system

$$\big(I - D_x f_\theta\big)v = \frac{\partial L}{\partial x^\star} \quad \text{for } v. \quad \text{Then:} \quad \frac{\partial L}{\partial \theta} = v\, D_\theta f_\theta. \tag{21}$$

That linear solve is itself done by a fixed point iteration or Krylov. It is efficient because its dimension is only that of the layer state. The computational cost of backward is therefore two fixed point solves: one in forward to get $x^\star$, one in backward to get v.