University of California
Santa Barbara

Contraction Theory in Control, Learning, and Optimization

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Mechanical Engineering

by

Alexander Davydov

Committee in charge:

Professor Francesco Bullo, Chair
Professor Bassam Bamieh
Professor Jeff Moehlis
Professor João Hespanha

June 2025

The Dissertation of Alexander Davydov is approved.

Professor Bassam Bamieh

Professor Jeff Moehlis

Professor João Hespanha

Professor Francesco Bullo, Committee Chair

June 2025

Contraction Theory in Control, Learning, and Optimization

To my friends and family who have walked beside me

Acknowledgements

> Don't pointless things have a place, too, in this far-from-perfect world? Remove everything pointless from an imperfect life, and it'd lose even its imperfection.
>
> *Haruki Murakami, Sputnik Sweetheart*

I once heard that the most read part of a person's Ph.D. thesis is the acknowledgments sections. For that reason, I provide a rather comprehensive list of acknowledgments.

First, I would like to thank my advisor, Francesco Bullo. Francesco welcomed me to Santa Barbara with open arms despite the pandemic. Francesco has always been a constant source of inspiration, has encouraged me to pursue my own research interests, and has given me a lot of flexibility in my scholarly pursuits. At the same time, he has always been available when I have needed guidance or advice. He has introduced me to countless colleagues and collaborators and words cannot express my gratitude to him.

Second, I would like to thank my committee, Bassam Bamieh, Jeff Moehlis, and João Hespanha. Bassam has always been generous with his time and has inspired me to think about fundamental tradeoffs and issues in control. He taught one of my favorite classes that I took at UCSB: Robust Control Theory. Jeff has always been willing to discuss anything technical or non-technical with me including Nobel Prizes and rock and metal music. He also provided me with key insights in my academic job search and what to look for in an academic department. João has an amazing ability in asking great clarifying questions. In addition to teaching me linear systems theory, he taught me clarity and how to ask good questions. Thank you.

Third, I would like to thank the researchers at the University of Maryland and Johns Hopkins University Applied Physics Laboratory I interacted with before UCSB who encouraged me to pursue a Ph.D. Specifically, thank you to Phillip, Greg, Xiaotian, and Sara. Special shout outs go to Derrick Yeo for introducing me to research and encouraging me to pursue controls as part of the First-Year Innovation & Research Experience (FIRE) program and to Yancy Diaz-Mercado for taking me on as an undergraduate researcher, introducing me to network controls, and recommending me to work with Francesco.

Fourth, I would like to thank friends and colleagues who have been part of Francesco's lab during my tenure at UCSB. Specifically, thank you Saber, Robin, Xiaoming, Liz H., Pedro, Kevin S., Francesco S., Sean J., Veronica, Giulia, Marco, Emilio, Jonas, Ryotaro, Ruben, Simone, Gilberto, Anand, Yohan, Abed, Madelyn, Arie, Arvind, Shiqi, Zahra M., Kittiphat, Wes, and Cole. My time in the lab would have been much less enjoyable without you all here.

Fifth, I would like to thank my colleagues who are part of the Center for Control, Dynamical Systems, and Computation. Namely, thank you Sean A., Bryce, Vade, Rohit, Keith, Michael B., Gowtham, Karthik, and Poorva. I have had insightful conversations with all of you and all of you made me look forward to the control seminars on Fridays.

Sixth, I would like to thank all of my friends at UCSB who have helped keep me sane outside of my time doing research. Whether it was trivia, running, surfing, tennis, or going wine tasting, you all have helped me stay grounded and reminded me to maintain a healthy work-life balance. Big thank you to Connor, Max E., Charlie, Nibodh, Mathias, PK, Asish, Spencer, Ozgur, Brandon, Kaylaa, Ursula, Alex L., Marianna, Alex Z., Sean B., Jared, Daniella, Gabby, Amber, Anderson, Juliette, Dixon, Kat, Liz S., Arturo, Caelin, Effrain, Kaley, Huck, Kana, and Ricky. Extra big shout outs go to my past and current housemates Will, Joe, Nick, Max L., Rajan, Michael M., Feifei, Katy, and Ben for tolerating me outside of work.

Seventh, I would like to thank everyone else in the Mechanical Engineering Department at UCSB who have made the logistics surrounding my Ph.D. as frictionless as possible. Thank you Katelynn, Lauren, and Laura for always being available when I had logistical questions or wanted to reserve a room for an interview. Thank you also to Bolin, Enoch, and Igor for insightful conversations along the way.

Eighth, I would like to thank the researchers at Toyota Research Institute for taking me on as a summer intern in 2024. I had limited experimental experience and was very lucky to have the opportunity to implement learning-based control with MPC on Keisuke, a full-size autonomously drifting Toyota Supra. Thank you to Thomas, Franck, Marcus, John, Makoto, and Michael T. for taking a chance on me.

Ninth, I would like to thank colleagues whom I met through collaborations or conferences who have provided me with insightful research and/or career advice. Specifically, thank you Emiliano, Matt H., Ambuj, Mihailo, Giovanni, Sam, Anton, Eduardo, Ian, César, Daniele, Zahra A., Daigo, Matt A., Ibrahim, Pol, Ahmed, Carmel, Leo, and Ellie.

Tenth, I would like to thank my friends from high school and University of Maryland who I continue to stay in contact with and have been rooting for me this whole way. Big thank you to Kevin Q., Benji, Janet, and Bryn. You all provide me with a constant source of inspiration and I'm lucky to stay in touch with you all.

Eleventh, a special thank you goes to Nathalie. A paragraph alone does not do justice for all the support you have provided over the course of my Ph.D. and job search. Thank you for all the countless hours we have shared, walks around the campus lagoon and Ellwood, Morro Bay, San Diego, and Houston :)

Last, but certainly not least, thank you Mama, Andrei (and Magnus and Coco) for providing me with constant support throughout the entirety of my education. This journey would not have been possible without you both in the background.

Curriculum Vitæ
Alexander Davydov

Education

2025	**Ph.D., Mechanical Engineering** University of California, Santa Barbara.
2020	**B.S., Mechanical Engineering** University of Maryland.
2020	**B.S., Mathematics** University of Maryland.

Research Experience

2020-2025	**UCSB Chancellor's Fellow** University of California, Santa Barbara.
Summer 2024	**ML & Controls Intern** Toyota Research Institute.
2021-2024	**NSF Graduate Research Fellow** University of California, Santa Barbara.
Summer 2020	**Technical Intern** Johns Hopkins University Applied Physics Laboratory.
2018-2020	**Undergraduate Researcher** University of Maryland.

Awards

2024	**IEEE Control Systems Letters Outstanding Paper Award**
2023	**AACC O. Hugo Schuck Best Paper Award**
2022	**ACC Best Student Paper Award Finalist**
2021	**NSF Graduate Research Fellowship**
2020	**UCSB Chancellor's Fellowship**

Professional Service

Reviewer	*IEEE Transactions on Automatic Control (T-AC)* *IEEE Transactions on Control of Network Systems (T-CNS)* *IEEE Open Journal of Control Systems (OJ-CSYS)* *Automatica* *SIAM Journal on Mathematics of Data Science (SIMODS)*

Journal of Machine Learning Research (JMLR)
IEEE Control Systems Letters (L-CSS)
European Journal of Control
IEEE Conference on Decision and Control (CDC)
American Control Conference (ACC)
IEEE International Conference on Robotics and Automation (ICRA)
Conference on Learning for Dynamics & Control (L4DC)
Modeling, Estimation and Control Conference (MECC)
ICLR Workshop on New Frontiers in Associative Memory (NFAM)

Invited Seminars

2025	**University of California, Riverside** Department of Electrical and Computer Engineering.
2025	**Rice University** Department of Mechanical Engineering.
2025	**McMaster University** Department of Mathematics.
2024	**University of Maryland** Department of Mechanical Engineering.
2024	**University of California, Berkeley** Department of Electrical Engineering and Computer Sciences.

Journal Publications

[1] 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, 2025. To Appear

[2] A. Davydov, A. V. Proskurnikov, and F. Bullo, "Non-Euclidean contraction analysis of continuous-time neural networks," IEEE Transactions on Automatic Control, vol. 70, no. 1, pp. 235–250, 2025.

[3] A. Davydov and F. Bullo, "Exponential stability of parametric optimization-based controllers via Lur'e contractivity," IEEE Control Systems Letters, vol. 8, pp. 1277–1282, 2024.

[4] A. Davydov and F. Bullo, "Perspectives on contractivity in control, optimization, and learning," IEEE Control Systems Letters, vol. 8, pp. 2087–2098, 2024.

[5] A. Davydov, S. Jafarpour, A. V. Proskurnikov, and F. Bullo, "Non-Euclidean monotone operator theory and applications," Journal of Machine Learning Research, vol. 25, no. 307, pp. 1–33, 2024.

[6] V. Centorrino, A. Davydov, A. Gokhale, G. Russo, and F. Bullo, “On weakly contracting dynamics for convex optimization,” IEEE Control Systems Letters, vol. 8, pp. 1745–1750, 2024.

[7] V. Centorrino, A. Gokhale, A. Davydov, G. Russo, and F. Bullo, “Positive competitive networks for sparse reconstruction,” Neural Computation, vol. 36, no. 6, pp. 1163–1197, 2024.

[8] A. Gokhale, A. Davydov, and F. Bullo, “Proximal gradient dynamics: Monotonicity, exponential convergence, and applications,” IEEE Control Systems Letters, pp. 2853–2858, 2024.

[9] V. Centorrino, A. Gokhale, A. Davydov, G. Russo, and F. Bullo, “Euclidean contractivity of neural networks with symmetric weights,” IEEE Control Systems Letters, vol. 7, pp. 1724–1729, 2023.

[10] A. Gokhale, A. Davydov, and F. Bullo, “Contractivity of distributed optimization and Nash seeking dynamics,” IEEE Control Systems Letters, vol. 7, pp. 3896–3901, 2023.

[11] S. Jafarpour, A. Davydov, and F. Bullo, “Non-Euclidean contraction theory for monotone and positive systems,” IEEE Transactions on Automatic Control, vol. 68, no. 9, pp. 5653–5660, 2023.

[12] A. V. Proskurnikov, A. Davydov, and F. Bullo, “The Yakubovich S-Lemma revisited: Stability and contractivity in non-Euclidean norms,” SIAM Journal on Control and Optimization, vol. 61, no. 4, pp. 1955–1978, 2023.

[13] A. Davydov, S. Jafarpour, and F. Bullo, “Non-Euclidean contraction theory for robust nonlinear stability,” IEEE Transactions on Automatic Control, vol. 67, no. 12, pp. 6667–6681, 2022.

[14] A. Davydov, P. Rivera-Ortiz, and Y. Diaz-Mercado, “Pursuer coordination in multiplayer reach-avoid games through control barrier functions,” IEEE Control Systems Letters, vol. 5, no. 6, pp. 1910–1915, 2021.

[15] A. Davydov and Y. Diaz-Mercado, “Sparsity structure and optimality of multi-robot coverage control,” IEEE Control Systems Letters, vol. 4, no. 1, pp. 13–18, 2020.

Conference Proceeding Publications

[1] S. Jaffe, A. Davydov, D. Lapsekili, A. K. Singh, and F. Bullo, “Learning neural contracting dynamics: Extended linearization and global guarantees,” in Advances in Neural Information Processing Systems, 2024

[2] X. Xu, A. Davydov, and Y. Diaz-Mercado, “On the equivalence of multi-agent 2d coverage control and leader-follower consensus network,” in American Control Conference, 2023, pp. 503–508.

[3] A. Davydov, S. Jafarpour, M. Abate, F. Bullo, and S. Coogan, "Comparative analysis of interval reachability for robust implicit and feedforward neural networks," in IEEE Conference on Decision and Control, 2022, pp. 2073–2078.

[4] A. Davydov, S. Jafarpour, A. V. Proskurnikov, and F. Bullo, "Non-Euclidean monotone operator theory with applications to recurrent neural networks," in IEEE Conference on Decision and Control, 2022, pp. 6332–6337.

[5] A. Davydov, A. V. Proskurnikov, and F. Bullo, "Non-Euclidean contractivity of recurrent neural networks," in American Control Conference, 2022, pp. 1527–1534.

[6] S. Jafarpour, M. Abate, A. Davydov, F. Bullo, and S. Coogan, "Robustness certificates for implicit neural networks: A mixed monotone contractive approach," in Learning for Dynamics and Control Conference, 2022, pp. 917–930.

[7] 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 Conference on Decision and Control, 2021, pp. 2923–2928.

[8] S. Jafarpour, A. Davydov, A. V. Proskurnikov, and F. Bullo, "Robust implicit networks via non-Euclidean contractions," in Advances in Neural Information Processing Systems, 2021.

[9] A. Davydov, D. W. Yeo, and Y. Diaz-Mercado, "Low-mobility atmospheric sensing via multi-vehicle adaptive coverage control," in AIAA Aviation Forum, 2020, p. 2821.

Publications Under Review

[1] A. Davydov, F. Djeumou, M. Greiff, M. Suminaka, M. Thompson, J. Subosits, and T. Lew, "First, learn what you don't know: Active information gathering for driving at the limits of handling," arXiv:2411.00107, 2025.

[2] R. Shima, A. Davydov, and F. Bullo, "Contractivity analysis and control design for Lur'e systems: Lipschitz, incrementally sector bounded, and monotone nonlinearities," arXiv:2503.20177, 2025.

Abstract

Contraction Theory in Control, Learning, and Optimization

by

Alexander Davydov

Contraction is a strong notion of stability for a dynamical system. It ensures that the distance between any two trajectories decays exponentially quickly in some metric. As a corollary of this property, contracting systems (i) are automatically input-to-state stable in the presence of disturbances, (ii) are robust to time-delays in the dynamics, (iii) entrain to a unique exponentially stable limit cycle when they are forced by a periodic input, and (iv) satisfy certain composition properties, among others. To be specific, in this thesis, we study the application of contraction theory in control, machine learning, and optimization.

In Chapter 2, we develop a novel theoretical framework for contraction analysis with respect to non-Euclidean norms based on weak pairings and one-sided Lipschitz constants. We show how we can use weak pairings to readily establish desirable consequences of contracting dynamical systems. In Chapter 3, we present a non-Euclidean monotone operator theory in analogy to the standard monotone operator theory on Hilbert spaces. This theory showcases how one can devise fixed point algorithms for computing zeros of non-Euclidean monotone operators including contracting dynamical systems. In Chapter 4, we present an application to machine learning where we design an implicit neural network and show how we can use contraction theory to compute fixed points and establish explicit Lipschitz estimates. In Chapter 5, we extend the work of Chapter 4 to more general neural dynamics including Hopfield and firing rate neural networks and provide necessary and sufficient conditions for them to be contracting with respect to a non-

Euclidean norm. In Chapter 6, we show how we can use contracting dynamics to derive novel tracking error bounds for time-varying dynamical systems and how to design feed-forward corrections to attain zero tracking error. We apply these results to time-varying convex optimization problems. Finally, in Chapter 7, we show how to leverage the virtual system method from contraction theory to establish when a linear time-invariant system with an optimization-based controller is exponentially stable.

Contents

Chapter 1

Introduction

"Begin at the beginning," the King said, very gravely, "and go on till you come to the end: then stop."

Lewis Carroll,
Alice in Wonderland

1.1 Motivation and outline of the thesis

Due to the underlying complexity of modern engineering systems, many engineers have begun relying upon optimization and machine learning-based methods for controlling them. For example, in autonomous driving and robot manipulation tasks, the underlying dynamics are complicated by the dependence on parameters that are difficult to estimate, e.g., tire forces in driving and complex contact geometries in robot manipulation. As a result, many state-of-the-art methods for controlling these complex systems rely upon methods such as model predictive control, reinforcement learning, imitation learning, or neural network feedback controllers. Despite the empirical successes of these methods, they are often quite costly in terms of data collection, require careful finetuning, and provide limited theoretical guarantees. Moreover, these methods are often described as “black-box,” i.e., there is limited interpretability in their predictions. Thus, these methods are prone to failures which ultimately violate stability and safety guarantees for the engineering system.

As a first step to establishing stability and safety guarantees for classes of engineering systems, we must define what properties we want our system to have. Namely, how do we know if our system has strayed too far from a nominal operating condition? How do we know that the model will be robust to exogenous disturbances? Traditionally, control theorists have studied global asymptotic stability or input-to-state stability as a desirable set of properties for engineering systems. Instead, in this thesis, I will argue that the system should be contracting in closed-loop. Contracting dynamical systems enjoy many desirable properties including incremental exponential stability, exponential input-to-state stability, robustness to uncertain and unmodeled dynamics, and they entrain to periodic inputs. As a consequence, contraction automatically implies the consequences of global asymptotic stability and input-to-state stability.

Motivated by the above, in this thesis, we study contraction theory, a mathematical framework for establishing the stability and robustness of nonlinear control systems. We say that a dynamical system is contracting if the distance between any two trajectories is decaying exponentially quickly in some metric. Although contraction is a property of a dynamical system, it has gained traction in the nonlinear control community as a methodology for establishing robustness bounds for closed-loop systems.

We focus our attention to systems which are contracting with respect to a metric induced by a norm. While this assumption may appear restrictive, we will show that there are many interesting classes of systems which satisfying this property. Namely, stable linear systems, certain dynamical neural networks, and dynamical systems solving convex optimization problems all satisfy this property. In this thesis, we make both novel theoretical contributions in understanding corollaries of contraction and in the application of contraction theory to neural networks, optimization, and control.

As an outline, in Chapter 2, we establish a mathematical framework for contraction in general, finite-dimensional normed vector spaces. We introduce the key concepts of weak pairings as a generalization of an inner product and the one-sided Lipschitz constant for arbitrary norms. With this language, we state 5 equivalent characterizations of contraction with respect to an arbitrary norm. Moreover, we prove incremental input-to-state stability for contracting dynamics and a sufficient condition for the interconnection of contracting dynamics to remain contracting.

In Chapter 3, we introduce a non-Euclidean monotone operator theory. While monotone operator theory on Hilbert spaces has been well-studied, the non-Euclidean extension has largely been missing. By leveraging weak pairings, we provide a natural generalization of monotone operator theory to vector spaced endowed with non-Euclidean norms.

In Chapter 4, we study a class of machine learning architectures referred to as deep equilibrium networks. These architectures are described via a fixed point equation. By leveraging a non-Euclidean contraction theory framework, we provide a novel sufficient condition for their well-posedness and establish an explicit estimate for their ℓ_∞ Lipschitz constant.

In Chapter 5, we extend the results of Chapter 4 to more general continuous-time neural networks. Specifically, we provide many sharp conditions for the non-Euclidean contrativity of neural networks. We primarily focus on firing rate and Hopfield neural networks but some additional classes of neural networks are considered as well.

In Chapter 6, we turn our attention to time-varying convex optimization. We demonstrate that many examples of dynamical systems solving convex optimization problems are in fact contracting with respect to a suitable Euclidean norm. We additionally prove equilibrium tracking error bounds for contracting dynamics and use them to establish rigorous tracking error bounds for time-varying convex optimization problems.

Finally, in Chapter 7, we study the interconnection of a linear system with a controller which is the solution to a parametric optimization problem. We provide a sufficient condition for exponential stability based on the contractivity of a Lur'e dynamical system and present two numerical experiments to support the claims.

1.2 Preliminaries and notation

In this section, we review some basic notation and preliminaries that will be used throughout the thesis.

The set of real numbers is denoted by $\mathbb{R}$, the set of positive numbers is denoted by $\mathbb{R}_{>0}$, the set of nonnegative numbers is denoted by $\mathbb{R}_{\geq 0}$ and the set of n-dimensional vectors with real (resp. positive or nonnegative) entries is denoted by $\mathbb{R}^n$ (resp. $\mathbb{R}^n_{>0}$ or $\mathbb{R}^n_{\geq 0}$). The set of $n \times m$ matrices with real entries is denoted by $\mathbb{R}^{n\times m}$. The extended real number line is denoted $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$, the set of integers is denoted by $\mathbb{Z}$, the set of nonnegative integers is denoted by $\mathbb{Z}_{\geq 0}$, and the set of complex number is denoted by $\mathbb{C}$. The real part of a complex number, z, is denoted $\mathrm{Re}(z)$. $\mathbb{0}_n \in \mathbb{R}^n$ is the n-dimensional vector with all entries equal to zero and $\mathbb{1}_n \in \mathbb{R}^n$ is the n-dimensional vector with all entries equal to one. For two vectors, v, w, we write $v \geq w$ if $v_i \geq w_i$ for all $i \in \{1, \dots, n\}$. We let I_n denote the $n \times n$ identity matrix. For a vector $\eta \in \mathbb{R}^n$, we will adopt the notations $\mathrm{diag}(\eta)$ or $[\eta]$ for the diagonal matrix whose diagonal entries equal the entries of η. For a vector or matrix, we use $(\cdot)^\top$ to denote its transpose. We say a symmetric matrix $A \in \mathbb{R}^{n\times n}$ is positive semidefinite (denoted $A \succeq 0$) if $x^\top Ax \geq 0$ for all $x \in \mathbb{R}^n$. A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is positive definite (denoted $A \succ 0$) if $x^\top Ax > 0$ for all $x \neq \mathbb{0}_n$. The matrix A is negative semidefinite (resp. negative definite) if $-A \succeq 0$ (resp. $-A \succ 0$). For two symmetric matrices $A, B \in \mathbb{R}^{n\times n}$ we write $A \preceq B$ (resp. $A \prec B$) if $B - A$ is positive semidefinite (resp. positive definite). For a square matrix $A \in \mathbb{R}^{n\times n}$, we let $\mathrm{spec}(A) \subseteq \mathbb{C}$ denote the spectrum of A, i.e., the set of its eigenvalues. For a square matrix A with all real eigenvalues, we let $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ denote the maximum and minimum eigenvalue, respectively. The *spectral abscissa* of a square matrix, A is $\alpha(A) = \max\{\mathrm{Re}(\lambda) \mid \lambda \in \mathrm{spec}(A)\}$. A square matrix is said to be *Metzler* if all of its off-diagonal entries are nonnegative.

For a set $\mathcal{S}$, let $2^{\mathcal{S}}$ denote its power set, $|\mathcal{S}|$ denote its cardinality, and, if $\mathcal{S} \subseteq \mathbb{R}^n$, $\mathrm{conv}(\mathcal{S})$ denote its convex hull.

Norms and logarithmic norms. We will let $\|\cdot\|$ denote an arbitrary norm on $\mathbb{R}^n$ and its corresponding induced norm on $\mathbb{R}^{n\times n}$, i.e., $\|A\| = \max_{\|x\|=1}\|Ax\|$. We say a norm $\|\cdot\|$ on $\mathbb{R}^n$ is *monotonic* if for all $x, y \in \mathbb{R}^n$, $|x| \leq |y| \implies \|x\| \leq \|y\|$. Of particular interest will be the ℓ_p norms with $p \in [1, \infty]$, denoted by $\|\cdot\|_p$ and are given by

$$\|x\|_p = \begin{cases} \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}, & p \neq \infty, \\ \max_{i\in\{1,\dots,n\}} |x_i|, & p = \infty. \end{cases} \tag{1.1}$$

We recall the standard ℓ_p induced norms for $p \in \{1, 2, \infty\}$:

$$\|A\|_2 = \sqrt{\lambda_{\max}(A^\top A)},$$
$$\|A\|_1 = \max_{j\in\{1,\dots,n\}} \sum_{i=1}^n |a_{ij}|, \quad \|A\|_\infty = \max_{i\in\{1,\dots,n\}} \sum_{j=1}^n |a_{ij}|.$$

where $\lambda_{\max}(A^\top A)$ is the largest eigenvalue of $A^\top A$.

We will also frequently consider *weighted norms*. Given a norm $\|\cdot\|$ and an invertible matrix $R \in \mathbb{R}^{n\times n}$, define the R-weighted ℓ_p norm, $\|\cdot\|_{p,R}$ by $\|x\|_{p,R} = \|Rx\|_p$. For $p \in \{1, \infty\}$, we will often restrict ourselves to diagonal weights with positive entries, i.e.,

$\|\cdot\|_{1,[\eta]}$ or $\|\cdot\|_{\infty,[\eta]^{-1}}$ where $\eta \in \mathbb{R}^n_{>0}$. For $p = 2$, we will frequently limit ourselves to positive definite weights and write $\|\cdot\|_{2,P^{1/2}}$ or $\|\cdot\|_P$ to mean the vector norm defined by $\|x\|_{2,P^{1/2}} = \|x\|_P = \sqrt{x^\top P x}$.

Let us now introduce the matrix logarithmic norm, sometimes referred to as matrix measure or log norm. For a given vector norm $\|\cdot\|$ on $\mathbb{R}^n$ and the induced matrix norm $\|\cdot\|$, the logarithmic norm is the map $\mu\colon \mathbb{R}^{n\times n} \to \mathbb{R}$ given by

$$\mu(A) = \lim_{h\to 0^+} \frac{\|I_n + hA\| - 1}{h}. \tag{1.2}$$

It is well known that this limit is well posed because the right-hand side is non-increasing in h, due to the convexity of the norm. For arbitrary $n \times n$ matrices A and B, the following properties hold:

sub-additivity: $$\mu(A+B) \le \mu(A) + \mu(B), \tag{1.3a}$$

weak homogeneity: $$\mu(\alpha A) = \alpha\mu(A), \ \ \forall \alpha \ge 0, \tag{1.3b}$$

convexity: $$\mu(\theta A + (1-\theta)B) \le \theta\mu(A) + (1-\theta)\mu(B), \ \ \forall \theta \in [0,1], \tag{1.3c}$$

norm/spectrum: $$-\|A\| \le -\mu(-A) \le \operatorname{Re}(\lambda) \le \mu(A) \le \|A\|, \ \ \forall \lambda \in \operatorname{spec}(A), \tag{1.3d}$$

translation: $$\mu(A + cI_n) = \mu(A) + c, \ \ \forall c \in \mathbb{R}, \tag{1.3e}$$

product: $$\max\{-\mu(A), -\mu(-A)\}\|x\| \le \|Ax\|, \ \ \forall x \in \mathbb{R}^n, \tag{1.3f}$$

norm of inverse: $$\mu(A) < 0 \implies \|A^{-1}\| \le -1/\mu(A). \tag{1.3g}$$

Note that convexity is an immediate consequence of sub-additivity and weak homogeneity. Additionally, by property (1.3d), the matrix measure is upper bounded by the matrix norm and may be negative. We refer to [1], and references therein, for the proof of these and additional properties enjoyed by matrix measures.

For ℓ_p norms, $p \in \{1, 2, \infty\}$, log norms admit simpler representations. Namely,

$$\mu_2(A) = \frac{1}{2}\lambda_{\max}\big(A + A^\top\big) \tag{1.4}$$

$$\mu_1(A) = \max_{j\in\{1,\dots,n\}} \Big(a_{jj} + \sum_{i=1,i\neq j}^{n} |a_{ij}|\Big), \;\; \mu_\infty(A) = \max_{i\in\{1,\dots,n\}} \Big(a_{ii} + \sum_{j=1,j\neq i}^{n} |a_{ij}|\Big). \tag{1.5}$$

For R invertible square, the log norm induced by the R-weighted ℓ_p norm is given by $\mu_{p,R}(A) = \mu_p(RAR^{-1})$. For a symmetric positive-definite matrix $P \in \mathbb{R}^{n\times n}$, the ℓ_2 weighted log norm satisfies $\mu_P(A) = \min\{b \in \mathbb{R} \mid PA + A^\top P \preceq 2bP\}$ [2, Lemma 2.7].

Lipschitz maps. Given two normed spaces $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$, $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ a map $F\colon \mathcal{X} \to \mathcal{Y}$ is Lipschitz from $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ to $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ with constant $\ell \geq 0$ if for all $x_1, x_2 \in \mathcal{X}$, it holds that

$$\|F(x_1) - F(x_2)\|_{\mathcal{Y}} \leq \ell\|x_1 - x_2\|_{\mathcal{X}}. \tag{1.6}$$

If $\mathcal{Y} = \mathcal{X}$ and $\|\cdot\|_{\mathcal{X}} = \|\cdot\|_{\mathcal{Y}}$, we instead say F is Lipschitz on $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ with constant $\ell \geq 0$. When the spaces $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ and $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ are clear from context, we will instead simply say F is Lipschitz. For a Lipschitz map F, we will adopt the notation $\mathsf{Lip}(F)$ to denote either the infimum value ℓ satisfying (1.6) or an upper bound thereof (which we are using should be clear from context). We will say a function F is locally Lipschitz if for each x in the domain of F, there exists an open neighborhood containing x such that F restricted to the open neighborhood is Lipschitz.

For a mapping $F\colon \mathcal{X} \to \mathcal{Y}$ where $\mathcal{X} \subseteq \mathbb{R}^n, \mathcal{Y} \subseteq \mathbb{R}^m$, let $\mathrm{Dom}(F)$ be its domain. If F is differentiable, let $DF(x) := \frac{\partial F(x)}{\partial x}$ denote its Jacobian evaluated at x. Further recall that if F is locally Lipschitz, then $DF(x)$ exists for almost every x in light of Rademacher's theorem. When $\mathcal{Y} \subseteq \mathbb{R}$, then $DF(x)^\top$ is denoted by $\nabla F(x) \in \mathbb{R}^n$ and is referred to as the gradient of F. When $\mathcal{Y} \subseteq \mathbb{R}$, the Jacobian of the map $\nabla F\colon \mathcal{X} \to \mathbb{R}^n$ is the Hessian of F and is denoted $\nabla^2 F$.

For a mapping $F\colon \mathbb{R}^n \to \mathbb{R}^n$, let $\mathrm{Zero}(F) := \{x \in \mathbb{R}^n \mid F(x) = \mathbb{0}_n\}$ and $\mathrm{Fix}(F) = \{x \in \mathbb{R}^n \mid F(x) = x\}$ be the sets of zeros of F and fixed points of F, respectively. We let $\mathsf{Id}\colon \mathbb{R}^n \to \mathbb{R}^n$ be the identity mapping. For a vector $y \in \mathbb{R}$, define the mappings $\mathrm{sign}\colon \mathbb{R} \to \mathbb{R}$ and $\mathrm{ReLU}\colon \mathbb{R} \to \mathbb{R}$ by $\mathrm{sign}(y) = y/|y|$ if $y \neq 0$ and $\mathrm{sign}(0) = 0$ and $\mathrm{ReLU}(y) = \max\{y, 0\}$. By extension, when $y \in \mathbb{R}^n$, by a slight abuse of notation, we define the vector-valued mappings $\mathrm{sign}\colon \mathbb{R}^n \to \mathbb{R}^n$ and $\mathrm{ReLU}\colon \mathbb{R}^n \to \mathbb{R}^n$ by $(\mathrm{sign}(y))_i = \mathrm{sign}(y_i)$ and $(\mathrm{ReLU}(y))_i = \mathrm{ReLU}(y_i)$.

We additionally recall the useful mean-value theorem for vector-valued mappings [3, Prop. 2.4.7]. Let $f\colon \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable. Then for every $x, y \in \mathbb{R}^n$,

$$f(x) - f(y) = \Big(\int_0^1 Df(\tau x + (1-\tau)y)d\tau\Big)(x - y). \tag{1.7}$$

Convex analysis and monotone operators. Let $\mathcal{C} \subseteq \mathbb{R}^n$ be convex, closed, and nonempty. The map $\iota_{\mathcal{C}}\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ is the *indicator function on* $\mathcal{C}$ and is defined by $\iota_{\mathcal{C}}(z) = 0$ if $z \in \mathcal{C}$ and $\iota_{\mathcal{C}}(z) = +\infty$ otherwise. The map $\mathrm{P}_{\mathcal{C}}\colon \mathbb{R}^n \to \mathcal{C}$ is the *projection on* $\mathcal{C}$ and is given by $\mathrm{P}_{\mathcal{C}}(x) = \mathrm{argmin}_{u \in \mathcal{C}} \|x - u\|_2$.

The epigraph of a map $g\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ is the set $\{(x, \xi) \in \mathbb{R}^{n+1} \mid g(x) \leq \xi\}$. The map g is (i) *convex* if its epigraph is a convex set, (ii) *proper* if its value is never $-\infty$ and is finite somewhere, and (iii) *closed* if it is proper and its epigraph is a closed set. The map $g\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ is (i) *strongly convex with parameter* $\rho > 0$ if the map $x \mapsto g(x) - \frac{\rho}{2}\|x\|_2^2$ is convex and (ii) strongly smooth with parameter $\ell \geq 0$ if it is differentiable and ∇g is Lipschitz on $(\mathbb{R}^n, \|\cdot\|_2)$ with constant ℓ.

Let $g\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex, closed, and proper (CCP). The *subdifferential of* g *at* $x \in \mathbb{R}^n$ is the set $\partial g(x) := \{z \in \mathbb{R}^n \mid g(x) - g(y) \geq z^\top(x - y) \text{ for all } y \in \mathbb{R}^n\}$. When g is differentiable at x, then $\partial g(x) = \{\nabla g(x)\}$. The *proximal operator of* g *with parameter*

$\gamma > 0$, $\operatorname{prox}_{\gamma g} \colon \mathbb{R}^n \to \mathbb{R}^n$, is defined by

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

The associated *Moreau envelope of g with parameter $\gamma > 0$*, $M_{\gamma g} \colon \mathbb{R}^n \to \mathbb{R}$, and its gradient are given by:

$$\begin{aligned} M_{\gamma g}(x) &= g(\operatorname{prox}_{\gamma g}(x)) + \frac{1}{2\gamma}\|x - \operatorname{prox}_{\gamma g}(x)\|_2^2, \\ \nabla M_{\gamma g}(x) &= \frac{1}{\gamma}(x - \operatorname{prox}_{\gamma g}(x)). \end{aligned} \tag{1.8}$$

The gradient of the Moreau envelope always exists and is Lipschitz on $(\mathbb{R}^n, \|\cdot\|_2)$ with constant $1/\gamma$.

A map $F \colon \mathbb{R}^n \to \mathbb{R}^n$ is (i) *monotone* if $(F(x) - F(y))^\top (x - y) \geq 0$ for all $x, y \in \mathbb{R}^n$ and (ii) *strongly monotone with parameter $m > 0$* if the map $F - m\mathsf{Id}$ is monotone. We refer to [4] for a comprehensive treatment of these tools.

Dynamical systems. For a time-varying vector field $f \colon \mathbb{R}_{\geq 0} \times \mathbb{R}^n \to \mathbb{R}^n$, we will frequently study the initial-value problem

$$\dot{x} = f(t, x), \qquad x(t_0) = x_0. \tag{1.9}$$

Under mild conditions on f, e.g., piecewise continuity in the first argument and local Lipschitzness in the second, then for each x_0, there exists $\tau(x_0) > t_0$ such that a solution of (1.9) exists on a time interval $[t_0, \tau(x_0))$ (note that $\tau(x_0) = +\infty$ is possible) and is unique on this interval. We will denote this solution at time $t \in [t_0, \tau(x_0))$ by $\phi(t, t_0, x_0)$. Namely, $\phi(t_0, t_0, x_0) = x_0$. At times, when it is clear, we will use $x(t)$ to denote this solution. Also, when f is autonomous, we will sometimes also adopt the notation $\phi_{x_0}(t)$

for the solution. At times, we will consider vector fields which are only continuous in their second argument. In this case, solutions may be nonunique and we will treat these cases separately to establish uniqueness of solutions.

Chapter 2

Non-Euclidean Contraction Theory for Robust Nonlinear Stability

This chapter was first published in the IEEE Transactions on Automatic Control [5].[1]

2.1 Introduction

Problem description and motivation: A vector field is contracting if its flow map is a contraction or, equivalently, if any two solutions approach one another exponentially fast. Contracting systems feature highly-ordered asymptotic behavior. First, initial conditions are forgotten. Second, a unique equilibrium is globally exponential stable when the vector field is time-invariant and two natural Lyapunov functions are automatically available (i.e., the distance to the equilibrium and the norm of the vector field). Third, a unique periodic solution is globally exponentially stable when the vector field is periodic; in other words, contracting system entrain to periodic inputs. Fourth and last, contracting systems enjoy natural robustness properties such as input-to-state stability and finite

Log norm condition	Demidovich condition	One-sided Lipschitz condition
$\mu_P(Df(x)) \leq b$	$PDf(x) + Df(x)^\top P \preceq 2bP$	$(x-y)^\top P\big(f(x) - f(y)\big) \leq b\\|x-y\\|^2_{2,P^{1/2}}$
$\mu_{1,R}(Df(x)) \leq b$	$\operatorname{sign}(Rv)^\top RDf(x)v \leq b\\|v\\|_{1,R}$	$\operatorname{sign}(Rx - Ry)^\top R(f(x) - f(y)) \leq b\\|x-y\\|_{1,R}$
$\mu_\infty(Df(x)) \leq b$	$\max_{i\in I_\infty(v)} (Df(x)v)_i v_i \leq b\\|v\\|^2_\infty$	$\max_{i\in I_\infty(x-y)} ((f(x) - f(y)))_i(x-y)_i \leq b\\|x-y\\|^2_\infty$

Table 2.1: Table of contraction equivalences, that is, equivalences between measure bounded Jacobians, Demidovich and one-sided Lipschitz conditions. $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is a continuously differentiable vector field with Jacobian Df. Each row contains three equivalent statements, to be understood for all $x, y \in \mathbb{R}^n$ and all $v \in \mathbb{R}^n$. We adopt the shorthand $I_\infty(v) = \{i \in \{1,\dots,n\} \mid |v_i| = \\|v\\|_\infty\}$. The matrix P is positive definite and the matrix R is invertible.

input-state gain in the presence of (Lipschitz continuous) unmodeled dynamics. Because of these highly-ordered and desirable behaviors, contracting systems are of great interest for engineering problems.

Contraction theory aims to combine, in a unified coherent framework, results from Lyapunov stability theory, incremental stability, fixed point theorems, monotone systems theory, and the geometry of Banach, Riemannian and Finsler spaces. Classical approaches primarily study contraction with respect to the ℓ_2 norm for continuously differentiable vector fields. However, recent works have shown that stability can be studied more systematically and efficiently using non-Euclidean norms (e.g., ℓ_1, ℓ_∞ and polyhedral norms) for large classes of network systems, including biological transcriptional systems [6], Hopfield neural networks [7], chemical reaction networks [8], traffic networks [9, 10, 11], multi-vehicle systems [12], and coupled oscillators [13, 14]. Moreover, for large-scale systems, error analysis based on the ℓ_∞ norm may more accurately capture the effect of bounded perturbations. As compared with the ℓ_2 norm, there is only limited work on non-Euclidean contraction theory.

It is well known that contraction with respect to the ℓ_2 norm is established via a test on the Jacobian of the vector field; it is also true however that an integral (derivative-free) test on the vector field itself is equivalent. While some differential tests are available for non-Euclidean norms, much less is known about the corresponding integral tests. We note that computing Jacobians for large-scale networks may be computationally intensive and so derivative-free contraction tests are desirable. In this chapter, we aim to characterize differential and integral tests for arbitrary norms, paying special attention to the ℓ_1 and ℓ_∞ norms, and provide a unifying framework for differential and integral tests.

Literature review: Contraction mappings in dynamical systems via logarithmic norms have been studied extensively and can be traced back to Lewis [15], Demidovich [16] and Krasovskiĭ [17]. Logarithmic norms and numerical methods for differential equations have been studied by Dahlquist [18] and Lozinskii [19] in as early as 1958; see also the influential survey by Ström [20]. Finally, logarithmic norms were applied to control problems by Desoer and Vidyasagar in [1, 21, 22] and contraction theory was first introduced by Lohmiller and Slotine in [23]. Since then, numerous generalizations to contraction theory have been proposed including partial contraction [24], contraction of stochastic differential equations [25], contraction in differential algebraic equations [26], contraction on Riemannian and Finsler manifolds, [27, 28], contraction for PDEs [29], transverse contraction [30], contraction after short transients [31], weak and semi-contraction [32], and k-contraction, i.e., contraction of k-dimensional bodies [33].

While the work of Lohmiller and Slotine explored differential conditions for contraction for the ℓ_2 norm, related integral conditions have been studied in the literature under such various names as the *one-sided Lipschitz condition* in [34], the *QUAD condition* in [35], the *nonlinear measure* [7], the *dissipative Lipschitz condition* [36], and *incremental quadratic stability* in [37]. A related unifying concept is the *logarithmic Lipschitz constant*, advocated in [38, 39]. Moreover, the key idea appears as early as [40], whereby minus the vector field is called *uniformly increasing* and in the work on discontinuous differential equations, see [41, Chapter 1, page 5] and references therein. Comparisons between the Lipschitz conditions, the QUAD condition, and contraction are detailed in [42], see also [34, Section 1.10, Exercise 6].

Tests for contraction with respect to non-Euclidean non-differentiable norms have not been widely studied. Early results on compartmental systems include [43, Theorem 2] and [44, Appendix 4]. The ℓ_1 integral test is used to study neural networks in [7] and traffic networks in [10]. Recent work [11] establishes that the ℓ_1 and ℓ_∞ norms are well suited to study contraction of monotone systems. A comprehensive understanding of connections between differential and integral conditions for these norms is desirable.

Aminzare and Sontag first drew connections between contraction theory and so-called "semi-inner products" in [39], where they give conditions for contraction in L^p spaces. Since then, they have explored contraction with respect to arbitrary norms in [14, 45], and [46], where arbitrary norms are used to study synchronization of diffusively coupled systems and contractivity of reaction diffusion PDEs. Most notably, in [45, Proposition 3] necessary and sufficient conditions for contraction are given using Deimling pairings (see [47, Chapter 3] for more details on Deimling pairings). This chapter builds upon these underappreciated works and underutilized connections.

Contributions: Our first contribution is the definition of *weak pairings* as a generalization of the classic Lumer pairings, as introduced in [48, 49]. We study various properties of weak pairings, including a useful *curve norm derivative formula* applicable to dynamical systems analysis. Additionally, we establish a key relationship between weak pairings and logarithmic norms, generalizing a result by Lumer in [48]; we refer to this relationship as *Lumer's equality.* For ℓ_p norms, $p \in \{1, \infty\}$, we present and characterize novel convenient choices for weak pairings: the sign pairing for the ℓ_1 norm and the max pairing for the ℓ_∞ norm. We argue that, due to their connection with logarithmic norms, weak pairings are a broadly-applicable tool for contraction analysis.

Our second contribution is proving five equivalent characterizations of contraction for continuously differentiable vector fields on $\mathbb{R}^n$ with respect to arbitrary norms. Using the language of weak pairings, we prove the equivalence between differential and integral tests for contraction; this result generalizes the known ℓ_2 norm results to non-differentiable norms such as the ℓ_1 and ℓ_∞ norms. We show that three of the five equivalences capture the logarithmic norm condition, the differential condition on the vector field (referred to as the *Demidovich condition*), and the integral condition (referred to as the *one-sided Lipschitz condition*). These results generalize [45, Proposition 3] in the sense that (i) we draw an additional connection between the logarithmic norm of the Jacobian and the weak pairing and (ii) use this connection and our sign and max pairings to write the explicit differential and integral conditions in Table 2.1.

Our third contribution is the extension of contraction theory to vector fields that are only continuous. This extension demonstrates that contraction can be understood as a property of the vector field, independent of its Jacobian. In other words, this extension establishes the importance of weak pairings over classical contraction approaches based on the Jacobian of the vector field and the stability of the linearized system.

Our fourth contribution is the formalization of *equilibrium contraction*, a weaker form of contraction where all trajectories exponentially converge to an equilibrium. This notion has been explored for example, in [22, Chapter 2, Theorem 22] and [24, Theorem 1]. These approaches establish global exponential convergence under two conditions: the vector field can be factorized as $f(t,x) = A(t,x)x$ and the logarithmic norm of $A(t,x)$ is uniformly negative. Our treatment of equilibrium contraction demonstrates that these two conditions are only sufficient, whereas we provide a necessary and sufficient characterization based upon the one-sided Lipschitz condition.

Our fifth contribution is proving novel robustness properties of contracting and equilibrium contracting vector fields. For strongly contracting systems we prove incremental input-to-state stability and provide novel input-state gain estimates. These results generalize [1, Theorem A] and [50, Theorem 1], where vector fields are required to have a special control-affine structure. We additionally prove a novel result for contraction under perturbations and are able to upper bound how far the unique equilibrium shifts.

Our sixth and final contribution is a general theorem about the contractivity of the interconnection of contracting systems. Motivated by applications to large scale systems, we provide a sufficient condition for contraction and establish optimal contraction rates for interconnected systems. This theorem is the counterpart for contracting system of the classic theorem about the interconnection of dissipative systems, e.g., see [51, Chapter 2]. This treatment generalizes the results in [20, Section 5], [13], and [27, Lemma 3.2], where optimal rates of contraction are not provided, vector fields are differentiable, and interconnections with inputs are not studied.

Chapter organization: Section 2.2 reviews Lumer pairings and Dini derivatives. Section 2.3 defines weak pairings and provides explicit formulas for ℓ_p norms. Section 2.4 proves contraction equivalences. Section 2.5 gives robustness results for contracting vector fields. Section 2.6 studies the interconnection of contracting systems. Section 2.7 provides conclusions.

2.2 A review of Lumer pairings and Dini derivatives

2.2.1 Norms and Lumer pairings

Definition 1 (Lumer pairings [48, 49])**.** *A* Lumer pairing *on* $\mathbb{R}^n$ *is a map* $[\cdot,\cdot] : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ *satisfying:*

(i) (Additivity in first argument) $[x_1 + x_2, y] = [x_1, y] + [x_2, y]$*, for all* $x_1, x_2, y \in \mathbb{R}^n$*,*

(ii) (Homogeneity) $[\alpha x, y] = [x, \alpha y] = \alpha[x, y]$*, for all* $x, y \in \mathbb{R}^n, \alpha \in \mathbb{R}$*,*

(iii) *(Positive definiteness)* $[x, x] > 0$, *for all* $x \neq \mathbb{0}_n$, *and*

(iv) *(Cauchy-Schwarz inequality)*

$|[x, y]| \leq [x, x]^{1/2}[y, y]^{1/2}$, *for all* $x, y \in \mathbb{R}^n$.

Lemma 2.2.1 (Norms and Lumer pairings [48]). *(i) If* $\mathbb{R}^n$ *is equipped with a Lumer pairing, it is also a normed space with norm* $\|x\| := [x, x]^{1/2}$, *for all* $x \in \mathbb{R}^n$.

(ii) Conversely, if $\mathbb{R}^n$ *is equipped with norm* $\|\cdot\|$, *then there exists an (not necessarily unique) Lumer pairing on* $\mathbb{R}^n$ *compatible with* $\|\cdot\|$ *in the sense that* $\|x\| = [x, x]^{1/2}$, *for all* $x \in \mathbb{R}^n$.

Lemma 2.2.2 (Lumer's equality [48, Lemma 12]). *Given a norm* $\|\cdot\|$ *on* $\mathbb{R}^n$, *a compatible Lumer pairing* $[\cdot, \cdot]$, *and a matrix* $A \in \mathbb{R}^{n \times n}$,

$$\mu(A) = \sup_{\|x\|=1} [Ax, x] = \sup_{x \neq \mathbb{0}_n} \frac{[Ax, x]}{\|x\|^2}. \tag{2.1}$$

Recall that a norm $\|\cdot\|$ on $\mathbb{R}^n$ is *differentiable* if, for all $x, y \in \mathbb{R}^n \setminus \{\mathbb{0}_n\}$, the following limit exists:

$$\lim_{h \to 0} \frac{\|x + hy\| - \|x\|}{h}.$$

The ℓ_p norm is differentiable for $p \in (1, \infty)$ and not differentiable for $p \in \{1, \infty\}$.

Lemma 2.2.3 (Gâteaux formula for the Lumer pairing [49]). *Let* $\|\cdot\|$ *be a norm on* $\mathbb{R}^n$. *If* $\|\cdot\|$ *is differentiable, then there exists a unique compatible Lumer pairing given by the* Gâteaux formula*:*

$$[x, y] = \|y\| \lim_{h \to 0} \frac{\|y + hx\| - \|y\|}{h}, \quad x, y \in \mathbb{R}^n \setminus \{\mathbb{0}_n\}. \tag{2.2}$$

Lemma 2.2.4 (Lumer pairing and log norm for weighted ℓ_p norms [52, Example 13.1(a)]). *For* $p \in (1, \infty)$ *and* $R \in \mathbb{R}^{n \times n}$ *invertible, let* $\|\cdot\|_{p,R}$, $[\cdot, \cdot]_{p,R}$ *and* $\mu_{p,R}(\cdot)$ *denote the weighted*

ℓ_p norm, its Lumer pairing and its log norm, respectively. Then, for $x, y \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n\times n}$,

$$\|x\|_{p,R} = \|Rx\|_p, \quad [x,y]_{p,R} = \frac{(Ry \circ |Ry|^{p-2})^\top Rx}{\|y\|_{p,R}^{p-2}},$$
$$\mu_{p,R}(A) = \max_{\|x\|_{p,R}=1} (Rx \circ |Rx|^{p-2})^\top RAx,$$

where $\circ$ is the entrywise product, $|\cdot|$ is the entrywise absolute value, and $(\cdot)^p$ is the entrywise power.

Corollary 2.2.5. *For $p = 2$ and $R = P^{1/2}$ where $P = P^\top \succ 0$, Lemma 2.2.4 implies*

$$\|x\|_{2,P^{1/2}}^2 = x^\top Px, \quad [x,y]_{2,P^{1/2}} = x^\top Py,$$
$$\mu_{2,P^{1/2}}(A) = \max_{\|x\|_{2,P^{1/2}}=1} x^\top A^\top Px = \lambda_{\max}\Big(\frac{PAP^{-1} + A^\top}{2}\Big).$$

2.2.2 Dini derivatives

Definition 2 (Upper right Dini derivative)**.** *The* upper right Dini derivative *of a function $\varphi\colon \mathbb{R}_{\geq 0} \to \mathbb{R}$ is*

$$D^+\varphi(t) := \limsup_{h\to 0^+} \frac{\varphi(t+h) - \varphi(t)}{h}. \tag{2.3}$$

Lemma 2.2.6 (Danskin's lemma [53])**.** *Given differentiable functions $f_1, \dots, f_m\colon (a,b) \to \mathbb{R}$, if $f(t) = \max_i f_i(t)$, then*

$$D^+ f(t) = \max\Big\{\frac{d}{dt} f_i(t) \;\Big|\; f_i(t) = f(t)\Big\}. \tag{2.4}$$

The following two lemmas are related to known results. We report them here for completeness sake.

Lemma 2.2.7 (Dini derivative of absolute value function)**.** *Let $x : (a,b) \to \mathbb{R}$ be differentiable. Then*

$$D^+|x(t)| = \dot{x}(t)\,\mathrm{sign}(x(t)) + |\dot{x}(t)|\chi_{\{0\}}(x(t)),$$

where $\chi_A(x)$ is the indicator function which is 1 when $x \in A$ and zero otherwise.

Proof. Since $|x(t)| = \max\{x(t), -x(t)\}$, Lemma 2.2.6 implies $D^+|x(t)| = \dot{x}(t)$ if $x(t) > 0$, $-\dot{x}(t)$ if $x(t) < 0$, and $|\dot{x}(t)|$ if $x(t) = 0$. □

Lemma 2.2.8 (Nonsmooth Grönwall inequality)**.** *Let $\varphi, r \colon [a,b] \to \mathbb{R}_{\geq 0}$ and $m \colon [a,b] \to \mathbb{R}$ be continuous. If $D^+\varphi(t) \leq m(t)\varphi(t) + r(t)$ for almost every $t \in (a,b)$, then, for every $t \in [a,b]$ and for $M(t) = \int_a^t m(\tau)d\tau$,*

$$\varphi(t) \leq e^{M(t)}\Big(\varphi(a) + \int_a^t r(\tau)e^{-M(\tau)}d\tau\Big). \tag{2.5}$$

Proof. Let $\psi(t) = \varphi(t)e^{-M(t)} \geq 0$ for all $t \in [a,b]$. Then

$$D^+\psi(t) \leq (D^+\varphi(t) - m(t)\varphi(t))e^{-M(t)} \leq r(t)e^{-M(t)}.$$

for almost every $t \in (a,b)$. Note that $r(t)e^{-M(t)}$ is continuous and satisfies $r(t)e^{-M(t)} \geq 0$ for all $t \in [a,b]$. Then by [54, Appendix A1, Proposition 2], for every $t \in [a,b]$, we have

$$\psi(t) \leq \psi(a) + \int_a^t r(\tau)e^{-M(\tau)}d\tau,$$

which, in turn, implies the claim. □

Lemma 2.2.9 (Dini comparison lemma [55, Lemma 3.4])**.** *Consider the initial value problem $\dot{\zeta} = f(t,\zeta)$, $\zeta(t_0) = \zeta_0$, where $f \colon \mathbb{R}_{\geq 0} \times \mathbb{R} \to \mathbb{R}$ is continuous in t and locally Lipschitz in ζ, for all $t \geq 0$ and $\zeta \in \mathbb{R}$. Let $[t_0, T)$ be the maximal interval of existence* *for* $\zeta(t)$ *and let* $v\colon \mathbb{R}_{\geq 0} \to \mathbb{R}$ *be continuous and satisfy*

$$D^+ v(t) \leq f(t, v(t)), \quad v(t_0) \leq \zeta_0.$$

Then $v(t) \leq \zeta(t)$ *for all* $t \in [t_0, T)$*.*

Lemma 2.2.10 (Coppel's differential inequality [56])**.** *Given a continuous map* $(t,x) \mapsto A(t,x) \in \mathbb{R}^{n\times n}$*, any solution* $x(\cdot)$ *of* $\dot{x} = A(t,x)x$ *satisfies*

$$D^+\|x(t)\| \leq \mu(A(t, x(t)))\|x(t)\|. \tag{2.6}$$

We conclude with a small useful result.

Lemma 2.2.11. *Consider the control system* $\dot{x} = f(t, x, u(t))$ *with* $f\colon \mathbb{R}_{\geq 0} \times \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n$ *continuous in* (t, x, u)*. Let* $\|\cdot\|_{\mathcal{X}}$ *be a norm on* $\mathbb{R}^n$ *and* $\|\cdot\|_{\mathcal{U}}$ *be a norm on* $\mathbb{R}^k$*. If there exists* $\ell \geq 0$ *such that, for each* $t \in \mathbb{R}_{\geq 0}, x \in \mathbb{R}^n, u, v \in \mathbb{R}^k$*,*

$$\|f(t,x,u) - f(t,x,v)\|_{\mathcal{X}} \leq \ell\|u - v\|_{\mathcal{U}},$$

then any two continuously differentiable solutions $x(\cdot), y(\cdot)$ *to the control system corresponding to continuous inputs* $u_x, u_y\colon \mathbb{R}_{\geq 0} \to \mathbb{R}^k$ *and with* $x(t) = y(t)$ *for some* $t \geq 0$ *satisfy*

$$D^+\|x(t) - y(t)\|_{\mathcal{X}} \leq \ell\|u_x(t) - u_y(t)\|_{\mathcal{U}}. \tag{2.7}$$

Proof. The result follows from the definition of Dini derivative and by Taylor expansions of $x(t + h)$ and $y(t + h)$. □

Symbol	Meaning
$\|\cdot\|_{p,R}$	ℓ_p norm weighted by R, $\|x\|_{p,R} = \|Rx\|_p$.
$[\cdot,\cdot]_{p,R}$	Lumer pairing compatible with $\|\cdot\|_{p,R}$
$[\![\cdot,\cdot]\!]_{p,R}$	Weak pairing compatible with $\|\cdot\|_{p,R}$
$\mu_{p,R}(\cdot)$	Log norm with respect to $\|\cdot\|_{p,R}$

Table 2.2: Table of symbols. We let $p \in [1,\infty]$ and $R \in \mathbb{R}^{n\times n}$ be invertible. If a norm, Lumer pairing, weak pairing, or log norm does not have a subscript, it is assumed to be arbitrary. If the subscript R is not included, $R = I_n$.

2.3 Weak pairings and calculus of non-Euclidean norms

2.3.1 Weak pairings definition and properties

We define the notion of a weak pairing which further weaken the conditions for a pairing to be a Lumer pairing.

Definition 3 (Weak pairing). *A* weak pairing *on $\mathbb{R}^n$ is a map $[\![\cdot,\cdot]\!] : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ such that the following properties hold:*

(i) (Subadditivity and continuity of first argument) $[\![x_1 + x_2, y]\!] \leq [\![x_1, y]\!] + [\![x_2, y]\!]$, for all $x_1, x_2, y \in \mathbb{R}^n$ and $[\![\cdot,\cdot]\!]$ is continuous in its first argument,

(ii) (Weak homogeneity) $[\![\alpha x, y]\!] = [\![x, \alpha y]\!] = \alpha\,[\![x, y]\!]$ and $[\![-x, -y]\!] = [\![x, y]\!]$, for all $x, y \in \mathbb{R}^n, \alpha \geq 0$,

(iii) (Positive definiteness) $[\![x, x]\!] > 0$, for all $x \neq \mathbb{0}_n$,

(iv) (Cauchy-Schwarz inequality) $|\,[\![x, y]\!]\,| \leq [\![x, x]\!]^{1/2}\,[\![y, y]\!]^{1/2}$, for all $x, y \in \mathbb{R}^n$.

From Definition 3, any Lumer pairing is a weak pairing, but not every weak pairing is a Lumer pairing. When necessary, we distinguish the symbols for Lumer pairings and weak pairings and make this clear in Table 2.2.

Theorem 2.3.1 (Compatibility of weak pairings with norms)**.** *If $[\![\cdot,\cdot]\!]$ is a weak pairing on $\mathbb{R}^n$, then $\|\cdot\| = [\![\cdot,\cdot]\!]^{1/2}$ is a norm. Conversely, if $\mathbb{R}^n$ is equipped with a norm $\|\cdot\|$, then there exists a weak pairing (but possibly many) such that $[\![\cdot,\cdot]\!] = \|\cdot\|^2$.*

Proof. First, we show that $\|\cdot\| = [\![\cdot,\cdot]\!]^{1/2}$ is a norm. Clearly it is positive definite by property (iii). For homogeneity,

$$\begin{aligned} \|\alpha x\|^2 &= [\![\alpha x, \alpha x]\!] = \alpha^2 [\![x,x]\!] \\ \implies \quad \|\alpha x\| &= |\alpha| [\![x,x]\!]^{1/2} = |\alpha|\|x\|, \end{aligned}$$

by weak homogeneity, property (ii). Finally, regarding the triangle inequality,

$$\begin{aligned} \|x+y\|^2 = [\![x+y, x+y]\!] &\leq [\![x, x+y]\!] + [\![y, x+y]\!] \\ &\leq (\|x\| + \|y\|)\|x+y\|, \end{aligned}$$

by subadditivity, property (i), and the Cauchy-Schwarz inequality, property (iv). This implies that $\|x+y\| \leq \|x\| + \|y\|$.

For the converse, the proof is identical to that in [48] since any Lumer pairing is a weak pairing. □

As a consequence, if $[\![x,y]\!]$ is a weak pairing compatible with the norm $\|\cdot\|$, then $[\![Rx, Ry]\!]$ is a weak pairing compatible with the R-weighted norm $\|\cdot\|_R$ for any invertible $R \in \mathbb{R}^{n\times n}$.

We now define two desirable properties of weak pairings.

Definition 4 (Additional weak pairing properties)**.** *Let $[\![\cdot,\cdot]\!]$ be compatible with the norm $\|\cdot\|$. Then $[\![\cdot,\cdot]\!]$ satisfies* *(i)* Deimling's inequality *if, for all $x, y \in \mathbb{R}^n$,*

$$[\![x, y]\!] \leq \|y\| \lim_{h \to 0^+} \frac{\|y + hx\| - \|y\|}{h}, \tag{2.8}$$

(ii) the curve norm derivative formula *if, for every differentiable $x : (a, b) \to \mathbb{R}^n$ and for almost every $t \in (a, b)$,*

$$\|x(t)\| D^+ \|x(t)\| = [\![\dot{x}(t), x(t)]\!] . \tag{2.9}$$

Note that any given weak pairing may or may not satisfy these properties. It is essentially known that any Lumer pairing satisfies Deimling's inequality; see Section 2.8.1 Lemma 2.8.2.

Theorem 2.3.2 (Lumer's equality for weak pairings)**.** *Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$ with compatible weak pairing $[\![\cdot, \cdot]\!]$ satisfying Deimling's inequality,* (2.8)*. Then for all $A \in \mathbb{R}^{n \times n}$*

$$\mu(A) = \sup_{\|x\|=1} [\![Ax, x]\!] = \sup_{x \neq \mathbb{0}_n} \frac{[\![Ax, x]\!]}{\|x\|^2}. \tag{2.10}$$

Proof. By Deimling's inequality, for every $x \in \mathbb{R}^n \setminus \{\mathbb{0}_n\}$,

$$\begin{aligned} [\![Ax, x]\!] &\leq \|x\| \lim_{h \to 0^+} \frac{\|x + hAx\| - \|x\|}{h} \\ &\leq \|x\|^2 \lim_{h \to 0^+} \frac{\|I_n + hA\| - 1}{h} = \|x\|^2 \mu(A). \end{aligned}$$

Thus, the inequality $\mu(A) \geq \sup_{x \neq \mathbb{0}_n} \frac{[\![Ax,x]\!]}{\|x\|^2}$ holds. For the other inequality, for $v \neq \mathbb{0}_n$, we define $\Omega(v) = \frac{[\![Av,v]\!]}{\|v\|^2}$ and note that for every $v \neq \mathbb{0}_n$ and $h > 0$,

$$\|(I_n - hA)v\| \geq \frac{1}{\|v\|} [\![(I_n - hA)v, v]\!] \geq (1 - h\Omega(v))\|v\| \geq (1 - h \sup_{\|v\|=1} \Omega(v))\|v\|,$$

where the first inequality holds by Cauchy-Schwarz, the second by subadditivity, and the final one since $-h < 0$ and by weak homogeneity of the weak pairing. Moreover, note that $\sup_{\|v\|=1} \Omega(v) \leq \|A\| \neq \infty$ by Cauchy-Schwarz for the weak pairing. Then for small enough $h > 0$, $I_n - hA$ is invertible and given by

$$\begin{aligned}
&(I_n - hA)^{-1} = I_n + hA + h^2A^2(I_n - hA)^{-1} \\
&\qquad\qquad \implies \|(I_n + hA)v\| \leq \|(I_n - hA)^{-1}v\| + h^2\|A^2(I_n - hA)^{-1}v\|,
\end{aligned}$$

where the last implication holds for all $v \in \mathbb{R}^n$ because of the triangle inequality. Moreover, defining $x = (I_n - hA)v$, for sufficiently small $h > 0$, we have

$$\frac{\|(I_n - hA)^{-1}x\|}{\|x\|} = \frac{\|v\|}{\|(I_n - hA)v\|} \leq \frac{1}{1 - h\sup_{\|v\|=1}\Omega(v)} \tag{2.11}$$

Then

$$\begin{aligned}
\mu(A) &= \lim_{h\to 0^+} \sup_{x \neq \mathbb{0}_n} \frac{\|(I_n + hA)x\|/\|x\| - 1}{h} \\
&\leq \lim_{h\to 0^+} \sup_{x \neq \mathbb{0}_n} \frac{\|(I_n - hA)^{-1}x\| + h^2\|A^2(I_n - hA)^{-1}x\| - \|x\|}{h\|x\|} \\
&\leq \lim_{h\to 0^+} \sup_{x \neq \mathbb{0}_n} \frac{\|(I_n - hA)^{-1}x\|/\|x\| - 1}{h} \\
&\leq \lim_{h\to 0^+} \frac{1}{h}\Big(\frac{1}{1 - h\sup_{\|v\|=1}\Omega(v)} - 1\Big) = \sup_{\|x\|=1} [\![Ax, x]\!] \,,
\end{aligned}$$

where the first line is the definition of the induced norm, the second line holds by the triangle inequality, the third line holds due to the subadditivity of the supremum, and the last line holds because the inequality in (2.11) holds for all $x \neq \mathbb{0}_n$. □

In the following subsections, we propose weak pairings for the ℓ_p norms in $\mathbb{R}^n$, $p \in [1, \infty]$, and show that they satisfy the two properties in Definition 4.

2.3.2 Weak pairings for differentiable norms

Each ℓ_p norm for $p \in (1, \infty)$ is differentiable. Therefore, the corresponding Lumer pairing is unique, given in Lemma 2.2.3, and satisfies Deimling's inequality, (2.8). Thus, we pick the weak pairing to be the unique compatible Lumer pairing from Lemma 2.2.4. Moreover, because of differentiability of the norm, in Section 2.8.1 Lemma 2.8.3 we show that the unique Lumer pairing satisfies the curve norm derivative formula in Definition 4(ii).

2.3.3 Non-differentiable norms: The ℓ_1 norm

The ℓ_1 norm given by $\|x\|_1 = \sum_{i=1}^n |x_i|$ fails to be differentiable at points where $x_i = 0$. Hence, we propose a pairing and show that it is a Lumer pairing compatible with the ℓ_1 norm.

Definition 5 (Sign pairing)**.** *For $R \in \mathbb{R}^{n\times n}$ invertible, let $\|\cdot\|_{1,R}$ be the weighted ℓ_1 norm given by $\|x\|_{1,R} = \|Rx\|_1$. The* sign pairing *$[\![\cdot,\cdot]\!]_{1,R} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is defined by*

$$[\![x, y]\!]_{1,R} := \|y\|_{1,R}\,\mathrm{sign}(Ry)^\top Rx. \tag{2.12}$$

Lemma 2.3.1. *The sign pairing is a Lumer pairing compatible with the weighted ℓ_1 norm.*

Proof. We verify the four properties of a Lumer pairing in Definition 1. Regarding property (i), for $x_1, x_2, y \in \mathbb{R}^n$,

$$\begin{aligned}[\![x_1 + x_2, y]\!]_{1,R} &= \|y\|_{1,R}\,\mathrm{sign}(Ry)^\top R(x_1 + x_2)\\ &= \|y\|_{1,R}\left(\mathrm{sign}(Ry)^\top Rx_1 + \mathrm{sign}(Ry)^\top Rx_2\right) = [\![x_1, y]\!]_{1,R} + [\![x_2, y]\!]_{1,R}.\end{aligned}$$

Regarding property (ii), for $\alpha \in \mathbb{R}$, $[\![\alpha x, y]\!]_{1,R} = \|y\|_{1,R}\,\mathrm{sign}(Ry)^\top R(\alpha x) = \alpha\,[\![x, y]\!]_{1,R}$. To check homogeneity in the second argument, we see that $\alpha = 0$ is trivial, so for $\alpha \neq 0$

$$[\![x, \alpha y]\!]_{1,R} = \|\alpha y\|_{1,R}\,\mathrm{sign}(\alpha R y)^\top R x = |\alpha|\frac{\alpha}{|\alpha|}\|y\|_{1,R}\,\mathrm{sign}(Ry)^\top Rx = \alpha\,[\![x, y]\!]_{1,R}\,.$$

Regarding property (iii), $[\![x, x]\!]_{1,R} = \|x\|_{1,R}\,\mathrm{sign}(Rx)^\top Rx = \|x\|_{1,R}^2 \geq 0$. This also proves compatibility. Regarding property (iv),

$$|\,[\![x, y]\!]_{1,R}\,| = [\![y, y]\!]_{1,R}^{1/2}\,|\,\mathrm{sign}(Ry)^\top Rx| \leq |\,\mathrm{sign}(Rx)^\top Rx|\,[\![y, y]\!]_{1,R}^{1/2} = [\![x, x]\!]_{1,R}^{1/2}\,[\![y, y]\!]_{1,R}^{1/2}\,.$$

□

Since the sign pairing is an Lumer pairing, it is a weak pairing that satisfies Deimling's inequality, (2.8). Finally, we separately establish the curve norm derivative formula, (2.9).

Theorem 2.3.3 (ℓ_1 curve norm derivative formula)**.** *Let $x : (a, b) \to \mathbb{R}^n$ be differentiable. Then*

(i) $D^+\|x(t)\|_{1,R} = \mathrm{sign}(Rx(t))^\top R\dot{x}(t)$, *for almost every* $t \in (a, b)$.

(ii) $\|x(t)\|_{1,R} D^+\|x(t)\|_{1,R} = [\![\dot{x}(t), x(t)]\!]_{1,R}$ *for almost every* $t \in (a, b)$.

Proof. Since $\|x(t)\|_{1,R} = \sum_{i=1}^n |(Rx(t))_i|$, it suffices to compute $D^+|x_i(t)|$. Then by Lemma 2.2.7:

$$\begin{aligned} D^+\|x(t)\|_{1,R} &= \sum_{i=1}^n D^+|(Rx(t))_i| = \sum_{i=1}^n \big((R\dot{x}(t))_i\,\mathrm{sign}((Rx)_i) + |(R\dot{x}(t))_i|\chi_{\{0\}}((Rx(t))_i)\big) \\ &= \mathrm{sign}(Rx(t))^\top R\dot{x}(t) + \sum_{i=1}^n |(R\dot{x}(t))_i|\chi_{\{0\}}((Rx(t))_i). \end{aligned}$$

Multiplying both sides by $\|x(t)\|_{1,R}$ gives

$$\|x(t)\|_{1,R}D^+\|x(t)\|_{1,R} = [\![\dot{x}(t), x(t)]\!]_{1,R} \\ + \|x(t)\|_{1,R}\sum\nolimits_{i=1}^n |(R\dot{x}(t))_i|\chi_{\{0\}}((Rx(t))_i).$$

To prove both results, it suffices to show that $\sum_{i=1}^n |(R\dot{x}(t))_i|\chi_{\{0\}}((Rx(t))_i) = 0$ for almost every $t \in (a,b)$. If $(Rx(t))_i \neq 0$, for all $i \in \{1,\dots,n\}$ and for all $t \in (a,b)$, the result holds since $\chi_{\{0\}}((Rx(t))_i) = 0$ for all $t \in (a,b), i \in \{1,\dots,n\}$. So suppose $(Rx(t))_i = 0$ for some i. Then either $(Rx(t))_i = 0$ for a single t, in which case the result holds. Otherwise $(Rx(t))_i = 0$ for all $t \in I \subseteq (a,b)$, where I is an interval. In this case, by differentiability of x, we have that $(R\dot{x}(t))_i = 0$ for almost every $t \in I$, so $|(R\dot{x}(t))_i|\chi_{\{0\}}((Rx)_i(t)) = 0$ for almost every $t \in I$ and hence almost every $t \in (a,b)$. □

2.3.4 Non-differentiable norms: The ℓ_∞ norm

The ℓ_∞ norm given by $\|x\|_\infty = \max_{i\in\{1,\dots,n\}} |x_i|$ fails to be differentiable at points where the infinity norm is achieved in more than one index. We propose a map and show that it is a weak pairing that satisfies the properties in Definition 4.

Definition 6 (Max pairing)**.** *For $R \in \mathbb{R}^{n\times n}$ invertible, let $\|\cdot\|_{\infty,R}$ be the weighted ℓ_∞ norm given by $\|x\|_{\infty,R} = \|Rx\|_\infty$. The* max pairing *$[\![\cdot,\cdot]\!]_{\infty,R} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is defined by*

$$[\![x, y]\!]_{\infty,R} := \max_{i\in I_\infty(Ry)} (Rx)_i(Ry)_i, \tag{2.13}$$

where $I_\infty(v) = \{j \in \{1,\dots,n\} \mid |v_j| = \|v\|_\infty\}$.

Lemma 2.3.2. *The max pairing is a weak pairing compatible with the weighted ℓ_∞ norm.*

Norm	Weak Pairing	Logarithmic norm
$\\|x\\|_{2,P^{1/2}} = \sqrt{x^\top P x}$	$[\![x,y]\!]_{2,P^{1/2}} = x^\top P y$	$\mu_{2,P^{1/2}}(A) = \frac{1}{2}\lambda_{\max}(PAP^{-1} + A^\top) = \max_{\\|x\\|_{2,P^{1/2}}=1} x^\top P A x$
$\\|x\\|_1 = \sum_i \|x_i\|$	$[\![x,y]\!]_1 = \\|y\\|_1 \operatorname{sign}(y)^\top x$	$\mu_1(A) = \max_{j\in\{1,\dots,n\}} \Big(a_{jj} + \sum_{i\neq j} \|a_{ij}\|\Big) = \sup_{\\|x\\|_1=1} \operatorname{sign}(x)^\top A x$
$\\|x\\|_\infty = \max_i \|x_i\|$	$[\![x,y]\!]_\infty = \max_{i\in I_\infty(y)} x_i y_i$	$\mu_\infty(A) = \max_{i\in\{1,\dots,n\}} \Big(a_{ii} + \sum_{j\neq i} \|a_{ij}\|\Big) = \max_{\\|x\\|_\infty=1} \max_{i\in I_\infty(x)} (Ax)_i x_i$

Table 2.3: Table of norms, weak pairings, and log norms for weighted ℓ_2, ℓ_1, and ℓ_∞ norms. We adopt the shorthand $I_\infty(x) = \{i \in \{1,\dots,n\} \mid |x_i| = \|x\|_\infty\}$. The matrix P is positive definite. Only the unweighted ℓ_1, ℓ_∞ norms, weak pairings, and log norms for $p \neq 2$ are included here since $\mu_{p,R}(A) = \mu_p(RAR^{-1})$ for any $p \in [1,\infty]$.

Proof. We verify the four properties of a weak pairing in Definition 3. Regarding property (i):

$$\begin{aligned}
[\![x_1 + x_2, y]\!]_{\infty,R} &= \max_{i\in I_\infty(Ry)} (R(x_1+x_2))_i (Ry)_i = \max_{i\in I_\infty(Ry)} (Rx_1)_i(Ry)_i + (Rx_2)_i(Ry)_i \\
&\leq \max_{i\in I_\infty(Ry)} (Rx_1)_i(Ry)_i + \max_{i\in I_\infty(Ry)} (Rx_2)_i(Ry)_i = [\![x_1,y]\!]_{\infty,R} + [\![x_2,y]\!]_{\infty,R}\,.
\end{aligned}$$

Further, for fixed $y \in \mathbb{R}^n$, the function $x \mapsto [\![x,y]\!]_{\infty,R}$ is continuous since $I_\infty(Ry)$ is fixed and the max of continuous functions is continuous. Regarding property (ii), for $\alpha \geq 0$,

$$\begin{aligned}
[\![\alpha x, y]\!]_{\infty,R} &= \max_{i\in I_\infty(Ry)} (R\alpha x)_i (Ry)_i = \alpha \max_{i\in I_\infty(Ry)} (Rx)_i(Ry)_i = \alpha\, [\![x,y]\!]_{\infty,R}\,, \\
[\![x, \alpha y]\!]_{\infty,R} &= \max_{i\in I_\infty(R\alpha y)} (Rx)_i (R\alpha y)_i = \alpha \max_{i\in I_\infty(Ry)} (Rx)_i(Ry)_i = \alpha\, [\![x,y]\!]_{\infty,R}\,, \\
[\![-x, -y]\!]_{\infty,R} &= \max_{i\in I_\infty(-Ry)} (-Rx)_i(-Ry)_i = [\![x,y]\!]_{\infty,R}\,.
\end{aligned}$$

Regarding property (iii)

$$[\![x, x]\!]_{\infty,R} = \max_{i \in I_\infty(Rx)} (Rx)_i (Rx)_i = \max_{i \in I_\infty(Rx)} \|x\|_{\infty,R}^2 = \|x\|_{\infty,R}^2 \geq 0.$$

This also shows that this weak pairing is compatible with the norm. Finally, regarding property (iv):

$$|\,[\![x, y]\!]_{\infty,R}\,| = \left| \max_{i \in I_\infty(Ry)} (Rx)_i (Ry)_i \right| \leq \left| \max_{i \in I_\infty(Rx)} \|y\|_{\infty,R} \|x\|_{\infty,R} \right| = [\![x, x]\!]_{\infty,R}^{1/2} \, [\![y, y]\!]_{\infty,R}^{1/2}.$$

□

We postpone to Section 2.8.2 the proof of the next lemma.

Lemma 2.3.3 (Deimling's inequality for the max pairing)**.** *The max pairing in Definition 6 satisfies Deimling's inequality,* (2.8)*.*

Theorem 2.3.4 (Derivative of ℓ_∞ norm along a curve)**.** *Let $x : (a, b) \to \mathbb{R}^n$ be differentiable. Then for all $t \in (a, b)$,*

(i) $D^+ \|x(t)\|_{\infty,R} = \max_{i \in I_\infty(Rx(t))} \operatorname{sign}((Rx(t))_i)(R\dot{x}(t))_i + \chi_{\{\mathbb{0}_n\}}(Rx(t)) \|\dot{x}(t)\|_{\infty,R}$*, and*

(ii) $\|x(t)\|_{\infty,R} D^+ \|x(t)\|_{\infty,R} = [\![\dot{x}(t), x(t)]\!]_{\infty,R}$*.*

Proof. From Danskin's lemma, Lemma 2.2.6, $f(t) = \max\{f_1(t), \dots, f_m(t)\}$ with differentiable f_i satisfies $D^+ f(t) = \max\{\frac{d}{dt} f_i(t) \mid f_i(t) = f(t)\}$. If the functions f_i are max functions themselves (e.g., absolute values in our case), a simple argument shows

$$D^+ f(t) = \max\{D^+ f_i(t) \mid f_i(t) = f(t)\}.$$

By the definition of $\|\cdot\|_{\infty,R}$, we have

$$D^+\|x(t)\|_{\infty,R} = \max_{i\in I_\infty(Rx(t))} D^+|(Rx(t))_i|.$$

Then by using the property for Dini derivatives of the absolute value function as in Lemma 2.2.7,

$$\begin{aligned} D^+\|x(t)\|_{\infty,R} &= \max_{i\in I_\infty(Rx(t))} \operatorname{sign}((Rx(t))_i)(R\dot{x}(t))_i + |(R\dot{x}(t))_i|\chi_{\{0\}}(Rx(t))_i \\ &= \max_{i\in I_\infty(Rx(t))} \operatorname{sign}((Rx(t))_i)(R\dot{x}(t))_i + \chi_{\{\mathbb{0}_n\}}(Rx(t))\|\dot{x}(t)\|_{\infty,R}. \end{aligned}$$

This proves the first result. To get the second result, multiply both sides by $\|x(t)\|_{\infty,R}$ to get

$$\|x(t)\|_{\infty,R}D^+\|x(t)\|_{\infty,R} = [\![\dot{x}(t), x(t)]\!]_{\infty,R} + \|x(t)\|_{\infty,R}\chi_{\{\mathbb{0}_n\}}(Rx(t))\|\dot{x}(t)\|_{\infty,R}.$$

Note that this second term is identically zero since if $\chi_{\{\mathbb{0}_n\}}(Rx(t)) = 1$, then $\|x(t)\|_{\infty,R} = 0$. □

Weak pairings, known expressions for log norms, and novel expressions for log norms from Lumer's equality for ℓ_p norms are summarized in Table 2.3.

2.4 Contraction theory via weak pairings

2.4.1 One-sided Lipschitz functions

Definition 1 (One-sided Lipschitz function). *Let $f\colon C \to \mathbb{R}^n$, where $C \subseteq \mathbb{R}^n$ is open and connected. We say f is* one-sided Lipschitz with respect to a weak pairing $[\![\cdot,\cdot]\!]$ *if* *the weak pairing satisfies Deimling's inequality,* (2.8)*, and there exists* $b \in \mathbb{R}$ *such that*

$$[\![f(x) - f(y), x - y]\!] \leq b\|x - y\|^2 \quad \textit{for all } x, y \in C. \tag{2.14}$$

We say b *is a* one-sided Lipschitz constant of f*. Moreover, the minimal one-sided Lipschitz constant of* f, $\mathrm{osL}(f)$*, is*

$$\mathrm{osL}(f) := \sup_{x \neq y} \frac{[\![f(x) - f(y), x - y]\!]}{\|x - y\|^2} \in \mathbb{R} \cup \{\infty\}. \tag{2.15}$$

We prove the following proposition in Section 2.8.3.

Proposition 2.4.1 (Properties of $\mathrm{osL}(f)$)**.** *Let* $f, g \colon C \to \mathbb{R}^n$ *be one-sided Lipschitz with respect to a weak pairing* $[\![\cdot, \cdot]\!]$*. Then for* $c \in \mathbb{R}$ *and* $\mathrm{Id} \colon C \to C$ *the identity map:*

(i) $\mathrm{osL}(f) \leq \sup_{x \neq y} \frac{\|f(x) - f(y)\|}{\|x - y\|}$,

(ii) $\mathrm{osL}(f + c\,\mathrm{Id}) = \mathrm{osL}(f) + c$,

(iii) $\mathrm{osL}(\alpha f) = \alpha\, \mathrm{osL}(f), \quad$ *for all* $\alpha \geq 0$,

(iv) $\mathrm{osL}(f + g) \leq \mathrm{osL}(f) + \mathrm{osL}(g)$.

Remark 2.4.2. *When* $f \colon C \to \mathbb{R}^n$ *is continuously differentiable and* C *is convex,* $\mathrm{osL}(f)$ *does not depend on the choice of weak pairing and instead depends only on the norm since*

$$\sup_{x \neq y} \frac{[\![f(x) - f(y), x - y]\!]}{\|x - y\|^2} = \sup_{x \in C} \mu(Df(x)),$$

which follows from the mean-value theorem for vector-valued functions in conjunction with Lumer's equality. $\triangle$

2.4.2 Contraction equivalences for continuously differentiable vector fields

Theorem 2.4.1 (Contraction equivalences for continuously differentiable vector fields)**.** *Consider the dynamics $\dot{x} = f(t,x)$, with f continuously differentiable in x and continuous in t. Let $C \subseteq \mathbb{R}^n$ be open, convex, and forward invariant and let $\|\cdot\|$ denote a norm with compatible weak pairing $[\![\cdot,\cdot]\!]$ satisfying Deimling's inequality, (2.8). Then, for $b \in \mathbb{R}$, the following statements are equivalent:*

(i) $\mathrm{osL}(f(t,\cdot)) \leq b$ *with respect to the weak pairing* $[\![\cdot,\cdot]\!]$*, for all* $t \geq 0$*,*

(ii) $[\![Df(t,x)v, v]\!] \leq b\|v\|^2$*, for all* $v \in \mathbb{R}^n, x \in C, t \geq 0$*,*

(iii) $\mu(Df(t,x)) \leq b$*, for all* $x \in C, t \geq 0$*,*

(iv) $D^+\|\phi(t,t_0,x_0) - \phi(t,t_0,y_0)\| \leq b\|\phi(t,t_0,x_0) - \phi(t,t_0,y_0)\|$*, for all* $x_0, y_0 \in C, 0 \leq t_0 \leq t$ *for which the solutions exist,*

(v) $\|\phi(t,t_0,x_0) - \phi(t,t_0,y_0)\| \leq e^{b(t-s)}\|\phi(s,t_0,x_0) - \phi(s,t_0,y_0)\|$*, for all* $x_0, y_0 \in C$ *and* $0 \leq t_0 \leq s \leq t$ *for which the solutions exist.*

Proof. Regarding (i) $\implies$ (ii), if $v = \mathbb{0}_n$, the result is trivial. By definition of $\mathrm{osL}(f(t,\cdot))$, $[\![f(t,x) - f(t,y), x-y]\!] \leq b\|x-y\|^2$, for all $x, y \in C, t \geq 0$. Fix $y \neq x$ and set $x = y + hv$ for an arbitrary $v \in \mathbb{R}^n$ and $h \in \mathbb{R}_{>0}$ sufficiently small. Then

$$\begin{aligned} & [\![f(t,y+hv) - f(t,y), hv]\!] \leq b\|hv\|^2 \\ \implies \quad & h\,[\![f(t,y+hv) - f(t,y), v]\!] \leq bh^2\|v\|^2, \end{aligned}$$

by the weak homogeneity of the weak pairing, property (ii). Dividing by h^2 and taking the limit as h goes to zero yields

$$\lim_{h\to 0^+} \left[\!\!\left[\frac{f(t,y+hv)-f(t,y)}{h}, v\right]\!\!\right] \leq b\|v\|^2 \qquad \Longrightarrow \qquad [\![Df(t,y)v,v]\!] \leq b\|v\|^2,$$

which follows from the continuity of the weak pairing in its first argument, property (i). Since y, v, and t were arbitrary, this completes the implication.

Regarding (ii) $\implies$ (iii), suppose $[\![Df(t,x)v,v]\!] \leq b\|v\|^2$ for all $x \in C, v \in \mathbb{R}^n, t \geq 0$. Let $v \neq \mathbb{0}_n$ and divide by $\|v\|^2$. Then take the sup over all $v \neq \mathbb{0}_n$ to get $\mu(Df(t,x)) = \sup_{v\neq\mathbb{0}_n} [\![Df(t,x)v,v]\!]/\|v\|^2 \leq b$, by Lumer's equality.

Regarding (iii) $\implies$ (iv), define $x(t) = \phi(t,t_0,x_0), y(t) = \phi(t,t_0,y_0)$, and $v(t) = x(t) - y(t)$ for $x_0, y_0 \in C, t_0 \geq 0$. Then by an application of the mean-value theorem for vector-valued functions,

$$\dot{v} = \left(\int_0^1 Df(t,y+sv)ds\right)v.$$

By an application of Coppel's differential inequality, Lemma 2.2.10, we have

$$\begin{aligned} D^+\|v(t)\| &\leq \mu\left(\int_0^1 Df(t,y(t)+sv(t))ds\right)\|v(t)\| \\ &\leq \int_0^1 \mu(Df(t,y(t)+sv(t)))ds\|v(t)\| \leq b\|v(t)\|, \end{aligned}$$

which follows from the subadditivity of log norms, [1]. Substituting back gives the inequality.

Regarding (iv) $\implies$ (v); this follows from an application of the nonsmooth Grönwall inequality, Lemma 2.2.8 on the interval $[s,t] \subseteq [t_0,t]$.

Regarding (v) $\implies$ (i), let $x_0, y_0 \in C$, $t_0 \geq 0$ be arbitrary. Then for $h \geq 0$,

$$\begin{aligned}\|\phi(t_0 + h, t_0, x_0) - \phi(t_0 + h, t_0, y_0)\| &= \|x_0 - y_0 + h(f(t_0, x_0) - f(t_0, y_0))\| + O(h^2) \\ &\leq e^{bh}\|x_0 - y_0\|.\end{aligned}$$

Subtracting $\|x_0 - y_0\|$ on both sides, dividing by $h > 0$ and taking the limit as $h \to 0^+$, we get

$$\lim_{h\to 0^+} \frac{\|x_0 - y_0 + h(f(t_0, x_0) - f(t_0, y_0))\| - \|x_0 - y_0\|}{h} \leq \lim_{h\to 0^+} \frac{e^{bh} - 1}{h}\|x_0 - y_0\|.$$

Evaluating the right hand side limit gives

$$\lim_{h\to 0^+} \frac{\|x_0 - y_0 + h(f(t_0, x_0) - f(t_0, y_0))\| - \|x_0 - y_0\|}{h} \leq b\|x_0 - y_0\|. \tag{2.16}$$

But by the assumption of Deimling's inequality,

$$\begin{aligned}&[\![f(t_0, x_0) - f(t_0, y_0), x_0 - y_0]\!] \\ &\leq \|x_0 - y_0\| \lim_{h\to 0^+} \frac{\|x_0 - y_0 + h(f(t_0, x_0) - f(t_0, y_0))\| - \|x_0 - y_0\|}{h}.\end{aligned}$$

Then multiplying both sides of (2.16) by $\|x_0 - y_0\|$ gives

$$[\![f(t_0, x_0) - f(t_0, y_0), x_0 - y_0]\!] \leq b\|x_0 - y_0\|^2.$$

Since t_0, x_0 and y_0 were arbitrary, the result holds.

□

Remark 2.4.3. *(i) A vector field f satisfying conditions (i), (ii) or (iii) with $b < 0$ is said to be* strongly contracting with rate $|b|$, *see [23]. Condition (ii) is referred to* *as the* Demidovich condition*, see [57]. A system whose trajectories satisfy conditions (iv) or (v) with* $b < 0$ *is said to be* incrementally exponentially stable without overshoot*, see [58].*

(ii) Theorem 2.4.1 holds for any choice of weak pairing satisfying Deimling's inequality, (2.8)*, (but not necessarily the curve norm derivative formula). Moreover, if a weak pairing does not satisfy Deimling's inequality, condition (ii) still implies (v) since* $\sup_{\|x\|=1} [\![Ax, x]\!] \geq \mu(A)$ *from Theorem 2.3.2.* △

Contraction equivalences (i), (ii), and (iii) are transcribed for the ℓ_p norms in Table 2.1 for the choices of weak pairings given in the previous section.

2.4.3 Contraction equivalences for continuous vector fields

Theorem 2.4.2 (Contraction equivalences for continuous vector fields)**.** *Consider the dynamics* $\dot{x} = f(t, x)$*, with* f *continuous in* (t, x)*. Let* $C \subseteq \mathbb{R}^n$ *be open, connected, and forward invariant and let* $\|\cdot\|$ *denote a norm with compatible weak pairing* $[\![\cdot,\cdot]\!]$ *satisfying Deimling's inequality,* (2.8)*, and the curve norm derivative formula,* (2.9)*. Then, for* $b \in \mathbb{R}$*, the following statements are equivalent:*

(i) $\mathrm{osL}(f(t,\cdot)) \leq b$ *with respect to the weak pairing* $[\![\cdot,\cdot]\!]$*, for all* $t \geq 0$*,*

(ii) $D^+\|\phi(t, t_0, x_0) - \phi(t, t_0, y_0)\| \leq b\|\phi(t, t_0, x_0) - \phi(t, t_0, y_0)\|$*, for all* $x_0, y_0 \in C, 0 \leq t_0 \leq t$ *for which the solutions exist,*

(iii) $\|\phi(t, t_0, x_0) - \phi(t, t_0, y_0)\| \leq e^{b(t-s)}\|\phi(s, t_0, x_0) - \phi(s, t_0, y_0)\|$*, for all* $x_0, y_0 \in C$ *and* $0 \leq t_0 \leq s \leq t$ *for which the solutions exist.*

Moreover, if statements (i), (ii), and (iii) hold, then solutions are unique. Finally, if statements (i), (ii), and (iii) hold with $b < 0$ *and there exists* $x^* \in \mathbb{R}^n$ *such that* $f(t, x^*) = 0$ *for all* $t \geq 0$*, then solutions exist uniquely for all time* $t \geq 0$*.*

Proof. Regarding (i) $\implies$ (iii), let $x_0, y_0 \in C, t_0 \geq 0$. If $\phi(t, t_0, x_0) = \phi(t, t_0, y_0)$ for some $t \geq t_0$, then Lemma 2.2.11 implies that the result holds. So suppose $\phi(t, t_0, x_0) \neq \phi(t, t_0, y_0)$. Let $v(t) = \phi(t, t_0, x_0)$ and $w(t) = \phi(t, t_0, y_0)$ and apply the curve norm derivative formula to $v(t) - w(t)$:

$$\|v(t) - w(t)\| D^+ \|v(t) - w(t)\| = [\![f(t, v(t)) - f(t, w(t)), v(t) - w(t)]\!] ,$$

for almost every $t \geq 0$. By the assumption of (i), dividing by $\|v(t) - w(t)\| \neq 0$ implies that

$$D^+ \|\phi(t, t_0, x_0) - \phi(t, t_0, y_0)\| \leq b \|\phi(t, t_0, x_0) - \phi(t, t_0, y_0)\|,$$

for almost every $t \geq t_0$. Then applying the nonsmooth Grönwall inequality, Lemma 2.2.8, gives (iii). Regarding (iii) $\implies$ (i), the proof is the same as in Theorem 2.4.1. Regarding (ii) $\implies$ (iii), the result follows from the nonsmooth Grönwall inequality, Lemma 2.2.8. Regarding (iii) $\implies$ (ii), we invoke a concept from Section 2.8.1. Following the proof of (iii) $\implies$ (i) gives the inequality $(f(t, x) - f(t, y), x - y)_+ \leq b\|x - y\|^2$. Then applying the curve norm derivative formula for Deimling pairings, Lemma 2.8.3, with $v(t) = \phi(t, t_0, x_0), w(t) = \phi(t, t_0, y_0)$ for $x_0, y_0 \in C, t_0 \geq 0$, implies that

$$\|v(t) - w(t)\| D^+ \|v(t) - w(t)\| \leq (f(t, v(t)) - f(t, w(t)), v(t) - w(t))_+,$$

for all $t \geq t_0$ for which $v(t), w(t)$ exist. Then substituting the previous inequality gives the result. To see uniqueness, note that if $\phi(t_0, t_0, x_0) = \phi(t_0, t_0, y_0)$, then $\|x_0 - y_0\| = 0$ and $\|\phi(t, t_0, x_0) - \phi(t, t_0, y_0)\| = 0$ for all $t \geq t_0$ for which the solutions exist by (iii). Regarding existence, if $b < 0$, consider the flow $\phi(t, t_0, x^*)$ which is constant for all $t \geq t_0$. Then any other solution exponentially converges to x^* and must exist for all $t \geq 0$. $\square$

Remark 2.4.4. *If $f(t, x)$ is only piecewise continuous in t, then both (i) and (ii) im-*

ply (iii), but the converse need not hold. We refer to [59] for contraction results for piecewise smooth vector fields in terms of their Jacobians. Theorem 2.4.2 does not require the computation of Jacobians and demonstrates that contraction is completely captured by the one-sided Lipschitz condition. △

2.4.4 Equilibrium contraction

Theorem 2.4.3 (Equilibrium contraction theorem)**.** *Consider the dynamics $\dot{x} = f(t,x)$, with f continuous in (t,x). Assume there exists x^* satisfying $f(t,x^*) = \mathbb{0}_n$ for all $t \geq 0$. Let $C \subseteq \mathbb{R}^n$ be open, connected, and forward invariant with $x^* \in C$ and let $\|\cdot\|$ denote a norm with compatible weak pairing $[\![\cdot,\cdot]\!]$ satisfying Deimling's inequality, (2.8), and the curve norm derivative formula, (2.9). Then, for $b \in \mathbb{R}$, the following statements are equivalent:*

(i) $[\![f(t,x), x - x^*]\!] \leq b\|x - x^*\|^2$*, for all* $x \in C, t \geq 0$,

(ii) $D^+\|\phi(t,t_0,x_0) - x^*\| \leq b\|\phi(t,t_0,x_0) - x^*\|$*, for all* $x_0 \in C, 0 \leq t_0 \leq t$.

(iii) $\|\phi(t,t_0,x_0) - x^*\| \leq e^{b(t-s)}\|\phi(s,t_0,x_0) - x^*\|$*, for all* $x_0 \in C, 0 \leq t_0 \leq s \leq t$.

Moreover, if C is convex and there exists a continuous map $(t,x) \mapsto A(t,x) \in \mathbb{R}^{n\times n}$ such that $f(t,x) = A(t,x)(x - x^*)$ for all t,x, then $\mu(A(t,x)) \leq b$ for all t,x implies (i), (ii), and (iii).

Proof. Regarding (i) $\implies$ (iii), let $x_0 \in C, t_0 \geq 0$. If $\phi(t,t_0,x_0) = x^*$, then the result holds. So assume $\phi(t,t_0,x_0) \neq x^*$, let $v(t) = \phi(t,t_0,x_0)$ and apply the curve norm derivative formula to $v(t) - x^*$. Then

$$\|v(t) - x^*\|D^+\|v(t) - x^*\| = [\![f(t,v(t)), v(t) - x^*]\!] \leq b\|v(t) - x^*\|^2,$$

for almost every $t \geq 0$. Dividing by $\|v(t) - x^*\| \neq 0$ implies $D^+\|v(t) - x^*\| \leq b\|v(t) - x^*\|$ for almost every $t \geq 0$. Then the nonsmooth Grönwall inequality, Lemma 2.2.8, gives (iii).

Regarding (ii) $\implies$ (iii), this result follows by applying the nonsmooth Grönwall inequality in Lemma 2.2.8.

Regarding (iii) $\implies$ (i), let $x_0 \in C, t_0 \geq 0$. Then for every $h > 0$,

$$\|\phi(t_0 + h, t_0, x_0) - x^*\| = \|x_0 - x^* + h(f(t_0, x_0))\| + O(h^2) \leq e^{bh}\|x_0 - x^*\|.$$

Subtracting $\|x_0 - x^*\|$ on both sides, dividing by $h > 0$ and taking the limit as $h \to 0^+$, we get

$$\lim_{h \to 0^+} \frac{\|x_0 - x^* + h(f(t_0, x_0))\| - \|x_0 - x^*\|}{h} \leq \lim_{h \to 0^+} \frac{e^{bh} - 1}{h}\|x_0 - x^*\|.$$

Evaluating the right hand side limit and multiplying both sides by $\|x_0 - x^*\|$ gives

$$\|x_0 - x^*\| \lim_{h \to 0^+} \frac{\|x_0 - x^* + h(f(t_0, x_0))\| - \|x_0 - x^*\|}{h} \leq b\|x_0 - x^*\|^2,$$

However, by the assumption of the weak pairing satisfying Deimling's inequality, (2.8), we get $[\![f(t_0, x_0), x_0 - x^*]\!] \leq b\|x_0 - x^*\|^2$. Since t_0, x_0 were arbitrary, the result holds.

Regarding (iii) $\implies$ (ii), we invoke a concept from Section 2.8.1. Following the proof of (iii) $\implies$ (i), we have

$$(f(t, x), x - x^*)_+ \leq b\|x - x^*\|^2, \quad \text{for all } x \in C, t \geq 0. \tag{2.17}$$

Then let $x_0 \in C, t_0 \geq 0$, let $v(t) = \phi(t, t_0, x_0)$, and apply the curve norm derivative formula for Deimling pairings, Lemma 2.8.3, to $v(t) - x^*$ to get

$$\|v(t) - x^*\| D^+ \|v(t) - x^*\| \leq (f(t, v(t)), v(t))_+,$$

for all $t \geq 0$. Using the inequality in (2.17) proves the result.

Now suppose that there exists a continuous map $(t, x) \mapsto A(t, x)$ such that $f(t, x) = A(t, x)(x - x^*)$ and $\mu(A(t, x)) \leq b$ for all $x \in C$ and all $t \geq 0$. Let $x_0 \in C, t_0 \geq 0$ and let $v(t) = \phi(t, t_0, x_0) - x^*$. Then

$$\dot{v} = A(t, \phi(t, t_0, x_0))v.$$

Applying Coppel's differential inequality, Lemma 2.2.10, implies

$$D^+\|v(t)\| \leq \mu(A(t, \phi(t, t_0, x_0)))\|v(t)\|.$$

Substituting $\mu(A(t, \phi(t, t_0, x_0))) \leq b$ gives (ii). □

Remark 2.4.5. *(i) A vector field f satisfying condition (i), with $b < 0$ is said to be* equilibrium contracting with respect to x^* and with rate $|b|$.

(ii) If $f(t, x)$ is continuously differentiable in x, the mean value theorem for vector-valued functions implies

$$f(t, x) = f(t, x) - f(t, x^*) = \left(\int_0^1 Df(t, x^* + (x - x^*)s)ds\right)(x - x^*).$$

One can then define the average Jacobian *of $f(t, x)$ about the equilibrium x^* to be*

$$\overline{Df}_{x^*}(t, x) := \int_0^1 Df(t, x^* + (x - x^*)s)ds. \tag{2.18}$$

Therefore, there always exists at least one matrix-valued map $A(t,x)$ such that $f(t,x) = A(t,x)(x - x^*)$.

(iii) *Condition (ii) implies that the choice of Lyapunov function $V(x) = \|x - x^*\|$ for $b < 0$ gives global exponential stability within C.* △

Example 1 (Counterexample)**.** *To see that (i) need not imply $\mu(A(t,x)) \leq b$, consider the dynamics in $\mathbb{R}^2$*

$$\dot{x} = A(x)x = \begin{bmatrix} -x_2^2 - 1 & 0 \\ 0 & x_1^2 - 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \tag{2.19}$$

with equilibrium point $x^* = \mathbb{0}_2$. For the unweighted ℓ_2 norm, $x \mapsto A(x)x$ satisfies Theorem 2.4.3(i) with $b = -1$ since $[\![A(x)x, x]\!]_2 = x^\top A(x)x = -\|x\|_2^2$. However, $\mu_2(A(x)) = x_1^2 - 1 \geq -1$. △

2.5 Robustness of contracting systems

We include a brief review of signal norms and system gains and refer the reader to [21, Chapter 2] for more details.

Definition 7 (Signal norms and system gains [21, Chapter 2])**.** *Given a norm $\|\cdot\|_{\mathcal{X}}$ on $\mathcal{X} = \mathbb{R}^n$, let $\mathcal{L}^q_{\mathcal{X}}$, $q \in [1,\infty]$, denote the vector space of continuous signals $x\colon \mathbb{R}_{\geq 0} \to \mathbb{R}^n$ with well-defined and bounded norm*

$$\|x(\cdot)\|_{\mathcal{X},q} := \begin{cases} \Big(\int_0^\infty \|x(t)\|_{\mathcal{X}}^q dt \Big)^{1/q}, & q \neq \infty, \\ \sup_{t \geq 0} \|x(t)\|_{\mathcal{X}}, & q = \infty. \end{cases} \tag{2.20}$$

A dynamical system with state $x \in \mathcal{X} = \mathbb{R}^n$ and input $u \in \mathcal{U} = \mathbb{R}^k$ has $\mathcal{L}^q_{\mathcal{X},\mathcal{U}}$ gain *bounded* *by* $\gamma > 0$ *if, for all* $u \in \mathcal{L}^q_{\mathcal{U}}$*, the state* x *from zero initial condition satisfies*

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

In what follows, for a control system $\dot{x} = f(t, x, u(t))$, we write $x(t)$ for the flow $\phi(t, t_0, x_0)$ subject to the vector field resulting from control input $u_x(t)$.

Theorem 2.5.1 (Input-to-state stability and gain of contracting systems)**.** *For a time and input-dependent vector field* f*, consider the dynamics*

$$\dot{x} = f(t, x, u(t)), \qquad x(0) = x_0 \in \mathcal{X} = \mathbb{R}^n, \tag{2.21}$$

where u *takes values in* $\mathcal{U} = \mathbb{R}^k$*. Assume there exists a norm* $\|\cdot\|_{\mathcal{X}} \colon \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ *with compatible weak pairing satisfying the curve norm derivative formula, (2.9), for all time* $[\![\cdot,\cdot]\!]_{\mathcal{X}}$*, a norm* $\|\cdot\|_{\mathcal{U}} \colon \mathbb{R}^k \to \mathbb{R}_{\geq 0}$*, and positive scalars* c *and* ℓ *such that*

(A1) $\mathrm{osL}(f(t,\cdot,u)) \leq -c$ *with respect to the weak pairing* $[\![\cdot,\cdot]\!]_{\mathcal{X}}$*, for all* $t \geq 0$*,* $u \in \mathbb{R}^k$*,*

(A2) $\|f(t,x,u) - f(t,x,v)\|_{\mathcal{X}} \leq \ell \|u - v\|_{\mathcal{U}}$*, for all* $t \in \mathbb{R}_{\geq 0}$*,* $x \in \mathbb{R}^n$*,* $u, v \in \mathbb{R}^k$*.*

Then

(i) any two solutions $x(t)$ *and* $y(t)$ *to (2.21) with continuous input signals* $u_x, u_y \colon \mathbb{R}_{\geq 0} \to \mathbb{R}^k$ *satisfy for all* $t \geq 0$*,*

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

(ii) f *is* incrementally input-to-state stable*, in the sense that, from any initial condi-*

tions $x_0, y_0 \in \mathbb{R}^n$,

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

(iii) f *has* incremental $\mathcal{L}^q_{\mathcal{X},\mathcal{U}}$ gain bounded by ℓ/c, for $q \in [1, \infty]$, *in the sense that solutions with* $x(0) = y(0)$ *satisfy*

$$\|x(\cdot) - y(\cdot)\|_{\mathcal{X},q} \leq \frac{\ell}{c} \|u_x(\cdot) - u_y(\cdot)\|_{\mathcal{U},q}. \tag{2.22}$$

Next, assume that f *satisfies the weaker Assumptions* (A1′)-(A2) *instead of* (A1)-(A2), *where*

(A1′) *there exists* $x^* \in \mathbb{R}^n$ *and continuous* $u^* \colon \mathbb{R}_{\geq 0} \to \mathbb{R}^k$ *such that* $f(t, x^*, u^*(t)) = \mathbb{0}_n$ *for all* t, *and* $[\![f(t,x,u) - f(t,x^*,u), x - y]\!]_{\mathcal{X}} \leq -c\|x - x^*\|^2_{\mathcal{X}}$ *for all* $t \geq 0, x \in \mathbb{R}^n, u \in \mathbb{R}^k$.

Then

(iv) the solution $x(t)$ *to* (2.21) *satisfies* $D^+\|x(t) - x^*\|_{\mathcal{X}} \leq -c\|x(t) - x^*\|_{\mathcal{X}} + \ell\|u(t) - u^*(t)\|_{\mathcal{U}}$, *for all* $t \geq 0$,

(v) f *is* input-to-state stable *in the sense that* $\|x(t) - x^*\|_{\mathcal{X}} \leq e^{-ct}\|x_0 - x^*\|_{\mathcal{X}} + \dfrac{\ell(1 - e^{-ct})}{c} \sup_{\tau \in [0,t]} \|u(\tau) - u^*(\tau)\|_{\mathcal{U}}$,

(vi) f *has* $\mathcal{L}^q_{\mathcal{X},\mathcal{U}}$ gain bounded by ℓ/c, for $q \in [1, \infty]$.

Proof. The result holds at times $t \geq 0$ when $x(t) = y(t)$ by Lemma 2.2.11. So assume

$x(t) \neq y(t)$. By the curve norm derivative formula, Assumptions (A1) and (A2) imply

$$\begin{aligned}\|x(t)-y(t)\|_{\mathcal{X}} D^+\|x(t)-y(t)\|_{\mathcal{X}} &= [\![f(t,x(t),u_x(t)) - f(t,y(t),u_y(t)), x(t)-y(t)]\!]_{\mathcal{X}} \\ &\leq [\![f(t,x(t),u_x(t)) - f(t,y(t),u_x(t)), x(t)-y(t)]\!]_{\mathcal{X}} \\ &\quad + [\![f(t,y(t),u_x(t)) - f(t,y(t),u_y(t)), x(t)-y(t)]\!]_{\mathcal{X}} \\ &\leq -c\|x(t)-y(t)\|_{\mathcal{X}}^2 + \ell\|x(t)-y(t)\|_{\mathcal{X}}\|u_x(t)-u_y(t)\|_{\mathcal{U}},\end{aligned}$$

which follows from the subadditivity and the Cauchy-Schwarz inequality for the weak pairing. This proves statement (i). Statement (ii) follows from the nonsmooth Grönwall inequality, Lemma 2.2.8. Regarding statement (iii), let $w(t) = \|u_x(t) - u_y(t)\|_{\mathcal{U}}$ and consider the scalar equation

$$\dot{\zeta} = -c\zeta + \ell w, \quad \zeta(0) = \|x(0)-y(0)\|_{\mathcal{X}} = 0. \tag{2.23}$$

Then by the Dini comparison lemma, Lemma 2.2.9, $\|x(t)-y(t)\|_{\mathcal{X}} \leq \zeta(t)$ for all $t \geq 0$. Let G be the linear operator given by $w \mapsto \zeta$ via the solution of (2.23). In other words, $\zeta = Gw$. Since G arises from a first-order scalar linear system, its induced $\mathcal{L}^q$ norm is ℓ/c for all $q \in [1,\infty]$, see [60, Proposition 2.3]. Thus,

$$\|x(\cdot)-y(\cdot)\|_{\mathcal{X},q} \leq \frac{\ell}{c}\|u_x(\cdot)-u_y(\cdot)\|_{\mathcal{U},q}, \text{ for all } q \in [1,\infty].$$

Regarding statement (iv), apply the curve norm derivative formula to $x(t) - x^*$ to get

$$\begin{aligned}\|x(t) - x^*\|_{\mathcal{X}} D^+\|x(t) - x^*\|_{\mathcal{X}} &= [\![f(t, x(t), u(t)) - f(t, x^*, u^*(t)), x(t) - x^*]\!]_{\mathcal{X}} \\ &\leq [\![f(t, x(t), u(t)) - f(t, x^*, u(t)), x(t) - x^*]\!]_{\mathcal{X}} \\ &\quad + [\![f(t, x^*, u(t)) - f(t, x^*, u^*(t)), x(t) - x^*]\!]_{\mathcal{X}} \\ &\leq -c\|x(t) - x^*\|_{\mathcal{X}}^2 + \ell\|u(t) - u^*(t)\|_{\mathcal{U}}\|x(t) - x^*\|_{\mathcal{X}}.\end{aligned}$$

Statement (v) then follows from the nonsmooth Grönwall inequality, Lemma 2.2.8. The proof of statement (vi) is identical to the proof of (iii). □

The next result studies contractivity under perturbations.

Theorem 2.5.2 (Contraction under perturbations)**.** *Consider the dynamics* $\dot{x} = f(t, x) + g(t, x)$. *If* $\mathrm{osL}(f(t, \cdot)) \leq -c < 0$ *and* $\mathrm{osL}(g(t, \cdot)) \leq d \in \mathbb{R}$ *with respect to the same weak pairing* $[\![\cdot, \cdot]\!]$ *for all* $t \geq 0$, *then*

(i) (contractivity under perturbations) *if* $d < c$, *then* $f + g$ *is strongly contracting with rate* $c - d$,

(ii) (equilibrium point under perturbations) *if additionally* f *and* g *are time-invariant, then the unique equilibrium point* x^* *of* f *and* x^{**} *of* $f + g$ *satisfy*

$$\|x^* - x^{**}\| \leq \frac{\|g(x^*)\|}{c - d}. \tag{2.24}$$

Proof. Statement (i) is an immediate consequence of Proposition 2.4.1(iv). Finally, we consider the two initial value problems $\dot{x} = f(x) + g(x)$ and $\dot{y} = f(y) + g(y) - g(x^*)$ with arbitrary initial conditions. Note $f(x^*) + g(x^*) - g(x^*) = 0$, that is, x^* is the unique equilibrium of the contracting system $f(y) + g(y) - g(x^*)$. Taking the limit as $t \to \infty$, Theorem 2.5.1(ii) implies statement (ii). □

2.6 Networks of contracting systems

We consider the *interconnection of* n *dynamical systems*

$$\dot{x}_i = f_i(t, x_i, x_{-i}), \quad \text{for } i \in \{1, \dots, n\}, \tag{2.25}$$

where $x_i \in \mathbb{R}^{N_i}$, $N = \sum_{i=1}^{n} N_i$, the subscript $-i = \{1, \dots, n\} \setminus \{i\}$ so that $x_{-i} \in \mathbb{R}^{N-N_i}$, and $f_i \colon \mathbb{R}_{\geq 0} \times \mathbb{R}^{N_i} \times \mathbb{R}^{N-N_i} \to \mathbb{R}^{N_i}$ is continuous. Let $\|\cdot\|_i$ and $[\![\cdot,\cdot]\!]_i$ denote a norm and a weak pairing on $\mathbb{R}^{N_i}$. Assume

(C1) at fixed x_{-i} and t, each map $x_i \mapsto f_i(t, x_i, x_{-i})$ satisfies $\mathrm{osL}(f_i(t, \cdot, x_{-i})) \leq -c_i < 0$ with respect to $[\![\cdot,\cdot]\!]_i$ which satisfies Deimling's inequality, (2.8), and the curve norm derivative formula, (2.9). If f_i is continuously differentiable in x_i, then this condition is equivalent to

$$\mu_i(Df_i(t, x_i, x_{-i})) \leq -c, \ \text{ for } x_i \in \mathbb{R}^{N_i},\ x_{-i} \in \mathbb{R}^{N-N_i}$$

(C2) at fixed x_i and t, each map $x_{-i} \mapsto f_i(t, x_i, x_{-i})$ satisfies a Lipschitz condition where, for all $j \in \{1, \dots, n\} \setminus \{i\}$, there exists $\gamma_{ij} \in \mathbb{R}_{\geq 0}$, such that for all $x_j, y_j \in \mathbb{R}^{N_j}$,

$$\|f_i(t, x_i, x_{-i}) - f_i(t, x_i, y_{-i})\|_i \leq \sum_{j=1, j\neq i}^{n} \gamma_{ij} \|x_j - y_j\|_j.$$

Lemma 2.6.1 (Efficiency of diagonally weighted norm [61, Lemma 3])**.** *Let* $M \in \mathbb{R}^{n\times n}$ *be Metzler and let* $\alpha(M)$ *denote its spectral abscissa. Then, for each* $\varepsilon > 0$*, there exists* $\xi \in \mathbb{R}^n_{>0}$ *satisfying the LMI*

$$\mathrm{diag}(\xi)M + M^\top \mathrm{diag}(\xi) \preceq 2(\alpha(M) + \varepsilon)\mathrm{diag}(\xi). \tag{2.26}$$

Moreover, if M is irreducible, the result holds with $\varepsilon = 0$.

Definition 8 (Diagonally weighted aggregation norm)**.** *For $\xi \in \mathbb{R}^n_{>0}$, define the ξ-weighted norm and corresponding weak pairing on $\mathbb{R}^N$ by*

$$\|(x_1, \dots, x_n)\|_\xi^2 = \sum\nolimits_{i=1}^n \xi_i \|x_i\|_i^2, \tag{2.27}$$

$$[\![(x_1, \dots, x_n), (y_1, \dots, y_n)]\!]_\xi = \sum\nolimits_{i=1}^n \xi_i [\![x_i, y_i]\!]_i \,. \tag{2.28}$$

It is easy to see that (2.27) defines a norm. We prove in Section 2.8.4 that (2.28) is a weak pairing that is compatible with (2.27) and satisfies Deimling's inequality, (2.8) and the curve norm derivative formula, (2.9).

Theorem 2.6.1 (Contractivity of interconnected system)**.** *Consider the interconnection of continuous systems (2.25) satisfying Assumptions (C1) and (C2) and define the* gain matrix

$$\Gamma := \begin{bmatrix} -c_1 & \dots & \gamma_{1n} \\ \vdots & & \vdots \\ \gamma_{n1} & \dots & -c_n \end{bmatrix}.$$

If Γ is Hurwitz, then

(i) for every $\varepsilon \in \,]0, |\alpha(\Gamma)|[$, there exists $\xi \in \mathbb{R}^n_{>0}$ such that the interconnected system is strongly contracting with respect to $\|\cdot\|_\xi$ in (2.27) with rate $|\alpha(\Gamma) + \varepsilon|$, and

(ii) if Γ is irreducible, the result (i) holds with $\varepsilon = 0$.

Proof. For $i \in \{1, \dots, n\}$, Assumptions (C1) and (C2) imply

$$\begin{aligned} [\![f_i(t, x_i, x_{-i}) - f_i(t, y_i, y_{-i}), x_i - y_i]\!]_i &\leq [\![f_i(t, x_i, x_{-i}) - f_i(t, y_i, x_{-i}), x_i - y_i]\!]_i \\ &\quad + [\![f_i(t, y_i, x_{-i}) - f_i(t, y_i, y_{-i}), x_i - y_i]\!]_i \\ &\leq -c_i\|x_i - y_i\|_i^2 + \sum\nolimits_{j=1, j\neq i}^n \gamma_{ij}\|x_j - y_j\|_j \|x_i - y_i\|_i, \end{aligned}$$

where we used the subadditivity and Cauchy-Schwarz inequality for the weak pairing. By Lemma 2.6.1, for $\varepsilon \in \,]0, |\alpha(\Gamma)|[$, select $\xi \in \mathbb{R}^n_{>0}$ satisfying (2.26). Next, we check the one-sided Lipschitz condition for the interconnected system on $\mathbb{R}^N$ with respect to norm (2.27) and weak pairing (2.28):

$$\begin{aligned}
&\sum\nolimits_{i=1}^{n} \xi_i \, [\![f_i(t, x_i, x_{-i}) - f_i(t, y_i, y_{-i}), x_i - y_i]\!]_i \\
&\leq -\sum\nolimits_{i=1}^{n} \xi_i c_i \|x_i - y_i\|_i^2 + \sum\nolimits_{i,j=1, j\neq i}^{n} \xi_i \gamma_{ij} \|x_j - y_j\|_j \|x_i - y_i\|_i \\
&= \begin{bmatrix} \|x_1 - y_1\|_1 \\ \vdots \\ \|x_n - y_n\|_n \end{bmatrix}^\top \operatorname{diag}(\xi)\Gamma \begin{bmatrix} \|x_1 - y_1\|_1 \\ \vdots \\ \|x_n - y_n\|_n \end{bmatrix} \\
&= \begin{bmatrix} \|x_1 - y_1\|_1 \\ \vdots \\ \|x_n - y_n\|_n \end{bmatrix}^\top \frac{\operatorname{diag}(\xi)\Gamma + \Gamma^\top \operatorname{diag}(\xi)}{2} \begin{bmatrix} \|x_1 - y_1\|_1 \\ \vdots \\ \|x_n - y_n\|_n \end{bmatrix},
\end{aligned}$$

so that the interconnected system is strongly contracting if the gain matrix Γ is diagonally stable. Moreover, using Lemma 2.6.1

$$\begin{aligned}
\sum\nolimits_{i=1}^{n} &\xi_i \, [\![f_i(t, x_i, x_{-i}) - f_i(t, y_i, y_{-i}), x_i - y_i]\!]_i \\
&\leq (\alpha(\Gamma) + \varepsilon) \begin{bmatrix} \|x_1 - y_1\|_1 \\ \vdots \\ \|x_n - y_n\|_n \end{bmatrix}^\top \operatorname{diag}(\xi) \begin{bmatrix} \|x_1 - y_1\|_1 \\ \vdots \\ \|x_n - y_n\|_n \end{bmatrix} \\
&= (\alpha(\Gamma) + \varepsilon) \sum\nolimits_{i=1}^{n} \xi_i \|x_i - y_i\|_i^2 = (\alpha(\Gamma) + \varepsilon) \|(x_1 - y_1, \dots, x_n - y_n)\|_\xi^2.
\end{aligned}$$

Then by Theorem 2.4.2, we have strong contraction with rate $|\alpha(\Gamma) + \varepsilon|$ and incremental exponential stability. Finally, if Γ is irreducible, we can take $\varepsilon = 0$ by Lemma 2.6.1. $\square$

An example interconnected system satisfying the Assumptions (C1) and (C2) is of the form $f_i(t, x_i, x_{-i}) = g_i(t, x_i) + \sum_{j=1, j\neq i}^{n} H_{ij} x_j$, where each vector field $g_i(t, x_i)$ has one-sided Lipschitz constant $-c_i$ and where Assumption (C2) is satisfied with γ_{ij} equal to the induced gain of H_{ij}.

Remark 2.6.2 (Input-to-state stability and gain of interconnected contracting systems)**.** *Consider interconnected subsystems of the form $\dot{x}_i = f(t, x_i, x_{-i}, u_i)$ with an input $u_i \in \mathbb{R}^{k_i}$. Assume each f_i satisfies Assumptions (C1) and (C2) at fixed input and, for fixed x_i, x_{-i}, t and all $u_i, v_i \in \mathbb{R}^{k_i}$,*

$$\|f(t, x_i, x_{-i}, u_i) - f(t, x_i, x_{-i}, v_i)\|_i \leq \ell_i \|u_i - v_i\|_{\mathcal{U}_i},$$

for some norm $\|\cdot\|_{\mathcal{U}_i}$ on $\mathbb{R}^{k_i}$. Then, with $c = |\alpha(\Gamma) + \varepsilon|$, Theorem 2.5.1 shows that the interconnected system is incrementally input-to-state stable with

$$\begin{aligned}\|x(t) - y(t)\|_\xi \leq e^{-ct}\|x_0 - y_0\|_\xi \\ + \frac{1 - e^{-ct}}{c} \sum_{i=1}^{n} \ell_i \xi_i \sup_{\tau\in[0,t]} \|u_{x,i}(\tau) - u_{y,i}(\tau)\|_{\mathcal{U}_i},\end{aligned}$$

and has finite incremental $\mathcal{L}^q_{\mathcal{X},\mathcal{U}}$ gain, for any $q \in [1, \infty]$. △

2.7 Conclusions

In this chapter, we present weak pairings as a novel tool to study contraction theory with respect to arbitrary norms. Through the language of weak pairings, we prove contraction equivalences for continuously differentiable vector fields, continuous vector fields, and for equilibrium contraction. For ℓ_p norms with $p \in [1, \infty]$, we present explicit formulas for the log norms, the Demidovich condition, and the one-sided Lipschitz condition, leading to novel contraction equivalences for $p \in \{1, \infty\}$. We then prove novel robustness results for contracting and equilibrium contracting systems including incremental input-to-state stability properties as well as finite incremental $\mathcal{L}^q_{\mathcal{X},\mathcal{U}}$ gain. Finally, we provide a main interconnection theorem for contracting subsystems, that provides a counterpart to similar theorems for dissipative subsystems.

Possible directions for future research include (i) leveraging our non-Euclidean conditions for control design akin to control contraction metrics [62], (ii) studying the generalization to nonsmooth Finsler Lyapunov functions and differentially positive systems [28, 63], (iii) exploring the additional structure in monotone systems [11], and finally (iv) studying generalizations of contraction including partial contraction [24], transverse contraction [30], and contraction after transients [31].

2.8 Auxiliary results

2.8.1 Deimling pairings

Definition 9 (Deimling pairing [52, Chapter 3])**.** *Given a norm $\|\cdot\|$ on $\mathbb{R}^n$, the Deimling pairing is the map $(\cdot,\cdot)_+ \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ defined by*

$$(x, y)_+ := \|y\| \lim_{h \to 0^+} \frac{\|y + hx\| - \|y\|}{h}. \tag{2.29}$$

This limit is known to exist for every $x, y \in \mathbb{R}^n$.

If a norm is differentiable, then its associated Lumer pairing coincides with the Deimling pairing. The Deimling pairing is also referred to as superior semi-inner product and right semi-inner product, [47, Chapter 3], [45, Remark 1].

Lemma 2.8.1 (Deimling pairing properties [47, Chapter 3, Proposition 5 and Corollary

5]). *Let* $\|\cdot\|$ *be a norm on* $\mathbb{R}^n$*. Then the following properties hold:*

(i) $(x_1 + x_2, y)_+ \leq (x_1, y)_+ + (x_2, y)_+$ *for all* $x_1, x_2, y \in \mathbb{R}^n$ *and* $(\cdot,\cdot)_+$ *is continuous in its first argument,*

(ii) $(\alpha x, y)_+ = (x, \alpha y)_+ = \alpha(x, y)_+$ *and* $(-x, -y)_+ = (x, y)_+$ *for all* $x, y \in \mathbb{R}^n, \alpha \geq 0$*,*

(iii) $(x, x)_+ = \|x\|^2$ *for all* $x \in \mathbb{R}^n$*,*

(iv) $|(x, y)_+| \leq \|x\|\|y\|$ *for all* $x, y \in \mathbb{R}^n$*.*

While a Deimling pairing need not be a Lumer pairing and a Lumer pairing need not be a Deimling pairing, both Deimling pairings and Lumer pairings are weak pairings.

Lemma 2.8.2 (Relationship between Deimling pairing and Lumer pairings for a norm [47, Chapter 3, Theorem 20])**.** *Let* $\|\cdot\|$ *be a norm on* $\mathbb{R}^n$ *and let* $[\cdot,\cdot]$ *be a compatible Lumer pairing. Then*

$$[x, y] \leq (x, y)_+, \quad \textit{for all } x, y \in \mathbb{R}^n. \tag{2.30}$$

Moreover, if $\mathcal{S}_\mathrm{p}$ *is the set of all Lumer pairings compatible with the norm, then* $(x, y)_+ = \sup_{[\cdot,\cdot]\in\mathcal{S}_\mathrm{p}}[x, y]$ *for all* $x, y \in \mathbb{R}^n$*.*

Lemma 2.8.3 (Deimling curve norm derivative formula [52, Proposition 13.1])**.** *Let* $x :]a, b[\to \mathbb{R}^n$ *be differentiable. Then*

$$\|x(t)\|D^+\|x(t)\| = (\dot{x}(t), x(t))_+, \quad \textit{for all } t \in]a, b[. \tag{2.31}$$

Hence, any Deimling pairing satisfies Deimling's inequality, (2.8), (by definition) and the curve norm derivative formula, (2.9), (with equality holding for all time).

Remark 2.8.4 (Logarithmic Lipschitz constant)**.** *In [38, Definition 5.2], the least upper* *bound logarithmic Lipschitz constant of a map* $f\colon C \to \mathbb{R}^n$ *is defined by*

$$M^+(f) := \sup_{x \neq y} \frac{(f(x) - f(y), x - y)_+}{\|x - y\|^2}. \tag{2.32}$$

In other words, $M^+(f)$ *is a special case of* $\mathrm{osL}(f)$*, where* osL *may be with respect to any weak pairing satisfying Deimling's inequality,* (2.8)*. Thus, we show that, in contraction theory, we are not restricted to using Deimling pairings for analysis.* △

Finally, for comparison's sake, we report from [52, Example 13.1(b)], the Deimling pairing for the ℓ_1 norm:

$$(x, y)_{+,1} = \|y\|_1 \big(\operatorname{sign}(y)^\top x + \sum\nolimits_{i=1}^n |x_i| \chi_{\{0\}}(y_i)\big). \tag{2.33}$$

2.8.2 Proof of Lemma 2.3.3

Before we prove Lemma 2.3.3, we first define a class of Lumer pairings called *single-index pairings.*

Lemma 2.8.5 (Single-index pairings)**.** *For* $R \in \mathbb{R}^{n\times n}$ *invertible, let* $\|\cdot\|_{\infty,R}$ *be the weighted* ℓ_∞ *norm. Let* $2^{[n]}$ *be the power set of* $\{1, \dots, n\}$*. A choice function on* $\{1, \dots, n\}$*,* $f : 2^{[n]} \setminus \{\emptyset\} \to \{1, \dots, n\}$ *satisfies for all* $S \in 2^{[n]} \setminus \{\emptyset\}$*,* $f(S) \in S$*. Let* $\mathcal{F}_{\text{choice}}$ *be the set of all choice functions on* $\{1, \dots, n\}$*. Then each* $f \in \mathcal{F}_{\text{choice}}$ *defines a Lumer pairing uniquely:*

$$[x, y]_{\infty,R} := (Rx)_{f(I_\infty(Ry))}(Ry)_{f(I_\infty(Ry))}. \tag{2.34}$$

Moreover, each Lumer pairing is compatible with the weighted ℓ_∞ *norm.*

Proof. First, we prove that any choice function defines a Lumer pairing. Let $f \in \mathcal{F}_{\text{choice}}$.

Regarding property (i), let $x_1, x_2, y \in \mathbb{R}^n$.

$$[x_1 + x_2, y]_{\infty,R} = R(x_1 + x_2)_{f(I_\infty(Ry))}(Ry)_{f(I_\infty(Ry))} = [x_1, y]_{\infty,R} + [x_2, y]_{\infty,R}.$$

For property (ii), let $\alpha \in \mathbb{R}$. Then

$$[\alpha x, y]_{\infty,R} = (R\alpha x)_{f(I_\infty(Ry))}(Ry)_{f(I_\infty(Ry))} = \alpha[x, y]_{\infty,R}.$$
$$[x, \alpha y]_{\infty,R} = (Rx)_{f(I_\infty(R\alpha y))}(R\alpha y)_{f(I_\infty(R\alpha y))} = \alpha[x, y]_{\infty,R}.$$

Regarding property (iii):

$$[x, x]_{\infty,R} = (Rx)_{f(I_\infty(Rx))}(Rx)_{f(I_\infty(Rx))} = \|x\|_{\infty,R}^2 \geq 0.$$

This also proves compatibility. Finally, for property (iv):

$$\begin{aligned}|[x, y]_{\infty,R}| &= |(Rx)_{f(I_\infty(Ry))}(Ry)_{f(I_\infty(Ry))}| = \|Ry\|_\infty |(Rx)_{f(I_\infty(Ry))}| \\ &\leq \|Ry\|_\infty \|Rx\|_\infty = [x, x]_{\infty,R}^{1/2}[y, y]_{\infty,R}^{1/2}.\end{aligned}$$

□

Corollary 2.8.6 (Relationship between max pairing and single-index pairings)**.** *Let $\mathcal{S}_{\text{index}}$ be the set of all single-index pairings on $\mathbb{R}^n$ compatible with norm $\|\cdot\|_{\infty,R}$. Then we have*

$$[\![x, y]\!]_{\infty,R} \geq [x, y]_{\infty,R}, \quad \textit{for all } [\cdot,\cdot]_{\infty,R} \in \mathcal{S}_{\text{index}}, x, y \in \mathbb{R}^n.$$

Moreover, for all $x, y \in \mathbb{R}^n$, there exists $[\cdot,\cdot]_{\infty,R} \in \mathcal{S}_{\text{index}}$ such that $[\![x, y]\!]_{\infty,R} = [x, y]_{\infty,R}$.

Proof. $[\![x, y]\!]_{\infty,R} \geq [x, y]_{\infty,R}$ follows by definition. Moreover, if x, y are fixed, let $i^* = \operatorname{argmax}_{i \in I_\infty(Ry)}(Ry)_i(Rx)_i$. Then any choice function satisfying $f(I_\infty(Ry)) = i^*$ defines a single-index pairing with $[\![x, y]\!]_{\infty,R} = [x, y]_{\infty,R}$. $\square$

Proof of Lemma 2.3.3. Let $x, y \in \mathbb{R}^n \setminus \{\mathbb{0}_n\}$. Then by Corollary 2.8.6, there exists $[\cdot,\cdot]_{\infty,R} \in \mathcal{S}_{\text{index}}$ such that $[\![x, y]\!]_{\infty,R} = [x, y]_{\infty,R}$. However, since $[\cdot,\cdot]_{\infty,R}$ is a Lumer pairing, it satisfies Deimling's inequality, (2.8). Thus,

$$[\![x, y]\!]_{\infty,R} = [x, y]_{\infty,R} \leq \|y\|_{\infty,R} \lim_{h\to 0^+} \frac{\|y + hx\|_{\infty,R} - \|y\|_{\infty,R}}{h}.$$

Since x, y were arbitrary, this proves the result. $\square$

2.8.3 Proof of Proposition 2.4.1

To prove Proposition 2.4.1, we will first prove one additional property of weak pairings.

Lemma 2.8.7. *Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$ with compatible weak pairing $[\![\cdot,\cdot]\!]$ satisfying Deimling's inequality, (2.8). Then for all $x \in \mathbb{R}^n, c \in \mathbb{R}$:*

$$[\![cx, x]\!] = c\|x\|^2. \tag{2.35}$$

Proof. If $c \geq 0$, the result is trivial, so without loss of generality, assume $c = -1$. Lumer's equality, Theorem 2.3.2, with $A = -I_n$ implies for all $x \in \mathbb{R}^n$

$$\sup_{x\neq\mathbb{0}_n} \frac{[\![-x, x]\!]}{\|x\|^2} = \mu(-I_n) = -1 \qquad \Longrightarrow \qquad [\![-x, x]\!] \leq -\|x\|^2.$$

Regarding the other inequality, observe that

$$[\![x, x]\!] = [\![x - x + x, x]\!] \leq [\![-x, x]\!] + 2\,[\![x, x]\!]$$
$$\Longrightarrow\ [\![-x, x]\!] \geq -\,[\![x, x]\!] = -\|x\|^2.$$

By weak homogeneity, this proves the result. $\square$

Proof of Proposition 2.4.1. Properties (i), (iii), and (iv) are consequences of Cauchy-Schwarz, weak homogeneity, and subadditivity of the weak pairing, respectively. Regarding property (ii), we will show the more general result that for any $x, y \in C, c \in \mathbb{R}$, $[\![x + cy, y]\!] = [\![x, y]\!] + c\|y\|^2$. The inequality

$$[\![x + cy, y]\!] \leq [\![x, y]\!] + c\|y\|^2,$$

follows from subadditivity and Lemma 2.8.7. Additionally,

$$[\![x, y]\!] = [\![x + cy - cy, y]\!] \leq [\![x + cy, y]\!] + [\![-cy, y]\!] = [\![x + cy, y]\!] - c\|y\|^2,$$

where the final equality holds by Lemma 2.8.7. Rearranging the inequality implies the result. $\square$

2.8.4 Sum decomposition of weak pairings

Lemma 2.8.8 (Weak pairing decomposition)**.** *For $N = \sum_{i=1}^n N_i$, let $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n) \in \mathbb{R}^N$ and $x_i, y_i \in \mathbb{R}^{N_i}$. Let $\|\cdot\|_i$ and $[\![\cdot,\cdot]\!]_i$ denote a norm and compatible weak pairing on $\mathbb{R}^{N_i}$ and let $\xi \in \mathbb{R}^n_{>0}$. Then the mapping $[\![\cdot,\cdot]\!]_\xi : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}^N$ defined by*

$$[\![x, y]\!]_\xi := \sum\nolimits_{i=1}^n \xi_i \, [\![x_i, y_i]\!]_i \,,$$

is a weak pairing compatible with the norm $\|x\|_\xi^2 = \sum_{i=1}^n \xi_i \|x_i\|_i^2$.

Proof. We verify the properties in Definition 3. Regarding property (i), let $x_1, x_2, y \in \mathbb{R}^N$. Then $[\![x_1 + x_2, y]\!]_\xi = \sum_{i=1}^n \xi_i \, [\![x_{1_i} + x_{2_i}, y_i]\!]_i \leq \sum_{i=1}^n \xi_i \, [\![x_{1_i}, y_i]\!]_i + \xi_i \, [\![x_{2_i}, y_i]\!]_i = [\![x_1, y]\!]_\xi + [\![x_2, y]\!]_\xi$. Continuity in the first argument follows from continuity of the first argument of each of the $[\![\cdot,\cdot]\!]_i$. Regarding property (ii), the result is straightforward because of weak homogeneity of each of the $[\![\cdot,\cdot]\!]_i$. Regarding property (iii),

$$[\![x, x]\!]_\xi = \sum_{i=1}^{n} \xi_i \, [\![x_i, x_i]\!]_i = \sum_{i=1}^{n} \xi_i \|x_i\|_i^2 > 0, \quad \text{for all } x \neq \mathbb{0}_N.$$

Regarding property (iv), since n is finite, by induction it suffices to check $n = 2$. For convenience, define $a_i = \xi_i \, [\![x_i, x_i]\!]_i \,, b_i = \xi_i \, [\![y_i, y_i]\!]_i$. Then

$$\begin{aligned}
[\![x, y]\!]_\xi^2 &= \Big(\sum\nolimits_{i=1}^{2} \xi_i \, [\![x_i, y_i]\!]_i \Big)^2 \leq \Big(\sum\nolimits_{i=1}^{2} a_i^{1/2} b_i^{1/2} \Big)^2 \\
&= (\sqrt{a_1 b_1} + \sqrt{a_2 b_2})^2 = a_1 b_1 + a_2 b_2 + 2\sqrt{a_1 b_1 a_2 b_2} \\
&\leq a_1 b_1 + a_2 b_2 + a_1 b_2 + a_2 b_1 = (a_1 + a_2)(b_1 + b_2) \\
&= \Big(\sum\nolimits_{i=1}^{2} a_i \Big) \Big(\sum\nolimits_{i=1}^{2} b_i \Big) = [\![x, x]\!]_\xi \, [\![y, y]\!]_\xi \,,
\end{aligned}$$

where we have used Cauchy-Schwarz for the $[\![\cdot,\cdot]\!]_i$ and the inequality $2\sqrt{\alpha\beta} \leq \alpha + \beta$ for $\alpha, \beta \geq 0$. Taking the square root of each side proves the result. □

Next we establish Deimling's inequality, (2.8), and the curve norm derivative formula, (2.9).

Lemma 2.8.9 (Deimling's inequality and curve norm derivative formula for weak pairing sum decomposition)**.** *Let* $[\![\cdot,\cdot]\!]_\xi$ *and* $\|\cdot\|_\xi$ *be defined as in Lemma 2.8.8. If*

(i) each $[\![\cdot,\cdot]\!]_i$ *satisfies Deimling's inequality, then* $[\![\cdot,\cdot]\!]_\xi$ *satisfies Deimling's inequality,*

(ii) each $[\![\cdot,\cdot]\!]_i$ *satisfies the curve norm derivative formula, then* $[\![\cdot,\cdot]\!]_\xi$ *satisfies the curve norm derivative formula.*

To prove Lemma 2.8.9(ii), we first prove a useful equivalent characterization of the curve norm derivative formula.

Proposition 2.8.10 (Equivalent curve norm derivative formula characterization)**.** *Let $x\colon (a,b) \to \mathbb{R}^n$ be differentiable and $[\![\cdot,\cdot]\!]$ be a weak pairing compatible with the norm $\|\cdot\|$ on $\mathbb{R}^n$. Then the following statements are equivalent*

(i) $\|x(t)\|D^+\|x(t)\| = [\![\dot{x}(t), x(t)]\!]$ *for almost every* $t \in (a,b)$,

(ii) $D^+\|x(t)\|^2 = 2\,[\![\dot{x}(t), x(t)]\!]$ *for almost every* $t \in (a,b)$.

Proof. We first prove (i) $\implies$ (ii). Initially, suppose that $t \in (a,b)$ is an instant in time at which $x(t) = \mathbb{0}_n$. Then we compute

$$\begin{aligned} D^+\|x(t)\|^2 &= \limsup_{h\to0^+} \frac{\|x(t+h)\|^2 - \|x(t)\|^2}{h} = \lim_{h\to0^+} \frac{\|x(t) + h\dot{x}(t)\|^2}{h} = \lim_{h\to0^+} \frac{\|h\dot{x}(t)\|^2}{h} \\ &= \lim_{h\to0^+} \frac{h^2\|\dot{x}(t)\|^2}{h} = 0, \end{aligned}$$

so the result holds for all $t \in (a,b)$ for which $x(t) = \mathbb{0}_n$. So alternatively suppose $x(t) \neq \mathbb{0}_n$. Then

$$\begin{aligned} &D^+\|x(t)\|^2 = \limsup_{h\to0^+} \frac{\|x(t+h)\|^2 - \|x(t)\|^2}{h} \\ &= \lim_{h\to0^+} \Big(\frac{\|x(t+h)\| - \|x(t)\|}{h}(\|x(t+h)\| + \|x(t)\|)\Big) \\ &= \big(D^+\|x(t)\|\big) \lim_{h\to0^+} \|x(t+h)\| + \|x(t)\| \overset{\text{a.e.}}{=} \frac{[\![\dot{x}(t), x(t)]\!]}{\|x(t)\|} \cdot 2\|x(t)\| = 2\,[\![\dot{x}(t), x(t)]\!]\,, \end{aligned}$$

where $\overset{\text{a.e.}}{=}$ denotes that the equality holds for almost every $t \in (a,b)$ by the assumption of (i). This proves (i) $\implies$ (ii). Regarding the other implication, note that the result holds trivially for all $t \in (a,b)$ for which $x(t) = \mathbb{0}_n$. Thus, we suppose that $x(t) \neq \mathbb{0}_n$ (note that this supposition implies that $\|x(t+h)\| + \|x(t)\| > 0$ for all $h > 0$). Then we compute

$$\begin{aligned}
D^+\|x(t)\| &= \limsup_{h\to 0^+} \frac{\|x(t+h)\| - \|x(t)\|}{h} \\
&= \lim_{h\to 0^+} \frac{\|x(t+h)\|^2 - \|x(t)\|^2}{h} \frac{1}{\|x(t+h)\| + \|x(t)\|} \\
&= \big(D^+\|x(t)\|^2\big) \lim_{h\to 0^+} \frac{1}{\|x(t+h)\| + \|x(t)\|} \overset{\text{a.e.}}{=} 2\,[\![\dot{x}(t), x(t)]\!]\, \frac{1}{2\|x(t)\|}.
\end{aligned}$$

Multiplying both sides by $\|x(t)\|$ proves the implication. □

We are now ready to prove Lemma 2.8.9.

Proof of Lemma 2.8.9. First we prove item (i). We prove the result for $n = 2$ and then by induction the result easily extends to arbitrary n. For $x = (x_1, x_2)$ and $y = (y_1, y_2)$, we compute

$$\begin{aligned}
[\![x, y]\!]_\xi &= \xi_1\,[\![x_1, y_1]\!]_1 + \xi_2\,[\![x_2, y_2]\!]_2 \\
&\leq \lim_{h\to 0^+} \Big(\xi_1\|y_1\|_1 \frac{\|y_1 + hx_1\|_1 - \|y_1\|_1}{h} + \xi_2\|y_2\|_2 \frac{\|y_2 + hx_2\|_2 - \|y_2\|_2}{h}\Big) \\
&= \lim_{h\to 0^+} \frac{\xi_1\|y_1\|_1\|y_1 + hx_1\|_1 + \xi_2\|y_2\|_2\|y_2 + hx_2\|_2 - \|y\|_\xi^2}{h},
\end{aligned}$$

where the first inequality holds by applying Deimling's inequality to each of $[\![\cdot,\cdot]\!]_i$ for $i \in \{1, 2\}$. Next we demonstrate that $\xi_1\|y_1\|_1\|y_1+hx_1\|_1+\xi_2\|y_2\|_2\|y_2+hx_2\|_2 \leq \|y\|_\xi\|y+hx\|_\xi$. Since both sides of the inequality are nonnegative, we square the left-hand side and compute

$$
\begin{aligned}
&\big(\xi_1\|y_1\|_1\|y_1+hx_1\|_1+\xi_2\|y_2\|_2\|y_2+hx_2\|_2\big)^2 = \\
&\xi_1^2\|y_1\|_1^2\|y_1+hx_1\|_1^2+\xi_2^2\|y_2\|_2^2\|y_2+hx_2\|_2^2+2\xi_1\xi_2\|y_1\|_1\|y_2\|_2\|y_1+hx_1\|_1\|y_2+hx_2\|_2 \\
&\leq \xi_1^2\|y_1\|_1^2\|y_1+hx_1\|_1^2+\xi_2^2\|y_2\|_2^2\|y_2+hx_2\|_2^2 \\
&+\xi_1\xi_2\|y_1\|_1^2\|y_2+hx_2\|_2^2+\xi_1\xi_2\|y_2\|_2^2\|y_1+hx_1\|_1^2 \\
&=\Big(\xi_1\|y_1\|_1^2+\xi_2\|y_2\|_2^2\Big)\Big(\xi_1\|y_1+hx_1\|_1^2+\xi_2\|y_2+hx_2\|_2^2\Big)=\|y\|_\xi^2\|y+hx\|_\xi^2,
\end{aligned}
$$

where the inequality holds due to $2\alpha\beta \leq \alpha^2+\beta^2$ for all $\alpha,\beta \in \mathbb{R}$ with $\alpha = \|y_1\|_1\|y_2 + hx_2\|_2, \beta = \|y_2\|_2\|y_2+hx_2\|_2$. This proves the desired inequality. As a consequence, we see

$$
\begin{aligned}
[\![x,y]\!]_\xi &\leq \lim_{h\to0^+}\frac{\xi_1\|y_1\|_1\|y_1+hx_1\|_1+\xi_2\|y_2\|_2\|y_2+hx_2\|_2-\|y\|_\xi^2}{h} \\
&\leq \lim_{h\to0^+}\frac{\|y\|_\xi\|y+hx\|_\xi-\|y\|_\xi^2}{h}=\|y\|_\xi\lim_{h\to0^+}\frac{\|y+hx\|_\xi-\|y\|_\xi}{h},
\end{aligned}
$$

which proves Deimling's inequality.

Regarding item (ii), let $x\colon (a,b)\to\mathbb{R}^N$ be differentiable. We apply Proposition 2.8.10 to prove that $D^+\|x(t)\|_\xi^2 = 2\,[\![\dot{x}(t),x(t)]\!]$ for almost every $t\in(a,b)$. We compute

$$
\begin{aligned}
D^+\|x(t)\|_\xi^2 &= D^+\Big(\sum_{i=1}^n \xi_i\|x_i(t)\|_i^2\Big)=\sum_{i=1}^n \xi_i D^+\|x_i(t)\|_i^2 \overset{\text{a.e.}}{=} 2\sum_{i=1}^n \xi_i\,[\![\dot{x}_i(t),x_i(t)]\!]_i \\
&= 2\,[\![\dot{x}(t),x(t)]\!]_\xi\,,
\end{aligned}
$$

where the third equality holds by the assumption that each $[\![\cdot,\cdot]\!]_i$ satisfies the curve norm derivative formula. Thus, the result is proved. □

Chapter 3

Non-Euclidean Monotone Operator Theory and Applications

This chapter was first published in the Journal of Machine Learning Research [64].[1]

3.1 Introduction

Problem description and motivation: Monotone operator theory is a fertile field of nonlinear functional analysis that extends the notion of monotone functions on $\mathbb{R}$ to mappings on Hilbert spaces. Monotone operator methods are widely used to solve problems in machine learning [65, 66], data science [67], optimization and control [68, 69], game theory [70], and systems analysis [71]. A crucial part of this theory is the design of algorithms for computing zeros of monotone operators. This problem is central in convex optimization since (i) the subdifferential of any convex function is monotone and (ii) minimizing a convex function is equivalent to finding a zero of its subdifferential.

To this end, there has been extensive research in the last decade in applying monotone operator methods to convex optimization; see, e.g., [72, 73, 74].

Existing monotone operator techniques are primarily based on inner-product spaces, while many problems are better-suited for analysis in more general normed spaces. For instance robustness analysis of artificial neural networks in machine learning often requires the use of the ℓ_∞ norm for high-dimensional input data such as images [75]. In distributed optimization, it is known that many natural conditions for the convergence of totally asynchronous algorithms are based upon contractions in an ℓ_∞ norm [76, Chapter 6, Section 3].

Motivated by problems in non-Euclidean spaces, we aim to extend monotone operator techniques for computing zeros of monotone operators to operators which are naturally "monotone" with respect to (w.r.t.) a non-Euclidean norm in a finite-dimensional space.

Literature review: The literature on monotone operators dates back to Minty and Browder [77, 78] and the connection to convex analysis was drawn upon by Minty and Rockafellar [79, 80]. Since these foundational works, the theory of monotone operators over Hilbert spaces and its connection with convex optimization continues to expand, especially in the last decade [4, 72, 74, 81]. Despite these connections between convex optimization and monotone operators, many problems in machine learning involve monotone operators beyond gradients of convex functions. Examples of such problems include generative adversarial networks, adversarially robust training of models, and training of models under fairness constraints. Instead of minimizing a convex function, to address these problems, one must solve for variational inequalities, monotone inclusions, and game-theoretic equilibria. In each of these more general cases, monotone operator theory has played an essential role in their analyses.

In machine learning, monotone operators have been used in the training of generative adversarial networks [82], in the design of novel neural network architectures [66], in the analysis of equilibrium behavior (infinite-depth limit) of neural networks [65], in the estimation of Lipschitz constants of neural networks [83, 84], and in normalizing flows [85]. Monotone operators have also been studied in the machine learning community in the context of variational inequality algorithms, stochastic monotone inclusions, and saddle-point problems; see e.g. [86, 87, 88, 89, 90, 91, 92, 93] for recent works in this direction. See also the recent survey [67] for applications in data science.

The theory of dissipative and accretive operators on Banach spaces largely parallels the theory of monotone operators on Hilbert spaces [52]. Despite these parallels, this theory has found far fewer direct applications to machine learning and data science; instead it is mainly applied for iterative solving integral equations and PDEs in L_p spaces for $p \neq 2$ (see the book [94] for iterative methods). Moreover, many works in Banach spaces focus on spaces that have a uniformly smooth or uniformly convex structure, which finite-dimensional ℓ_1 and ℓ_∞ spaces do not possess. In a similar vein, methods based on Bregman divergences utilize smoothness and strict convexity of the distance-generating convex functions [95]. Connections between logarithmic norms and dissipative and accretive operators may be found in [96, 38].

A concept similar to a monotone operator in a Hilbert space is that of a contracting vector field in dynamical systems theory [23]. If the metric with respect to which the vector field is contracting is the standard Euclidean distance, the vector field, F, is strongly infinitesimally contracting if and only if the negative vector field $-\mathsf{F}$ is strongly monotone when thought of as on operator on $\mathbb{R}^n$. However, vector fields need not be contracting with respect to a Euclidean distance. Indeed, a vector field may be contracting w.r.t. a non-Euclidean norm but not a Euclidean one [45]. Due to the connection between monotone operators and contracting vector fields, it is of interest to explore the properties of operators that may be thought of as monotone w.r.t. a non-Euclidean norm. In this spirit, preliminary connections between contracting vector fields and monotone operators were made in [97].

Contributions: Our contributions are as follows. First, to address the gap in applying monotone operator strategies to problems that arise in finite-dimensional non-Euclidean spaces, we propose a non-Euclidean monotone operator framework that is based on the theory of weak pairings [5] and logarithmic norms. We use weak pairings as a substitute for inner products and we demonstrate that many classic results from monotone operator theory are applicable to its non-Euclidean counterpart. In particular, we show that the resolvent and reflected resolvent operators of a non-Euclidean monotone mapping exhibit properties similar to those arising in Hilbert spaces. To ensure that the resolvent and reflected resolvents have full domain, we prove an extension of the classic Minty-Browder theorem [77, 78] in Theorem 3.3.1.

Second, leveraging the non-Euclidean monotone operator framework, we show that traditional iterative algorithms such as the forward step method and proximal point method can be used to compute zeros of non-Euclidean monotone mappings. We provide convergence rate estimates for these iterative algorithms and the Cayley method in Theorems 3.4.1, 3.4.2, and 3.4.3 and demonstrate that for diagonally-weighted ℓ_1 and ℓ_∞ norms, they exhibit improved convergence rates compared to their Euclidean counterparts. Notably, we prove that for a Lipschitz mapping which is monotone w.r.t. a diagonally-weighted ℓ_1 or ℓ_∞ norm, the forward step method is guaranteed to converge for a sufficiently small step size, whereas convergence cannot be guaranteed if the mapping is monotone with respect to a Euclidean norm.

Third, we study operator splitting methods for mappings which are monotone w.r.t. diagonally-weighted ℓ_1 or ℓ_∞ norms. In Theorems 3.5.1 and 3.5.2 we prove that the forward-backward, Peaceman-Rachford, and Douglas-Rachford splitting algorithms are all guaranteed to converge, with some key differences compared to the classical theory. For instance, in the classical setting where two operators, F and G, are monotone w.r.t.

a Euclidean norm, the forward-backward splitting algorithm will only converge if F is cocoercive. In contrast, when considering ℓ_1 or ℓ_∞ norms, Lipschitzness of F is sufficient for convergence.

Fourth, we present new insights into non-Euclidean properties of proximal operators and their impact on the study of special set-valued operator inclusions. Specifically, in Proposition 3.6.4, we demonstrate that when F is the subdifferential of a separable, proper, lower semicontinuous, convex function, its resolvent and reflected resolvent are nonexpansive with respect to an ℓ_∞ norm. To showcase the practical relevance of this result, we apply our non-Euclidean monotone operator theory to the equilibrium computation of a recurrent neural network (RNN). We extend the recent work of [98] and show that our theory provides novel iterations and convergence criteria for RNN equilibrium computation.

Finally, we study the robustness of the RNN via its ℓ_∞ norm Lipschitz constant. In Theorem 3.6.1, we generalize the results from [84] to non-Euclidean norms and provide sharper estimates for the ℓ_∞ Lipschitz constant than were provided in the previous work [98].

A preliminary version of this work appeared in [99]. Compared to this preliminary version, this version (i) provides novel theoretical results on the analysis of nonsmooth operators which are monotone with respect to general norms, (ii) proves a novel generalization of the classical Minty-Browder theorem for these non-Euclidean monotone mappings, (iii) study special classes of set-valued inclusions by providing novel non-Euclidean properties of proximal operators, (iv) includes a more comprehensive application to RNNs, allowing for more general activation functions and studies the robustness of the neural network by providing a tighter Lipschitz estimate, and (v) includes proofs of all technical results. Finally, we provide further comparisons to monotone operator theory on Hilbert spaces. Other prior work, [5, 100], focuses on continuous-time contracting dynamical systems with respect to non-Euclidean norms and their robustness properties. In contrast, this work instead uses weak pairings, developed in [5], to establish monotonicity properties of maps with respect to non-Euclidean norms and how we can find zeros of these maps using iterative methods. The prior works [5, 100] do not consider these discrete-time iterations.

3.2 Preliminaries

3.2.1 Weak Pairings

We briefly review the notion of a weak pairing (WP) on $\mathbb{R}^n$ from [5] which generalizes inner products to non-Euclidean spaces.

Definition 2 (Weak pairing). *A* weak pairing *is a map* $[\![\cdot,\cdot]\!] : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ *satisfying:*

(i) (sub-additivity and continuity of first argument) $[\![x_1 + x_2, y]\!] \leq [\![x_1, y]\!] + [\![x_2, y]\!]$*, for all* $x_1, x_2, y \in \mathbb{R}^n$ *and* $[\![\cdot,\cdot]\!]$ *is continuous in its first argument,*

(ii) (weak homogeneity) $[\![\alpha x, y]\!] = [\![x, \alpha y]\!] = \alpha\,[\![x, y]\!]$ *and* $[\![-x, -y]\!] = [\![x, y]\!]$*, for all* $x, y \in \mathbb{R}^n, \alpha \geq 0$*,*

(iii) (positive definiteness) $[\![x, x]\!] > 0$*, for all* $x \neq \mathbb{0}_n$*,*

(iv) (Cauchy-Schwarz inequality) $|\,[\![x, y]\!]\,| \leq [\![x, x]\!]^{1/2}\,[\![y, y]\!]^{1/2}$*, for all* $x, y \in \mathbb{R}^n$.

For every norm $\|\cdot\|$ on $\mathbb{R}^n$, there exists a (possibly not unique) compatible WP $[\![\cdot,\cdot]\!]$ such that $\|x\|^2 = [\![x, x]\!]$, for every $x \in \mathbb{R}^n$. If the norm is induced by an inner product, the WP coincides with the inner product.

Definition 3 (Deimling's inequality and curve norm derivative formula). *Let* $\|\cdot\|$ *be a norm on* $\mathbb{R}^n$ *with compatible WP* $[\![\cdot,\cdot]\!]$.

(i) The WP $[\![\cdot,\cdot]\!]$ *satisfies* Deimling's inequality *if*

$$[\![x,y]\!] \leq \|y\| \lim_{h\to 0^+} h^{-1}(\|y+hx\|-\|y\|), \qquad \textit{for all } x,y\in\mathbb{R}^n. \tag{3.1}$$

(ii) The WP $[\![\cdot,\cdot]\!]$ *satisfies the* curve norm derivative formula *if for all differentiable* $x\colon\,]a,b[\, \to \mathbb{R}^n$, $\|x(t)\|D^+\|x(t)\| = [\![\dot{x}(t),x(t)]\!]$ *holds for almost every* $t\in\,]a,b[$*, where* D^+ *denotes the upper right Dini derivative.*[2]

For every norm, there exists at least one WP that satisfies the properties in Definition 3.[3] Thus, going forward, we assume that WPs satisfy these additional properties.

We will focus on WPs corresponding to diagonally-weighted ℓ_1 and ℓ_∞ norms. Specifically, from [5, Table III], we introduce the WPs $[\![\cdot,\cdot]\!]_{1,[\eta]}, [\![\cdot,\cdot]\!]_{\infty,[\eta]^{-1}} : \mathbb{R}^n\times\mathbb{R}^n \to \mathbb{R}$ defined by

$$[\![x,y]\!]_{1,[\eta]} = \|y\|_{1,[\eta]}\operatorname{sign}(y)^\top[\eta]x \qquad \text{and} \qquad [\![x,y]\!]_{\infty,[\eta]^{-1}} = \max_{i\in I_\infty([\eta]^{-1}y)} \eta_i^{-2}y_ix_i. \tag{3.2}$$

where $I_\infty(x) = \{i\in\{1,\dots,n\} \mid |x_i| = \|x\|_\infty\}$. One can show that both of these WPs satisfy Deimling's inequality and the curve-norm derivative formula. Formulas for more general ℓ_p norms are available in [5].

3.2.2 Contractions, Nonexpansive Maps, and Iterations

Definition 4 (Lipschitz continuity)**.** *Let* $\|\cdot\|$ *be a norm and* $\mathsf{F}\colon\mathbb{R}^n\to\mathbb{R}^n$ *be a mapping.* F *is* Lipschitz continuous *with constant* $\ell\in\mathbb{R}_{\geq 0}$ *if*

$$\|\mathsf{F}(x_1)-\mathsf{F}(x_2)\| \leq \ell\|x_1-x_2\| \qquad \textit{for all } x_1,x_2\in\mathbb{R}^n. \tag{3.3}$$

[2] The definition and properties of Dini derivatives are presented in [101].

[3] Indeed, given a norm, the map $[\![\cdot,\cdot]\!]:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$ given by $[\![x,y]\!] = \|y\|\lim_{h\to 0^+}h^{-1}(\|y+hx\|-\|y\|)$ defines a WP that satisfies all of these properties. For more discussions about properties of this pairing, we refer to [52, Section 13] and [5, Appendix A].

Moreover we define $\mathsf{Lip}(\mathsf{F})$ *to be the minimal (or infimum) constant which satisfies* (3.3).

If two mappings $\mathsf{F}, \mathsf{G} \colon \mathbb{R}^n \to \mathbb{R}^n$ are Lipschitz continuous w.r.t. the same norm, then the composition $\mathsf{F} \circ \mathsf{G}$ has Lipschitz constant $\mathsf{Lip}(\mathsf{F} \circ \mathsf{G}) \leq \mathsf{Lip}(\mathsf{F}) \, \mathsf{Lip}(\mathsf{G})$.

Definition 5 (One-sided Lipschitz mappings, [5])**.** *Given a norm* $\|\cdot\|$ *with compatible WP* $[\![\cdot,\cdot]\!]$*, a map* $\mathsf{F} \colon \mathbb{R}^n \to \mathbb{R}^n$ *is* one-sided Lipschitz *with constant* $c \in \mathbb{R}$ *if*

$$[\![\mathsf{F}(x_1) - \mathsf{F}(x_2), x_1 - x_2]\!] \leq c\|x_1 - x_2\|^2 \qquad \textit{for all } x_1, x_2 \in \mathbb{R}^n. \tag{3.4}$$

Moreover we define $\mathrm{osL}(\mathsf{F})$ *to be the minimal (or infimum) constant which satisfies* (3.4).

As was proved in [5, Theorem 27], if $\mathsf{F}, \mathsf{G} \colon \mathbb{R}^n \to \mathbb{R}^n$ are one-sided Lipschitz w.r.t. the same WP, then $\mathrm{osL}(\alpha\mathsf{F}) = \alpha \, \mathrm{osL}(\mathsf{F})$, $\mathrm{osL}(\mathsf{F} + \mathsf{G}) \leq \mathrm{osL}(\mathsf{F}) + \mathrm{osL}(\mathsf{G})$, and $\mathrm{osL}(\mathsf{F} + c\mathsf{Id}) = \mathrm{osL}(\mathsf{F}) + c$ for all $\alpha \geq 0$, $c \in \mathbb{R}$. Note that (i) the one-sided Lipschitz constant is upper bounded by the Lipschitz constant, (ii) a Lipschitz continuous map is always one-sided Lipschitz, and (iii) the one-sided Lipschitz constant may be negative. Moreover, if F is locally Lipschitz *continuous*, we have an alternative characterization of $\mathrm{osL}(\mathsf{F})$.

Lemma 3.2.1 ($\mathrm{osL}(\mathsf{F})$ for locally Lipschitz continuous F, [100, Theorem 16])**.** *Suppose the map* $\mathsf{F} \colon \mathbb{R}^n \to \mathbb{R}^n$ *is locally Lipschitz continuous. Then* F *is one-sided Lipschitz with constant* $c \in \mathbb{R}$ *if and only if* [4]

$$\mu(D\mathsf{F}(x)) \leq c \qquad \textit{for almost every } x \in \mathbb{R}^n. \tag{3.5}$$

Definition 6 (Contractions and nonexpansive maps)**.** *Let* $\mathsf{T} \colon \mathbb{R}^n \to \mathbb{R}^n$ *be Lipschitz continuous w.r.t.* $\|\cdot\|$*. We say* T *is a* contraction *if* $\mathsf{Lip}(\mathsf{T}) < 1$*, and* T *is* nonexpansive *if* $\mathsf{Lip}(\mathsf{T}) \leq 1$.

[4]Note that for locally Lipschitz continuous F, $D\mathsf{F}(x)$ exists for almost every x by Rademacher's theorem.

Definition 7 (Picard iteration)**.** *Let* $\mathsf{T}\colon \mathbb{R}^n \to \mathbb{R}^n$ *be a contraction w.r.t. a norm* $\|\cdot\|$ *with* $\mathsf{Lip}(\mathsf{T}) < 1$*. The* Picard iteration *applied to* T *with initial condition* x^0 *defines the sequence* $\{x^k\}_{k=0}^{\infty}$ *by*

$$x^{k+1} = \mathsf{T}(x^k). \tag{3.6}$$

By the Banach fixed-point theorem T has a unique fixed point, x^*, and the Picard iteration applied to T satisfy $\|x^k - x^*\| \leq \mathsf{Lip}(\mathsf{T})^k \|x^0 - x^*\|$, for any initial condition x^0.

If T is nonexpansive with $\mathrm{Fix}(\mathsf{T}) \neq \emptyset$, Picard iteration may fail to find a fixed point of T. Such situations can be addressed by the following iteration and convergence result, initially proved in [102] and with rate given in [103].

Definition 8 (Krasnosel'skii–Mann iteration)**.** *Let* $\mathsf{T}\colon \mathbb{R}^n \to \mathbb{R}^n$ *be nonexpansive w.r.t. a norm* $\|\cdot\|$*. The* Krasnosel'skii–Mann iteration[5] *applied to* T *with initial condition* x^0 *and* $\theta \in \,]0,1[$ *defines the sequence* $\{x^k\}_{k=0}^{\infty}$ *by*

$$x^{k+1} = (1-\theta)x^k + \theta\mathsf{T}(x^k). \tag{3.7}$$

Lemma 3.2.2 (Asymptotic regularity and convergence of Krasnosel'skii–Mann iteration, [103, 102])**.** *Let* $\mathsf{T}\colon \mathbb{R}^n \to \mathbb{R}^n$ *be nonexpansive w.r.t. a norm* $\|\cdot\|$ *and consider the Krasnosel'skii–Mann iteration as in* (3.7)*. Suppose* $\mathrm{Fix}(\mathsf{T}) \neq \emptyset$*. Then for any initial condition* x^0*,*

$$\|x^k - \mathsf{T}(x^k)\| \leq \frac{2\inf_{x^*\in\mathrm{Fix}(\mathsf{T})} \|x^0 - x^*\|}{\sqrt{k\pi\theta(1-\theta)}} = \mathcal{O}(1/\sqrt{k}). \tag{3.8}$$

Moreover, the sequence of iterates, $\{x_k\}_{k=0}^{\infty}$*, converges to a fixed point of* T*.*

[5] The Krasnosel'skii–Mann iteration may be defined with step sizes $\theta_k \in \,]0,1[$ which vary for each iteration. In this document, we will only work with constant step sizes.

3.3 Non-Euclidean monotone operators

3.3.1 Definitions and properties

Definition 9 (Non-Euclidean monotone mapping)**.** *A mapping* $\mathsf{F}\colon \mathbb{R}^n \to \mathbb{R}^n$ *is strongly monotone with monotonicity parameter* $c > 0$ *w.r.t. a norm* $\|\cdot\|$ *on* $\mathbb{R}^n$ *if there exists a compatible WP* $[\![\cdot,\cdot]\!]$ *and if for all* $x, y \in \mathbb{R}^n$*,*

$$-[\![-(\mathsf{F}(x) - \mathsf{F}(y)), x - y]\!] \geq c\|x - y\|^2. \tag{3.9}$$

If the inequality holds with $c = 0$*, we say* F *is monotone w.r.t.* $\|\cdot\|$*.*

In the language of Banach spaces, such a function F is called strongly accretive [94, Definition 8.10]. Note that Definition 9 is equivalent to $-\operatorname{osL}(-\mathsf{F}) \geq c$.

In the case of a Euclidean norm, the WP corresponds to the inner product and Definition 9 corresponds to the usual definition of a monotone operator as in [77] and [4, Definition 20.1].

By properties of osL, if $\mathsf{F}, \mathsf{G}\colon \mathbb{R}^n \to \mathbb{R}^n$ are both monotone w.r.t. the same norm (and WP), then $-\operatorname{osL}(-\mathsf{F} - \mathsf{G}) \geq -\operatorname{osL}(-\mathsf{F}) - \operatorname{osL}(-\mathsf{G})$ and thus a sum of mappings which are monotone w.r.t. the same norm are monotone. Additionally, if F is monotone with monotonicity parameter $c \geq 0$, then for any $\alpha \geq 0$, $-\operatorname{osL}(-\mathsf{Id} - \alpha\mathsf{F}) = 1 - \alpha \operatorname{osL}(-\mathsf{F})$ and thus $\mathsf{Id} + \alpha\mathsf{F}$ is strongly monotone with monotonicity parameter $1 + \alpha c$.

Remark 3.3.1 (Connection with contracting vector fields)**.** *A mapping* $\mathsf{F}\colon \mathbb{R}^n \to \mathbb{R}^n$ *is strongly infinitesimally contracting with rate* $c > 0$ *w.r.t. a norm* $\|\cdot\|$ *on* $\mathbb{R}^n$ *provided* $\operatorname{osL}(\mathsf{F}) \leq -c$ *[5]. If* $c = 0$*, we say* F *is weakly infinitesimally contracting w.r.t.* $\|\cdot\|$*. Clearly* F *is strongly monotone if and only if* $-\mathsf{F}$ *is strongly infinitesimally contracting. Vector fields which are strongly infinitesimally contracting w.r.t. a norm generate flows* *which are contracting with respect to the same norm. In the case of weakly infinitesimally contracting vector fields, their flows are nonexpansive.*

Lemma 3.3.2 (Monotonicity for locally Lipschitz continuous mappings)**.** *Let* $\mathsf{F}\colon \mathbb{R}^n \to \mathbb{R}^n$ *be locally Lipschitz continuous.* F *is (strongly) monotone with monotonicity parameter* $c \geq 0$ *w.r.t. a norm* $\|\cdot\|$ *if and only if* $-\mu(-D\mathsf{F}(x)) \geq c$ *for almost every* $x \in \mathbb{R}^n$*.*

Proof. Lemma 3.3.2 is a straightforward application of Lemma 3.2.1. □

We can see the application of Lemma 3.3.2 more explicitly in the context of continuously differentiable monotone operators in Euclidean norms. To be specific, for an operator $\mathsf{F}\colon \mathbb{R}^n \to \mathbb{R}^n$, let $\|\cdot\|_2$ be the Euclidean norm with corresponding inner product $\langle\!\langle \cdot, \cdot \rangle\!\rangle$. Then, following [77], F is monotone with respect to $\|\cdot\|_2$ if

$$\langle\!\langle \mathsf{F}(x) - \mathsf{F}(y), x - y \rangle\!\rangle \geq 0, \qquad \text{for all } x, y \in \mathbb{R}^n.$$

If F is continuously differentiable, this condition is known to be equivalent to (see e.g., [72]) $D\mathsf{F}(x) + D\mathsf{F}(x)^\top \succeq 0$, or equivalently $-\mu_2(-D\mathsf{F}(x)) \geq 0$ or $\frac{1}{2}\lambda_{\min}(D\mathsf{F}(x) + D\mathsf{F}(x)^\top) \geq 0$, where $\mu_2(A) = \frac{1}{2}\lambda_{\max}(A + A^\top)$ is the log norm corresponding to the norm $\|\cdot\|_2$. This result coincides with what was demonstrated in Lemma 3.3.2.

Example 2. *An affine function* $\mathsf{F}(x) = Ax + b$ *is monotone if and only if* $-\mu(-A) \geq 0$ *and strongly monotone with parameter* c *if and only if* $-\mu(-A) \geq c$*. This condition implies that the spectrum of* A *lies in the portion of the complex plane given by* $\{z \in \mathbb{C} \mid \mathrm{Re}(z) \geq c\}$*.*

3.3.2 Resolvent, reflected resolvents, forward step operators, and Lipschitz estimates

Monotone operator theory transforms the problem of finding a zero of a monotone operator into finding a fixed point of a suitably defined operator. Monotone operator theory on Hilbert spaces studies the resolvent and reflected resolvent, operators dependent on the original operator, with fixed points corresponding to zeros of the original monotone operator. In this subsection we study these same two operators and also the forward step operator in the context of operators which are monotone w.r.t. a non-Euclidean norm. In particular, we characterize the Lipschitz constants of these operators, first providing Lipschitz upper bounds for arbitrary norms and then specializing to diagonally-weighted ℓ_1 and ℓ_∞ norms.

Definition 10 (Resolvent and reflected resolvent)**.** *Let $\mathsf{F}\colon \mathbb{R}^n \to \mathbb{R}^n$ be a monotone mapping w.r.t. some norm. The resolvent of F with parameter $\alpha > 0$ denoted by $\mathsf{J}_{\alpha\mathsf{F}}\colon \mathrm{Dom}(\mathsf{J}_{\alpha\mathsf{F}}) \to \mathbb{R}^n$ and defined by*

$$\mathsf{J}_{\alpha\mathsf{F}} = (\mathsf{Id} + \alpha\mathsf{F})^{-1}. \tag{3.10}$$

The reflected resolvent of F with parameter $\alpha > 0$ is denoted by $\mathsf{R}_{\alpha\mathsf{F}}\colon \mathrm{Dom}(\mathsf{R}_{\alpha\mathsf{F}}) \to \mathbb{R}^n$ and defined by

$$\mathsf{R}_{\alpha\mathsf{F}} = 2\mathsf{J}_{\alpha\mathsf{F}} - \mathsf{Id}. \tag{3.11}$$

Definition 11 (Forward step operator)**.** *Let $\mathsf{F}\colon \mathbb{R}^n \to \mathbb{R}^n$ be a mapping and $\alpha \in \mathbb{R}$. The forward step of F with parameter $\alpha > 0$ is denoted by $\mathsf{S}_{\alpha\mathsf{F}}\colon \mathbb{R}^n \to \mathbb{R}^n$ and defined by*

$$\mathsf{S}_{\alpha\mathsf{F}} = \mathsf{Id} - \alpha\mathsf{F}. \tag{3.12}$$

Note that for any $\alpha > 0$, we have $\mathsf{F}(x) = \mathbb{0}_n$ if and only if $x = \mathsf{J}_{\alpha\mathsf{F}}(x) = \mathsf{R}_{\alpha\mathsf{F}}(x) =$

$\mathsf{S}_{\alpha\mathsf{F}}(x)$, i.e., $\mathrm{Zero}(\mathsf{F}) = \mathrm{Fix}(\mathsf{J}_{\alpha\mathsf{F}}) = \mathrm{Fix}(\mathsf{R}_{\alpha\mathsf{F}}) = \mathrm{Fix}(\mathsf{S}_{\alpha\mathsf{F}})$. Note that under the assumption that F is monotone, both $\mathsf{J}_{\alpha\mathsf{F}}$ and $\mathsf{R}_{\alpha\mathsf{F}}$ are single-valued mappings.

We have deliberately not been specific with the domains of the resolvent and reflected resolvent operators. As we will show in the following theorem, under mild assumptions (continuity and monotonicity), both of their domains are all of $\mathbb{R}^n$.

Theorem 3.3.1 (A non-Euclidean Minty-Browder theorem)**.** *Suppose* $\mathsf{F}\colon \mathbb{R}^n \to \mathbb{R}^n$ *is continuous and monotone. Then for every* $\alpha > 0$, $\mathrm{Dom}(\mathsf{J}_{\alpha\mathsf{F}}) = \mathrm{Dom}(\mathsf{R}_{\alpha\mathsf{F}}) = \mathbb{R}^n$.

Proof. Note that $\mathrm{Dom}(\mathsf{J}_{\alpha\mathsf{F}}) = \mathbb{R}^n$ provided that for every $u \in \mathbb{R}^n$, there exists $x \in \mathbb{R}^n$ such that $(\mathsf{Id} + \alpha\mathsf{F})(x) = u$. To establish this fact, consider the differential equation

$$\dot{x} = -x - \alpha\mathsf{F}(x) + u =: G(x). \tag{3.13}$$

Note that any equilibrium, x^*, of (3.13) satisfies $(\mathsf{Id} + \alpha\mathsf{F})(x^*) = u$. Thus it suffices to show that the differential equation (3.13) has an equilibrium. First we note that for all $x, y \in \mathbb{R}^n$,

$$[\![G(x) - G(y), x - y]\!] \leq [\![-(x-y), x-y]\!] + \alpha\,[\![-(\mathsf{F}(x) - \mathsf{F}(y)), x - y]\!] \leq -\|x-y\|^2. \tag{3.14}$$

Thus, we conclude that $\mathrm{osL}(G) \leq -1$. In line with Remark 3.3.1, we conclude that G is strongly infinitesimally contracting which ensures uniqueness of solutions to (3.13) (see [5, Theorem 31]). Let $\phi(t, x_0)$ denote the flow of the dynamics (3.13) at time $t \geq 0$ from initial condition $x(0) = x_0$. Then by [5, Theorem 31], we conclude that

$$\|\phi(t, x_0) - \phi(t, y_0)\| \leq e^{-t}\|x_0 - y_0\|$$

for all $x_0, y_0 \in \mathbb{R}^n$ and for all $t \geq 0$. In other words, for a fixed $t > 0$, the map $x \mapsto \phi(t, x)$

is a contraction. By the Banach fixed point theorem, for $\tau > 0$, there exists unique x^* such that $x^* = \phi(\tau, x^*)$. Then either x^* is an equilibrium point of (3.13) or it is part of a periodic orbit with period τ. If x^* were part of a periodic orbit, then every other point on the periodic orbit would be a fixed point of $\phi(\tau, \cdot)$, contradicting the uniqueness of the fixed point from the Banach fixed point theorem. Thus, we conclude that x^* is an equilibrium point of (3.13) and thus verifies $(\mathsf{Id} + \alpha\mathsf{F})(x^*) = u$. This proves that $\mathrm{Dom}(\mathsf{J}_{\alpha\mathsf{F}}) = \mathbb{R}^n$. The proof for $\mathsf{R}_{\alpha\mathsf{F}}$ is a consequence of $\mathrm{Dom}(\mathsf{J}_{\alpha\mathsf{F}}) = \mathbb{R}^n$. $\square$

We have the following corollary about inverses of strongly monotone mappings.

Corollary 3.3.3 (Lipschitz constants of inverses of strongly monotone operators)**.** *Suppose $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ is continuous and strongly monotone with monotonicity parameter $c > 0$. Then $\mathsf{F}^{-1}\colon \mathbb{R}^n \to \mathbb{R}^n$ is a Lipschitz continuous mapping with Lipschitz constant estimate $\mathsf{Lip}(\mathsf{F}^{-1}) \leq 1/c$.*

Proof. To see this fact, note that

$$\|\mathsf{F}(x) - \mathsf{F}(y)\|\|x - y\| \geq -\left[\!\left[-(\mathsf{F}(x) - \mathsf{F}(y)), x - y\right]\!\right] \geq c\|x - y\|^2, \tag{3.15}$$

where the left hand inequality is the Cauchy-Schwarz inequality for WPs. So if $\mathsf{F}(x) = \mathsf{F}(y)$, then necessarily $x = y$, which implies that F^{-1} is a single-valued mapping. The fact that $\mathrm{Dom}(\mathsf{F}^{-1}) = \mathbb{R}^n$ follows the same argument as in Theorem 3.3.1 instead studying the differential equation $\dot{x} = -\mathsf{F}(x)$. Choosing $u, v \in \mathbb{R}^n$ and substituting $x = \mathsf{F}^{-1}(u), y = \mathsf{F}^{-1}(v)$ into (3.15), we conclude

$$\|u - v\| \geq c\|x - y\| = c\|\mathsf{F}^{-1}(u) - \mathsf{F}^{-1}(v)\|, \tag{3.16}$$

which shows that $\mathsf{Lip}(\mathsf{F}^{-1}) \leq 1/c$. $\square$

For each of $\mathsf{J}_{\alpha\mathsf{F}}, \mathsf{R}_{\alpha\mathsf{F}}$ and, $\mathsf{S}_{\alpha\mathsf{F}}$ we have now established that each of their domains is all of $\mathbb{R}^n$ and that fixed points of these operators correspond to zeros of F. In order to compute zeros of F, we aim to provide estimates of the Lipschitz constants of $\mathsf{J}_{\alpha\mathsf{F}}, \mathsf{R}_{\alpha\mathsf{F}}$, and $\mathsf{S}_{\alpha\mathsf{F}}$ as a function of α and the norm and show when these maps are either contractions or nonexpansive. The following lemmas characterize these Lipschitz estimates.

Lemma 3.3.4 (Lipschitz estimates of the forward step operator)**.** *Let $\mathsf{F}\colon \mathbb{R}^n \to \mathbb{R}^n$ be Lipschitz continuous w.r.t. the norm $\|\cdot\|$ with constant $\mathsf{Lip}(\mathsf{F}) = \ell$.*

(i) Suppose F is monotone w.r.t. $\|\cdot\|$ with monotonicity parameter $c \geq 0$, then

$$\mathsf{Lip}(\mathsf{S}_{\alpha\mathsf{F}}) \leq e^{-\alpha c} + e^{\alpha\ell} - 1 - \alpha\ell, \qquad \text{for all } \alpha > 0. \tag{3.17}$$

(ii) Alternatively suppose $\|\cdot\|$ is a diagonally weighted ℓ_1 or ℓ_∞ norm and F is monotone w.r.t. $\|\cdot\|$ with monotonicity parameter $c \geq 0$, then

$$\mathsf{Lip}(\mathsf{S}_{\alpha\mathsf{F}}) \leq 1 - \alpha c \leq 1, \qquad \text{for all } \alpha \in \Big(0, \frac{1}{\mathrm{diagL}(\mathsf{F})}\Big], \tag{3.18}$$

where $\mathrm{diagL}(\mathsf{F}) := \sup_{x\in\mathbb{R}^n\setminus\Omega_{\mathsf{F}}} \max_{i\in\{1,\dots,n\}} (D\mathsf{F}(x))_{ii} \leq \ell$, where Ω_{F} is the measure zero set of points where F is not differentiable.

Proof. Regarding item (i), we recall the inequality [18, pp. 14], [38, Prop. 2.1]

$$\|e^{\alpha A}\| \leq e^{\alpha\mu(A)}, \qquad \text{for all } \alpha \geq 0, A \in \mathbb{R}^{n\times n}. \tag{3.19}$$

We additionally note that since F is Lipschitz continuous, $\mathsf{S}_{\alpha\mathsf{F}}$ is as well and $\mathsf{S}_{\alpha\mathsf{F}}$ has $\mathsf{Lip}(\mathsf{S}_{\alpha\mathsf{F}}) \leq L$ if and only if $\|D\mathsf{S}_{\alpha\mathsf{F}}(x)\| \leq L$ for almost every $x \in \mathbb{R}^n$. Also we have that $D\mathsf{S}_{\alpha\mathsf{F}}(x) = I_n - \alpha D\mathsf{F}(x)$ everywhere it exists and that $D\mathsf{F}(x)$ satisfies $-\mu(-D\mathsf{F}(x)) \geq c$

and $\|D\mathsf{F}(x)\| \leq \ell$. In what follows, when we write $D\mathsf{F}(x)$, we mean for all x for which the Jacobian exists.

To derive an upper bound on $\|I_n - \alpha D\mathsf{F}(x)\|$, we define

$$S(x) := \sum_{i=2}^{\infty} \frac{(-\alpha)^i D\mathsf{F}(x)^i}{i!} = e^{-\alpha D\mathsf{F}(x)} - I_n + \alpha D\mathsf{F}(x)$$

and it is straightforward to see that $\|S(x)\| \leq \sum_{i=2}^{\infty} \frac{\alpha^i \|D\mathsf{F}(x)\|^i}{i!} \leq e^{\alpha\ell} - 1 - \alpha\ell$. Moreover,

$$\begin{aligned}\|I_n - \alpha D\mathsf{F}(x)\| &\leq \|e^{-\alpha D\mathsf{F}(x)}\| + \|S(x)\| \leq e^{\alpha\mu(-D\mathsf{F}(x))} + e^{\alpha\ell} - 1 - \alpha\ell \\ &\leq e^{-\alpha c} + e^{\alpha\ell} - 1 - \alpha\ell.\end{aligned} \tag{3.20}$$

Since this bound holds for all x for which $D\mathsf{S}_{\alpha\mathsf{F}}(x)$ exists, the result is proved.

Regarding item (ii), for every $x \in \mathbb{R}^n$ for which $D\mathsf{F}(x)$ exists,

$$\|I_n - \alpha D\mathsf{F}(x)\|_{\infty,[\eta]^{-1}} = \max_{i\in\{1,\dots,n\}} |1 - \alpha(D\mathsf{F}(x))_{ii}| + \sum_{j=1,j\neq i}^{n} |-\alpha(D\mathsf{F}(x))_{ij}|\frac{\eta_j}{\eta_i} \tag{3.21}$$

$$= \max_{i\in\{1,\dots,n\}} 1 - \alpha(D\mathsf{F}(x))_{ii} + \sum_{j=1,j\neq i}^{n} |-\alpha(D\mathsf{F}(x))_{ij}|\frac{\eta_j}{\eta_i} \tag{3.22}$$

$$= 1 + \alpha\mu(-D\mathsf{F}(x)) \leq 1 - \alpha c, \tag{3.23}$$

where (3.22) holds because $0 < \alpha \leq \frac{1}{\mathrm{diagL}(\mathsf{F})}$ so that $1 - \alpha(D\mathsf{F}(x))_{ii} \geq 0$ for all $x \in \mathbb{R}^n, i \in \{1, \dots, n\}$ and (3.23) is due to the formula for $\mu_{\infty,[\eta]^{-1}}$. The proof for $\mu_{1,[\eta]}$ is analogous, replacing row sums by column sums, and is omitted. □

Remark 3.3.5. *If $c > 0$, then for small enough $\alpha > 0$, one can make the upper bound on* $\mathsf{Lip}(\mathsf{S}_{\alpha\mathsf{F}})$ *in* (3.17) *less than unity. In particular, one can show that minimizing the upper bound* (3.17) *yields the optimal step size $\alpha_{\mathrm{opt}} = \frac{1}{\ell}\ln(s(\gamma))$ and contraction factor $s(\gamma) + s(\gamma)^{-\gamma} - 1 - \ln(s(\gamma))$, where $\gamma = c/\ell \leq 1$ and $s(\gamma)$ is the unique solution to the* *transcendental equation* $s - 1 - \gamma s^{-\gamma} = 0$.

Remark 3.3.6. *Note that for general norms, if* F *is monotone, but not strongly monotone, then* $\mathsf{S}_{\alpha\mathsf{F}}$ *need not be nonexpansive for any* $\alpha > 0$. *Indeed, consider* $\mathsf{F}(x) = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right) x$, *which is monotone w.r.t. the* ℓ_2 *norm, but* $\mathsf{S}_{\alpha\mathsf{F}}$ *is not nonexpansive for any* $\alpha > 0$. *On the other hand, Lemma 3.3.4((ii)) implies that if* F *is monotone w.r.t. a diagonally weighted* ℓ_1 *or* ℓ_∞ *norm, then* $\mathsf{S}_{\alpha\mathsf{F}}$ *is nonexpansive for sufficiently small* α.

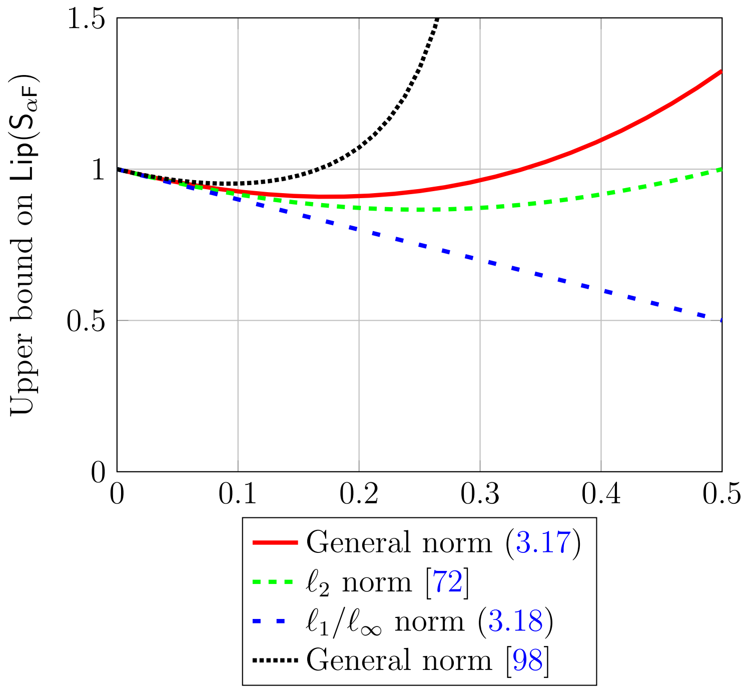


Figure 3.1: Plots of upper bounds of $\mathsf{Lip}(\mathsf{S}_{\alpha\mathsf{F}})$ with respect to different norms. We fix parameters $c = 1, \ell = 2$ and vary the choice of norm. The solid red curve corresponds to the Lipschitz estimate (3.17) for arbitrary norms, the densely dashed green curve corresponds to the estimate $\mathsf{Lip}(\mathsf{S}_{\alpha\mathsf{F}}) \leq \sqrt{1 - 2\alpha c + \alpha^2\ell^2}$ from [72, pp. 16] for the ℓ_2 norm, the loosely dashed blue curve corresponds to the estimate (3.18) for diagonally-weighted ℓ_1/ℓ_∞ norms which is valid on the interval $]0, \frac{1}{\mathrm{diagL}(\mathsf{F})}]$. Finally, the dotted black curve corresponds to the estimate $\mathsf{Lip}(\mathsf{S}_{\alpha\mathsf{F}}) \leq \left(1 + \alpha c - \frac{\alpha^2\ell^2}{1-\alpha\ell}\right)^{-1}$ previously established in [98, Theorem 1]. We see that the estimate (3.17) is a tighter estimate than the estimate from [98] and that Lipschitz upper bounds are least conservative in the case of diagonally-weighted ℓ_1/ℓ_∞ norms.

We plot the upper bounds on the estimates of $\mathsf{Lip}(\mathsf{S}_{\alpha\mathsf{F}})$ as a function of α and choice of norm for fixed parameters c and ℓ in Figure 3.1.

Lemma 3.3.7 (Lipschitz constant of the resolvent operator)**.** *Suppose* $\mathsf{F}:\mathbb{R}^n \to \mathbb{R}^n$ *is continuous and monotone with monotonicity parameter* $c \geq 0$*. Then,*

$$\mathsf{Lip}(\mathsf{J}_{\alpha\mathsf{F}}) \leq \frac{1}{1+\alpha c}, \qquad \textit{for all } \alpha > 0. \tag{3.24}$$

Proof. We observe that $\mathsf{Id} + \alpha\mathsf{F}$ is strongly monotone with parameter $1+\alpha c$. Then by Corollary 3.3.3, the result holds. □

Lemma 3.3.8 (Lipschitz constant of the reflected resolvent)**.** *Suppose* $\mathsf{F}:\mathbb{R}^n \to \mathbb{R}^n$ *is Lipschitz continuous with constant* ℓ *w.r.t. a norm* $\|\cdot\|$*.*

(i) Suppose F *is monotone w.r.t.* $\|\cdot\|$ *with monotonicity parameter* $c \geq 0$*. Then*

$$\mathsf{Lip}(\mathsf{R}_{\alpha\mathsf{F}}) \leq \frac{e^{-\alpha c} + e^{\alpha\ell} - 1 - \alpha\ell}{1+\alpha c}, \qquad \textit{for all } \alpha > 0. \tag{3.25}$$

(ii) Alternatively suppose $\|\cdot\|$ *is a diagonally weighted* ℓ_1 *or* ℓ_∞ *norm. Moreover, suppose* F *is monotone w.r.t.* $\|\cdot\|$ *with monotonicity parameter* $c \geq 0$*. Then,*

$$\mathsf{Lip}(\mathsf{R}_{\alpha\mathsf{F}}) \leq \frac{1-\alpha c}{1+\alpha c} \leq 1, \qquad \textit{for all } \alpha \in \Big(0, \frac{1}{\mathrm{diagL}(\mathsf{F})}\Big]. \tag{3.26}$$

Proof. Recall from [72, pp. 21] that since F is monotone and continuous, we have that $\mathsf{R}_{\alpha\mathsf{F}} = \mathsf{S}_{\alpha\mathsf{F}} \circ \mathsf{J}_{\alpha\mathsf{F}}$. Both results then follow from $\mathsf{Lip}(\mathsf{R}_{\alpha\mathsf{F}}) \leq \mathsf{Lip}(\mathsf{S}_{\alpha\mathsf{F}})\,\mathsf{Lip}(\mathsf{J}_{\alpha\mathsf{F}})$ and the bounds on $\mathsf{Lip}(\mathsf{S}_{\alpha\mathsf{F}})$ from Lemma 3.3.4 and on $\mathsf{Lip}(\mathsf{J}_{\alpha\mathsf{F}})$ from Lemma 3.3.7. □

Lemma 3.3.8 stands in striking contrast with results on monotone operators in Hilbert spaces which says that for any maximally monotone operator,[6] F, the reflected resolvent

[6]Recall that in monotone operator theory on Hilbert spaces, a set-valued mapping $\mathsf{F}\colon \mathbb{R}^n \to 2^{\mathbb{R}^n}$ is maximally monotone if it is monotone and there does not exist another monotone operator, G, whose graph properly contains the graph of F. See [4, Sec. 20.2] for more details.

of F with parameter $\alpha > 0$ is nonexpansive for every $\alpha > 0$. Indeed, in the non-Euclidean case, this property cannot be recovered as is demonstrated in the following example.

Example 3. *Consider the linear mapping* $\mathsf{F}(x) = Ax = \left(\begin{smallmatrix} 2 & -2 \\ 1 & 1 \end{smallmatrix}\right) x$. F *is monotone w.r.t. the* ℓ_∞ *norm since* $-\mu_\infty(-A) = -\mu_\infty \left(\begin{smallmatrix} -2 & 2 \\ -1 & -1 \end{smallmatrix}\right) = 0$. *For* $\alpha = 2$, *we compute*

$$\mathsf{J}_{\alpha\mathsf{F}}(x) = \begin{pmatrix} 3/23 & 4/23 \\ -2/23 & 5/23 \end{pmatrix} x, \qquad \mathsf{R}_{\alpha\mathsf{F}}(x) = \begin{pmatrix} -17/23 & 8/23 \\ -4/23 & -13/23 \end{pmatrix} x.$$

Thus, $\mathsf{Lip}(\mathsf{J}_{\alpha\mathsf{F}}) = 7/23$ *and* $\mathsf{Lip}(\mathsf{R}_{\alpha\mathsf{F}}) = 25/23$. *In other words, for* $\alpha = 2$, $\mathsf{J}_{\alpha\mathsf{F}}$ *is a contraction and* $\mathsf{R}_{\alpha\mathsf{F}}$ *is not nonexpansive.*

Despite this key divergence from the classical theory, we will still be able to prove convergence of iterative algorithms involving the reflected resolvent under suitable assumptions on the parameter $\alpha > 0$.

3.4 Finding zeros of non-Euclidean monotone operators

For a mapping $\mathsf{F} \colon \mathbb{R}^n \to \mathbb{R}^n$ which is continuous and monotone, consider the problem of finding an $x \in \mathbb{R}^n$ that satisfies

$$\mathsf{F}(x) = \mathbb{0}_n. \tag{3.27}$$

Without further assumptions on F, this problem may have no solutions or nonunique solutions. First we provide a preliminary sufficient condition for existence and uniqueness of a solution.

Lemma 3.4.1 (Uniqueness of zeros of strongly monotone maps)**.** *Suppose* $\mathsf{F} \colon \mathbb{R}^n \to \mathbb{R}^n$ *is continuous and strongly monotone. Then* $\mathrm{Zero}(\mathsf{F})$ *is a singleton.*

Proof. We have that $\mathrm{Zero}(\mathsf{F}) = \mathrm{Fix}(\mathsf{J}_{\alpha\mathsf{F}})$ for $\alpha > 0$. By Lemma 3.3.7, we have that $\mathsf{Lip}(\mathsf{J}_{\alpha\mathsf{F}}) \leq 1/(1+\alpha c) < 1$, where $c > 0$ is the monotonicity parameter of F. Then by the Banach fixed point theorem, $\mathsf{J}_{\alpha\mathsf{F}}$ has a unique fixed point and thus $\mathrm{Zero}(\mathsf{F})$ is a singleton. □

Alternatively, if F is continuous and monotone, then we study fixed points of the nonexpansive map $\mathsf{J}_{\alpha\mathsf{F}}$, which may or may not exist and may or may not be unique. In what follows, we will study the case where it is known a priori that zeros of F exist but need not be unique.

We show that the most known algorithms for finding zeros of monotone operators on Hilbert spaces (see, e.g., [72]) can be generalized to non-Euclidean monotone operators using our framework and, furthermore, explicitly estimate the convergence rate of these methods.

3.4.1 The forward step method

Algorithm 1 (Forward step method)**.** *The* forward step method *corresponds to the fixed point iteration*

$$x^{k+1} = \mathsf{S}_{\alpha\mathsf{F}}(x^k) = x^k - \alpha\mathsf{F}(x^k). \tag{3.28}$$

Theorem 3.4.1 (Convergence guarantees for the forward step method)**.** *Let $\mathsf{F}\colon \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous with constant ℓ w.r.t. a norm $\|\cdot\|$ and let $x^0 \in \mathbb{R}^n$ be arbitrary.*

(i) Suppose F is strongly monotone w.r.t. $\|\cdot\|$ with monotonicity parameter $c > 0$. Then the iteration (3.28) converges to the unique zero, x^*, of F for every $\alpha \in (0, \alpha^*)$. Moreover, for every $k \in \mathbb{Z}_{\geq 0}$,

$$\|x^{k+1} - x^*\| \leq (e^{-\alpha c} + e^{\alpha\ell} - 1 - \alpha\ell)\|x^k - x^*\|,$$

where α^* is the unique positive value of α that satisfies $e^{-\alpha^* c} + e^{\alpha^* \ell} = 2 + \alpha^* \ell$.

(ii) *Alternatively suppose $\|\cdot\|$ is a diagonally-weighted ℓ_1 or ℓ_∞ norm and F is strongly monotone w.r.t. $\|\cdot\|$ with monotonicity parameter $c > 0$. Then the iteration (3.28) converges to the unique zero, x^*, of F for every $\alpha \in (0, \frac{1}{\operatorname{diagL}(\mathsf{F})}]$. Moreover, for every $k \in \mathbb{Z}_{\geq 0}$,*

$$\|x^{k+1} - x^*\| \leq (1 - \alpha c)\|x^k - x^*\|,$$

with the convergence rate optimized at $\alpha = 1/\operatorname{diagL}(\mathsf{F})$.

(iii) *Alternatively suppose $\|\cdot\|$ is a diagonally weighted ℓ_1 or ℓ_∞ norm and F is monotone w.r.t. $\|\cdot\|$. Then $\operatorname{Zero}(\mathsf{F}) \neq \emptyset$ implies the iteration (3.28) converges to an element of $\operatorname{Zero}(\mathsf{F})$ for every $\alpha \in \,]0, \frac{1}{\operatorname{diagL}(\mathsf{F})}[$.*

Proof. Regarding statement (i), from Lemma 3.3.4(i), we have that $\mathsf{Lip}(\mathsf{S}_{\alpha\mathsf{F}}) \leq e^{-\alpha c} + e^{\alpha \ell} - 1 - \alpha \ell$. It is straightforward to compute that at $\alpha = \alpha^*$, $\mathsf{Lip}(\mathsf{S}_{\alpha\mathsf{F}}) \leq 1$ and for $\alpha \in (0, \alpha^*)$ we have that $\mathsf{Lip}(\mathsf{S}_{\alpha\mathsf{F}}) < 1$. Thus, $\mathsf{S}_{\alpha\mathsf{F}}$ is a contraction and fixed points of $\mathsf{S}_{\alpha\mathsf{F}}$ correspond to zeros of F. Then by the Banach fixed point theorem, the result follows.

Regarding statement (ii), Lemma 3.3.4(ii) implies that $\mathsf{Lip}(\mathsf{S}_{\alpha\mathsf{F}}) = 1 - \alpha c < 1$ for all $\alpha \in (0, 1/\operatorname{diagL}(\mathsf{F}))$. The result is then a consequence of the Banach fixed point theorem.

Regarding statement (iii), since F is monotone w.r.t. a diagonally weighted ℓ_1 or ℓ_∞ norm, $\mathsf{S}_{\alpha\mathsf{F}}$ is nonexpansive for $\alpha \in (0, 1/\operatorname{diagL}(\mathsf{F})]$ by Lemma 3.3.4(ii). Moreover, for every $\alpha \in (0, 1/\operatorname{diagL}(\mathsf{F})[$, there exists $\theta \in (0,1)$ such that $\mathsf{S}_{\alpha\mathsf{F}} = (1-\theta)\mathsf{Id} + \theta \mathsf{S}_{\tilde{\alpha}\mathsf{F}}$, for some $\tilde{\alpha} \in (0, 1/\operatorname{diagL}(\mathsf{F})]$. Therefore the iteration (3.28) is the Krasnosel'skii–Mann iteration of the nonexpansive operator $\mathsf{S}_{\tilde{\alpha}\mathsf{F}}$ and Lemma 3.2.2 implies the result. □

Note that Theorem 3.4.1(iii) is a direct consequence of the fact that the forward step operator is nonexpansive for suitable $\alpha > 0$ when the mapping is monotone w.r.t. a diagonally-weighted ℓ_1 or ℓ_∞ norm, a fact which need not hold when the mapping is monotone w.r.t. a different norm, e.g., a Hilbert one. See the relevant discussion in Remark 3.3.6 for an example of a mapping, F, which is monotone with respect to the ℓ_2 norm but $\mathsf{S}_{\alpha\mathsf{F}}$ is not nonexpansive for any $\alpha > 0$.

3.4.2 The proximal point method

Algorithm 2 (Proximal point method)**.** *The* proximal point method *corresponds to the fixed point iteration*

$$x^{k+1} = \mathsf{J}_{\alpha\mathsf{F}}(x^k) = (\mathsf{Id} + \alpha\mathsf{F})^{-1}(x^k). \tag{3.29}$$

Theorem 3.4.2 (Convergence guarantees for the proximal point method)**.** *Suppose* $\mathsf{F}\colon \mathbb{R}^n \to \mathbb{R}^n$ *is continuous and let* $x^0 \in \mathbb{R}^n$ *be arbitrary.*

(i) Suppose F *is strongly monotone w.r.t. a norm* $\|\cdot\|$ *with monotonicity parameter* $c > 0$*. Then the iteration* (3.29) *converges to the unique zero,* x^**, of* F *for every* $\alpha \in (0,\infty)$*. Moreover, for every* $k \in \mathbb{Z}_{\geq 0}$*, the iteration satisfies*

$$\|x^{k+1} - x^*\| \leq \frac{1}{1+\alpha c}\|x^k - x^*\|.$$

(ii) Alternatively suppose F *is monotone and globally Lipschitz continuous w.r.t. a diagonally weighted* ℓ_1 *or* ℓ_∞ *norm* $\|\cdot\|$ *and* $\mathrm{diagL}(\mathsf{F}) \neq 0$*. Then if* $\mathrm{Zero}(\mathsf{F}) \neq \emptyset$*, the iteration* (3.29) *converges to an element of* $\mathrm{Zero}(\mathsf{F})$ *for every* $\alpha \in (0,\infty)$*.*

Proof. Regarding statement (i), Lemma 3.3.7 provides the Lipschitz estimate $\mathsf{Lip}(\mathsf{J}_{\alpha\mathsf{F}}) \leq \frac{1}{1+\alpha c} < 1$ for all $\alpha > 0$. Thus $\mathsf{J}_{\alpha\mathsf{F}}$ is a contraction and since fixed points of $\mathsf{J}_{\alpha\mathsf{F}}$ correspond to zeros of F, the Banach fixed point theorem implies the result.

Regarding statement (ii), we will demonstrate that the iteration (3.29) is a Krasnosel'skii–Mann iteration of a suitably-defined nonexpansive mapping. To this end, let

$\theta \in (0,1)$ be arbitrary and consider the auxiliary mapping $\overline{\mathsf{R}}^{\theta}_{\alpha\mathsf{F}} \colon \mathbb{R}^n \to \mathbb{R}^n$ given by $\overline{\mathsf{R}}^{\theta}_{\alpha\mathsf{F}} := \frac{\mathsf{J}_{\alpha\mathsf{F}}}{\theta} - \frac{1-\theta}{\theta}\mathsf{Id}$. Then it is straightforward to compute

$$\begin{aligned}\overline{\mathsf{R}}^{\theta}_{\alpha\mathsf{F}} &= \frac{(\mathsf{Id}+\alpha\mathsf{F})^{-1}}{\theta} - \frac{1-\theta}{\theta}(\mathsf{Id}+\alpha\mathsf{F})\circ(\mathsf{Id}+\alpha\mathsf{F})^{-1} \\ &= \Big(\frac{\mathsf{Id}}{\theta} - \frac{1-\theta}{\theta}(\mathsf{Id}+\alpha\mathsf{F})\Big)\circ(\mathsf{Id}+\alpha\mathsf{F})^{-1} = \Big(\mathsf{Id} - \frac{(1-\theta)\alpha}{\theta}\mathsf{F}\Big)\circ \mathsf{J}_{\alpha\mathsf{F}} = \mathsf{S}_{\frac{1-\theta}{\theta}\alpha\mathsf{F}}\circ\mathsf{J}_{\alpha\mathsf{F}}.\end{aligned}$$

Moreover, $\mathsf{J}_{\alpha\mathsf{F}}$ is nonexpansive by Lemma 3.3.7, and by Lemma 3.3.4(ii), $\mathsf{Lip}(\mathsf{S}_{\frac{1-\theta}{\theta}\alpha\mathsf{F}}) \leq 1$, for all $\alpha \in (0, \frac{1-\theta}{\theta\,\mathrm{diagL}(\mathsf{F})}]$. We conclude that $\mathsf{Lip}(\overline{\mathsf{R}}^{\theta}_{\alpha\mathsf{F}}) \leq \mathsf{Lip}(\mathsf{S}_{\frac{1-\theta}{\theta}\alpha\mathsf{F}})\,\mathsf{Lip}(\mathsf{J}_{\alpha\mathsf{F}}) \leq 1$ for $\alpha \in (0, \frac{1-\theta}{\theta\,\mathrm{diagL}(\mathsf{F})}]$ which implies that $\overline{\mathsf{R}}^{\theta}_{\mathsf{F}}$ is nonexpansive for all α in this range.

Let $\alpha > 0$ be arbitrary. Then for any[7] $\theta \leq \frac{1}{1+\alpha\,\mathrm{diagL}(\mathsf{F})} \in (0,1)$, we have that $\mathsf{J}_{\alpha\mathsf{F}} = (1-\theta)\mathsf{Id} + \theta\overline{\mathsf{R}}^{\theta}_{\alpha\mathsf{F}}$, and $\overline{\mathsf{R}}^{\theta}_{\mathsf{F}}$ is nonexpansive since $\alpha \in (0, \frac{1-\theta}{\theta\,\mathrm{diagL}(\mathsf{F})}]$. Thus, the iteration (3.29) is the Krasnosel'skii–Mann iteration for $\overline{\mathsf{R}}^{\theta}_{\alpha\mathsf{F}}$ and the result follows from Lemma 3.2.2. □

Remark 3.4.2. *Theorem 3.4.2(ii) is an analog of the classical result in monotone operator theory on Hilbert spaces which states that the resolvent of a maximally monotone operator is firmly nonexpansive [77] and [4, Prop. 23.8]. This firm nonexpansiveness is a consequence of the fact that the reflected resolvent of a maximally monotone operator with respect to a Euclidean norm is always nonexpansive and* $\mathsf{J}_{\alpha\mathsf{F}} = \frac{1}{2}\mathsf{Id} + \frac{1}{2}\mathsf{R}_{\alpha\mathsf{F}}$. *Note that this property need not hold when* F *is monotone with respect to a non-Euclidean norm but we are able to show that in the case of diagonally-weighted* ℓ_1/ℓ_∞ *norms, a similar result holds.*

[7]Note that $\frac{1}{1+\alpha\,\mathrm{diagL}(\mathsf{F})} \in (0,1)$ holds under the assumption $\mathrm{diagL}(\mathsf{F}) \neq 0$ since $\mathrm{diagL}(\mathsf{F}) \geq 0$ for any monotone F.

3.4.3 The Cayley method

Algorithm 3 (Cayley method)**.** *The* Cayley method *corresponds to the iteration*

$$x^{k+1} = \mathsf{R}_{\alpha\mathsf{F}}(x^k) = 2(\mathsf{Id} + \alpha\mathsf{F})^{-1}(x^k) - x^k. \tag{3.30}$$

Theorem 3.4.3 (Convergence guarantees for the Cayley method)**.** *Suppose* $\mathsf{F}\colon \mathbb{R}^n \to \mathbb{R}^n$ *is Lipschitz continuous with constant* ℓ *w.r.t. a norm* $\|\cdot\|$ *and let* $x^0 \in \mathbb{R}^n$ *be arbitrary.*

(i) Suppose F *is strongly monotone w.r.t.* $\|\cdot\|$ *with monotonicity parameter* $c > 0$*. Then the iteration* (3.30) *converges to the unique zero,* x^**, of* F *for sufficiently small* $\alpha > 0$*. Moreover, for every* $k \in \mathbb{Z}_{\geq 0}$*, the iteration satisfies*

$$\|x^{k+1} - x^*\| \leq \frac{e^{-\alpha c} + e^{\alpha\ell} - 1 - \alpha\ell}{1 + \alpha c}\|x^k - x^*\|.$$

(ii) Alternatively suppose $\|\cdot\|$ *is a diagonally weighted* ℓ_1 *or* ℓ_∞ *norm and* F *is strongly monotone w.r.t.* $\|\cdot\|$ *with monotonicity parameter* $c > 0$*. Then the iteration* (3.30) *converges to the unique zero,* x^**, of* F *for every* $\alpha \in (0, \frac{1}{\operatorname{diagL}(\mathsf{F})}]$*. Moreover, for every* $k \in \mathbb{Z}_{\geq 0}$*, the iteration satisfies*

$$\|x^{k+1} - x^*\| \leq \frac{1 - \alpha c}{1 + \alpha c}\|x^k - x^*\|,$$

with the convergence rate being optimized at $\alpha = 1/\operatorname{diagL}(\mathsf{F})$*.*

(iii) Alternatively suppose $\|\cdot\|$ *is a diagonally weighted* ℓ_1 *or* ℓ_∞ *norm and* F *is monotone w.r.t.* $\|\cdot\|$*. Then if* $\mathrm{Zero}(\mathsf{F}) \neq \emptyset$*, the Krasnosel'skii–Mann iteration with* $\theta = 1/2$

$$x^{k+1} = \frac{1}{2}x^k + \frac{1}{2}\mathsf{R}_{\alpha\mathsf{F}}(x^k)$$

<table>
<tr><td rowspan="3">Algorithm, Iterated map</td><td colspan="6">F strongly monotone and globally Lipschitz continuous</td></tr>
<tr><td colspan="2">ℓ_2</td><td colspan="2">General norm</td><td colspan="2">Diagonally weighted ℓ_1 or ℓ_∞</td></tr>
<tr><td>α range</td><td>Optimal Lip</td><td>α range</td><td>Optimal Lip</td><td>α range</td><td>Optimal Lip</td></tr>
<tr><td>Forward step, $\mathsf{S}_{\alpha\mathsf{F}}$</td><td>$\left]0, \frac{2c}{\ell^2}\right[$</td><td>$1 - \frac{1}{2\kappa^2} + \mathcal{O}\left(\frac{1}{\kappa^3}\right)$</td><td>$]0, \alpha^*[$</td><td>$1 - \frac{1}{2\kappa^2} + \mathcal{O}\left(\frac{1}{\kappa^3}\right)$</td><td>$\left]0, \frac{1}{\mathrm{diagL}(\mathsf{F})}\right]$</td><td>$1 - \frac{1}{\kappa_\infty}$</td></tr>
<tr><td>Proximal point, $\mathsf{J}_{\alpha\mathsf{F}}$</td><td>$]0, \infty[$</td><td>A.S.</td><td>$]0, \infty[$</td><td>A.S.</td><td>$]0, \infty[$</td><td>A.S.</td></tr>
<tr><td>Cayley, $\mathsf{R}_{\alpha\mathsf{F}}$</td><td>$]0, \infty[$</td><td>$1 - \frac{1}{2\kappa} + \mathcal{O}\left(\frac{1}{\kappa^2}\right)$</td><td>$]0, \alpha^*[$</td><td>$1 - \frac{2}{\kappa^2} + \mathcal{O}\left(\frac{1}{\kappa^3}\right)$</td><td>$\left]0, \frac{1}{\mathrm{diagL}(\mathsf{F})}\right]$</td><td>$1 - \frac{2}{\kappa_\infty} + \mathcal{O}\left(\frac{1}{\kappa_\infty^2}\right)$</td></tr>
</table>

Table 3.1: Table of step size ranges and Lipschitz constants for three algorithms for finding zeros of monotone operators with respect to arbitrary norms. For F strongly monotone and Lipschitz continuous, let c be its monotonicity parameter (with respect to the appropriate norm), ℓ its appropriate Lipschitz constant, and $\mathrm{diagL}(\mathsf{F}) := \sup_{x \in \mathbb{R}^n \setminus \Omega_\mathsf{F}} \max_{i \in \{1,\dots,n\}} (D\mathsf{F}(x))_{ii} \leq \ell$. Additionally, $\kappa := \ell/c \geq 1$, $\kappa_\infty := \mathrm{diagL}(\mathsf{F})/c \in [1, \kappa]$, and α^* is the unique positive solution to $e^{-\alpha^* c} + e^{\alpha^* \ell} = 2 + \alpha^* \ell$. *A.S.* means the Lipschitz constant can be made arbitrarily small. We do not assume that the strongly monotone F is the gradient of a strongly convex function.

correspond to the proximal point iteration (3.29)*, which is guaranteed to converge to an element of* $\mathrm{Zero}(\mathsf{F})$ *for every* $\alpha \in (0, \infty)$.

Proof. Regarding statement (i), from Lemma 3.3.8(i), we have that $\mathsf{Lip}(\mathsf{R}_{\alpha\mathsf{F}}) \leq (e^{-\alpha c} + e^{\alpha \ell} - 1 - \alpha \ell)/(1 + \alpha c)$ which is less than unity for small enough $\alpha > 0$. Thus, for small enough α, $\mathsf{R}_{\alpha\mathsf{F}}$ is a contraction and fixed points of $\mathsf{R}_{\alpha\mathsf{F}}$ correspond to zeros of F. Thus, by the Banach fixed point theorem, the result follows.

Regarding statement (ii), Lemma 3.3.8(ii) implies that $\mathsf{Lip}(\mathsf{R}_{\alpha\mathsf{F}}) \leq (1 - \alpha c)/(1 + \alpha c) < 1$ for $\alpha \in (0, 1/\mathrm{diagL}(\mathsf{F})]$. The result is then a consequence of the Banach fixed point theorem.

Statement (iii) holds since $\frac{1}{2}\mathsf{Id} + \frac{1}{2}(2\mathsf{J}_{\alpha\mathsf{F}} - \mathsf{Id}) = \mathsf{J}_{\alpha\mathsf{F}}$, and convergence follows by Theorem 3.4.2(ii) since $\mathrm{Zero}(\mathsf{F}) \neq \emptyset$. □

Table 3.1 summarizes and compares the range of step sizes and Lipschitz estimates as provided by the classical monotone operator theory for the ℓ_2 norm [72, pp. 16 and 20] and by Theorems 3.4.1, 3.4.2, and 3.4.3 for general and diagonally-weighted ℓ_1/ℓ_∞ norms.

3.5 Finding zeros of a sum of non-Euclidean monotone operators

In many instances, one may wish to execute the proximal point method, Algorithm 2, to compute a zero of a continuous monotone mapping $\mathsf{N}\colon \mathbb{R}^n \to \mathbb{R}^n$. However, the implementation of the iteration (3.29) may be hindered by the difficulty in evaluating $\mathsf{J}_{\alpha\mathsf{N}}$. To remedy this issue, it is often assumed that N can be expressed as the sum of two monotone mappings F and G where $\mathsf{J}_{\alpha\mathsf{G}}$ may be easy to compute and F satisfies some regularity condition. Alternatively, in some situations, decomposing $\mathsf{N} = \mathsf{F} + \mathsf{G}$ and finding $x \in \mathbb{R}^n$ such that $(\mathsf{F} + \mathsf{G})(x) = \mathbb{0}_n$ provides additional flexibility in choice of algorithm and may improve convergence rates.

Motivated by the above, we consider the problem of finding an $x \in \mathbb{R}^n$ such that

$$(\mathsf{F} + \mathsf{G})(x) = \mathbb{0}_n, \tag{3.31}$$

where $\mathsf{F}, \mathsf{G}\colon \mathbb{R}^n \to \mathbb{R}^n$ are continuous and monotone w.r.t. a diagonally weighted ℓ_1 or ℓ_∞ norm.[8] In particular, we focus on the forward-backward, Peaceman-Rachford, and Douglas-Rachford splitting algorithms. For some extensions of the theory to set-valued mappings, we refer to Section 3.6.1.

3.5.1 Forward-backward splitting

Algorithm 4 (Forward-backward splitting)**.** *Assume $\alpha > 0$. Then in [72, Section 7.1] it is shown that*

$$(\mathsf{F} + \mathsf{G})(x) = \mathbb{0}_n \quad \iff x = (\mathsf{J}_{\alpha\mathsf{G}} \circ \mathsf{S}_{\alpha\mathsf{F}})(x).$$

[8] The results that follow can also be generalized to arbitrary norms using the Lipschitz estimates derived for $\mathsf{J}_{\alpha\mathsf{F}}, \mathsf{R}_{\alpha\mathsf{F}}$, and $\mathsf{S}_{\alpha\mathsf{F}}$ in Section 3.3.2.

The forward-backward splitting method *corresponds to the fixed point iteration*

$$x^{k+1} = \mathsf{J}_{\alpha\mathsf{G}}(x^k - \alpha\mathsf{F}(x^k)). \tag{3.32}$$

If both F *and* G *are monotone, define the* averaged forward-backward splitting iteration

$$x^{k+1} = \frac{1}{2}x^k + \frac{1}{2}\mathsf{J}_{\alpha\mathsf{G}}(x^k - \alpha\mathsf{F}(x^k)). \tag{3.33}$$

Theorem 3.5.1 (Convergence guarantees for forward-backward splitting)**.** *Let* $\mathsf{F}\colon \mathbb{R}^n \to \mathbb{R}^n$ *be Lipschitz continuous w.r.t. a diagonally weighted* ℓ_1 *or* ℓ_∞ *norm* $\|\cdot\|$, $\mathsf{G}\colon \mathbb{R}^n \to \mathbb{R}^n$ *be continuous and monotone w.r.t. the same norm, and let* $x^0 \in \mathbb{R}^n$ *be arbitrary.*

(i) Suppose F *is strongly monotone w.r.t.* $\|\cdot\|$ *with monotonicity parameter* $c > 0$, *then the iteration* (3.32) *converges to the unique zero,* x^*, *of* $\mathsf{F}+\mathsf{G}$ *for every* $\alpha \in (0, \frac{1}{\operatorname{diagL}(\mathsf{F})}]$. *Moreover, for every* $k \in \mathbb{Z}_{\geq 0}$, *the iteration satisfies*

$$\|x^{k+1} - x^*\| \leq (1 - \alpha c)\|x^k - x^*\|,$$

with the convergence rate being optimized at $\alpha = 1/\operatorname{diagL}(\mathsf{F})$.

(ii) If F *is monotone w.r.t.* $\|\cdot\|$ *and* $\operatorname{Zero}(\mathsf{F}+\mathsf{G}) \neq \emptyset$, *then the iteration* (3.33) *converges to an element of* $\operatorname{Zero}(\mathsf{F}+\mathsf{G})$ *for every* $\alpha \in (0, \frac{1}{\operatorname{diagL}(\mathsf{F})}]$.

Proof. Regarding statement (i), Lemmas 3.3.4(ii) and 3.3.7 together imply that $\mathsf{Lip}(\mathsf{J}_{\alpha\mathsf{G}} \circ \mathsf{S}_{\alpha\mathsf{F}}) \leq \mathsf{Lip}(\mathsf{J}_{\alpha\mathsf{G}})\,\mathsf{Lip}(\mathsf{S}_{\alpha\mathsf{F}}) \leq 1 - \alpha c < 1$ for all $\alpha \in]0, 1/\operatorname{diagL}(\mathsf{F})]$. Then since $\operatorname{Fix}(\mathsf{J}_{\alpha\mathsf{F}} \circ \mathsf{S}_{\alpha\mathsf{F}}) = \operatorname{Zero}(\mathsf{F}+\mathsf{G})$, the result is then a consequence of the Banach fixed point theorem.

Statement (ii) follows from $\mathsf{Lip}(\mathsf{J}_{\alpha\mathsf{G}} \circ \mathsf{S}_{\alpha\mathsf{F}}) \leq 1$ and that the iteration (3.33) is the Krasnosel'skii–Mann iteration of the nonexpansive mapping $\mathsf{J}_{\alpha\mathsf{G}} \circ \mathsf{S}_{\alpha\mathsf{F}}$ with $\theta = 1/2$. $\square$

Remark 3.5.1 (Comparison with convergence criteria)**.** *The following comparisons are in order.*

- *Compared to the Hilbert case, in the non-Euclidean setting, if both* F *and* G *are monotone, then iteration* (3.33) *must be applied to compute a zero of* $\mathsf{F}+\mathsf{G}$*. In the Hilbert case, iteration* (3.32) *may be used instead since the composition of averaged operators is averaged. For non-Hilbert norms, the composition of two averaged operators need not be averaged.*

- *In monotone operator theory on Hilbert spaces, Lipschitz continuity of* F *is not sufficient for the convergence of the iteration* (3.32)*. Instead, a standard sufficient condition for convergence is cocoercivity of* F*, see [104] and [4, Theorem 26.14]. In the case of diagonally-weighted* ℓ_1/ℓ_∞ *norms, Lipschitz continuity is sufficient for convergence. This fact is due to the nonexpansiveness of* $\mathsf{S}_{\alpha\mathsf{F}}$ *for* ℓ_1/ℓ_∞ *monotone* F *and small enough* $\alpha > 0$ *as discussed in Remark 3.3.6.*

3.5.2 Peaceman-Rachford and Douglas-Rachford splitting

Algorithm 5 (Peaceman-Rachford and Douglas-Rachford splitting)**.** *Let* $\alpha > 0$*. Then in [72, Section 7.3], it is shown that*

$$(\mathsf{F}+\mathsf{G})(x) = \mathbb{0}_n \quad \iff \quad (\mathsf{R}_{\alpha\mathsf{F}} \circ \mathsf{R}_{\alpha\mathsf{G}})(z) = z \textit{ and } x = \mathsf{J}_{\alpha\mathsf{G}}(z). \tag{3.34}$$

The Peaceman-Rachford splitting method *corresponds to the fixed point iteration*

$$\begin{aligned} x^{k+1} &= \mathsf{J}_{\alpha\mathsf{G}}(z^k), \\ z^{k+1} &= z^k + 2\mathsf{J}_{\alpha\mathsf{F}}(2x^{k+1} - z^k) - 2x^{k+1}. \end{aligned} \tag{3.35}$$

If both F *and* G *are monotone, the term* $\mathsf{R}_{\alpha\mathsf{F}} \circ \mathsf{R}_{\alpha\mathsf{G}}$ *in* (3.34) *is averaged to yield*

$$(\mathsf{F}+\mathsf{G})(x) = \mathbb{0}_n \quad \iff \quad \Big(\frac{1}{2}\mathsf{Id} + \frac{1}{2}\mathsf{R}_{\alpha\mathsf{F}} \circ \mathsf{R}_{\alpha\mathsf{G}}\Big)(z) = z \text{ and } x = \mathsf{J}_{\alpha\mathsf{G}}(z). \tag{3.36}$$

The fixed point iteration corresponding to (3.36) *is called the* Douglas-Rachford splitting method *and is given by*

$$\begin{aligned} x^{k+1} &= \mathsf{J}_{\alpha\mathsf{G}}(z^k), \\ z^{k+1} &= z^k + \mathsf{J}_{\alpha\mathsf{F}}(2x^{k+1} - z^k) - x^{k+1}. \end{aligned} \tag{3.37}$$

Theorem 3.5.2 (Convergence guarantees for Peaceman-Rachford and Douglas-Rachford splitting)**.** *Let both* $\mathsf{F}\colon \mathbb{R}^n \to \mathbb{R}^n$ *and* $\mathsf{G}\colon \mathbb{R}^n \to \mathbb{R}^n$ *be Lipschitz continuous w.r.t. a diagonally weighted* ℓ_1 *or* ℓ_∞ *norm* $\|\cdot\|$*, let* G *be monotone w.r.t. the same norm, and let* $x^0 \in \mathbb{R}^n$*.*

(i) Suppose F *is strongly monotone w.r.t.* $\|\cdot\|$ *with monotonicity parameter* $c > 0$*. Then the sequence of* $\{x_k\}_{k=0}^\infty$ *generated by the iteration* (3.35) *converges to the unique zero,* x^**, of* $\mathsf{F}+\mathsf{G}$ *for every* $\alpha \in \Big(0, \min\Big\{\frac{1}{\mathrm{diagL}(\mathsf{F})}, \frac{1}{\mathrm{diagL}(\mathsf{G})}\Big\}\Big]$*. Moreover, for every* $k \in \mathbb{Z}_{\geq 0}$*, the iteration satisfies*

$$\|x^{k+1} - x^*\| \leq \frac{1-\alpha c}{1+\alpha c}\|x^k - x^*\|,$$

with the convergence rate being optimized at $\alpha = \min\Big\{\frac{1}{\mathrm{diagL}(\mathsf{F})}, \frac{1}{\mathrm{diagL}(\mathsf{G})}\Big\}$*.*

(ii) Alternatively suppose F *is monotone w.r.t.* $\|\cdot\|$ *and* $\mathrm{Zero}(\mathsf{F}+\mathsf{G}) \neq \emptyset$*. Then the sequence of* $\{x_k\}_{k=0}^\infty$ *generated by the iteration* (3.37) *converges to an element of* $\mathrm{Zero}(\mathsf{F}+\mathsf{G})$ *for every* $\alpha \in \Big(0, \min\Big\{\frac{1}{\mathrm{diagL}(\mathsf{F})}, \frac{1}{\mathrm{diagL}(\mathsf{G})}\Big\}\Big]$*.*

Proof. Regarding statement (i), by Lemma 3.3.8(ii), we have that

$$\mathsf{Lip}(\mathsf{R}_{\alpha\mathsf{F}} \circ \mathsf{R}_{\alpha\mathsf{G}}) \leq \mathsf{Lip}(\mathsf{R}_{\alpha\mathsf{F}})\,\mathsf{Lip}(\mathsf{R}_{\alpha\mathsf{G}}) \leq \frac{1-\alpha c}{1+\alpha c} < 1$$

for $\alpha \in (0, \min\{1/\operatorname{diagL}(\mathsf{F}), 1/\operatorname{diagL}(\mathsf{G})\}]$. Then since $\mathsf{Lip}(\mathsf{J}_{\alpha\mathsf{G}})$ is nonexpansive, the Banach fixed point theorem implies the result.

Statement (ii) holds because Lemma 3.3.8(ii) implies $\mathsf{Lip}(\mathsf{R}_{\alpha\mathsf{F}} \circ \mathsf{R}_{\alpha\mathsf{G}}) \leq 1$. Then the iteration (3.37) converges because of Lemma 3.2.2. $\square$

Compared to classical criteria for the convergence of the Douglas-Rachford iteration, Theorem 3.5.2 requires Lipschitz continuity of F and G in order to utilize the Lipschitz estimates for the reflected resolvents $\mathsf{R}_{\alpha\mathsf{F}}$ and $\mathsf{R}_{\alpha\mathsf{G}}$. Moreover, the parameter $\alpha > 0$ must be chosen small enough in the non-Euclidean setting whereas convergence is guaranteed for any choice of α in the Hilbert case. This is because the reflected resolvent is only nonexpansive for a certain range of α when the norm is not a Hilbert one, see Example 3.

3.6 Set-valued inclusions and an application to recurrent neural networks

3.6.1 Set-valued inclusions and non-Euclidean properties of proximal operators

In many instances one may wish to solve an inclusion problem of the form

$$\text{Find } x \in \mathbb{R}^n \text{ such that } \quad \mathbb{0}_n \in (\mathsf{F} + \mathsf{G})(x), \tag{3.38}$$

where $\mathsf{F}\colon \mathbb{R}^n \to \mathbb{R}^n$ is a single-valued continuous monotone mapping but $\mathsf{G}\colon \mathbb{R}^n \to 2^{\mathbb{R}^n}$ is a set-valued mapping. In monotone operator theory on Hilbert spaces, leveraging the fact that $\mathsf{J}_{\alpha\mathsf{G}}$ is single-valued and nonexpansive for every $\alpha > 0$ when G is maximally monotone, algorithms such as the forward-backward splitting and Douglas-Rachford splitting may be used to solve (6.4) under suitable assumptions on F.

In this section we aim to prove similar results in the non-Euclidean case. We will specialize to the case that G is the subdifferential of a *separable, proper lower semicontinuous (l.s.c.), convex function*. To start we must recall the proximal operator of a l.s.c. convex function.

Definition 12 (Proximal operator, [105] and [4, Def. 12.23]). *Let $g\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper l.s.c. convex function. The* proximal operator *of g evaluated at $x \in \mathbb{R}^n$ is the map $\mathrm{prox}_g\colon \mathbb{R}^n \to \mathbb{R}^n$ defined by*

$$\mathrm{prox}_g(x) = \operatorname*{argmin}_{z\in\mathbb{R}^n} \frac{1}{2}\|x - z\|_2^2 + g(z). \tag{3.39}$$

Since $g\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ is proper, l.s.c., and convex, we can see that for $\alpha > 0$ and fixed $x \in \mathbb{R}^n$, the map $z \mapsto \frac{1}{2}\|x - z\|_2^2 + \alpha g(z)$ is strongly convex and thus has a unique minimizer, so for each $x \in \mathbb{R}^n$, $\mathrm{prox}_{\alpha g}(x)$ is single-valued. Moreover, we have the following connection between proximal operators and resolvents of subdifferentials.

Proposition 3.6.1 ([106] and [4, Example 23.3]). *Suppose $g\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ is proper, l.s.c., and convex. Then for every $\alpha > 0$, $\mathsf{J}_{\alpha\partial g}(x) = \mathrm{prox}_{\alpha g}(x)$.*

In the case of scalar functions, one can exactly capture the set of functions which are proximal operators of some proper l.s.c. convex functions.

Proposition 3.6.2 ([4, Proposition 24.31]). *Let $\phi\colon \mathbb{R} \to \mathbb{R}$. Then ϕ is the proximal operator of a proper l.s.c. convex function $g\colon \mathbb{R} \to \overline{\mathbb{R}}$, i.e., $\phi = \mathrm{prox}_g$ if and only if ϕ*

satisfies

$$0 \leq \frac{\phi(x) - \phi(y)}{x - y} \leq 1, \quad \text{for all } x, y \in \mathbb{R}, x \neq y. \tag{3.40}$$

A list of examples of scalar functions satisfying (3.40) and their corresponding proper l.s.c. convex function is provided in [107, Table 1].

To prove non-Euclidean properties of proximal operators, we will leverage a well-known property, which we highlight in the following proposition.

Proposition 3.6.3 (Proximal operator of separable convex functions, [108, Section 2.1])**.** *For $i \in \{1, \dots, n\}$, let $g_i \colon \mathbb{R} \to \overline{\mathbb{R}}$, be proper, l.s.c., and convex. Define $g \colon \mathbb{R}^n \to \overline{\mathbb{R}}$ by $g(x) = \sum_{i=1}^n g_i(x_i)$. Then g is proper, l.s.c., and convex and for all $\alpha > 0$,*

$$\mathrm{prox}_{\alpha g}(x) = (\mathrm{prox}_{\alpha g_1}(x_1), \dots, \mathrm{prox}_{\alpha g_n}(x_n)) \in \mathbb{R}^n.$$

If g satisfies $g(x) = \sum_{i=1}^n g_i(x_i)$ with each g_i proper, l.s.c., and convex, we call g separable.

In the following novel proposition, we showcase that when g is separable, $\mathrm{prox}_{\alpha g}$ and $2\mathrm{prox}_{\alpha g} - \mathsf{Id}$ are nonexpansive w.r.t. non-Euclidean norms.

Proposition 3.6.4 (Nonexpansiveness of proximal operators of separable convex maps)**.** *For $i \in \{1, \dots, n\}$, let each $g_i \colon \mathbb{R} \to \overline{\mathbb{R}}$ be proper, l.s.c., and convex. Define $g \colon \mathbb{R}^n \to \overline{\mathbb{R}}$ by $g(x) = \sum_{i=1}^n g_i(x_i)$. For every $\alpha > 0$ and for any $\eta \in \mathbb{R}^n_{>0}$, both $\mathsf{J}_{\alpha\partial g} = \mathrm{prox}_{\alpha g}$ and $\mathsf{R}_{\alpha\partial g} = 2\mathrm{prox}_{\alpha g} - \mathsf{Id}$ are nonexpansive w.r.t. $\|\cdot\|_{\infty,[\eta]^{-1}}$.*[9]

Proof. By Proposition 3.6.3 we have $\mathrm{prox}_{\alpha g}(x) = (\mathrm{prox}_{\alpha g_1}(x_1), \dots, \mathrm{prox}_{\alpha g_n}(x_n))$. Moreover, each $\mathrm{prox}_{\alpha g_i}$ is nonexpansive and monotone by Proposition 3.6.2 and thus satisfies

$$0 \leq (\mathrm{prox}_{\alpha g_i}(x_i) - \mathrm{prox}_{\alpha g_i}(y_i))(x_i - y_i) \leq (x_i - y_i)^2, \qquad \text{for all } x_i, y_i \in \mathbb{R}. \tag{3.41}$$

[9]More generally, $\mathrm{prox}_{\alpha g}$ and $2\mathrm{prox}_{\alpha g} - \mathsf{Id}$ are nonexpansive with respect to any monotonic norm.

We then conclude

$$\begin{aligned}\|\mathrm{prox}_{\alpha g}(x) - \mathrm{prox}_{\alpha g}(y)\|_{\infty,[\eta]^{-1}} &= \max_{i\in\{1,\dots,n\}} \frac{1}{\eta_i}|\mathrm{prox}_{\alpha g_i}(x_i) - \mathrm{prox}_{\alpha g_i}(y_i)| \\ &\le \max_{i\in\{1,\dots,n\}} \frac{1}{\eta_i}|x_i - y_i| = \|x-y\|_{\infty,[\eta]^{-1}}.\end{aligned}$$

Regarding $\mathsf{R}_{\alpha\partial g}$, we note that (3.41) implies for all $x_i, y_i \in \mathbb{R}$

$$\begin{aligned}-(x_i-y_i)^2 \le ((2\mathrm{prox}_{\alpha g_i}(x_i) - x_i) - (2\mathrm{prox}_{\alpha g_i}(y_i) - y_i))(x_i - y_i) \le (x_i-y_i)^2, \\ \implies |(2\mathrm{prox}_{\alpha g_i}(x_i) - x_i) - (2\mathrm{prox}_{\alpha g_i}(y_i) - y_i)| \le |x_i - y_i|.\end{aligned}$$

Following the same reasoning as for $\mathrm{prox}_{\alpha g}$, we conclude that $2\mathrm{prox}_{\alpha g} - \mathsf{Id}$ is nonexpansive w.r.t. $\|\cdot\|_{\infty,[\eta]^{-1}}$. □

We recall from monotone operator theory on Hilbert spaces that if $\mathsf{F}\colon \mathbb{R}^n \to \mathbb{R}^n$ is continuous and $\mathsf{G}\colon \mathbb{R}^n \to 2^{\mathbb{R}^n}$ satisfies $\mathrm{Dom}(\mathsf{J}_{\alpha\mathsf{G}}) = \mathbb{R}^n$ and $\mathsf{J}_{\alpha\mathsf{G}}(x)$ is single-valued for all $x \in \mathbb{R}^n, \alpha > 0$, then the following equivalences hold: (i) $\mathbb{0}_n \in (\mathsf{F}+\mathsf{G})(x)$, (ii) $x = (\mathsf{J}_{\alpha\mathsf{G}} \circ \mathsf{S}_{\alpha\mathsf{F}})(x)$, and (iii) $z = (\mathsf{R}_{\alpha\mathsf{F}} \circ \mathsf{R}_{\alpha\mathsf{G}})(z)$ and $x = \mathsf{J}_{\alpha\mathsf{G}}(z)$ [72, pp. 25 and 28]. In other words, even if G is a set-valued mapping, forward-backward and Peaceman-Rachford splitting methods may be applied to compute zeros of the inclusion problem (6.4).

When $\mathsf{F}\colon \mathbb{R}^n \to \mathbb{R}^n$ in (6.4) is Lipschitz continuous and strongly monotone w.r.t. $\|\cdot\|_{\infty,[\eta]^{-1}}$ with monotonicity parameter $c > 0$ and $\mathsf{G} = \partial g$ for a separable proper, l.s.c., convex mapping $g\colon \mathbb{R}^n \to \overline{\mathbb{R}}$, by Proposition 3.6.4, the composition $\mathrm{prox}_{\alpha g} \circ \mathsf{S}_{\alpha\mathsf{F}}$ is a contraction w.r.t. $\|\cdot\|_{\infty,[\eta]^{-1}}$ for small enough $\alpha > 0$. Therefore, the forward-backward splitting method, Algorithm 4, may be applied to find a zero of the splitting problem (6.4). Analogously, for small enough $\alpha > 0$, $\mathsf{R}_{\alpha\mathsf{F}} \circ \mathsf{R}_{\alpha\mathsf{G}}$ is a contraction w.r.t. $\|\cdot\|_{\infty,[\eta]^{-1}}$ and Peaceman-Rachford splitting, Algorithm 5, may be applied to find a zero of the problem (6.4). In the following section, we present an application of the above theory to recurrent neural networks.

3.6.2 Iterations for recurrent neural network equilibrium computation

Consider the continuous-time recurrent neural network

$$\dot{x} = -x + \Phi(Ax + Bu + b) =: F(x, u), \tag{3.42}$$

where $x \in \mathbb{R}^n, u \in \mathbb{R}^m, A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times m}, b \in \mathbb{R}^n$, and $\Phi\colon \mathbb{R}^n \to \mathbb{R}^n$ is a separable activation function, i.e., it acts entry-wise in the sense that $\Phi(x) = (\phi(x_1), \dots, \phi(x_n))^\top$. In this section we consider activation functions $\phi\colon \mathbb{R} \to \mathbb{R}$ satisfying slope bounds of the form

$$d_1 = \inf_{x,y\in\mathbb{R},x\neq y} \frac{\phi(x) - \phi(y)}{x - y} \geq 0, \qquad d_2 = \sup_{x,y\in\mathbb{R},x\neq y} \frac{\phi(x) - \phi(y)}{x - y} \leq 1. \tag{3.43}$$

Most standard activation functions used in machine learning satisfy these bounds. In [100, Theorem 23], it was shown that a sufficient condition for the strong infinitesimal contractivity of the map $x \mapsto F(x, u)$ is the existence of weights $\eta \in \mathbb{R}^n_{>0}$ such that $\mu_{\infty,[\eta]^{-1}}(A) < 1$; if this condition holds, the recurrent neural network (3.42) is strongly infinitesimally contracting w.r.t. $\|\cdot\|_{\infty,[\eta]^{-1}}$ with rate $1 - \max\{d_1\gamma, d_2\gamma\}$, where we define $\gamma = \mu_{\infty,[\eta]^{-1}}(A) < 1$.

Suppose that, for fixed u, we are interested in efficiently computing the unique equilibrium point x^*_u of $F(x, u)$. Note that equilibrium points x^*_u satisfy $x^*_u = \Phi(Ax^*_u + Bu + b)$ which corresponds to an implicit neural network (INN), which have recently gained significant attention in the machine learning community [109, 66, 110]. In this regard, computing equilibrium points of (3.42) corresponds to computing the forward pass of an

INN.

Since the map $x \mapsto F(x,u)$ is strongly infinitesimally contracting w.r.t. $\|\cdot\|_{\infty,[\eta]^{-1}}$, the map $x \mapsto -F(x,u)$ is strongly monotone with monotonicity parameter $1-\max\{d_1\gamma, d_2\gamma\}$ (see Remark 3.3.1). As a consequence, applying the forward step method, Algorithm 1, to compute x_u^* yields the iteration

$$x^{k+1} = (1-\alpha)x^k + \alpha\Phi(Ax^k + Bu + b), \tag{3.44}$$

which is the iteration proposed in [98]. This iteration is guaranteed to converge for every $\alpha \in (0, \frac{1}{1-\min_{i\in\{1,\dots,n\}}\min\{d_1\cdot(A)_{ii}, d_2\cdot(A)_{ii}\}}]$ with contraction factor $1-\alpha(1-\max\{d_1\gamma, d_2\gamma\})$ by Theorem 3.4.1(ii).

However, rather than viewing finding an equilibrium of (3.42) as finding a zero of a non-Euclidean monotone operator, it is also possible to view it as a monotone inclusion problem of the form (6.4).

Proposition 3.6.5 ([66, Theorem 1]). *Suppose ϕ satisfies the bounds (3.43). Then finding an equilibrium point x_u^* of (3.42) is equivalent to the (set-valued) operator splitting problem $\mathbb{0}_n \in (\mathsf{F}+\mathsf{G})(x_u^*)$, with*

$$\mathsf{F}(z) = (I_n - A)z - (Bu+b), \qquad \mathsf{G}(z) = \partial g(z), \tag{3.45}$$

where we denote $g(z) = \sum_{i=1}^n f(z_i)$ and $f\colon \mathbb{R} \to \overline{\mathbb{R}}$ is proper, l.s.c., convex, and satisfies $\phi = \mathrm{prox}_f$.

Proof. By Proposition 3.6.2, since ϕ satisfies the bounds (3.43), there exists a proper, l.s.c., convex f with $\phi = \mathrm{prox}_f$. The remainder of the proof is equivalent to that in [66, Thm 1]. □

While Proposition 3.6.5 was leveraged in [66] for monotonicity w.r.t. the ℓ_2 norm, we will use it for F which is monotone w.r.t. a diagonally-weighted ℓ_∞ norm.[10]

Checking that F is strongly monotone w.r.t. $\|\cdot\|_{\infty,[\eta]^{-1}}$ is straightforward under the assumption that $\gamma < 1$. As a consequence of Propositions, 3.6.4 and 3.6.5, we can consider different operator splitting algorithms to compute the equilibrium of (3.42). First, the forward-backward splitting method, Algorithm 4, as applied to this problem is

$$x^{k+1} = \mathrm{prox}_{\alpha g}((1-\alpha)x^k + \alpha(Ax^k + Bu + b)). \tag{3.46}$$

Since F is Lipschitz continuous, this iteration is guaranteed to converge to the unique fixed point of (3.42) by Theorem 3.5.1(i). Moreover, the contraction factor for this iteration is $1-\alpha(1-\gamma)$ for $\alpha \in (0, \frac{1}{1-\min_i (A)_{ii}}]$, with contraction factor being minimized at $\alpha^* = \frac{1}{1-\min_i (A)_{ii}}$. Note that compared to the iteration (3.44), iteration (3.46) has a larger allowable range of step sizes and improved contraction factor at the expense of computing a proximal operator at each iteration.

Alternatively, the fixed point may be computed by means of the Peaceman-Rachford splitting method, Algorithm 5, which can be written

$$\begin{aligned} x^{k+1} &= (I_n + \alpha(I_n - A))^{-1}(z^k + \alpha(Bu + b)), \\ z^{k+1} &= z^k + 2\mathrm{prox}_{\alpha g}(2x^{k+1} - z^k) - 2x^{k+1}. \end{aligned} \tag{3.47}$$

Since F is Lipschitz continuous and $\mathsf{R}_{\alpha\mathsf{G}}$ is nonexpansive for every $\alpha > 0$, this iteration converges to the unique fixed point of (3.42) for α in a suitable range by Theorem 3.5.2(i). Moreover, the contraction factor is $\frac{1-\alpha(1-\gamma)}{1+\alpha(1-\gamma)}$ for $\alpha \in (0, \frac{1}{1-\min_i (A)_{ii}}]$, which comes from the Lipschitz constant of F. In other words, the contraction factor is improved for Peaceman-Rachford compared to forward-backward splitting and the range of allowable step sizes is

[10]Unless $A = A^\top$, the monotone inclusion problem (3.45) does not arise from a convex minimization problem.

identical. For RNNs where $(I_n + \alpha(I_n - A))$ may be easily inverted, this splitting method may be preferred.

3.6.3 Numerical implementations

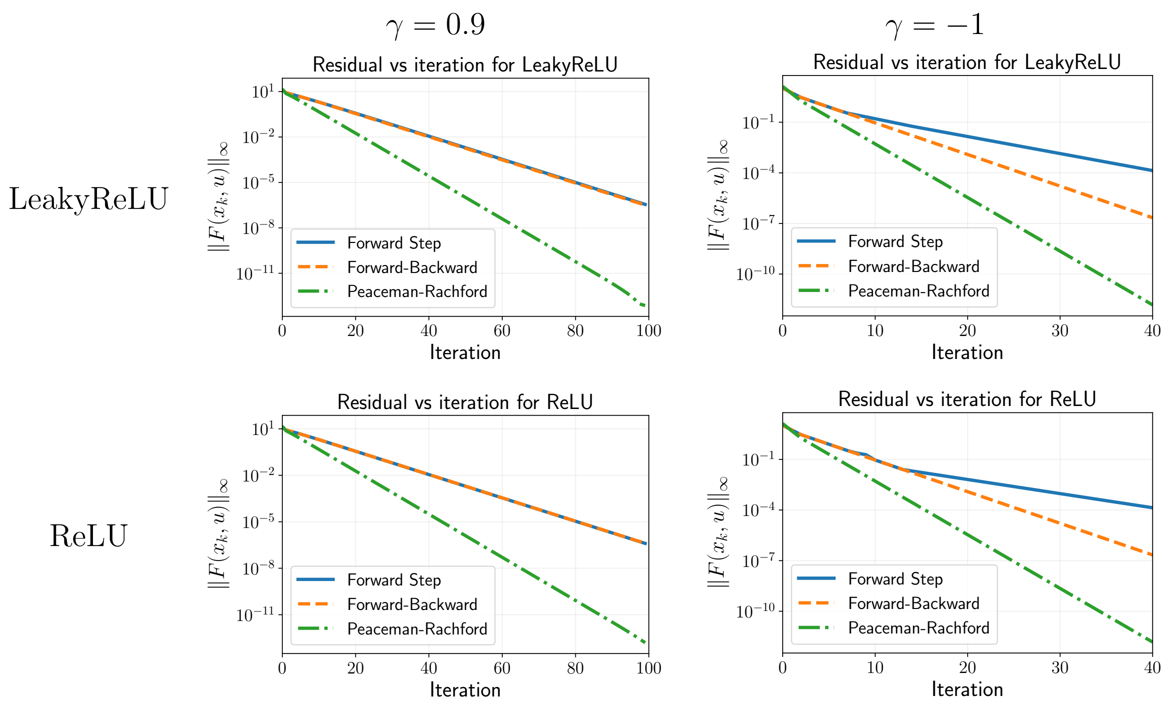


Figure 3.2: Residual versus number of iterations for forward-step method (3.44), forward-backward splitting (3.46), and Peaceman-Rachford splitting (3.47) for computing the equilibrium of the recurrent neural network (3.42). The top two plots correspond to ϕ = LeakyReLU with $a = 0.1$ and the bottom two plots correspond to ϕ = ReLU. The left two plots correspond to $\gamma = 0.9$ and the right two plots correspond to $\gamma = -1$. Curves for the forward-step method and forward-backward splitting are directly on top of one another in the left two plots. Note the difference in the number of iterations with respect to the parameter γ.

To assess the efficacy of the iterations in (3.44), (3.46), and (3.47), we generated A, B, b, u in (3.42) and applied the iterations to compute the equilibrium. We generate $A \in \mathbb{R}^{200\times200}, B \in \mathbb{R}^{200\times50}, u \in \mathbb{R}^{50}, b \in \mathbb{R}^{200}$ with entries normally distributed as $A_{ij}, B_{ij}, b_i, u_i \sim \mathcal{N}(0,1)$. To ensure that $A \in \mathbb{R}^{200\times200}$ satisfies the constraint $\mu_{\infty,[\eta]^{-1}}(A) \leq$

γ for some $\eta \in \mathbb{R}^{200}_{>0}$, we pick $[\eta] = I_{200}$ and orthogonally project A onto the convex polytope $\{A \in \mathbb{R}^{200\times 200} \mid \mu_\infty(A) \leq \gamma\}$ using CVXPY [111]. In experiments, we consider $\gamma \in \{-1, 0.9\}$ and consider activation functions $\phi(x) = \mathrm{ReLU}(x) = \max\{x, 0\}$ and $\phi(x) = \mathrm{LeakyReLU}(x) = \max\{x, ax\}$ with $a = 0.1$.[11] The proper, l.s.c., convex f corresponding to these activation functions are available in [107, Table 1].

For all iterations, we initialize x^0 at the origin and for the Peaceman-Rachford iteration, we initialize z^0 at the origin. For each iteration we pick the largest theoretically allowable step size, which in all cases was $\frac{1}{1-\min_i (A)_{ii}}$ (since $\min_{i\in\{1,\dots,n\}}(A)_{ii}$ was negative in all cases). For the case of $\gamma = 0.9$, we found that the largest theoretically allowable step size was $\alpha \approx 0.182$ and for $\gamma = -1$ the largest step size was $\alpha \approx 0.175$. The plots of the residual $\|x_k - \Phi(Ax_k + Bu + b)\|_\infty = \|F(x_k, u)\|_\infty$ versus the number of iterations for all different cases is shown in Figure 3.2.[12]

We see that, when $\gamma = 0.9$, both forward-step and forward-backward splitting methods for computing the equilibrium of (3.42) converge at the same rate. This result agrees with the theory since $\gamma > 0$, so that $\max\{d_1\gamma, d_2\gamma\} = \gamma$ for both ReLU and LeakyReLU and the estimated contraction factor for both the forward step method and forward-backward splitting is $1 - \alpha(1-\gamma) \approx 0.982$. For the Peaceman-Rachford splitting method and $\gamma = 0.9$, the estimated contraction factor is $\frac{1-\alpha(1-\gamma)}{1+\alpha(1-\gamma)} \approx 0.964$, which justifies the improved rate of convergence. When $\gamma = -1$, the forward-backward splitting method converges faster than the forward step method. This result agrees with the theory since the estimated contraction factor for the forward step method is $1-\alpha(1-\phi(\gamma)) \approx 0.807$ in the case of LeakyReLU and ≈ 0.825 in the case of ReLU while the estimated contraction factor for forward-backward splitting is $1 - \alpha(1-\gamma) \approx 0.649$ independent of activation

[11] Note that the slope bounds from (3.40) are $d_1 = 0, d_2 = 1$ for ReLU and $d_1 = a, d_2 = 1$ for LeakyReLU with $a \in [0, 1[$.

[12] All iterations and graphics were run and generated in Python. Code to reproduce experiments is available at https://github.com/davydovalexander/RNN-Equilibrium-NonEucMonotone.

function. On the other hand, for the Peaceman-Rachford splitting method and $\gamma = -1$, the estimated contraction factor is $\frac{1-\alpha(1-\gamma)}{1+\alpha(1-\gamma)} \approx 0.481$, which justifies the improved rate of convergence.

3.6.4 Tightened Lipschitz constants for continuous-time RNNs

We are interested in studying the robustness of the RNN (3.42) to input perturbations. In other words, given a nominal input, u, and its corresponding equilibrium output, x_u^*, we aim to upper-bound the deviation of the output due to a change in the input. The Lipschitz constant of a neural network is one common metric used to evaluate its robustness, as discussed in works such as [112, 83, 113]. In the context of implicit neural networks, Lipschitz constants have been studied in [114, 84, 98], with [84] unrolling forward-backward splitting iterations to provide ℓ_2 Lipschitz estimates. In what follows, we generalize the procedure in [84] using techniques from non-Euclidean monotone operator theory to provide novel and tighter ℓ_∞ Lipschitz estimates.

Theorem 3.6.1 (Lipschitz estimate of equilibrium points of (3.42))**.** *Suppose that A satisfies $\mu_{\infty,[\eta]^{-1}}(A) = \gamma < 1$ for some $\eta \in \mathbb{R}^n_{>0}$ and that $\phi = \mathrm{prox}_f$ for some proper, l.s.c., convex $f\colon \mathbb{R} \to \overline{\mathbb{R}}$. Define $f_{\mathsf{N}}\colon \mathbb{R}^m \to \mathbb{R}^n$ by $f_{\mathsf{N}}(u) = x_u^*$ where x_u^* solves the fixed point problem $x_u^* = \Phi(Ax_u^* + Bu + b)$.[13] Then for $\eta_{\max} = \max_{i\in\{1,\dots,n\}} \eta_i$, $\eta_{\min} = \min_{i\in\{1,\dots,n\}} \eta_i$, and $\mathsf{Lip}_\infty(f_{\mathsf{N}})$ denoting the minimal ℓ_∞ Lipschitz constant of f_{N},*

$$\mathsf{Lip}_\infty(f_{\mathsf{N}}) \leq \frac{\eta_{\max}}{\eta_{\min}} \frac{\|B\|_\infty}{1-\gamma}. \tag{3.48}$$

Proof. We consider the forward-backward splitting iteration given input u as $x_u^{k+1} = \mathrm{prox}_{\alpha g}((1-\alpha)x_u^k + \alpha(Ax_u^k + Bu + b))$ with initial condition $x_u^0 = \mathbb{0}_n$ which is guaranteed

[13] Note that if x_u^* solves the fixed point problem $x_u^* = \Phi(Ax_u^* + Bu + b)$, then it is an equilibrium point of the RNN (3.42).

to converge for $\alpha \in (0, \frac{1}{1-\min_i(A)_{ii}}]$ since $\mathrm{prox}_{\alpha g}$ is nonexpansive and $\mathsf{S}_{\alpha\mathsf{F}}$ is a contraction w.r.t. $\|\cdot\|_{\infty,[\eta]^{-1}}$ for every α in this range where F is defined as in (3.45). We find

$$\begin{aligned}
\|x_u^k - x_v^k\|_{\infty,[\eta]^{-1}} &= \|\mathrm{prox}_{\alpha g}((1-\alpha)x_u^{k-1} + \alpha(Ax_u^{k-1} + Bu + b)) \\
&\quad - \mathrm{prox}_{\alpha g}((1-\alpha)x_v^{k-1} + \alpha(Ax_v^{k-1} + Bv + b))\|_{\infty,[\eta]^{-1}} \\
&\leq \|\mathsf{S}_{\alpha(\mathsf{Id}-A)}(x_u^{k-1} - x_v^{k-1})\|_{\infty,[\eta]^{-1}} + \alpha\|B(u-v)\|_{\infty,[\eta]^{-1}} \qquad (3.49) \\
&\leq \mathsf{Lip}(\mathsf{S}_{\alpha(\mathsf{Id}-A)})^k \|x_u^0 - x_v^0\|_{\infty,[\eta]^{-1}} + \alpha\|B(u-v)\|_{\infty,[\eta]^{-1}} \sum_{i=0}^{k-1} \mathsf{Lip}(\mathsf{S}_{\alpha(\mathsf{Id}-A)})^i \\
&= \alpha\|B(u-v)\|_{\infty,[\eta]^{-1}} \sum_{i=0}^{k-1} \mathsf{Lip}(\mathsf{S}_{\alpha(\mathsf{Id}-A)})^i, \qquad (3.50)
\end{aligned}$$

where (3.49) holds because of nonexpansiveness of $\mathrm{prox}_{\alpha g}$ and the triangle inequality and (3.50) is a consequence of $x_u^0 = x_v^0 = \mathbb{0}_n$.

Since the forward-backward splitting iteration converges for every α in the desired range, we can take the limit as $k \to \infty$ and find that $x_u^k \to x_u^*$ and $x_v^k \to x_v^*$ as $k \to \infty$. Then

$$\begin{aligned}
\|x_u^* - x_v^*\|_{\infty,[\eta]^{-1}} &\leq \alpha\|B(u-v)\|_{\infty,[\eta]^{-1}} \sum_{i=0}^{\infty} \mathsf{Lip}(\mathsf{S}_{\alpha(\mathsf{Id}-A)})^i \qquad (3.51) \\
&= \frac{\alpha\|B(u-v)\|_{\infty,[\eta]^{-1}}}{1 - \mathsf{Lip}(\mathsf{S}_{\alpha(\mathsf{Id}-A)})} \leq \frac{\alpha\|B(u-v)\|_{\infty,[\eta]^{-1}}}{1-(1-\alpha(1-\gamma))} = \frac{\|B(u-v)\|_{\infty,[\eta]^{-1}}}{1-\gamma}, \qquad (3.52)
\end{aligned}$$

which implies the result because $\eta_{\max}^{-1}\|z\|_\infty \leq \|z\|_{\infty,[\eta]^{-1}} \leq \eta_{\min}^{-1}\|z\|_\infty$ for every $z \in \mathbb{R}^n$. ☐

Remark 3.6.6. *In [98, Corollary 5], the following Lipschitz estimate is given:*

$$\mathsf{Lip}_\infty(f_\mathsf{N}) \leq \frac{\eta_{\max}}{\eta_{\min}} \frac{\|B\|_\infty}{1-\max\{\gamma, 0\}}. \qquad (3.53)$$

The Lipschitz estimate in Theorem 3.6.1 is always a tighter bound than the estimate (3.53) *and allows the choice of negative γ to further lower the Lipschitz constant of the RNN. Indeed, one way to make the neural network more robust to uncertainties in its input would be to ensure that γ is a large negative number.*

3.7 Conclusion

In this chapter, we introduce a non-Euclidean version of classical results in monotone operator theory with a focus on mappings that are monotone with respect to diagonally-weighted ℓ_1 or ℓ_∞ norms. Our results show that the resolvent and reflected resolvent maintain many useful properties from the Hilbert case, and we prove that commonly used algorithms for finding zeros of monotone operators and their sums remain effective in the non-Euclidean setting. We applied our theory to the problem of equilibrium computation and Lipschitz constant estimation of recurrent neural networks, yielding novel iterations and tighter upper bounds on Lipschitz constants via forward-backward splitting.

Topics of future research include (i) extending results to more general Banach spaces with a focus on L_1 and L_∞ spaces, (ii) studying the convergence of additional operator splitting methods such as forward-backward-forward [115] and Davis-Yin [116] splittings, (iii) extending the theory to variable step size methods, and (iv) considering additional machine learning applications such as ℓ_∞ robustness of deep neural networks as a non-Euclidean analog of [83] or reinforcement learning and dynamic programming, where ℓ_∞ contractive and nonexpansive operators are prevalent; see the recent work [117] for preliminary ideas in this direction.

Chapter 4

Robust Implicit Networks via Non-Euclidean Contractions

This chapter was first published in the Proceedings of the Conference on Advances in Neural Information Processing Systems [98].[1]

4.1 Introduction

Implicit neural networks are infinite-depth learning models with layers defined implicitly through a fixed-point equation. Examples of implicit neural networks include deep equilibrium models [109] and implicit deep learning models [110]. Implicit networks can be considered as generalizations of feedforward neural networks with input-injected weight tying, i.e., training parameters are transferable between layers. Indeed, in implicit networks, function evaluation is executed by solving a fixed-point equation and backpropagation is implemented by computing gradients using implicit differentiation.

Due to these unique features, implicit models enjoy more flexibility and improved memory efficiency compared to traditional neural networks. At the same time, implicit networks can suffer from instability in their training due to the nonlinear nature of their fixed-point equations and can show brittle input-output behaviors due to their model flexibility.

It is known that implicit neural networks require careful tuning and initialization to avoid ill-posed training procedures. Indeed, without additional assumptions, their fixed-point equation may not have a unique solution and the numerical algorithms for finding their solutions might not converge. Several recent works in the literature have focused on studying well-posedness and convergence of the fixed-point equations of implicit networks using frameworks such as monotone operator theory [66], contraction theory [110], and a mixture of both [114]. Despite several insightful results, important questions about conditions for well-posedness of implicit networks and efficient algorithms that converge to their solutions are still open.

One of the key features of implicit neural networks is their flexibility, which might come at the cost of low input-output robustness. As first noted in [118], the input-output behavior of deep neural networks can be vulnerable to perturbations; close enough input data can lead to completely different outputs. This lack of robustness can lead to unreliable performance of neural networks in safety-critical applications. Among several notions of robustness, the Lipschitz constant of a neural network is a coarse but rigorous measure which can be used to estimate input-output sensitivity of the network [118]. For this reason, there has been a growing interest in estimating the input-output Lipschitz constant of deep neural networks with respect to the ℓ_2-norm [112, 83]. However, it turns out that in some applications, the input-output Lipschitz constants with respect to non-Euclidean norms are more informative measures for studying robustness. One such application appears in the robustness analysis of neural networks with large-scale inputs under widely-distributed adversarial perturbations (examples of these adversarial perturbations can be found in [118]). For these examples, the input-output ℓ_2-Lipschitz constant does not provide complete information about robustness of the network; a neural network with small input-output ℓ_2-Lipschitz constant can be very sensitive to widespread entrywise-small perturbations of the input signal. On the other hand, the input-output ℓ_∞-Lipschitz constant provides a different metric which appears to be well-suited for the analysis of widespread distributed perturbations. Another application is the estimation of input signal confidence intervals from output deviations, where the input-output ℓ_∞-Lipschitz constant of the network provides more scalable bounds than its ℓ_2 counterpart.

Related works

Implicit learning models. Numerous works in learning theory have shown the power of deep learning models with implicit layers. In these learning models, the notion of layers are replaced by a composition rule, which can be either a fixed-point iteration or a solution to a differential equation. Well-known frameworks for deep learning using implicit infinite-depth layers include deep equilibrium networks [109], implicit deep learning [110], and Neural ODEs [119]. In [120], a class of implicit recurrent neural networks is considered and it is demonstrated that, with this architecture, the models do not suffer from vanishing nor exploding gradients. Implicit layers have also been used to study convex optimization problems [121] and to design control strategies [122]. Convergence to global minima of certain classes of implicit networks is studied in [123].

Well-posedness and numerical algorithms for fixed-point equations. There has been a recent interest in studying well-posedness and numerical stability of implicit-depth learning models. [110] proposes a sufficient spectral condition for well-posedness and for convergence of the Picard iterations associated with the fixed-point equation of implicit networks. In [66, 114], using monotone operator theory, a suitable parametrization of the weight matrix is proposed which guarantees the stable convergence of suitable fixed-point iterations. A recent influential survey on monotone operators is [72]. A recent survey on fixed point strategies in data science is given by [67].

Robustness of learning models It is known that neural networks can be vulnerable to adversarial input perturbations [118]. A large body of literature is devoted to improve robustness of neural networks using various defense strategies against adversarial examples [75, 124]. While these strategies are effective in many scenarios, they do not provide formal guarantees for robustness [125]. However, there has been a recent interest in designing classifiers that are provably robust with respect to adversarial perturbations [126, 127]. The input-output Lipschitz constant of a neural network is a rigorous metric for its worst-case sensitivity with respect to input perturbations. Several recent works have focused on estimating the Lipschitz constant and enforcing its boundedness. For example, [112, 128] propose a convex optimization framework using quadratic constraints and semidefinite programming to obtain upper bounds on Lipschitz constants of deep neural networks. In [113], a training algorithm is designed to ensure boundedness of the Lipschitz constant of the neural network via a semidefinite program. Other methods for estimating the Lipschitz constant of deep neural networks include [129, 130, 83]. For implicit neural networks, a sensitivity-based robustness analysis is proposed in [110]. Lipschitz constants of deep equilibrium networks have also been studied in [84, 114] using monotone operator theory.

Contributions

In this chapter, using non-Euclidean contraction theory with respect to the ℓ_∞-norm, we propose our novel framework, Non-Euclidean Monotone Operator Network (NEMON), to design implicit neural networks and study their well-posedness, stability, and robustness. First, we develop elements of a novel non-Euclidean monotone operator theory akin to the frameworks in [4, 72]. Using the concept of the logarithmic norm (henceforth log norm), we introduce the essential notion of one-sided Lipschitz constant of a map. Based upon this notion, we prove a general fixed-point theorem with weaker requirements than classical results on Picard and Krasnosel'skii–Mann iterations. For maps with one-sided Lipschitz constant less than unity, we show that an average iteration converges for sufficiently small step sizes and optimize its rate of convergence. For the special case of the weighted ℓ_∞-norm, we show that this average iteration can be accelerated by choosing a larger step size. Additionally, we study perturbed fixed-point equations and establish a bound on the distance between perturbed and nominal equilibrium points as a function of one-sided Lipschitz condition. Second, for implicit neural networks, we use our new fixed-point theorem to (i) establish ℓ_∞-norm conditions for their well-posedness, (ii) design accelerated numerical algorithms for computing their solutions, and (iii) provide upper bounds on their input-output ℓ_∞-Lipschitz constants. Third, we propose a parametrization for matrices with appropriate bound on their one-sided Lipschitz constants and use this parametrization with the average iteration to design a training optimization problem. Finally, we perform several numerical experiments illustrating improved performance of NEMON in image classification on the MNIST and the CIFAR-10 datasets compared to the state-of-the-art models in [110, 66]. Additionally, by adding the input-output Lipschitz constant as regularizer in the training problem, we observe improved robustness to some classes of adversarial perturbations. We include all relevant proofs in Section 4.7.3.

4.2 Fixed-point equations and one-sided Lipschitz constants

In this section, we show that the notion of one-sided Lipschitz constant can be used to study solvability of fixed-point equation:

$$x = \mathsf{F}(x), \tag{4.1}$$

where $\mathsf{F}\colon \mathbb{R}^n \to \mathbb{R}^n$ is a differentiable map. Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$, then in view of the Banach fixed-point theorem, a simple sufficient condition for existence of a unique solution for the fixed-point equation (4.1) is $\mathsf{Lip}(\mathsf{F}) < 1$. We note that the sufficient condition $\mathsf{Lip}(\mathsf{F}) < 1$ depends on the specific form of the fixed-point equation (4.1) and can be relaxed by a suitable rewriting of this fixed-point equation. Given an averaging parameter $\alpha \in (0, 1]$ we define the *average map* $\mathsf{F}_\alpha\colon \mathbb{R}^n \to \mathbb{R}^n$ by $\mathsf{F}_\alpha := (1-\alpha)\mathsf{Id} + \alpha\mathsf{F}$, where Id is the identity map. Using this notion, an equivalent reformulation of the fixed-point equation (4.1) is:

$$x = (1-\alpha)x + \alpha\mathsf{F}(x) = \mathsf{F}_\alpha(x). \tag{4.2}$$

For $\alpha = 1$, we have $\mathsf{F}_\alpha(x) = \mathsf{F}(x)$ and equation (4.2) coincides with equation (4.1). For every $\alpha \in (0, 1)$, the map F_α is different from F but equations (4.1) and (4.2) are equivalent. Hence, if $\mathsf{Lip}(\mathsf{F}_\alpha) < 1$, then by the Banach fixed-point theorem, the fixed point equation (4.2) (and therefore the fixed point equation (4.1)) has a unique solution x^* and the sequence $\{y_k\}_{k=1}^{\infty}$ defined by

$$y_{k+1} = (1-\alpha)y_k + \alpha\mathsf{F}(y_k), \qquad \text{for all } k \in \mathbb{Z}_{\geq 0} \tag{4.3}$$

converges geometrically to x^* with rate $\mathsf{Lip}(\mathsf{F}_\alpha)$. As a result of the parametrization (4.2), the condition $\mathsf{Lip}(\mathsf{F}) < 1$ for existence and uniqueness of the fixed-point can be relaxed to sufficient conditions

$$\mathsf{Lip}(\mathsf{F}_\alpha) < 1, \tag{4.4}$$

parametrized by $\alpha \in (0,1]$. Additionally, if condition (4.4) is satisfied, then algorithm (4.3) computes the fixed point x^*. It can be shown that the condition (4.4) becomes less conservative as α decreases. The next theorem shows that in the limit as $\alpha \to 0^+$, condition (4.4) approaches the condition $\mathrm{osL}(\mathsf{F}) < 1$.

Theorem 4.2.1 (Fixed points via one-sided Lipschitz conditions)**.** *Let $\mathsf{F}\colon \mathbb{R}^n \to \mathbb{R}^n$ be differentiable and Lipschitz with constant $\ell > 0$ with respect to a norm $\|\cdot\|$. Define the average map $\mathsf{F}_\alpha = (1-\alpha)\mathsf{Id} + \alpha\mathsf{F}$ and, for $c > 0$, the function $\gamma_{\ell,c}\colon (0, \frac{c}{(c+\ell+1)(\ell+1)}) \to \mathbb{R}$ by:*

$$\gamma_{\ell,c}(\alpha) := \Big(1 + \alpha c - \frac{\alpha^2(\ell+1)^2}{1-\alpha(\ell+1)}\Big)^{-1}.$$

Then the following statements are equivalent:

(i) $\mathrm{osL}(\mathsf{F}) < 1 - c$,

(ii) $\mathsf{Lip}(\mathsf{F}_\alpha) = \gamma_{\ell,c}(\alpha)$, *for* $0 < \alpha < \frac{c}{(c+\ell+1)(\ell+1)}$.

Moreover, if the equivalent conditions (i) or (ii) hold, then, for condition number $\kappa = \frac{1+\mathsf{Lip}(\mathsf{F})}{1-\mathrm{osL}(\mathsf{F})}$,

(iii) F *has a unique fixed point x^*;*

(iv) for $0 < \alpha < \frac{1}{\kappa(\kappa+1)}$, F_α is a contraction mapping with contraction factor $\gamma_{\ell,c}(\alpha) < 1$;

(v) *the* minimum *contraction factor* $\gamma^*_{\ell,c} = 1 - \frac{1}{4\kappa^2} + \frac{1}{8\kappa^3} + \mathcal{O}\left(\frac{1}{\kappa^4}\right)$ *and the minimizing averaging parameter* α^* *of* F_α *is*

$$\alpha^* = \frac{\kappa}{1-\mathrm{osL}(\mathsf{F})}\left(1 - \frac{1}{\sqrt{1+1/\kappa}}\right) = \frac{1}{1-\mathrm{osL}(\mathsf{F})}\left(\frac{1}{2\kappa^2} - \frac{3}{8\kappa^3} + \mathcal{O}\left(\frac{1}{\kappa^4}\right)\right).$$

The average iteration (4.2) is often referred to as the Krasnosel'skii–Mann iteration or the damped iteration [4]. Compared to [4, Theorem 5.15], Theorem 4.2.1(iv) studies convergence of the Krasnosel'skii–Mann iteration for arbitrary norms, proposes a weaker convergence condition of the form $\mathrm{osL}(\mathsf{F}) < 1$ (hence, F need not be non-expansive). However, it ensures convergence for only sufficiently small $\alpha > 0$ and assumes that F is differentiable (as will be shown, however, the latter assumption can be relaxed).

4.2.1 Accelerated convergence for weighted ℓ_∞ norms

For diagonally weighted ℓ_∞ norms, one can strengthen Theorem 4.2.1(iv) to prove the convergence of the average iteration (4.2) on a larger domain of the parameter α.

Theorem 4.2.2 (Accelerated fixed point algorithm for ℓ_∞ norms)**.** *Let* $\mathsf{F}\colon \mathbb{R}^n \to \mathbb{R}^n$ *be differentiable and Lipschitz with respect to the weighted non-Euclidean norm* $\|\cdot\|_{\infty,[\eta]^{-1}}$*. Define the average map* $\mathsf{F}_\alpha = (1-\alpha)\mathsf{Id} + \alpha\mathsf{F}$ *and pick* $\mathrm{diagL}(\mathsf{F}) \in [-\mathsf{Lip}(\mathsf{F}), \mathrm{osL}(\mathsf{F})]$ *to satisfy*

$$\mathrm{diagL}(\mathsf{F}) \le \min_{i\in\{1,\dots,n\}} \inf_{x\in\mathbb{R}^n} D\mathsf{F}_{ii}(x). \tag{4.5}$$

If $\mathrm{osL}(\mathsf{F}) < 1$*, then* F *has a unique fixed-point* x^* *and*

(i) *for* $0 < \alpha \le \dfrac{1}{1-\mathrm{diagL}(\mathsf{F})}$, F_α *is a contraction mapping with the contraction factor* $1 - \alpha(1-\mathrm{osL}(\mathsf{F})) < 1$*;*

(ii) *the minimum contraction factor and minimizing averaging parameter of* F_α *are,* *respectively,*

$$\begin{aligned}\mathsf{Lip}(\mathsf{F}_{\alpha^*}) &= 1 - \frac{1-\mathrm{osL}(\mathsf{F})}{1-\mathrm{diagL}(\mathsf{F})} = 1 - \frac{1}{\kappa_\infty}, \qquad \textit{for } \kappa_\infty = \frac{1-\mathrm{diagL}(\mathsf{F})}{1-\mathrm{osL}(\mathsf{F})} \leq \frac{1+\mathsf{Lip}(\mathsf{F})}{1-\mathrm{osL}(\mathsf{F})},\\ \alpha^* &= \frac{1}{1-\mathrm{diagL}(\mathsf{F})}.\end{aligned}$$

Note that $\mathrm{diagL}(\mathsf{F})$ is well-defined because of the Lipschitz continuity assumption.

It is instructive to compare the minimum contraction factor in the general Theorem 4.2.1 with the minimum contraction factor for ℓ_∞ norms in Theorem 4.2.2 and how they depend upon the corresponding condition numbers κ and κ_∞. We note that (i) the relevant condition number diminishes $\kappa \geq \kappa_\infty$, and (ii) the minimum contraction factor $\mathsf{Lip}(\mathsf{F}_{\alpha^*}) = 1 - \frac{1}{4\kappa^2} + \mathcal{O}(1/\kappa^4)$ improves to $\mathsf{Lip}(\mathsf{F}_{\alpha^*}) = 1 - \frac{1}{\kappa_\infty}$. This acceleration justifies the title of this section.

4.2.2 Perturbed fixed-point problems

In this subsection, we focus on solvability of the perturbed fixed-point equation:

$$x = \mathsf{F}(x,u), \tag{4.6}$$

where $\mathsf{F}\colon \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}^n$ is differentiable in x. We define $\mathsf{F}_u(x) = \mathsf{F}(x,u)$ and $\mathsf{F}_x(u) = \mathsf{F}(x,u)$. Given a norm $\|\cdot\|_{\mathcal{X}}$ in $\mathbb{R}^n$ and $\|\cdot\|_{\mathcal{U}}$ in $\mathbb{R}^r$, F is Lipschitz in its first argument with constant $\mathsf{Lip}_x(\mathsf{F}) \in \mathbb{R}_{\geq 0}$ if

$$\|\mathsf{F}(x_1,u) - \mathsf{F}(x_2,u)\|_{\mathcal{X}} \leq \mathsf{Lip}_x(\mathsf{F})\|x_1 - x_2\|_{\mathcal{X}} \quad \text{for all } x_1, x_2 \in \mathbb{R}^n \text{ and } u \in \mathbb{R}^r,$$

and it is Lipschitz in its second argument with constant $\mathsf{Lip}_u(\mathsf{F}) \in \mathbb{R}_{\geq 0}$ if

$$\|\mathsf{F}(x, u_1) - \mathsf{F}(x, u_2)\|_{\mathcal{X}} \leq \mathsf{Lip}_u(\mathsf{F})\|u_1 - u_2\|_{\mathcal{U}} \quad \text{for all } x \in \mathbb{R}^n \text{ and } u_1, u_2 \in \mathbb{R}^r,$$

and it is one-sided Lipschitz in its first argument with constant $\mathrm{osL}_x(\mathsf{F}) \in \mathbb{R}$ if

$$\mu(D_x\mathsf{F}(x, u)) \leq \mathrm{osL}_x(\mathsf{F}) \quad \text{for all } x_1, x_2 \in \mathbb{R}^n \text{ and } u \in \mathbb{R}^r.$$

The following result, which is in the spirit of Lim's Lemma [131], provides an upper bound on the distance between fixed-points of the perturbed equation (4.6).

Theorem 4.2.3 (Perturbed fixed-points)**.** *Given a norm $\|\cdot\|_{\mathcal{X}}$ in $\mathbb{R}^n$ and a norm $\|\cdot\|_{\mathcal{U}}$ in $\mathbb{R}^r$, consider a map $\mathsf{F}\colon \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}^n$ differentiable in the first argument and Lipschitz in both arguments. If F is one-sided Lipschitz with constant $\mathrm{osL}_x(\mathsf{F}) < 1$, then*

(i) for every $u \in \mathbb{R}^m$, the map F_u has a unique fixed point x_u^*;

(ii) for every $u, v \in \mathbb{R}^m$, $\|x_u^* - x_v^*\|_{\mathcal{X}} \leq \dfrac{\mathsf{Lip}_u(\mathsf{F})}{1 - \mathrm{osL}_x(\mathsf{F})}\|u - v\|_{\mathcal{U}}$.

Finally, Theorems 4.2.1, 4.2.2, and 4.2.3 are not directly applicable to activation function that are not differentiable. In Section 4.7.6, we show that for specific form of the fixed-point equation (4.1), where $\mathsf{F} = \Phi \circ \mathsf{H}$ and $\Phi\colon \mathbb{R}^n \to \mathbb{R}^n$ is a weakly increasing, non-expansive, diagonal activation function and $\mathsf{H}\colon \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}^n$ is a differentiable function, all of the conclusions of Theorems 4.2.1, 4.2.2, and 4.2.3 hold by requiring equation (4.5) to be true almost everywhere.

4.3 Contraction analysis of implicit neural networks

In this section, we use contraction theory to lay the foundation for our Non-Euclidean Monotone Operator Network (NEMON) model of implicit neural networks. Given $A \in$

$\mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times r}$, $C \in \mathbb{R}^{q\times n}$, and $D \in \mathbb{R}^{q\times r}$, we consider the implicit neural network

$$x = \Phi(Ax + Bu) := \mathsf{N}(x,u), \qquad y = Cx + Du, \tag{4.7}$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^r$, $y \in \mathbb{R}^q$, and $\Phi\colon \mathbb{R}^n \to \mathbb{R}^n$ is defined by $\Phi(x) = (\phi_1(x_1), \dots, \phi_n(x_n))$. For every $i \in \{1,\dots,n\}$, we assume the activation function $\phi_i\colon \mathbb{R} \to \mathbb{R}$ is weakly increasing, i.e., $\phi_i(x_i) \geq \phi_i(z_i)$ for $x_i \geq z_i$, and non-expansive, i.e., $|\phi_i(x_i) - \phi_i(z_i)| \leq |x_i - z_i|$ for all x_i and z_i; if ϕ_i is differentiable, these conditions are equivalent to $0 \leq \phi_i'(x_i) \leq 1$ for all $x_i \in \mathbb{R}$.

We are able to provide the following estimates on all relevant Lipschitz constants.

Theorem 4.3.1 (Lipschitz and one-sided Lipschitz constants for the implicit neural network)**.** *Consider the implicit neural network in equation (4.7) with weakly increasing and non-expansive activation functions Φ. With respect to $\|\cdot\|_{\infty,[\eta]^{-1}}$, $\eta \in \mathbb{R}^n_{>0}$, on $\mathbb{R}^n$ and $\|\cdot\|_{\mathcal{U}}$ on the input space $\mathbb{R}^r$, the map $\mathsf{N}\colon \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}^n$ is one-sided Lipschitz continuous in the first variable and Lipschitz continuous in both variables with constants:*

$$\mathrm{osL}_x(\mathsf{N}) = \mathrm{ReLU}(\mu_{\infty,[\eta]^{-1}}(A)) \ , \qquad \mathsf{Lip}_x(\mathsf{N}) = \|A\|_{\infty,[\eta]^{-1}} \ , \tag{4.8}$$

$$\mathsf{Lip}_u(\mathsf{N}) = \|B\|_{(\infty,[\eta]^{-1}),\mathcal{U}} \ , \qquad \mathrm{diagL}(\mathsf{N}) = \min_{i\in\{1,\dots,n\}}(A_{ii})_- \ , \tag{4.9}$$

where $(z)_- = 0$ if $z \geq 0$ and $(z)_- = z$ if $z < 0$.

We now use these estimates to establish multiple properties of the implicit neural network.

Corollary 4.3.1 (Well posedness, input-state Lipschitz constant, and computation)**.** *Consider the model (4.7), with parameters (A,B,C,D) and with weakly increasing and non-expansive activation functions Φ. Define the average map $\mathsf{N}_\alpha := (1-\alpha)\mathsf{Id} + \alpha\mathsf{N}$ and* *consider the norms* $\|\cdot\|_{\infty,[\eta]^{-1}}$, $\eta \in \mathbb{R}^n_{>0}$, *on* $\mathbb{R}^n$, $\|\cdot\|_{\mathcal{U}}$ *on the input space* $\mathbb{R}^r$ *and* $\|\cdot\|_{\mathcal{Y}}$ *on the output space* $\mathbb{R}^q$. *Then*

(i) if $\mu_{\infty,[\eta]^{-1}}(A) < 1$, *then* (4.7) *is well posed, i.e., there exists a unique fixed point,*

(ii) the map N_α *is a contraction mapping for* $0 < \alpha \leq \alpha^* := \big(1 - \min_{i\in\{1,\dots,n\}}(A_{ii})_-\big)^{-1}$ *with minimum contraction factor* $\mathsf{Lip}(\mathsf{N}_{\alpha^*}) = 1 - \frac{1-\mathrm{ReLU}(\mu_{\infty,[\eta]^{-1}}(A))}{1-\min_{i\in\{1,\dots,n\}}(A_{ii})_-}$.

(iii) the Lipschitz constants from input u *to fixed point* x^*_u *and to the output* $y = Cx^*_u + Du$ *are*

$$\mathsf{Lip}_{u\to x^*} := \frac{\mathsf{Lip}_u(\mathsf{N})}{1 - \mathrm{osL}_x(\mathsf{N})} = \frac{\|B\|_{(\infty,[\eta]^{-1}),\mathcal{U}}}{1 - \mathrm{ReLU}(\mu_{\infty,[\eta]^{-1}}(A))}, \tag{4.10}$$

$$\mathsf{Lip}_{u\to y} := \frac{\|B\|_{(\infty,[\eta]^{-1}),\mathcal{U}}\|C\|_{\mathcal{Y},(\infty,[\eta]^{-1})}}{1 - \mathrm{ReLU}(\mu_{\infty,[\eta]^{-1}}(A))} + \|D\|_{\mathcal{Y},\mathcal{U}}. \tag{4.11}$$

4.4 Training implicit neural networks

Problem setup Given an input data matrix $U = [u_1,\dots,u_m] \in \mathbb{R}^{r\times m}$ and a corresponding output data matrix $Y = [y_1,\dots,y_m] \in \mathbb{R}^{q\times m}$, we aim to learn matrices A, B, C, D so that the neural network (4.7) approximates the input-output relationship. We rewrite the model for matrix inputs as $\widehat{Y} = CX + DU$, where $X = \Phi(AX + BU)$. From Corollary 4.3.1(i), if each ϕ_i is weakly increasing and non-expansive, the fixed point problem is well-posed when $\mu_{\infty,[\eta]^{-1}}(A) < 1$ for some $\eta \in \mathbb{R}^n_{>0}$. We consider a training problem of the form

$$\begin{aligned} \min_{A,B,C,D,X} \quad & \mathcal{L}(Y, CX + DU) + \mathcal{P}(A,B,C,D) \\ & X = \Phi(AX + BU), \quad \mu_{\infty,[\eta]^{-1}}(A) \leq \gamma, \end{aligned} \tag{4.12}$$

where $\mathcal{L}$ is a loss function, $\mathcal{P}$ is a penalty function, and $\gamma < 1$ is a hyperparameter ensuring the fixed point problem is well-posed. For $\eta = \mathbb{1}_n$, we can remove the constraint $\mu_\infty(A) \leq \gamma$ in the training optimization problem (4.12) using the following parametrization of weight matrix A:

$$A = T - \mathrm{diag}(|T|\mathbb{1}_n) + \gamma I_n. \tag{4.13}$$

In Section 4.7.2, we show that parametrization (4.13) characterizes the set of matrices in $\mathbb{R}^{n\times n}$ satisfying $\mu_\infty(A) \leq \gamma$. Using the parametrization (4.13) in the training problem not only improves the computational efficiency of the optimization but also allows for the design of implicit neural networks with additional structure such as convolutions. Suppose $u \in \mathbb{R}^{rs^2}$ is a r-channel input of size $s \times s$ and $x \in \mathbb{R}^{ns^2}$ is an n-channel hidden layer. To define our implicit CNN, we select the weight matrix $A \in \mathbb{R}^{ns^2 \times ns^2}$ as the matrix form of a 2D convolutional operator. If we consider a circular convolution operator, then A is a circulant matrix. Using the parametrization (4.13), A is circulant if and only if T is circulant. Therefore, the training problem for implicit CNNs can be cast as an unconstrained optimization problem using the above parametrization with a circulant T.

Improving robustness via Lipschitz regularization We now focus on learning robust implicit neural networks with bounded Lipschitz constants via a regularization strategy. Setting both $\|\cdot\|_{\mathcal{U}}$ and $\|\cdot\|_{\mathcal{Y}}$ as $\|\cdot\|_\infty$ in the input-output Lipschitz bound (4.11), we get

$$\begin{aligned}\mathsf{Lip}_{u\to y} &= \frac{\|B\|_{(\infty,[\eta]^{-1}),(\infty)}\|C\|_{(\infty),(\infty,[\eta]^{-1})}}{1 - \mathrm{ReLU}(\mu_{\infty,[\eta]^{-1}}(A))} + \|D\|_{\infty,\infty} \\ &\leq \frac{1}{2}\frac{\|B\|^2_{(\infty,[\eta]^{-1}),(\infty)} + \|C\|^2_{(\infty),(\infty,[\eta]^{-1})}}{1 - \mathrm{ReLU}(\mu_{\infty,[\eta]^{-1}}(A))} + \|D\|_{\infty,\infty},\end{aligned}$$

where the inequality provides a convex upper bound for the input-output Lipschitz constant. Therefore, using the hyperparameter $\lambda > 0$, the regularized optimization problem is written as

$$\begin{aligned} \min_{A,B,C,D,X} \quad & \mathcal{L}(Y, CX + DU) + \lambda\Big(\frac{1}{2}\frac{\|B\|^2_{(\infty,[\eta]^{-1}),(\infty)} + \|C\|^2_{(\infty),(\infty,[\eta]^{-1})}}{1 - \mathrm{ReLU}(\mu_{\infty,[\eta]^{-1}}(A))} + \|D\|_{\infty,\infty}\Big) \\ & X = \Phi(AX + BU), \quad \mu_{\infty,[\eta]^{-1}}(A) \leq \gamma. \end{aligned} \tag{4.14}$$

Certified adversarial robustness via Lipschitz bounds Given a nominal input $u \in \mathbb{R}^r$, we consider any perturbed input v within an ℓ_∞-ball of radius ε around u. In this case, we have

$$\|y_u - y_v\|_\infty \leq \mathsf{Lip}_{u\to y} \|u - v\|_\infty \leq \mathsf{Lip}_{u\to y}\, \varepsilon. \tag{4.15}$$

Then we define $\mathrm{margin}(u) = (y_u)_i - \max_{j\neq i}(y_u)_j$, where $(y_u)_i$ is the logit corresponding to the (correct) label i for the input u. Then provided $L\varepsilon \leq \frac{1}{2}\mathrm{margin}(u)$, NEMON is certifiably robust to any perturbed input v within an ℓ_∞-ball of radius ε centered at u.

Backpropagation of gradients via average iteration From [110] we now show how the average iteration can be used to perform backpropagation via the implicit function theorem. For simplicity, we assume that each activation function ϕ_i is differentiable and consider mini-batches of size 1, i.e., we have $X = x \in \mathbb{R}^n$, $U = u \in \mathbb{R}^r$ and $\widehat{Y} = \widehat{y} \in \mathbb{R}^q$. Let x^* be the unique solution of the fixed-point equation (4.7). Then the chain rule implies

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial A} &= (\nabla_{x^*}\mathcal{L})x^\top, \qquad & \frac{\partial \mathcal{L}}{\partial B} &= (\nabla_{x^*}\mathcal{L})u^\top, \\ \frac{\partial \mathcal{L}}{\partial C} &= (\nabla_{\widehat{y}}\mathcal{L})x^\top, \qquad & \frac{\partial \mathcal{L}}{\partial D} &= (\nabla_{\widehat{y}}\mathcal{L})u^\top. \end{aligned}$$

Since $\mathcal{L}$ depends explicitly on $\widehat{y}$, computing $\nabla_{\widehat{y}}\mathcal{L}$ is straightforward. Computing $\nabla_{x^*}\mathcal{L}$ is more complicated since X^* is defined only implicitly. However, it be shown that

$$\nabla_{x^*}\mathcal{L} = (C(I - D\Phi A)^{-1}D\Phi)^\top \nabla_{\widehat{y}}\mathcal{L}.$$

Since $\mu_{\infty,[\eta]^{-1}}(A) < 1$, by Lemma 4.7.2 we get that $\mu_{\infty,[\eta]^{-1}}(D\Phi A) < 1$. This implies that the matrix $G := (I_n - D\Phi A)^{-1}D\Phi \in \mathbb{R}^{n\times n}$ exists and is the solution to the following fixed-point equation [110, Section 6.2]

$$G = D\Phi(AG + I_n). \tag{4.16}$$

Moreover, $\mu_{\infty,[\eta]^{-1}}(D\Phi A) < 1$ and Theorem 4.2.2 together imply that the fixed-point equation (4.16) has a unique solution G^* and, for every $0 < \alpha \leq \alpha^* := \big(1 - \min_i (A_{ii})_-\big)^{-1}$, the average iterations

$$G_{k+1} = (1-\alpha)G_k + \alpha D\Phi(AG_k + I_n), \qquad \text{for all } k \in \mathbb{Z}_{\geq 0}$$

are contracting with the minimum contraction factor $1 - \alpha^*\big(1 - \mathrm{ReLU}(\mu_{\infty,[\eta]^{-1}}(A))\big)$ at step size α^*.

4.5 Theoretical and numerical comparisons

In this section, we provide a comprehensive comparison of our framework with the state-of-the-art implicit neural networks[2].

[2]All models were trained using Google Colab with a Tesla P100-PCIE-16GB GPU.

4.5.1 Implicit network models

We start by reviewing the existing models for implicit networks in the literature.

Implicit deep learning model. [110] proposes a class of implicit neural networks with input-output behavior described by (4.7). It is shown that a sufficient condition for existence and uniqueness of a solution and convergence of the Picard iterations for the fixed point equation $x = \Phi(Ax + Bu)$ is $\lambda_{\mathrm{pf}}(|A|) < 1$, where $|A|$ denotes the entrywise absolute value of the matrix A and λ_{pf} denotes the Perron-Frobenius eigenvalue. For training, the optimization problem (4.12) is used where the constraint $\mu_{\infty,[\eta]^{-1}}(A) \leq \gamma$ is replaced by $\|A\|_\infty \leq \gamma$ [110, Equation 6.3].[3] It is easy to see that our well-posedness condition in Corollary 4.3.1(i) is less conservative than $\lambda_{\mathrm{pf}}(|A|) < 1$ and its convex relaxation $\|A\|_\infty < 1$.

Monotone operator deep equilibrium network (MON). [66] proposes to use monotone operator theory to guarantee well-posedness of the fixed-point equation as well as its convergence to the solutions. The input-output behavior of the network is described by (4.7). For training, the optimization problem (4.12) is used where the constraint $\mu_{\infty,[\eta]^{-1}}(A) \leq \gamma$ is replaced by $I_n - \frac{1}{2}(A + A^\top) \succeq (1-\gamma)I_n$. In order to ensure that this constraint is always satisfied in the training procedure, the weight matrix A is parametrized as $A = \gamma I_n - W^\top W - Z + Z^\top$, for arbitrary $W, Z \in \mathbb{R}^{n\times n}$ [66, Appendix D].[4] In the context of contraction theory,

$$I_n - \tfrac{1}{2}(A + A^\top) \succeq (1-\gamma)I_n \quad \Longleftrightarrow \quad \mu_2(A) \leq \gamma,$$

[3] The implicit deep learning implementation is available at `https://github.com/beeperman/idl`.

[4] The MON implementation is available at `https://github.com/locuslab/monotone_op_net`.

which is shown in Section 4.7. Thus, the parametrization $A = \gamma I_n - W^\top W - Z + Z^\top$ can be considered as the ℓ_2-norm version of the parametrization described by equation (4.13). In other words, the monotone operator network formulation is a Euclidean transcription of the framework we propose in this chapter.

4.5.2 MNIST experiments

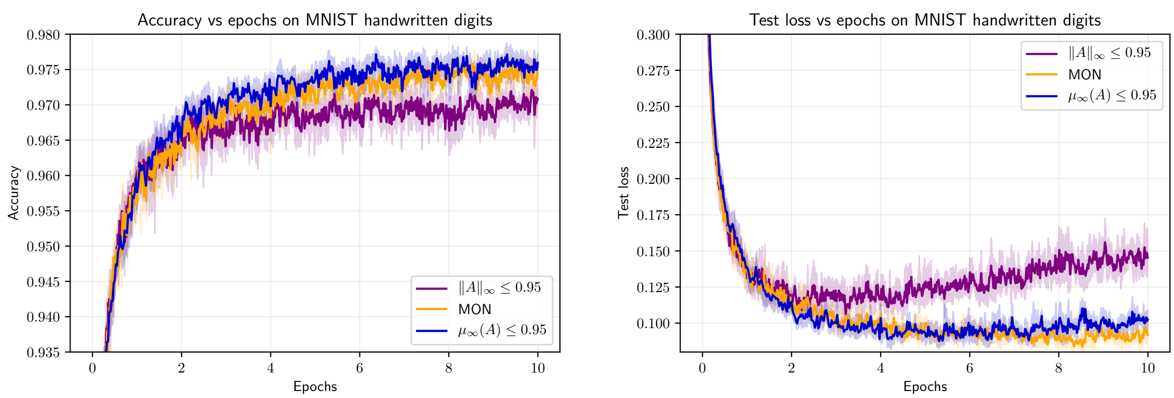


Figure 4.1: Performance comparison between the NEMON model with $\mu_\infty(A) \leq 0.95$, the implicit deep learning model with $\|A\|_\infty \leq 0.95$, and MON with $I_n - \frac{1}{2}(A + A^\top) \succeq 0.05 I_n$ on the MNIST dataset. The curves are generated by mean accuracy and mean loss over 5 different runs while light envelopes around the curves correspond to the standard deviation over the runs. Average best accuracy for the NEMON model is 0.9772, while it is 0.9721 for implicit deep learning model and 0.9762 for the MON model.

In the digit classification dataset MNIST, input data are 28×28 pixel images of handwritten digits between 0-9. There are 60000 training images and 10000 test images. For training, images are reshaped into 784 dimensional column vectors and entries are scaled into the range $[0, 1]$. As a loss function, we use the cross-entropy. All models are of order $n = 100$, used the ReLU activation function, and are trained with a batch size of 300 over 10 epochs with a learning rate of 1.5×10^{-2}. Curves for accuracy and loss versus epochs for the three models are shown in Figure 4.1. Regarding training times, using the average iteration, NEMON took, on average, 12 forward iterations, 13 backward iterations, and 9.8 seconds to train per epoch. Using the Peaceman-Rachford iteration, MON took, on average, 17 forward iterations, 16 backward iterations, and 9.5 seconds to train per epoch. Using the Picard iteration, the implicit deep learning model took, on average, 10 forward iterations, 5 backward iterations, and 5.8 seconds to train per epoch. We observe that the NEMON model performs better than the implicit deep learning model and has a comparable performance to MON.

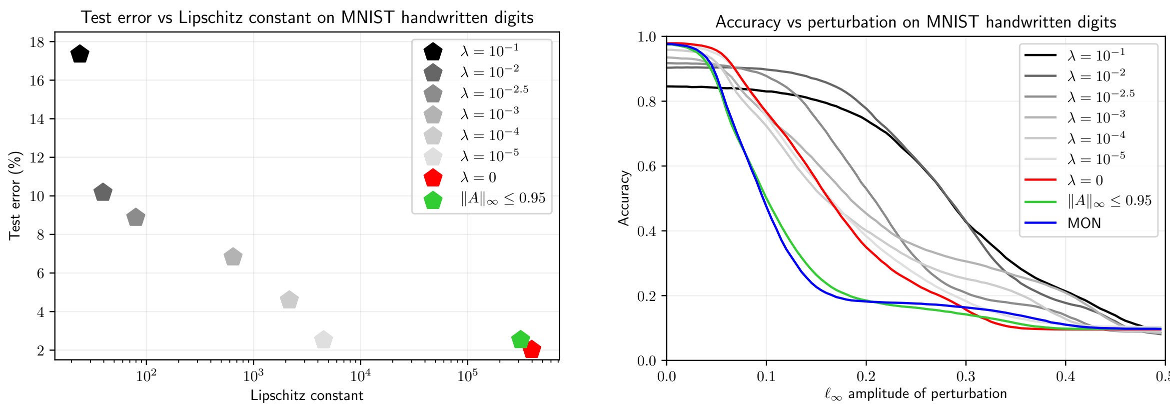


Figure 4.2: On the left is a plot of test error versus Lipschitz constant for the implicit deep learning model with $\|A\|_\infty \leq 0.95$ and for NEMON with $\mu_\infty(A) \leq 0.95$ and parametrized by the regularization hyperparameter λ. We define the test error as 1 minus the accuracy. On the right is a plot of accuracy versus ℓ_∞ perturbation of a deterministic adversarial image inversion attack where we additionally include the MON model with $I_n - \frac{1}{2}(A + A^\top) \succeq 0.05 I_n$.

We also study the robustness of the NEMON model compared to the implicit deep learning model and the MON model on the MNIST dataset. We train various models regularized by the input-output Lipschitz constant as in (4.14). Additionally, to verify robustness of the different models, we consider several adversarial attacks and plot the accuracy versus perturbation of such an attack. In Figure 4.2, we consider a continuous image inversion attack [132], where each pixel is perturbed in the direction of pixel value inversion with amplitude given by the ℓ_∞ perturbation. For more details on this and other types of adversarial perturbations, we refer to Section 4.8. We observe that for $\lambda = 10^{-5}$, the regularized NEMON model achieves a two order of magnitude decrease in its input-output Lipschitz constant compared to the un-regularized NEMON models. In addition, we see that the implicit deep learning model and the MON model are more sensitive to the continuous image inversion attack than NEMON. Moreover, as the regularization parameter λ increases, the NEMON model becomes increasingly robust to this attack.

4.5.3 CIFAR-10 experiments

In the image classification dataset CIFAR-10, input data are 32×32 color images in 10 classes. There are 50000 training images and 10000 test images. We compare our proposed NEMON model with a convolutional structure to a single convolutional layer MON model. Each model used 81 channels. We train both models with a batch size of 256 and a learning rate of 10^{-3} for 40 epochs. For training, using the average iteration, NEMON took, on average, 10 forward iterations, 10 backward iterations, and 75.0 seconds per epoch to train. Using the Peaceman-Rachford iteration, MON took, on average, 5 forward iterations, 5 backward iterations, and 101.8 seconds per epoch to train.

We focus primarily on the robustness of NEMON and MON with respect to ℓ_∞-norm bounded perturbations on CIFAR-10. To this end, we additionally trained two NEMON models with regularization parameters $\lambda \in \{10^{-4}, 10^{-5}\}$. In Figure 4.3, on the left is a plot of the certified robustness of each of the models via their ℓ_∞-Lipschitz constants. For MON, we got the ℓ_∞-Lipschitz bound using the method in [84] for the ℓ_2-Lipschitz bound and using the upper bound $\|u\|_2 \leq \sqrt{rs^2}\|u\|_\infty$. On the right is a plot of the accuracy of different models with respect to the projected gradient descent attack. We observe that the un-regularized and regularized NEMON models are more robust to ℓ_∞-norm bounded perturbations than is MON.

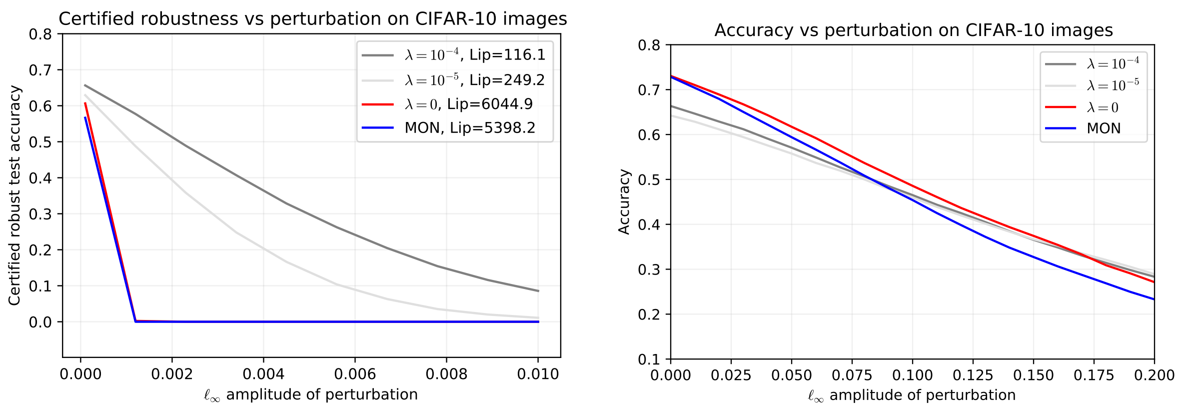


Figure 4.3: On the left is a plot of certified robustness via the Lipschitz constants of MON with the constraint $I_n - \frac{1}{2}(A + A^\top) \succeq I_n$ and NEMON with the constraint $\mu_\infty(A) \leq 0$. On the right is a plot of accuracy versus ℓ_∞ perturbation of the projected gradient descent attack.

4.6 Conclusion

Using non-Euclidean contraction theory, we propose a framework to study stability of fixed-point equations. We apply this framework to analyze well-posedness and convergence of implicit neural networks and to design an efficient training algorithm to incorporate robustness guarantees. For future research, we envision that our framework is applicable to study stability and robustness of implicit learning models with additional structure such as graph neural networks.

4.7 Proofs and auxiliary results

4.7.1 Graphical comparison between Lipschitz and one-sided Lipschitz estimates

In the following example, we compare the regions $\mathsf{Lip}(A) < 1$ and $\mathrm{osL}(A) < 1$ for a matrix $A \in \mathbb{R}^{2\times 2}$ with respect to the ℓ_∞-norm.

Example 4. *Let* $A = \begin{bmatrix} a & b \\ b & a \end{bmatrix}$*, it is easy to see that condition* $\mathsf{Lip}(A) < 1$ *for* ℓ_∞*-norm can be written as* $\|A\|_\infty = |a| + |b| < 1$*. One can also define the average operator* A_α *using parameter* $\alpha \in (0, 1]$ *as follows:*

$$A_\alpha = (1 - \alpha)I_2 + \alpha A.$$

Figure 4.4 compares the regions $\mathsf{Lip}(A) < 1$*,* $\mathsf{Lip}(A_\alpha) < 1$*, and* $\mathrm{osL}(A) < 1$ *based on the parameters* a *and* b*. It can be shown that as* $\alpha \to 0^+$*, the condition* $\mathsf{Lip}(A_\alpha) < 1$ *converges to* $\mathrm{osL}(A) < 1$*.*

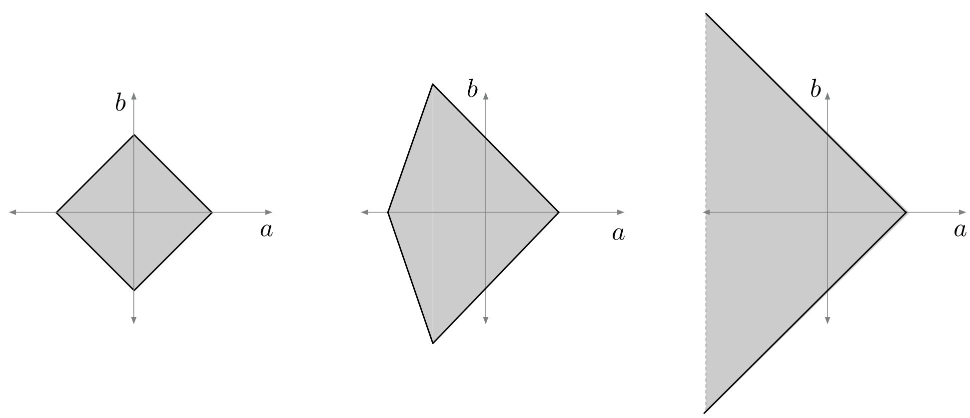


Figure 4.4: The left figure shows the region $\mathsf{Lip}(A) \leq 1$, the middle figure shows the region $\mathsf{Lip}(A_\alpha) \leq 1$ for $\alpha = \frac{1}{2}$, and the right figure shows $\mathrm{osL}(A) \leq 1$. Both Lip and osL are with respect to the ℓ_∞-norm

4.7.2 Novel results about non-Euclidean log norms

In this section we provide some results regarding the log norm and matrix norm for weighted ℓ_1 and ℓ_∞-norms.

Lemma 4.7.1 (Non-Euclidean contraction estimates)**.** *Let* $A = [a_{ij}] \in \mathbb{R}^{n\times n}$ *and* $\eta \in \mathbb{R}^n_{>0}$*,* *(i) For every* $\alpha \in \mathbb{R}$ *such that* $|\alpha| \leq (\max_i |a_{ii}|)^{-1}$,

$$\|I_n + \alpha A\|_{1,[\eta]} = 1 + \alpha\mu_{1,[\eta]}(A),$$
$$\|I_n + \alpha A\|_{\infty,[\eta]^{-1}} = 1 + \alpha\mu_{\infty,[\eta]^{-1}}(A).$$

(ii) the minimizer and minimum value of $\min_{\alpha\geq 0} \|I_n + \alpha A\|_{\infty,[\eta]^{-1}}$ *can be computed via the linear program:*

$$\begin{aligned}
\min_{\alpha,t} \quad & t \\
& 1 + \alpha(a_{ii} + r_i) \leq t, \qquad i \in \{1,\dots,n\}, \\
& -1 + \alpha(-a_{ii} + r_i) \leq t, \qquad i \in \{1,\dots,n\}, \\
& \alpha \geq 0.
\end{aligned}$$

where $r_i = \sum_{j\neq i} \frac{\eta_j}{\eta_i}|a_{ij}|$.

Proof. Regarding part (i), we compute

$$\|I_n + \alpha A\|_{\infty,[\eta]^{-1}} = \max_{i\in\{1,\dots,n\}} \big\{|1 + \alpha a_{ii}| + \alpha \sum_{j=1,j\neq i}^{n} \frac{\eta_j}{\eta_i}|a_{ij}|\big\}. \tag{4.17}$$

Since $|\alpha| \leq (\max_i |a_{ii}|)^{-1}$, we know $|\alpha||a_{ii}| \leq 1$ for all $i \in \{1,\dots,n\}$. Therefore $1+\alpha a_{ii} \geq 0$ and $|1 + \alpha a_{ii}| = 1 + \alpha a_{ii}$, for every $i \in \{1,\dots,n\}$. In summary, replacing in (4.17),

$$\|I_n + \alpha A\|_{\infty,[\eta]^{-1}} = \max_{i\in\{1,\dots,n\}} \big\{1 + \alpha a_{ii} + \alpha \sum_{j=1,j\neq i}^{n} \frac{\eta_j}{\eta_i}|a_{ij}|\big\} = 1 + \alpha\mu_{\infty,[\eta]^{-1}}(A).$$

The proof of the formula relating the weighted ℓ_1-norm and the weighted ℓ_1 log norm will follow mutatis mutandis to the above proof for ℓ_∞-norm and we omit it in the interest of brevity.

Regarding part (ii), using formula (4.17), we get

$$\begin{aligned}\|I_n + \alpha A\|_{\infty,[\eta]^{-1}} &= \max_{i\in\{1,\dots,n\}} \big\{|1+\alpha a_{ii}| + \alpha r_i\big\} \\ &= \max_{i\in\{1,\dots,n\}} \big\{1+\alpha a_{ii} + \alpha r_i, -1-\alpha a_{ii} + \alpha r_i\big\}.\end{aligned}$$

The result then follows. □

The following results are related to [133, Theorem 3.8] and [134, Lemma 3] and, indirectly, to [7]. In comparison with [133, 134], we prove sharper bounds for a more general setting.

Lemma 4.7.2 (Log norm inequalities under multiplicative scalings)**.** *For each $A \in \mathbb{R}^{n\times n}$, $C \in \mathbb{R}^{n\times n}$ diagonal positive, and $\eta \in \mathbb{R}^n_{>0}$,*

(i) $\max_{d\in[0,1]^n} \mu_{\infty,[\eta]}(-C + [d]A) = \max\big\{\mu_{\infty,[\eta]}(-C), \mu_{\infty,[\eta]}(-C+A)\big\}$, *and*

(ii) $\max_{d\in[0,1]^n} \mu_{1,[\eta]}(-C + A[d]) = \max\big\{\mu_{1,[\eta]}(-C), \mu_{1,[\eta]}(-C+A)\big\}$.

Proof. Define the short-hand $r_i = a_{ii} + \sum_{j=1,j\neq i}^n |a_{ij}|\eta_i/\eta_j$ and note

$$\mu_{\infty,[\eta]}(-C) = \max_{i\in\{1,\dots,n\}} \{-c_i\}, \quad \mu_{\infty,[\eta]}(-C+A) = \max_{i\in\{1,\dots,n\}} \{-c + r_i\}, \quad \text{and}$$
$$\mu_{\infty,[\eta]}(-C+[d]A) = \max_{i\in\{1,\dots,n\}} \{-c_i + d_i r_i\}.$$

Since $0 \le d_i \le 1$, we note

$$\begin{aligned} r_i \le 0 \quad &\Longrightarrow \quad d_i r_i \le 0 \quad \Longrightarrow \quad -c_i + d_i r_i \le -c_i, \\ r_i > 0 \quad &\Longrightarrow \quad d_i r_i \ge 0 \quad \Longrightarrow \quad -c_i + d_i r_i \le -c_i + r_i. \end{aligned}$$

Therefore

$$\max_{d\in[0,1]^n}\max_{i\,:\,r_i\le 0}\{-c_i+d_ir_i\} = \max_{i\,:\,r_i\le 0}\max_{d_i\in[0,1]}\{-c_i+d_ir_i\} = \max_{i\,:\,r_i\le 0}\{-c_i\} \le \mu_{\infty,[\eta]}(-C),$$

$$\max_{d\in[0,1]^n}\max_{i\,:\,r_i>0}\{-c_i+d_ir_i\} = \max_{i\,:\,r_i>0}\max_{d_i\in[0,1]}\{-c_i+d_ir_i\} = \max_{i\,:\,r_i\le 0}\{-c_i+r_i\} \le \mu_{\infty,[\eta]}(-C+A).$$

In summary

$$\begin{aligned}\max_{d\in[0,1]^n}\mu_{\infty,[\eta]}(-C+[d]A) &= \max_{d\in[0,1]^n}\max_{i\in\{1,\dots,n\}}\{-c_i+d_ir_i\}\\ &= \max_{d\in[0,1]^n}\max\Big\{\max_{i\,:\,r_i\le 0}\{-c_i+d_ir_i\},\max_{i\,:\,r_i>0}\{-c_i+d_ir_i\}\Big\}\\ &\le \max\big\{\mu_{\infty,[\eta]}(-C),\mu_{\infty,[\eta]}(-C+A)\big\}.\end{aligned}$$

On the other hand, we note that

$$\begin{aligned}\max_{d\in[0,1]^n}\mu_{\infty,[\eta]}([d]A-C) &\ge \max\big\{\mu_{\infty,[\eta]}([\mathbb{0}_n]A-C),\mu_{\infty,[\eta]}([\mathbb{1}_n]A-C)\big\}\\ &= \max\big\{\mu_{\infty,[\eta]}(-C),\mu_{\infty,[\eta]}(-C+A)\big\},\end{aligned}$$

thereby proving the equality in statement (i). Next, recall $\mu_{1,[\eta]}(B)=\mu_{\infty,[\eta]}(B^\top)$ for all B and compute

$$\begin{aligned}\max_{d\in[0,1]^n}\mu_{1,[\eta]}(-C+A[d]) &= \max_{d\in[0,1]^n}\mu_{\infty,[\eta]}(-C+[d]A^\top)\\ &= \max\big\{\mu_{\infty,[\eta]}(-C),\mu_{\infty,[\eta]}(-C+A^\top)\big\}\\ &= \max\big\{\mu_{1,[\eta]}(-C),\mu_{\infty,[\eta]}(-C+A)\big\}.\end{aligned}$$

This concludes the proof of statement (ii). $\square$

In the same style as [66, Proposition 1] and [114, Theorems 1 and 2], the next lemma provides a parametrization of all matrices satisfying a μ_∞ constraint.

Lemma 4.7.3 (Parametrization of matrices with bounded ℓ_∞ measure)**.** *For any $\gamma \in \mathbb{R}$,*

(i) given any $A \in \mathbb{R}^{n\times n}$ with $\mu_\infty(A) \leq \gamma$, there exists a $T \in \mathbb{R}^{n\times n}$ such that $A = T - \operatorname{diag}(|T|\mathbb{1}_n) + \gamma I_n$,

(ii) given any $T \in \mathbb{R}^{n\times n}$, the matrix $A = T - \operatorname{diag}(|T|\mathbb{1}_n) + \gamma I_n \in \mathbb{R}^{n\times n}$ satisfies $\mu_\infty(A) \leq \gamma$,

where we let $|T|$ denote the entry-wise absolute value of T.

Proof. Regarding statement (i), define

$$\begin{aligned} t_{ij} &= a_{ij} && \text{for all } i \neq j \in \{1,\dots,n\}, \\ t_{ii} &= \frac{1}{2}\big(a_{ii} + \sum_{j=1, j\neq i}^{n} |a_{ij}| - \gamma\big), && \text{for } i \in \{1,\dots,n\}. \end{aligned}$$

Because $\mu_\infty(A) \leq \gamma$, we know $a_{ii} + \sum_{j=1,j\neq i}^{n} |a_{ij}| \leq \gamma$ for each i. This implies that $t_{ii} \leq 0$ and therefore $t_{ii} - |t_{ii}| = a_{ii} + \sum_{j=1,j\neq i}^{n} |a_{ij}| - \gamma$. It is an easy transcription now to show that this equality and the off-diagonal equality $t_{ij} = a_{ij}$ together imply $A = T - \operatorname{diag}(|T|\mathbb{1}_n) + \gamma I_n$.

Regarding statement (ii), note that $a_{ij} = t_{ij}$ for all $j \neq i$, and $a_{ii} = t_{ii} - \sum_{j=1}^{n} |t_{ij}| + \gamma$. Then, for all i,

$$\begin{aligned} a_{ii} + \sum_{j=1, j\neq i}^{n} |a_{ij}| &= \big(t_{ii} - \sum_{j=1}^{n} |t_{ij}| + \gamma\big) + \sum_{j=1, j\neq i}^{n} |t_{ij}| \\ &= t_{ii} - |t_{ii}| + \gamma = \begin{cases} \gamma, & \text{if } t_{ii} \geq 0, \\ -2|t_{ii}| + \gamma, & \text{if } t_{ii} < 0. \end{cases} \end{aligned}$$

Therefore, $a_{ii} + \sum_{j=1,j\neq i}^{n} |a_{ij}| \leq \gamma$ for all i and, in turn, $\mu_\infty(A) \leq \gamma$. $\square$

We conclude with a simple graph-theoretical interpretation of the main well-posedness condition $\mu_\infty(A) < 1$. Loosely speaking, we call $-a_{ii}$ the self-attenuation of neuron i and $\sum_{j=1,j\neq i}^n |a_{ij}|$ the strength of its outgoing synapses. Then

$$\begin{aligned}\mu_\infty(A) < 1 \quad &\iff \quad a_{ii} + \sum_{j=1,j\neq i}^{n} |a_{ij}| < 1 \quad \text{for all } i \\ &\iff \quad \text{for each neuron, strength of outgoing synapses} < 1+ \text{ self-attenuation.}\end{aligned} \tag{4.18}$$

4.7.3 Proofs and additional results on non-differentiable activation functions

4.7.4 Proofs of Theorems 4.2.1 and 4.2.2

Proof of Theorem 4.2.1. Regarding (ii) $\implies$ (i), note that, for every $x \in \mathbb{R}^n$ and every $0 < \alpha \leq \alpha^*$,

$$\mu(D\mathsf{F}_\alpha(x)) \leq \|D\mathsf{F}_\alpha(x)\| \leq \gamma_{\ell,c}(\alpha).$$

As a result, $\alpha\mu(D\mathsf{F}(x)) = \mu(D\mathsf{F}_\alpha(x)) - 1 + \alpha \leq -1 + \alpha + \gamma_{\ell,c}(\alpha)$. Thus,

$$\mu(D\mathsf{F}(x)) \leq 1 - \frac{1-\gamma_{\ell,c}(\alpha)}{\alpha}, \qquad \text{for all } x \in \mathbb{R}^n.$$

By choosing $\alpha = \widehat{\alpha} = \frac{2c}{(2c+\ell+1)(\ell+1)} < \frac{c}{(c+\ell+1)(\ell+1)}$, we get

$$\mu(D\mathsf{F}(x)) \leq 1 - \frac{1-\gamma_{\ell,c}(\widehat{\alpha})}{\widehat{\alpha}} = 1 - \frac{1-(1-\widehat{\alpha}c)}{\widehat{\alpha}} = 1-c, \qquad \text{for all } x \in \mathbb{R}^n.$$

Thus, $\sup_{x\in\mathbb{R}^n} \mu(D\mathsf{F}(x)) \leq 1 - c$. This implies that $\mathrm{osL}(\mathsf{F}) \leq 1 - c$.

Regarding (i) $\implies$ (ii), using the mean value theorem for vector valued functions, we compute

$$\|\mathsf{F}_\alpha(x) - \mathsf{F}_\alpha(y)\| = \left\| \int_0^1 D\mathsf{F}_\alpha(tx + (1-t)y)dt(x-y) \right\| \leq \|\overline{D\mathsf{F}}_\alpha(x,y)\| \|x-y\|,$$

where $\overline{D\mathsf{F}}_\alpha(x,y) = \int_0^1 D\mathsf{F}_\alpha(tx+(1-t)y)dt$, for every $x, y \in \mathbb{R}^n$.

Next, to obtain an upper bound on $\|\overline{D\mathsf{F}}_\alpha(x,y)\|$, we first derive a lower bound on $\|\overline{D\mathsf{F}}_\alpha^{-1}(x,y)\|$. We start by noting that, the product property (1.3f) implies $\|Av\| \geq -\mu(-A)\|v\|$, for every $v \in \mathbb{R}^n$ and every $A \in \mathbb{R}^{n\times n}$. Therefore, for every $v \in \mathbb{R}^n$,

$$\|\overline{D\mathsf{F}}_\alpha^{-1}(x,y)v\| \geq -\mu(-\overline{D\mathsf{F}}_\alpha^{-1}(x,y))\|v\|. \tag{4.19}$$

Since $\overline{D\mathsf{F}}_\alpha(x,y) = I_n + \alpha(-I_n + \overline{D\mathsf{F}}(x,y))$ and $\alpha < \frac{c}{(c+\ell+1)(\ell+1)} \leq \frac{1}{\ell+1}$, we can use the Neumann series to get

$$\overline{D\mathsf{F}}_\alpha^{-1}(x,y) = \sum_{i=0}^{\infty}(-1)^i\alpha^i(-I_n + \overline{D\mathsf{F}}(x,y))^i. \tag{4.20}$$

We first compute an upper bound for $\mu(\overline{D\mathsf{F}}(x))$. Since $\mathrm{osL}(\mathsf{F}) \leq 1-c$, by the subadditive property (1.3a) of the log norms, we get

$$\begin{aligned}\mu(-I_n + \overline{D\mathsf{F}}(x,y)) &= \mu\left(\int_0^1 (-I_n + D\mathsf{F}(tx+(1-t)y))dt\right) \\ &\leq \int_0^1 \mu\big(-I_n + D\mathsf{F}(tx+(1-t)y)\big)dt \leq -c.\end{aligned} \tag{4.21}$$

Now, we use equation (4.20) to obtain

$$\begin{aligned}
\|\overline{D\mathsf{F}}_\alpha^{-1}(x,y)v\| &\geq -\mu\Big(\sum_{i=0}^{\infty}(-1)^{i+1}\alpha^i(-I_n+\overline{D\mathsf{F}}(x,y))^i\Big)\|v\| \\
&\geq -\Big(\mu(-I_n)+\alpha\mu(-I_n+\overline{D\mathsf{F}}(x,y)) \\
&\quad +\sum_{i=2}^{\infty}\alpha^i\mu\big((-1)^{i+1}(-I_n+\overline{D\mathsf{F}}(x,y))^i\big)\Big)\|v\| \\
&\geq (1+\alpha c-\sum_{i=2}^{\infty}(\alpha(\ell+1))^i)\|v\| = \Big(1+\alpha c-\frac{\alpha^2(\ell+1)^2}{1-\alpha(\ell+1)}\Big)\|v\|, \qquad (4.22)
\end{aligned}$$

where the first inequality holds by (4.19), the second inequality holds by subadditive property of the log norm (1.3a), and the third inequality holds because, using (4.21) and (1.3d), we obtain the upper bound:

$$\mu\big((-1)^{i+1}(-I_n+\overline{D\mathsf{F}}(x,y))^i\big) \leq \|(-I_n+\overline{D\mathsf{F}}(x,y))^i\| \leq (1+\ell)^i, \qquad \text{for all } i\in\mathbb{Z}_{\geq 0}.$$

Note that $\alpha \in (0, \frac{c}{(c+\ell+1)(\ell+1)})$. Equation (4.22) implies that, for each $w\in\mathbb{R}^n$ and $v=\overline{D\mathsf{F}}_\alpha(x,y)w$,

$$\frac{\|\overline{D\mathsf{F}}_\alpha(x,y)w\|}{\|w\|} = \frac{\|v\|}{\|\overline{D\mathsf{F}}_\alpha^{-1}(x,y)v\|} \leq \gamma_{\ell,c}(\alpha).$$

As a result, $\|\overline{D\mathsf{F}}_\alpha(x,y)\| \leq \gamma_{\ell,c}(\alpha)$ and

$$\|\mathsf{F}_\alpha(x)-\mathsf{F}_\alpha(y)\| \leq \gamma_{\ell,c}(\alpha)\|x-y\|, \qquad \text{for all } x,y\in\mathbb{R}^n.$$

Regarding parts (iii) and (iv), a straightforward calculation shows that, if $0<\alpha<\frac{c}{(c+\ell+1)(\ell+1)}$, then $1/\Big(1+\alpha c-\frac{\alpha^2(\ell+1)^2}{1-\alpha(\ell+1)}\Big)<1$. The result then follows from the Banach fixed-point theorem. Regarding part (v), we define the function $\xi :]0,\frac{c}{(c+\ell+1)(\ell+1)}[\,\to\mathbb{R}_{>0}$

by $\xi(\alpha) = 1 + \alpha c - \frac{\alpha^2(\ell+1)^2}{1-\alpha(\ell+1)}$. Then it is clear that $\xi(\alpha) = 1/\gamma_{\ell,c}(\alpha)$. Note that

$$\begin{aligned}\frac{d\xi}{d\alpha} &= (c+\ell+1) - \frac{\ell+1}{(1-\alpha(\ell+1))^2},\\ \frac{d^2\xi}{d\alpha^2} &= -\frac{2(\ell+1)^2}{(1-\alpha(\ell+1))^3}.\end{aligned}$$

Since $\frac{d^2\xi}{d\alpha^2} \leq 0$, we conclude that ξ is a concave function on $(0, \frac{c}{(c+\ell+1)(\ell+1)})$ and its maximum is achieved at α^* for which $\frac{d\xi}{d\alpha}(\alpha^*) = 0$. By a straightforward calculation, we get

$$\alpha^* = \frac{\kappa}{c}\left(1 - \frac{1}{\sqrt{1+1/\kappa}}\right)$$

and it is easy to see that the optimal value is as claimed in the theorem statement. □

Proof of Theorem 4.2.2. We restrict ourselves to the norm $\|\cdot\|_{\infty,[\eta]^{-1}}$; the proof for $\|\cdot\|_{1,[\eta]}$ is similar and omitted in the interest of brevity.

Regarding part (i), first we note that $\mathrm{diagL}(\mathsf{F}) \leq \mathrm{osL}(\mathsf{F}) < 1$, since for every $i \in \{1\ldots,n\}$ and every $x \in \mathbb{R}^n$

$$D\mathsf{F}_{ii}(x) \leq D\mathsf{F}_{ii}(x) + \sum_{j\neq i} |D\mathsf{F}_{ij}(x)|\frac{\eta_i}{\eta_j} = \mu_{\infty,[\eta]^{-1}}(D\mathsf{F}(x)) \leq \mathrm{osL}(\mathsf{F}) < 1. \tag{4.23}$$

This implies that $\frac{1}{1-\mathrm{diagL}(\mathsf{F})} > 0$ and $(1-\mathrm{osL}(\mathsf{F}))/(1-\mathrm{diagL}(\mathsf{F})) \leq 1$. Moreover, for every $x \in \mathbb{R}^n$,

$$\|(1-\alpha)I_n + \alpha D\mathsf{F}(x)\|_{\infty,[\eta]^{-1}} = \|I_n + \alpha(-I_n + D\mathsf{F}(x))\|_{\infty,[\eta]^{-1}}.$$

Next, we study the diagonal entries of $-I_n + D\mathsf{F}(x)$. By the definition of $\mathrm{diagL}(\mathsf{F})$ and by equation (4.23),

$$\begin{aligned}
& -1 + \mathrm{diagL}(\mathsf{F}) \leq -1 + D\mathsf{F}_{ii}(x) < 0 && \text{(for every } i \in \{1, \dots, n\} \text{ and } x\text{)} \\
\implies \quad & |1 - \mathrm{diagL}(\mathsf{F})| \geq |-1 + D\mathsf{F}_{ii}(x)| \\
\implies \quad & 1 - \mathrm{diagL}(\mathsf{F}) \geq \max_i |-1 + D\mathsf{F}_{ii}(x)| \\
\implies \quad & \frac{1}{1 - \mathrm{diagL}(\mathsf{F})} \leq \frac{1}{\max_i |-1 + D\mathsf{F}_{ii}(x)|}.
\end{aligned}$$

Therefore, $\alpha \leq \frac{1}{\max_i |-1+D\mathsf{F}_{ii}(x)|}$ and we can use Lemma 4.7.1(i) to deduce that

$$\begin{aligned}
\|(1-\alpha)I_n + \alpha D\mathsf{F}(x)\|_{\infty,[\eta]^{-1}} &= 1 + \alpha\mu_{\infty,[\eta]^{-1}}(-I_n + D\mathsf{F}(x)) \\
&= 1 + \alpha(-1 + \mu_{\infty,[\eta]^{-1}}(D\mathsf{F}(x))) && \text{for all } x \in \mathbb{R}^n \\
&\leq 1 + \alpha(-1 + \mathrm{osL}(\mathsf{F})) = 1 - \alpha(1 - \mathrm{osL}(\mathsf{F})) < 1.
\end{aligned}$$

where the second equality follows from the translation property (1.3e) of log norms, and the inequality holds because $\mu_{\infty,[\eta]^{-1}}(D\mathsf{F}(x)) \leq \mathrm{osL}(\mathsf{F})$ for all x, and the last inequality holds because $\mathrm{osL}(\mathsf{F}) < 1$. This means that $\mathsf{Lip}(\mathsf{F}_\alpha) < 1$, for every $0 < \alpha \leq \frac{1}{1-\mathrm{diagL}(\mathsf{F})}$ and the result follows from the Banach fixed-point theorem.

Regarding part (ii), we note the contraction factor is a strictly decreasing function of α. At $\alpha = 0$ the factor is 1 and at the maximum of value of α that is, at $\alpha^* = (1 - \mathrm{diagL}(\mathsf{F}))^{-1}$ the contraction factor is still positive since $(1 - \mathrm{osL}(\mathsf{F}))/(1 - \mathrm{diagL}(\mathsf{F})) \leq 1$. Hence the minimum contraction factor is achieved at α^*. □

4.7.5 Proof of Theorem 4.2.3 and comparison with the literature

Before we prove Theorem 4.2.3, it is useful to compare it with similar results in the literature. The result in [131, Lemma 1] is more general than Theorem 4.2.3 by allowing F to be a multi-valued map defined on a metric space. However, Theorem 4.2.3(ii) uses the one-side Lipschitz constant and provides a tighter upper bound on the distance between fixed-points of F compared to its counterpart in [131, Lemma 1].

Proof of Theorem 4.2.3. Let $[\![\cdot,\cdot]\!]$ be a WP for the norm $\|\cdot\|_{\mathcal{X}}$ on $\mathbb{R}^n$.

Regarding part (i), for every $u \in \mathbb{R}^m$, we note that by definition of $\mathrm{osL}_x(\mathsf{F})$, for every $u \in \mathbb{R}^r$,

$$[\![\mathsf{F}(x,u) - \mathsf{F}(y,u), x - y]\!] \leq \mathrm{osL}_x(\mathsf{F})\|x - y\|_{\mathcal{X}}^2,$$

This implies that $\mathrm{osL}(\mathsf{F}_u) \leq \mathrm{osL}_x(\mathsf{F}) < 1$, for every $u \in \mathbb{R}^r$. Thus, by Theorem 4.2.1(iii), F_u has a unique fixed-point x_u^*.

Regarding part (ii), let $[\![\cdot,\cdot]\!]$ be a WP for the norm $\|\cdot\|_{\mathcal{X}}$ on $\mathbb{R}^n$ and compute

$$\begin{aligned}
\|x_u^* - x_v^*\|_{\mathcal{X}}^2 &= [\![x_u^* - x_v^*, x_u^* - x_v^*]\!] && \text{(by compatibility)}\\
&= [\![\mathsf{F}_u(x_u^*) - \mathsf{F}_v(x_v^*), x_u^* - x_v^*]\!]\\
&\leq [\![\mathsf{F}_u(x_u^*) - \mathsf{F}_u(x_v^*), x_u^* - x_v^*]\!] + [\![\mathsf{F}_u(x_v^*) - \mathsf{F}_v(x_v^*), x_u^* - x_v^*]\!] && \text{(by sub-additivity)}\\
&\leq \mathrm{osL}_x(\mathsf{F})\|x_u^* - x_v^*\|_{\mathcal{X}}^2 + \|\mathsf{F}_u(x_v^*) - \mathsf{F}_v(x_v^*)\|_{\mathcal{X}}\|x_u^* - x_v^*\|_{\mathcal{X}} && \text{(by Cauchy-Schwarz)}\\
&\leq \mathrm{osL}_x(\mathsf{F})\|x_u^* - x_v^*\|_{\mathcal{X}}^2 + \mathsf{Lip}_u(\mathsf{F})\|u - v\|_{\mathcal{U}}\|x_u^* - x_v^*\|_{\mathcal{X}}.
\end{aligned}$$

This implies that $(1 - \mathrm{osL}_x(\mathsf{F}))\|x_u^* - x_v^*\|_{\mathcal{X}} \leq \mathsf{Lip}_u(\mathsf{F})\|u - v\|_{\mathcal{U}}$ and the result of part (ii) follows. □

4.7.6 Non-differentiable fixed-point problems

In many machine learning applications, the activation functions are continuous but non-differentiable and thus our results in Sections 4.2 do not directly apply to these problems. In this subsection, we focus on a specific form of the fixed-point equation (4.1), where $\mathsf{F} = \Phi \circ \mathsf{H}$ and $\Phi\colon \mathbb{R}^n \to \mathbb{R}^n$ is a diagonal activation function with absolutely continuous components and $\mathsf{H}\colon \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}^n$ is a differentiable function. It can be shown that, for this class of systems, conclusions of Theorems 4.2.1, 4.2.2, and 4.2.3 still hold with respect to weighted ℓ_∞-norms. Here, we present a result which extends Theorems 4.2.2 and 4.2.3 for $\mathsf{H}(x,u) = \mathsf{G}(x) + Bu$ given some $B \in \mathbb{R}^{n\times r}$ and with respect to the norm $\|\cdot\|_{\infty.[\eta]^{-1}}$.

Theorem 4.7.1 (Fixed points for non-differentiable activation functions)**.** *Consider the norm $\|\cdot\|_{\infty,[\eta]^{-1}}$ on $\mathbb{R}^n$ for some $\eta \in \mathbb{R}^n_{>0}$ and the norm $\|\cdot\|_{\mathcal{U}}$ on $\mathbb{R}^r$. Additionally, consider the following perturbed fixed point problem:*

$$x = \Phi(\mathsf{G}(x) + Bu) := \Phi^{\mathsf{G}}(x,u),$$

where $\Phi\colon \mathbb{R}^n \to \mathbb{R}^n$ is a diagonal function given by $(\phi_1(x_1), \dots, \phi_n(x_n))$ with non-expansive and weakly increasing ϕ_i, $\mathsf{G}\colon \mathbb{R}^n \to \mathbb{R}^n$ is a continuously differentiable function, and $B \in \mathbb{R}^{n\times r}$. Define the average map $\Phi^{\mathsf{G}}_\alpha(x,u) := (1-\alpha)x + \Phi^{\mathsf{G}}(x,u)$ and pick $\mathrm{diagL}(\mathsf{G})_- \in [-\mathsf{Lip}(\mathsf{G}), \mathrm{osL}(\mathsf{G})]$ *such that*

$$\mathrm{diagL}(\mathsf{G})_- \leq \min_i \inf_{x\in\mathbb{R}^n} D\mathsf{G}_{ii}(x)_-.$$

Assume that $\mathrm{osL}(\mathsf{G}) < 1$*. Then,*

(i) for every $u \in \mathbb{R}^n$, the map $\Phi^{\mathsf{G}}(\cdot,u)$ has a unique fixed-point x^*_u;

(ii) for every $0 < \alpha \leq \frac{1}{1-\mathrm{diagL}(\mathsf{G})_-}$ *and every* $u \in \mathbb{R}^r$, $\Phi_\alpha^{\mathsf{G}}(\cdot, u)$ *is a contraction map with contraction factor* $1 - \alpha(1 - \mathrm{ReLU}(\mathrm{osL}(\mathsf{G})))$*;*

(iii) for every $u, v \in \mathbb{R}^r$*, we have* $\|x_u^* - x_v^*\|_{\infty,[\eta]^{-1}} \leq \frac{\mathsf{Lip}_u \Phi^{\mathsf{G}}}{1-\mathrm{ReLU}(\mathrm{osL}\,\mathsf{G})}\|u - v\|_{\mathcal{U}}$.

Proof of Theorem 4.7.1. Regarding part (i), the assumptions on each scalar activation function imply that (i) $\Phi\colon \mathbb{R}^n \to \mathbb{R}^n$ is non-expansive with respect to $\|\cdot\|_{\infty,[\eta]^{-1}}$ and (ii) for every $p, q \in \mathbb{R}$, there exists $\theta_i \in [0,1]$ such that $\phi_i(p) - \phi_i(q) = \theta_i(p - q)$ or in the matrix form $\Phi(\mathbf{p}) - \Phi(\mathbf{q}) = \Theta(\mathbf{p} - \mathbf{q})$ where Θ is a diagonal matrix with diagonal elements $\theta_i \in [0,1]$ and $\mathbf{p}, \mathbf{q} \in \mathbb{R}^n$. As a result, we have

$$\begin{aligned}\|\Phi_\alpha^{\mathsf{G}}(x_1, u) - \Phi_\alpha^{\mathsf{G}}(x_2, u)\|_{\infty,[\eta]^{-1}} &= \|(1-\alpha)(x_1 - x_2) + \alpha\Theta(\mathsf{G}(x_1) - \mathsf{G}(x_2))\|_{\infty,[\eta]^{-1}} \\ &\leq \sup_{y\in\mathbb{R}^n} \|I_n + \alpha(-I_n + \Theta D\mathsf{G}(y))\|_{\infty,[\eta]^{-1}}\|x_1 - x_2\|_{\infty,[\eta]^{-1}}.\end{aligned}$$

where the inequality holds by the mean value theorem. Then, for every $\alpha \in \left]0, \frac{1}{1-\mathrm{diagL}(\Theta D\mathsf{G})}\right]$,

$$\begin{aligned}\|I_n + \alpha(-I_n + \Theta D\mathsf{G}(y))\|_{\infty,[\eta]^{-1}} &= 1 + \alpha\mu_{\infty,[\eta]^{-1}}\big(-I_n + \Theta D\mathsf{G}(y)\big) \\ &\leq 1 + \alpha\big(-1 + \mu_{\infty,[\eta]^{-1}}(\Theta D\mathsf{G}(y))\big) \\ &\leq 1 + \alpha\big(-1 + \mu_{\infty,[\eta]^{-1}}(D\mathsf{G}(y))_+\big) \\ &\leq 1 - \alpha(1 - \mathrm{osL}(\mathsf{G})_+) < 1,\end{aligned}$$

where the first equality holds by Lemma 4.7.1(i), the second inequality holds by subadditive property of matrix measures (1.3a), and the third inequality holds by Lemma 4.7.2. Moreover, since $\theta_i \in [0,1]$, we have $\theta_i D\mathsf{G}_{ii} \geq (D\mathsf{G}_{ii})_-$, for every $i \in \{1, \ldots, n\}$. This means that

$$\mathrm{diagL}(\Theta D\mathsf{G}) = \min_i \inf_{y\in\mathbb{R}^n} (\Theta D\mathsf{G}(y))_{ii} \geq \min_i \inf_{y\in\mathbb{R}^n} (D\mathsf{G}_{ii}(y))_- = \mathrm{diagL}(\mathsf{G})_-.$$

This implies that, for every $\alpha \in (0, \frac{1}{1-\mathrm{diagL}(\mathsf{G})_-}]$,

$$\|\Phi_\alpha^{\mathsf{G}}(x_1, u) - \Phi_\alpha^{\mathsf{G}}(x_2, u)\|_{\infty,[\eta]^{-1}} \leq (1 - \alpha(1 - \mathrm{ReLU}(\mathrm{osL}(\mathsf{G})))\|x_1 - x_2\|_{\infty,[\eta]^{-1}}.$$

Since $1 - \alpha(1 - \mathrm{ReLU}(\mathrm{osL}(\mathsf{G}))) < 1$, the map $\Phi_\alpha^{\mathsf{G}}(\cdot, u)$ is a contraction for every $\alpha \in (0, \frac{1}{1-\mathrm{diagL}(\mathsf{G})_-}]$. This concludes the proof of parts (i) and (ii),

Regarding part (iii), from formula (2.15) for the one-sided Lipschitz constant and formula (3.2) for the relevant WP, we obtain that, for all $x_1, x_2 \in \mathbb{R}^n$,

$$\begin{aligned}
&[\![\Phi(\mathsf{G}(x_1) + Bu) - \Phi(\mathsf{G}(x_2) + Bu), x_1 - x_2]\!]_{\infty,[\eta]^{-1}} \\
&\qquad = \max_{i \in I_\infty([\eta]^{-1}(x_1 - x_2))} \eta_i^{-2}(x_1 - x_2)_i(\phi_i((\mathsf{G}(x_1) + Bu)_i) - \phi_i((\mathsf{G}(x_2) + Bu)_i)) \\
&\qquad = \max_{i \in I_\infty([\eta]^{-1}(x_1 - x_2))} \theta_i \eta_i^{-2}(x_1 - x_2)_i((\mathsf{G}(x_1) + Bu)_i - (\mathsf{G}(x_2) + Bu)_i) \\
&\qquad = \max_{i \in I_\infty([\eta]^{-1}(x_1 - x_2))} \theta_i \eta_i^{-2}(x_1 - x_2)_i(\mathsf{G}(x_1) - \mathsf{G}(x_2))_i,
\end{aligned}$$

Next, we recall Lumer's equality and write it as

$$\mathrm{osL}(\mathsf{G}) = \sup_{x_1 \neq x_2} \max_{i \in I_\infty([\eta]^{-1}(x_1 - x_2))} \eta_i^{-2}(x_1 - x_2)_i(\mathsf{G}(x_1) - \mathsf{G}(x_2))_i.$$

Next, we consider two cases. Suppose that $\mathrm{osL}(\mathsf{G}) \leq 0$. Since $\theta_i \in [0, 1]$ for all i, we obtain

$$[\![\Phi(\mathsf{G}(x_1) + Bu) - \Phi(\mathsf{G}(x_2) + Bu), x_1 - x_2]\!]_{\infty,[\eta]^{-1}} \leq 0,$$

since the maximum value is achieved at $\theta_i = 0$ for all i. Alternatively, suppose that

$\mathrm{osL}(\mathsf{G}) > 0$. Then

$$\begin{aligned}
&[\![\Phi(\mathsf{G}(x_1) + Bu) - \Phi(\mathsf{G}(x_2) + Bu), x_1 - x_2]\!]_{\infty,[\eta]^{-1}} \\
&\qquad = \max_{i \in I_\infty([\eta]^{-1}(x_1 - x_2))} \theta_i \eta_i^{-2} (x_1 - x_2)_i (\mathsf{G}(x_1) - \mathsf{G}(x_2))_i \\
&\qquad \leq \max_{i \in I_\infty([\eta]^{-1}(x_1 - x_2))} \eta_i^{-2} (x_1 - x_2)_i (\mathsf{G}(x_1) - \mathsf{G}(x_2))_i \leq \mathrm{osL}(\mathsf{G}) \|x_1 - x_2\|^2_{\infty,[\eta]^{-1}},
\end{aligned}$$

since the maximum value is achieved at $\theta_i = 1$ for all i. This means that $\mathrm{osL}(\Phi^{\mathsf{G}}) = \mathrm{ReLU}(\mathrm{osL}(\mathsf{G}))$. Now we compute

$$\begin{aligned}
\|x_u^* - x_v^*\|^2_{\infty,[\eta]^{-1}} &= [\![x_u^* - x_v^*, x_u^* - x_v^*]\!]_{\infty,[\eta]^{-1}} \\
&= \left[\!\left[\Phi_u^{\mathsf{G}}(x_u^*) - \Phi_v^{\mathsf{G}}(x_v^*), x_u^* - x_v^*\right]\!\right]_{\infty,[\eta]^{-1}} \\
&\leq \left[\!\left[\Phi_u^{\mathsf{G}}(x_u^*) - \Phi_u^{\mathsf{G}}(x_v^*), x_u^* - x_v^*\right]\!\right]_{\infty,[\eta]^{-1}} + \left[\!\left[\Phi_u^{\mathsf{G}}(x_v^*) - \Phi_v^{\mathsf{G}}(x_v^*), x_u^* - x_v^*\right]\!\right]_{\infty,[\eta]^{-1}} \\
&\leq \mathrm{osL}(\mathsf{G})_+ \|x_u^* - x_v^*\|^2_{\infty,[\eta]^{-1}} + \|\Phi_u^{\mathsf{G}}(x_v^*) - \Phi_v^{\mathsf{G}}(x_v^*)\|_{\infty,[\eta]^{-1}} \|x_u^* - x_v^*\|_{\infty,[\eta]^{-1}} \\
&\leq \mathrm{osL}(\mathsf{G})_+ \|x_u^* - x_v^*\|^2_{\infty,[\eta]^{-1}} + \mathsf{Lip}_u(\Phi^{\mathsf{G}}) \|u - v\|_{\mathcal{U}} \|x_u^* - x_v^*\|_{\infty,[\eta]^{-1}}.
\end{aligned}$$

This implies that $(1 - \mathrm{ReLU}(\mathrm{osL}(\mathsf{G}))) \|x_u^* - x_v^*\|_{\infty,[\eta]^{-1}} \leq \mathsf{Lip}_u(\Phi^{\mathsf{G}}) \|u - v\|_{\mathcal{U}}$ and the result follows. □

4.7.7 Proofs of results in Section 4.3

Proof of Theorem 4.3.1. The assumptions on each scalar activation function imply that (i) $\Phi\colon \mathbb{R}^n \to \mathbb{R}^n$ is non-expansive with respect to $\|\cdot\|_{\infty,[\eta]^{-1}}$, and (ii) for every $p, q \in \mathbb{R}$, there exists $\theta_i \in [0,1]$ such that $\phi_i(p) - \phi_i(q) = \theta_i(p - q)$. Regarding the equality $\mathrm{osL}_x(\mathsf{N}) = \mathrm{ReLU}(\mu_{\infty,[\eta]^{-1}}(A))$, from formula (2.15) for the one-sided Lipschitz constant and formula (3.2) for the relevant WP, we obtain that, for all $x_1, x_2 \in \mathbb{R}^n$,

$$\begin{aligned}
&[\![\Phi(Ax_1 + Bu) - \Phi(Ax_2 + Bu), x_1 - x_2]\!]_{\infty,[\eta]^{-1}} \\
&\qquad = \max_{i \in I_\infty([\eta]^{-1}(x_1 - x_2))} \eta_i^{-2}(x_1 - x_2)_i(\phi_i((Ax_1 + Bu)_i) - \phi_i((Ax_2 + Bu)_i)) \\
&\qquad = \max_{i \in I_\infty([\eta]^{-1}(x_1 - x_2))} \theta_i \eta_i^{-2}(x_1 - x_2)_i((Ax_1 + Bu)_i - (Ax_2 + Bu)_i) \\
&\qquad = \max_{i \in I_\infty([\eta]^{-1}(x_1 - x_2))} \theta_i \eta_i^{-2}(x_1 - x_2)_i(Ax_1 - Ax_2)_i,
\end{aligned}$$

Next, we recall Lumer's equality and write it as

$$\mu_{\infty,[\eta]^{-1}}(A) = \sup_{x_1 \neq x_2} \max_{i \in I_\infty([\eta]^{-1}(x_1 - x_2))} \eta_i^{-2}(x_1 - x_2)_i((Ax_1)_i - (Ax_2)_i).$$

Next, we consider two cases. Suppose that $\mu_{\infty,[\eta]^{-1}}(A) \leq 0$. Since $\theta_i \in [0, 1]$ for all i, we obtain

$$[\![\Phi(Ax_1 + Bu) - \Phi(Ax_2 + Bu), x_1 - x_2]\!]_{\infty,[\eta]^{-1}} \leq 0,$$

since the maximum value is achieved at $\theta_i = 0$ for all i. Alternatively, suppose that $\mu_{\infty,[\eta]^{-1}}(A) > 0$. Then

$$\begin{aligned}
&[\![\Phi(Ax_1 + Bu) - \Phi(Ax_2 + Bu), x_1 - x_2]\!]_{\infty,[\eta]^{-1}} \\
&\qquad = \max_{i \in I_\infty([\eta]^{-1}(x_1 - x_2))} \theta_i \eta_i^{-2}(x_1 - x_2)_i(Ax_1 - Ax_2)_i \\
&\qquad \leq \max_{i \in I_\infty([\eta]^{-1}(x_1 - x_2))} \eta_i^{-2}(x_1 - x_2)_i(Ax_1 - Ax_2)_i \leq \mu_{\infty,[\eta]^{-1}}(A)\|x_1 - x_2\|^2_{\infty,[\eta]^{-1}},
\end{aligned}$$

since the maximum value is achieved at $\theta_i = 1$ for all i. This concludes the proof of formula $\mathrm{osL}_x(\mathsf{N}) = \mu_{\infty,[\eta]^{-1}}(A)_+$. Next, since Φ is non-expansive, we compute

$$\begin{aligned}\|\mathsf{N}(x_1,u) - \mathsf{N}(x_2,u)\|_{\infty,[\eta]^{-1}} &= \|\Phi(Ax_1+Bu) - \Phi(Ax_2+Bu)\|_{\infty,[\eta]^{-1}} \\ &\le \|(Ax_1+Bu) - (Ax_2+Bu)\|_{\infty,[\eta]^{-1}} \\ &\le \|A(x_1-x_2)\|_{\infty,[\eta]^{-1}} \le \|A\|_{\infty,[\eta]^{-1}}\|x_1-x_2\|_{\infty,[\eta]^{-1}},\end{aligned}$$

proving the formula for $\mathsf{Lip}_x(\mathsf{N}) = \|A\|_{\infty,[\eta]^{-1}}$. The proof of the formula $\mathsf{Lip}_u(\mathsf{N}) = \|B\|_{(\infty,[\eta]^{-1}),\mathcal{U}}$ is essentially identical. Finally, if each ϕ_i is differentiable then we compute

$$\begin{aligned}\mathrm{diagL}(\mathsf{N}) &= \min_{i\in\{1,\dots,n\}} \inf_{x\in\mathbb{R}^n u\in\mathbb{R}^r} D\mathsf{N}_{ii}(x,u) = \min_{i\in\{1,\dots,n\}} \inf_{x\in\mathbb{R}^n u\in\mathbb{R}^r} \phi_i'((Ax+Bu)_i)A_{ii} \\ &\le \min_{i\in\{1,\dots,n\}} \begin{cases} 0, & \text{if } A_{ii} > 0 \\ A_{ii}, & \text{if } A_{ii} \le 0 \end{cases} = \min_{i\in\{1,\dots,n\}} (A_{ii})_-,\end{aligned} \tag{4.24}$$

because of the properties of the activation functions. Now suppose that there exists $i \in \{1,\dots,n\}$ such that ϕ_i is not differentiable. Using Theorem 4.7.1(ii) with $\mathsf{G} = A$, $\mathrm{diagL}(\mathsf{N})$ is chosen to be equal to be $\mathrm{diagL}(A)_-$ which in turn is equal to $\min_{i\in\{1,\dots,n\}}(A_{ii})_-$. □

Proof of Corollary 4.3.1. The results are immediate consequences of Theorem 4.2.2 (or more generally Theorem 4.7.1 for non-differentiable activation functions) and of the Lipschitz estimates in Theorem 4.3.1. □

4.8 Adversarial attacks on implicit neural networks

In this section, we study the effect of different adversarial attacks on the existing implicit network models as well as to the NEMON model.

4.8.1 Attack models

First, we review several attack models that are used in the literature to study the input-output resilience of neural networks. Each attack consists of a model for generating suitable perturbations of the test input data. Perturbations with respect to these attacks were generated using the Foolbox software package[5].

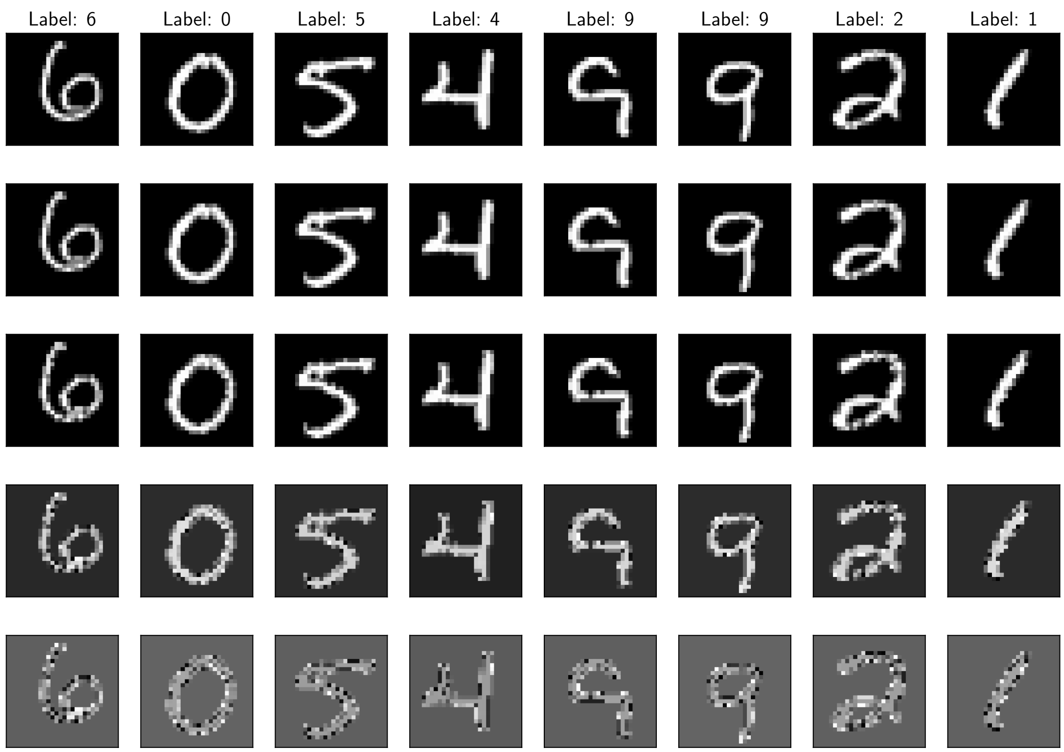


Figure 4.5: Images of MNIST handwritten digits perturbed by the continuous image inversion attack. For $i \in \{1, \dots, 5\}$, row i corresponds to an ℓ_∞ perturbation amplitude $\varepsilon = 0.1 \times (i-1)$. In other words, the top row has unperturbed images, the second row has images that is perturbed by an ℓ_∞ amplitude $\varepsilon = 0.1$, etc.

Continuous image inversion. The continuous image inversion attack is defined by:

$$U_{\text{adversarial}} = U + \varepsilon \operatorname{sign}\left(\tfrac{1}{2}\mathbb{1}_r \mathbb{1}_m^\top - U\right). \tag{4.25}$$

[5]The Foolbox implementation is licensed under the MIT License and is available at `https://github.com/bethgelab/foolbox.`

It is clear that this attack is independent of the neural network model. Plots of perturbed MNIST images under the continuous image inversion attack are shown in Figure 4.6. In Figure 4.2, the right plot compares the accuracy of the NEMON model, the implicit deep learning model [110], and the MON model [66] for $\varepsilon \in [0.0.5]$.

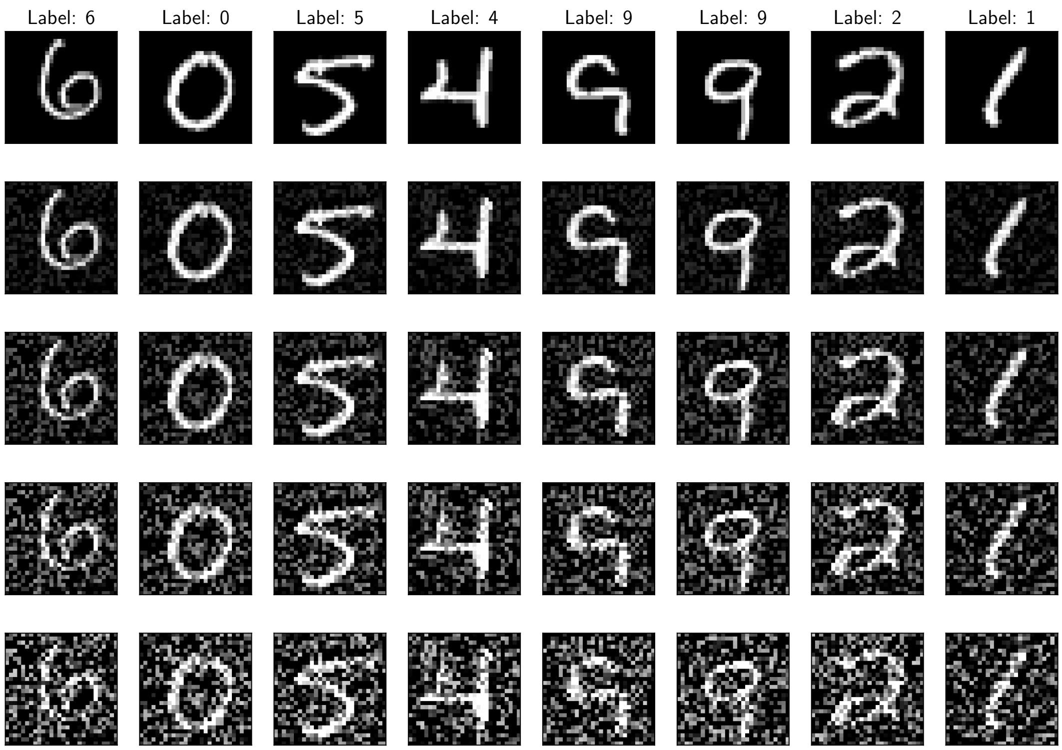


Figure 4.6: Images of MNIST handwritten digits as perturbed by uniform additive ℓ_∞ noise. For $i \in \{1, \dots, 5\}$, row i corresponds to an ℓ_∞ perturbation amplitude $\varepsilon = 0.2 \times (i-1)$. In other words, the top row has unperturbed images, the second row has images that is perturbed by an ℓ_∞ amplitude $\varepsilon = 0.2$, etc.

Uniform additive ℓ_∞-noise. For this attack, the test images are perturbed by an additive noise with ℓ_∞ magnitude sampled uniformly from the interval $[0,1]$. Plots of perturbed MNIST images under uniform additive ℓ_∞-noise are shown in Figure 4.6. Figure 4.7 shows scatter plots of the accuracy of the NEMON model, the implicit deep learning model, and the MON model over 1000 sample attacks.

Fast gradient sign method. Given input data $U \in \mathbb{R}^{r\times m}$ and output labels $Y \in \mathbb{R}^{q\times m}$, the fast gradient sign method (FGSM) generates adversarial inputs via the formula

$$U_{\text{adversarial}} = U + \varepsilon \, \text{sign}\Big(\frac{\partial \mathcal{L}}{\partial U}(Y, CX + DU)\Big), \tag{4.26}$$

where $\mathcal{L}$ is the loss function used to train the network and ε provides the ℓ_∞ amplitude of the perturbation. Plots of perturbed MNIST images under the FGSM are shown in Figure 4.8. Plots of accuracy versus ℓ_∞ perturbation under the FGSM are shown in Figure 4.9.

Projected gradient descent method. The projected gradient descent method (PGDM) can be thought of as perturbing the input with several steps of the FGSM. The PGDM attack can be defined for any norm, but for consistency, we reproduce it only for the ℓ_∞-norm. For the input data $U \in \mathbb{R}^{r\times m}$ and outputs $Y \in \mathbb{R}^{q\times m}$, PGDM defines the finite sequence of perturbations $\{\delta_k\}_{k=1}^M$ by

$$\delta_{k+1} = \mathrm{P}_{\overline{\mathcal{B}(\varepsilon)}}\left(\delta_k + \alpha \, \text{sign}\Big(\frac{\partial \mathcal{L}}{\partial U}(Y, CX + D(U + \delta_k)\Big)\right), \tag{4.27}$$

where M is some prescribed maximum number of steps, α is a stepsize, and $\mathrm{P}_{\overline{\mathcal{B}(\varepsilon)}}$ is the ℓ_2 orthogonal projection operator onto the entrywise ℓ_∞ closed ball with radius ε. This projection operator corresponds to clipping each entry of the matrix so that it is in the range $[-\varepsilon, \varepsilon]$. Then, the perturbed input is simply

$$U_{\text{adversarial}} = U + \delta_M.$$

Plots of perturbed MNIST images under the PGDM are shown in Figure 4.10. Plots of accuracy versus ℓ_∞ perturbation under the PGDM are shown in Figure 4.11.

4.8.2 Other methods to decrease the ℓ_∞ Lipschitz constant

Recall that the input-output Lipschitz constant of the model (4.7) with both $\|\cdot\|_{\mathcal{U}}$ and $\|\cdot\|_{\mathcal{Y}}$ equal to the ℓ_∞-norm is given by

$$\mathsf{Lip}_{u\to y} = \frac{\|B\|_{(\infty,[\eta]^{-1}),(\infty)}\|C\|_{(\infty),(\infty,[\eta]^{-1})}}{1-\mathrm{ReLU}(\mu_{\infty,[\eta]^{-1}}(A))} + \|D\|_{\infty,\infty}.$$

When input data, U, is perturbed, the perturbation is directly fed into the output Y via the output equation $Y = CX + DU$. For this reason, a simple change to attempt to minimize the effect of input perturbations on the output is to replace the DU term in the output equation by a static bias, i.e.,

$$Y = CX + b\mathbb{1}_m^\top,$$

where $b \in \mathbb{R}^q$. This simple modification to the model changes the input-output Lipschitz constant to

$$\mathsf{Lip}_{u\to y} = \frac{\|B\|_{(\infty,[\eta]^{-1}),(\infty)}\|C\|_{(\infty),(\infty,[\eta]^{-1})}}{1-\mathrm{ReLU}(\mu_{\infty,[\eta]^{-1}}(A))}.$$

Finally, another degree of freedom is the parameter $\gamma < 1$ in the constraint $\mu_{\infty,[\eta]^{-1}}(A) \leq \gamma$. In all previously shown experiments on MNIST, we selected $\gamma = 0.95$. From the expression for the input-output Lipschitz constant of the network (4.11), $\mu_{\infty,[\eta]^{-1}}(A) = 0.95$ leads to a small denominator, resulting in a relatively large input-output Lipschitz constant. A simple modification to moderate the Lipschitz constant is to impose $\mu_{\infty,[\eta]^{-1}}(A) \leq \epsilon$ for some small $\epsilon \geq 0$. This attempts to maximize the denominator in the expression for the Lipschitz constant.

For these modifications to the models, plots of accuracy versus ℓ_∞-perturbation generated by the FGSM are shown in Figure 4.12. In this figure, we set $\epsilon = 0.05$ for the NEMON models. For comparison, the well-posedness condition for MON is set to be

$\mu_2(A) \leq \epsilon$. We do not modify the condition $\|A\|_\infty \leq 0.95$ as imposing the constraint $\|A\|_\infty \leq \epsilon$ is overly restrictive and would result in a significant drop in accuracy.

4.8.3 Robustness of implicit neural networks on the MNIST dataset

In this section, we compare the performance of the NEMON model with $\mu_\infty(A) \leq 0.95$ to the implicit deep learning model with $\|A\|_\infty \leq 0.95$ and to the monotone operator equilibrium network (MON) with $I_n - \frac{1}{2}(A + A^\top) \succeq 0.05 I_n$ with respect to the attacks described in the previous subsection on the MNIST dataset.

For the continuous image inversion attack, Figure 4.2 shows the curves for accuracy versus ℓ_∞-amplitude of the perturbation. We observe that, compared to the NEMON model, the implicit deep learning model and the monotone operator equilibrium network (MON) have larger drops in accuracy for small perturbations. For the NEMON model, as λ increases, the accuracy at zero perturbation decreases. However, as λ increases, the overall robustness of NEMON improves as its accuracy does not decrease substantially even for large amplitudes of perturbation.

For uniform additive ℓ_∞-noise, scatter plots with accuracy versus ℓ_∞ amplitude of the perturbation are shown in Figure 4.7. We see that the NEMON model with $\lambda = 0$, the implicit deep learning model, and the MON model all perform comparably. The NEMON models with $\lambda = 10^{-1}$ and $\lambda = 10^{-2.5}$ both see improved robustness as their accuracy does not drop as noticeably with ℓ_∞ amplitude of the perturbation. Surprisingly, the NEMON model with $\lambda = 10^{-5}$ seems to be less robust than the NEMON model with $\lambda = 0$.

For the FGSM, Figure 4.9 shows the curves for accuracy versus ℓ_∞ amplitude of the perturbation. We see that the NEMON models with $\lambda = 10^{-5}$ and $\lambda = 10^{-4}$ are the least robust, followed by the NEMON model with $\lambda = 0$ and the MON. Only for $\lambda \in \{10^{-2.5}, 10^{-2}, 10^{-1}\}$ do we see an improvement in robustness for the NEMON model at the price of a decrease in nominal accuracy. Note that for the FGSM, each model experiences different perturbations.

For the PGDM, Figure 4.11 shows the curves for accuracy versus ℓ_∞ amplitude of the perturbation. We see that the results are comparable with the perturbation generated by the FGSM, with the exception that the implicit deep learning model now performs comparably with the monotone operator equilibrium model. Note that for the PGDM, each model experiences different perturbations.

Finally, we compare the performance of the models with the modification that the output equation is $Y = CX + b\mathbb{1}_m^\top$. Figure 4.12 shows the curves for accuracy versus ℓ_∞ amplitude of the FGSM perturbation for the NEMON model with $\mu_\infty(A) \leq 0.05$, the implicit deep learning model with $\|A\|_\infty \leq 0.95$, and the monotone operator equilibrium model with $I_n - \frac{1}{2}(A + A^\top) \succeq 0.05I_n$. For these modifications in the models, we see improvement in overall accuracy compared to original models of implicit networks (4.7) shown in Figure 4.9. Additionally, we observe comparable performance in the NEMON model with $\lambda = 0$ and the implicit deep learning model, with the MON performing slightly better than both. For the NEMON model with $\lambda = 10^{-4}$, the accuracy at zero perturbation is comparable to the NEMON model with $\lambda = 0$ and the overall robustness of the NEMON model to the FGSM attack is significantly improved. However, as λ increases, we see that the nominal accuracy and overall robustness of the NEMON models deteriorate.

4.8.4 Robustness of implicit neural networks on the CIFAR-10 dataset

In this section, we compare the performance of the NEMON model with $\mu_\infty(A) \leq 0$ to MON with $I_n - \frac{1}{2}(A + A^\top) \succeq I_n$ with respect to the FGSM attack described in the previous subsection on the CIFAR-10 dataset.

For the FGSM attack on the CIFAR-10 dataset, Figure 4.13 shows the accuracy versus the ℓ_∞ amplitude of the perturbation for the regularized and un-regularized NEMON model and the MON model. We observe that un-regularized NEMON model is more accurate than MON for all amplitudes of perturbation. For example, at ℓ_∞-perturbation equal to 0.1, the accuracy of un-regularized NEMON is 39% whereas the accuracy of MON at this attack amplitude is 35%. Moreover, the regularized NEMON with the regularization parameter $\lambda = 10^{-4}$ has a clean performance accuracy of 66% which is lower than the clean accuracy of both MON and the un-regularized NEMON. However, the regularized NEMON demonstrates a consistent improvement in accuracy for sizeable ℓ_∞-perturbations. For example, at an ℓ_∞-perturbation equal to 0.15, the accuracy of the regularized NEMON model is 29% whereas the accuracy of MON at this attack amplitude is 24%.

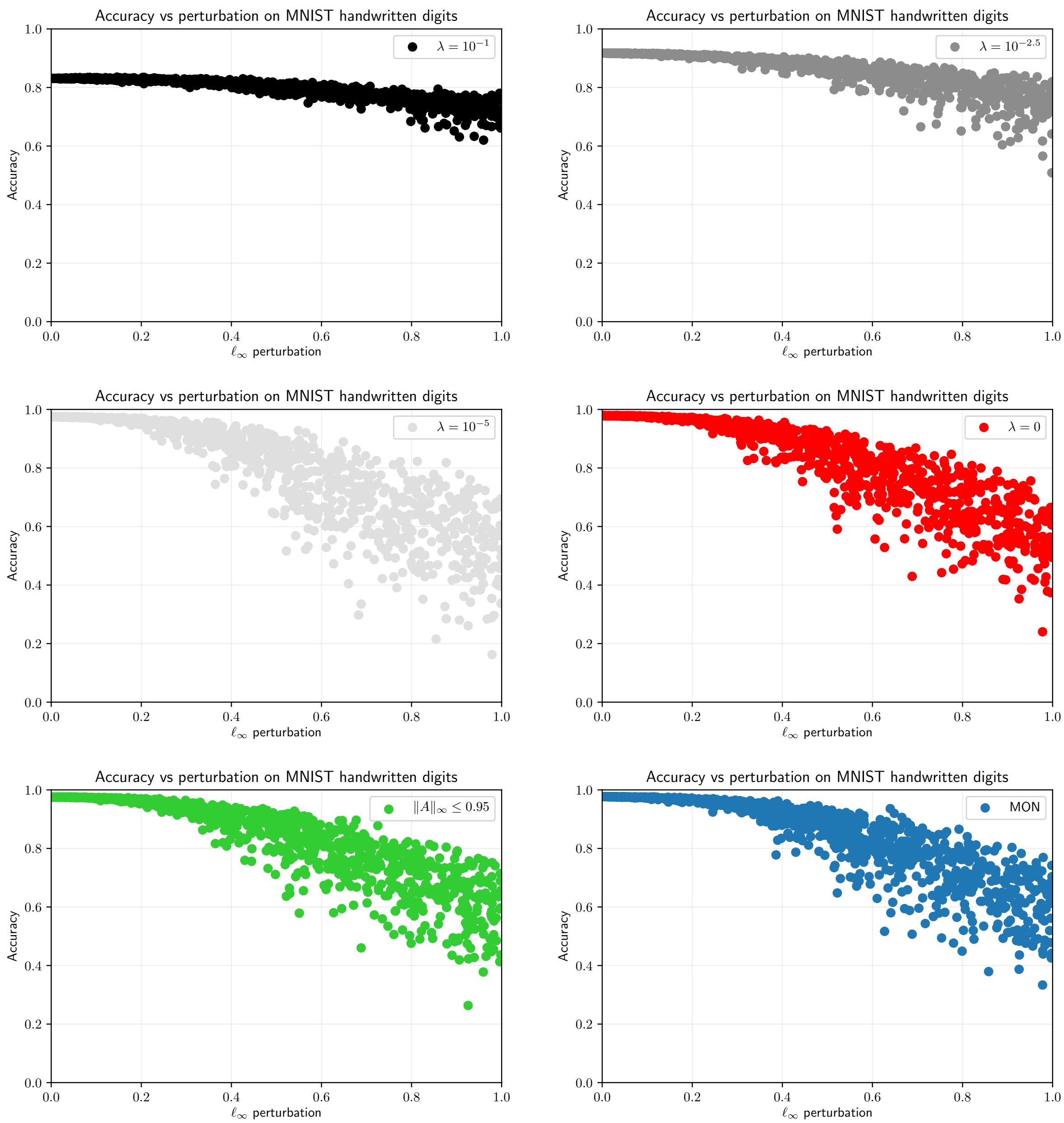


Figure 4.7: Scatter plots of accuracy versus ℓ_∞ perturbation as generated by uniform additive ℓ_∞ noise over 1000 trials. Plots are shown for the NEMON model $\mu_\infty(A) \leq 0.95$ with $\lambda \in \{10^{-1}, 10^{-2.5}, 10^{-5}, 0\}$, the implicit deep learning model $\|A\|_\infty \leq 0.95$, and the monotone operator equilibrium network (MON) with $I_n - \frac{1}{2}(A + A^\top) \succeq 0.05 I_n$.

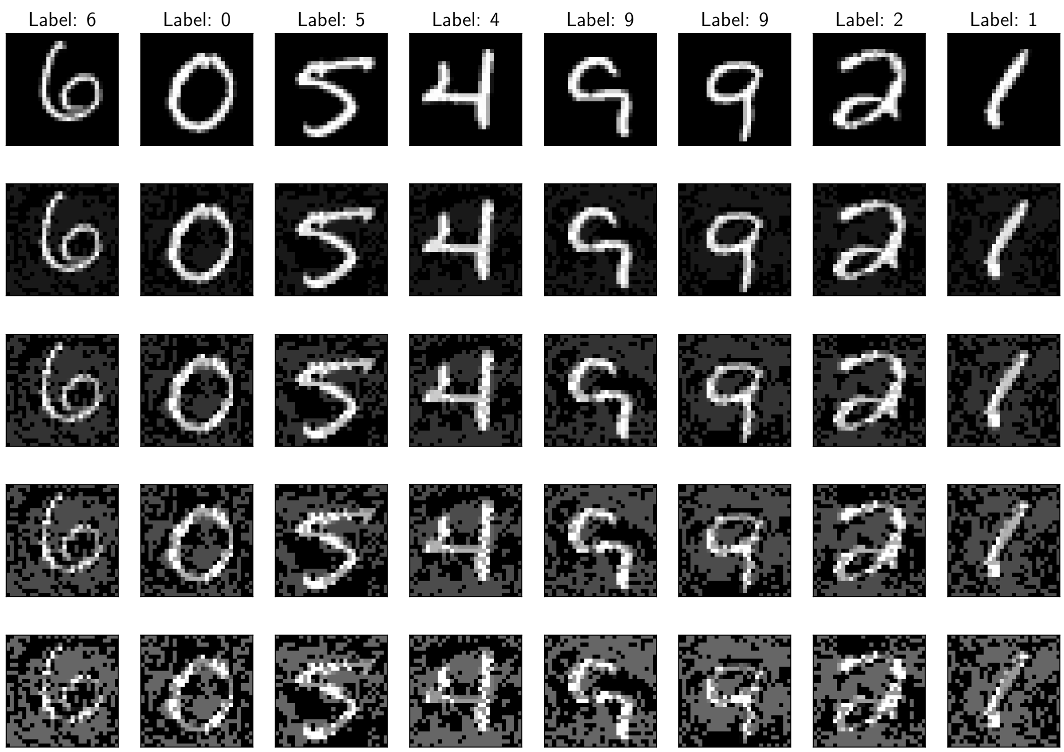


Figure 4.8: Images of MNIST handwritten digits as perturbed by the FGSM. For $i \in \{1, \ldots, 5\}$, row i corresponds to an ℓ_∞ perturbation amplitude $\varepsilon = 0.1 \times (i-1)$. In other words, the top row has unperturbed images, the second row has images that is perturbed by an ℓ_∞ amplitude $\varepsilon = 0.1$, etc.

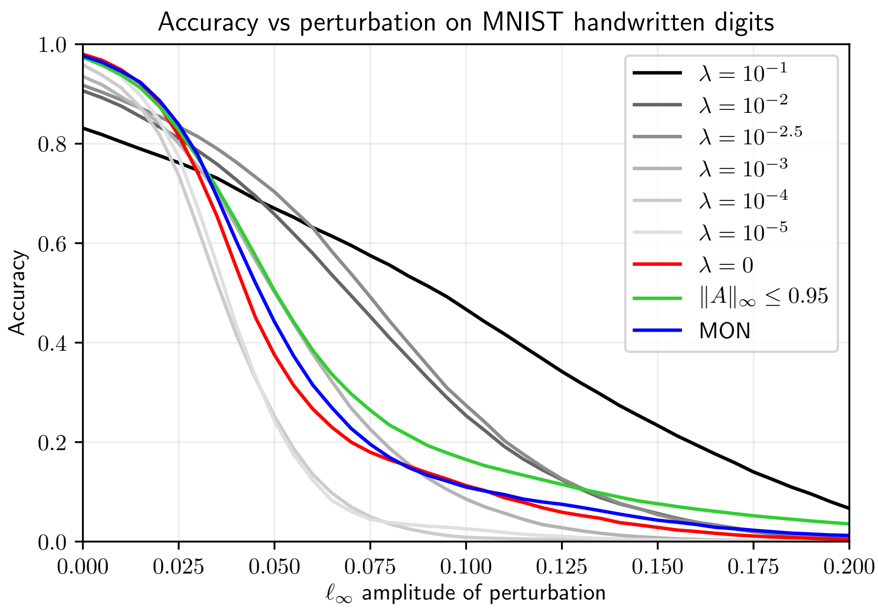


Figure 4.9: Plot of accuracy versus ℓ_∞ perturbation as generated by the FGSM for the NEMON model with $\mu_\infty(A) \leq 0.95$, the implicit deep learning model with $\|A\|_\infty \leq 0.95$, and MON with $I_n - \frac{1}{2}(A + A^\top) \succeq 0.05 I_n$.

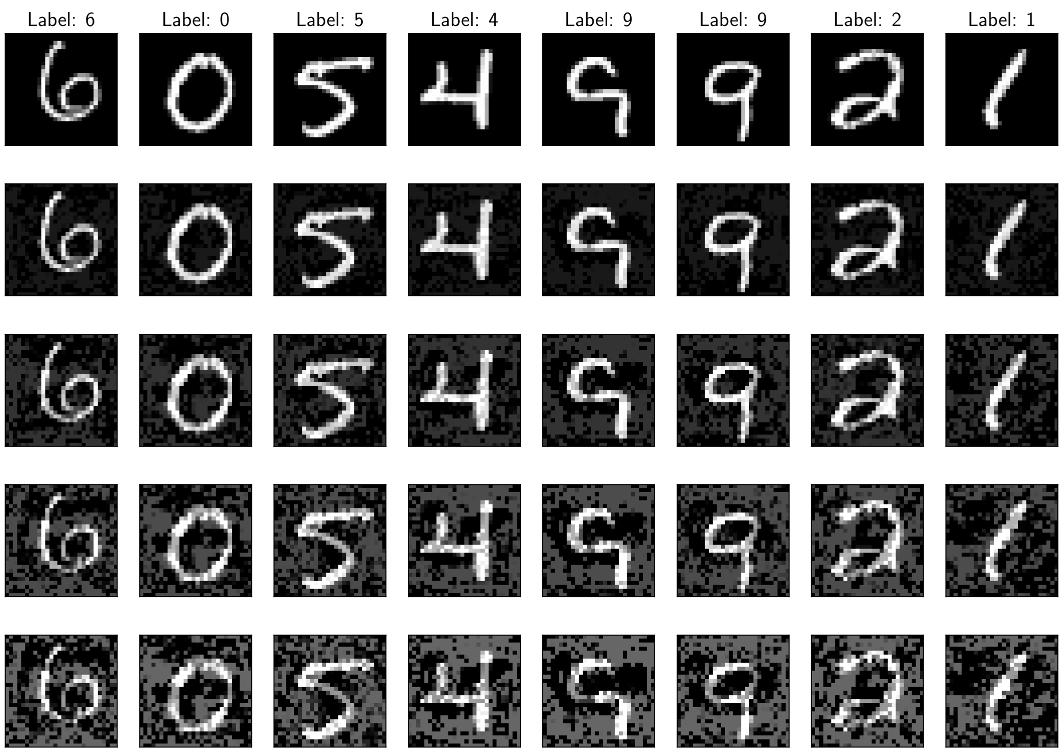


Figure 4.10: Images of MNIST handwritten digits as perturbed by the PGDM. For $i \in \{1, \dots, 5\}$, row i corresponds to an ℓ_∞ perturbation amplitude $\varepsilon = 0.1 \times (i-1)$. In other words, the top row has unperturbed images, the second row has images that is perturbed by an ℓ_∞ amplitude $\varepsilon = 0.1$, etc.

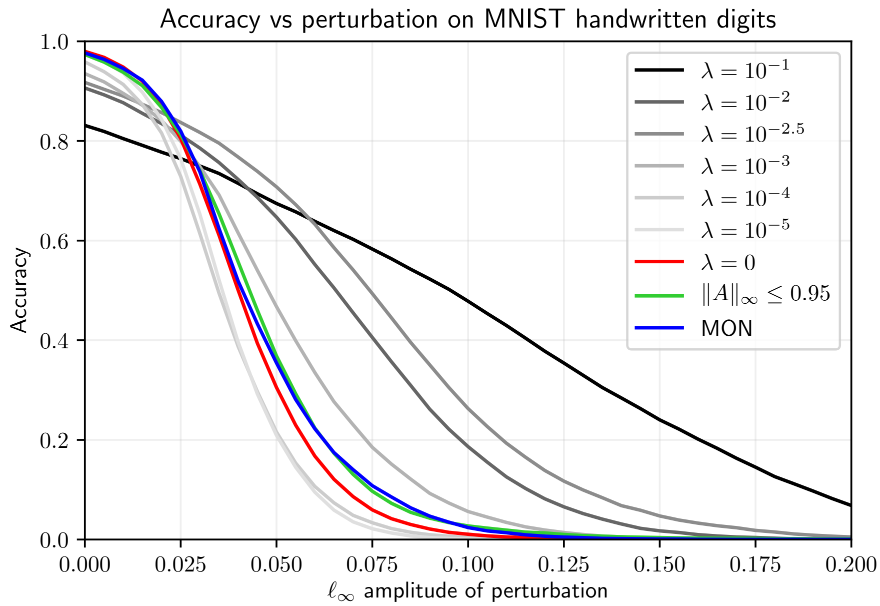


Figure 4.11: Plot of accuracy versus ℓ_∞ perturbation as generated by the PGDM for the NEMON model with $\mu_\infty(A) \leq 0.95$, the implicit deep learning model with $\|A\|_\infty \leq 0.95$, and MON with $I_n - \frac{1}{2}(A + A^\top) \succeq 0.05 I_n$.

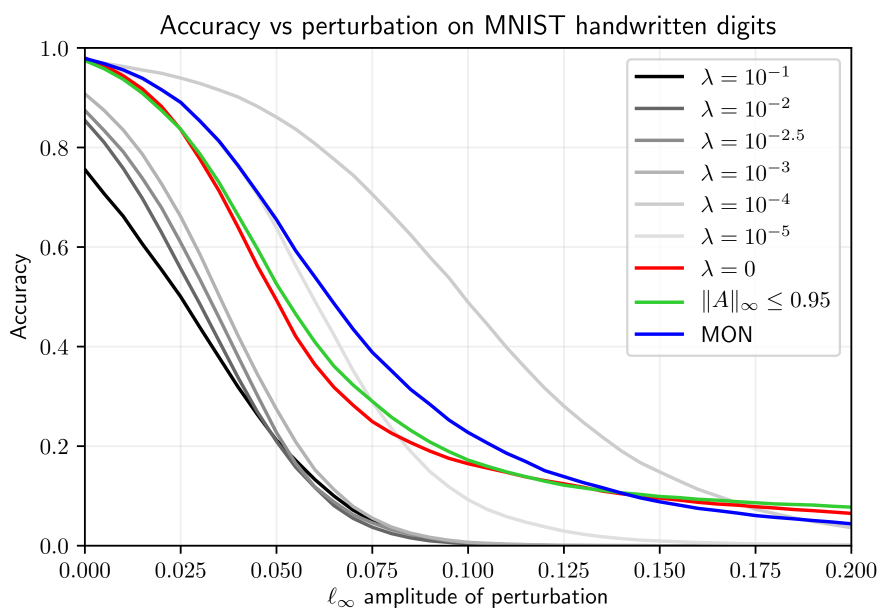


Figure 4.12: Plot of accuracy versus ℓ_∞ perturbation as generated by the FGSM for the NEMON model with $\mu_\infty(A) \leq 0.05$, the implicit deep learning model with $\|A\|_\infty \leq 0.95$, and MON with $I_n - \frac{1}{2}(A + A^\top) \succeq 0.05 I_n$. The output equation is $Y = CX + b\mathbb{1}_m^\top$.

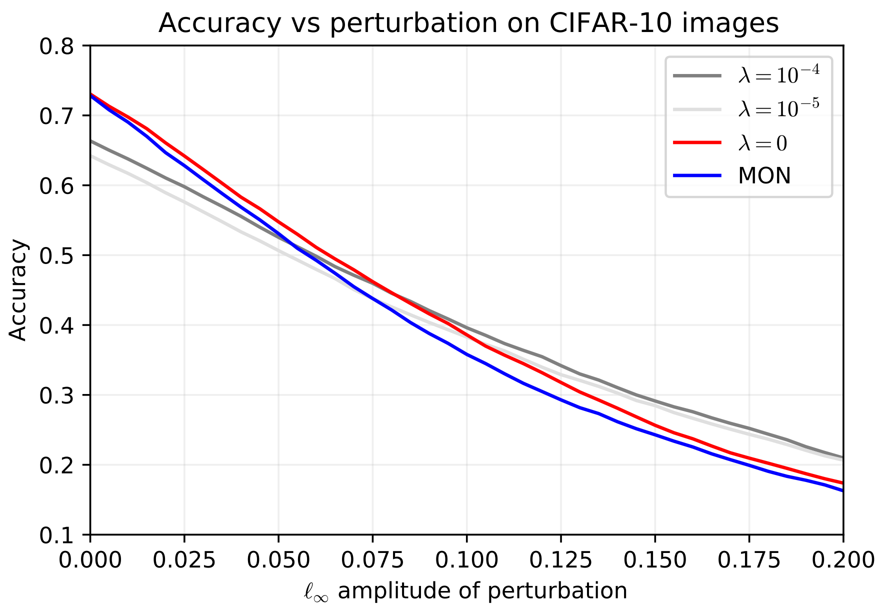


Figure 4.13: Plot of accuracy versus ℓ_∞ perturbation as generated by the FGSM for the NEMON model with $\mu_\infty(A) \leq 0$ and MON with $I_n - \frac{1}{2}(A + A^\top) \succeq I_n$.

Chapter 5

Non-Euclidean Contraction Analysis of Continuous-Time Neural Networks

An early version of this chapter was first published in the Proceedings of the American Control Conference, 2022 [135]. The chapter, as it appears, was first published in IEEE Transactions on Automatic Control [100].[1]

5.1 Introduction

Motivation from dynamical neuroscience and machine learning. Tremendous progress made in neuroscience research has produced new understanding of biological neural processes. Similarly, machine learning has become a key technology in modern society, with remarkable progress in numerous computational tasks. Much ongoing research focuses on artificial learning systems inspired by neuroscience that (i) generalize better, (ii) learn from fewer examples, and (iii) are increasingly energy-efficient. We argue that further progress in these disciplines hinges upon modeling, analysis, and computational challenges, some of which we highlight in what follows.

In **dynamical neuroscience**, several continuous-time neural network (NN) models are widely studied, including membrane potential models such as the Hopfield neural network [136] and firing-rate models [137]. Clearly, such models are simplifications of complex neural dynamics. For example, if $f(x)$ is an NN model of a neural circuit, the true dynamics may be better described by

$$\dot{x}(t) = f(x(t)) + g(x(t), x(t - \tau(t))), \tag{5.1}$$

where g captures model uncertainty and time-delays. In other words, to account for uncertainty in the system, the nominal dynamics $f(x)$ must exhibit robust stability with respect to unmodeled dynamics and delays. Additionally, central pattern generators (CPGs) are biological neural circuits that generate periodic signals and are the source of rhythmic motor behaviors such as walking and swimming. To properly model CPGs in NNs, a computational neuroscientist would need to ensure that, if an NN is interconnected with a CPG, then all trajectories of the NN converge to a unique stable limit cycle.

Machine learning scientists have widely adopted discrete-time NNs for pattern recognition and analysis of sequential data and much recent interest [109, 120, 114, 98] has focused on the closely-related class of implicit NNs. In particular, training implicit networks corresponds to solving fixed-point problems of the form

$$x = \Phi(Ax + Bu + b), \tag{5.2}$$

where x is the neural state variable, Φ is an activation function, A and B are synaptic weights, u is the input stimulus, and b is a bias term. Note that (i) the fixed point in equation (5.2) is the equilibrium point of the continuous-time NN $\dot{x} = -x + \Phi(Ax + Bu + b)$, (ii) the training problem requires the efficient computation of gradients of a given loss function with respect to model parameters; in turn, this computation can be cast again as a fixed-point problem. In other words, in the design of implicit NNs, it is essential to pick model weights in such a way that fixed-point equations have unique solutions for all possible inputs and activation functions, and fixed-points and corresponding gradients can be computed efficiently.

Finally, an additional challenge facing machine learning scientists is robustness to adversarial perturbations. Indeed, it is well-known [118] that artificial deep NNs are sensitive to adversarial perturbations: small input changes may lead to large output changes and loss in pattern recognition accuracy. One proposed remedy is to characterize the Lipschitz constants of these networks and use them as regularizers in the training process. This remedy leads to certifiable robustness bounds with respect to adversarial perturbations [130, 128]. In short, the input/output Lipschitz constants of NNs need to be tightly estimated, e.g., in the context of the fixed-point equation (5.2).

A contraction theory for neural networks. Motivated by the challenges arising in neuroscience and machine learning, this chapter aims to perform a *robust stability analysis* of continuous-time NNs and develop *optimization methods* for discrete-time NN models. Serendipitously, both these objectives can be simultaneously achieved through a contraction analysis for the NN dynamics.

For concreteness' sake, we briefly review how the aforementioned challenges are addressed by a contraction analysis. Infinitesimally contracting dynamics enjoy highly ordered *transient* and *asymptotic* behaviors: (i) initial conditions are forgotten and a certain distance between trajectories is monotonically vanishing [23], (ii) time-invariant systems admit a unique globally exponentially stable equilibrium with two natural Lyapunov functions (distance from the equilibrium and norm of the vector field) [23], (iii) periodic systems admit a unique globally exponentially stable periodic solution or, for systems with periodic inputs, each solution entrains to the periodic input [6], (iv) contracting vector fields enjoy highly robust behavior, e.g., see [138, 5], including (a) input-to-state stability, (b) finite input-state gain, (c) contraction margin with respect to unmodeled dynamics, and (d) input-to-state stability under delayed dynamics. Hence, the contraction rate is a natural measure/indicator of robust stability.

Regarding computational efficiency, our recent work [97, 98] shows how to design efficient fixed-point computation schemes for contracting systems (with respect to arbitrary and non-Euclidean ℓ_1/ℓ_∞ norms) in the style of monotone operator theory [74]. Specifically, for contracting dynamics with respect to a diagonally-weighted ℓ_1/ℓ_∞ norm, optimal step-sizes and convergence factors are given in [98, Theorem 2]. These results are directly applicable to the computation of fixed-points in implicit neural networks, as in equation (5.2). These step-sizes, however, depend on the contraction rate. Therefore, optimizing the contraction rate of the dynamics directly improves the convergence factor of the corresponding discrete algorithm.

Literature review. The dynamical properties of continuous-time NN models have been studied for several decades. Shortly after Hopfield's original work [136], control-theoretic ideas were proposed in [139]. Later, [140, 141, 142] obtained various version of the following result: Lyapunov diagonal stability of the synaptic matrix is sufficient, and in some cases necessary, for the existence, uniqueness, and global asymptotic stability of the equilibrium. More recently, [143] studies linear-threshold rate neural dynamics, where activation functions are piecewise-affine; it is shown that the dynamics have a unique equilibrium if and only if the synaptic matrix is a $\mathcal{P}$-matrix, a weaker condition than Lyapunov diagonal stability. Since checking this condition is NP-hard, more conservative conditions are provided as well. Beyond Lyapunov diagonal stability and $\mathcal{P}$-matrices, [133] is the earliest reference on the application of logarithmic norms and contraction-theoretic principles to Hopfield neural networks and provides results on ℓ_p logarithmic norms of the Jacobian for networks with smooth activation functions. Alternatively, [144] proposes a quasi-dominance condition on the synaptic matrix (in lieu of Lyapunov diagonal stability). Finally, similar to non-Euclidean contraction, [7] proposes the notion of the nonlinear measure of a map to study global asymptotic stability; this notion is closely related to the ℓ_1 one-sided Lipschitz constant of the Hopfield neural network vector field. A comprehensive survey on stability criteria for continuous-time NNs is available in [145].

The importance of non-Euclidean log norms in contraction theory is highlighted, for example, in [6, 14]. In the spirit of these works, the non-Euclidean contractivity of monotone Hopfield neural networks is studied in [146]; see also [147] for the non-Euclidean contractivity of Hopfield neural networks undergoing Hebbian learning.

Finally, Euclidean contractivity of continuous-time NNs has been studied, e.g., see the early reference [133], the related discussion in [114], and the recent work [148].

Contributions. This chapter contributes fundamental control-theoretic understanding to the study of artificial neural networks in machine learning and neuronal circuits in neuroscience, thereby building a hopefully useful bridge among these three disciplines.

Specifically, the chapter develops a comprehensive contraction theory for classes of continuous-time NN models. In order to develop this theory, we make several technical contributions on non-Euclidean logarithmic norms and nonsmooth contraction theory. To be specific, first, we obtain novel logarithmic norm results including (i) the quasiconvexity of the ℓ_1 and ℓ_∞ logarithmic norms with respect to diagonal weights and provide novel optimization techniques to compute optimal weights which yield larger contraction rates, (ii) logarithmic norm properties of principal submatrices of a matrix with respect to monotonic norms, and (iii) explicit formulas for the ℓ_1 and ℓ_∞ logarithmic norms under multiplicatively-weighted uncertainty, resulting in a maximization of the logarithmic norm over a matrix polytope. The matrix polytopes described in (iii) are of special interest since the Jacobian matrix of the Hopfield or firing-rate neural network vector field always lies inside this polytope. The formulas in (iii) generalize previous results [133, Theorem 3.8], [134, Lemma 3] and [98, Lemma 8].

Motivated by our non-Euclidean logarithmic norm results, we define M-Hurwitz matrices, i.e., matrices whose Metzler majorant is Hurwitz. We compare M-Hurwitz matrices with other classes of matrices including quasidominant, totally Hurwitz, and Lyapunov diagonally stable matrices.

Second, we provide a nonsmooth extension to contraction theory. We show that, for locally Lipschitz vector fields, the one-sided Lipschitz constant is equal to the essential supremum of the logarithmic norm of the Jacobian. This equality allows us to use our novel logarithmic norm results and apply them to NNs that have nonsmooth activation functions.

Finally, we apply our theoretical developments as we establish conditions for the non-Euclidean contractivity of multiple classes of recurrent neural circuits and nonlinear dynamical models, including Hopfield, firing rate, Persidskii, Lur'e, and others. We consider locally Lipschitz activation functions that satisfy an inequality of the form $d_1 \leq \dfrac{\phi(x)-\phi(y)}{x-y} \leq d_2$, for all $x \neq y \in \mathbb{R}$, where d_1 may be negative and d_2 may be infinite. Indeed, the importance of nonmonotonic activation functions is discussed in [149]. This class of activation functions is more general than all of the continuous activation functions mentioned in [145, Section II.B]. Thus, our non-Euclidean contraction framework allows for a more systematic framework for the analysis of these classes of NNs with fewer restrictions on the activation functions. For each model, we propose a linear program to characterize the optimal contraction rate and corresponding weighted non-Euclidean ℓ_1

or ℓ_∞ norm. In some special cases, we show that the linear program reduces to checking an M-Hurwitz condition. Our results simplify the computation of a common Lyapunov function over a polytope with 2^n vertices to a simple condition involving just 2 of its vertices or, in some cases, all the way to a closed form expression.

For each model, we demonstrate that the dynamics enjoy strong, absolute and total contractivity properties. In the spirit of absolute and connective stability, absolute contractivity means that the dynamics are contracting independently of the choice of activation function and connective stability means that the dynamics remain contracting whenever edges between neurons are removed. Total contractivity means that if the synaptic matrix is replaced by any principal submatrix, the resulting dynamics remain contracting. The process of replacing the nominal NN with a subsystem NN is referred to as "pruning" both in neuroscience and in machine learning.

A preliminary version of this work appeared in [135]. Compared to [135], this version (i) includes proofs of all technical results, (ii) provides closed-form worst-case log norms over a larger class of matrix polytopes in Lemma 5.4.3, (iii) studies a more general class of locally Lipschitz activation functions in Section 5.6, allowing for both nonmonotonic activation functions as well as activations that have unbounded derivative, (iv) has a complete characterization of contractivity of Hopfield and firing-rate neural networks with respect to both ℓ_1 and ℓ_∞ norms, (v) provides a novel sufficient (and nearly necessary) condition for the non-Euclidean contractivity of a Lur'e model with multiple nonlinearities in Theorem 5.6.11, and (vi) includes additional comparisons to Euclidean contractivity conditions in Remark 5.6.1 and to Lyapunov diagonal stability in Section 5.2.

Notation unique to this chapter. For two matrices A, B, we let A_{ij} be the entry in the i-th row and j-th column of A, $A \circ B$ be entrywise multiplication and $|A|$ be the entrywise absolute value. Vector (and matrix) inequalities of the form $x \leq y$ are entrywise. The *Metzler majorant* of a square matrix A is $\lceil A \rceil_{\mathrm{Mzr}} \in \mathbb{R}^{n \times n}$ is defined by

$$(\lceil A \rceil_{\mathrm{Mzr}})_{ij} = \begin{cases} A_{ii}, & \text{if } i = j \\ |A_{ij}|, & \text{if } i \neq j \end{cases}.$$

5.2 Preview of main contractivity results and advantages of a non-Euclidean analysis

To motivate the mathematical tools and analysis in Sections 5.4-5.6, we will showcase the main contractivity results for Hopfield and firing-rate neural networks under simplifying assumptions to provide a baseline for comparison to other standard stability conditions for these classes of neural networks.

The continuous-time Hopfield and firing-rate neural networks are the following two dynamical systems:

$$\dot{x} = -Cx + A\Phi(x) + u =: f_{\mathrm{H}}(x), \tag{5.3}$$

$$\dot{x} = -Cx + \Phi(Ax + u) =: f_{\mathrm{FR}}(x), \tag{5.4}$$

where $x \in \mathbb{R}^n$ is the state of the neural network (either a vector of membrane potentials or firing rates), $C \in \mathbb{R}^{n\times n}$ is a positive semidefinite diagonal matrix of dissipation rates, $A \in \mathbb{R}^{n\times n}$ is the synaptic matrix , $u \in \mathbb{R}^n$ is a constant external stimulus, and $\Phi\colon \mathbb{R}^n \to \mathbb{R}^n$ is an activation function which satisfies $\Phi(x) = (\phi_1(x_1), \dots, \phi_n(x_n))$. In the machine learning literature, such continuous-time NNs have been given the name neural ODEs [119].

For exposition's sake, we make the following standing assumptions throughout the remainder of this section:

Assumption 1. *(i)* $C = I_n$,

(ii) *the matrix* $\lceil A \rceil_{\mathrm{Mzr}}$ *is irreducible, and*

(iii) *each* ϕ_i *is continuously differentiable and satisfies* $0 \leq \phi_i'(x) \leq 1$ *for all* $x \in \mathbb{R}$.

Under these assumptions, we can state our main results compactly:

Proposition 5.2.1. *Consider the Hopfield and firing-rate neural networks* (5.3) *and* (5.4) *satisfying Assumption 1, suppose* $\alpha(\lceil A \rceil_{\mathrm{Mzr}}) < 1$, *and define* $c = 1 - \max\{\alpha(\lceil A \rceil_{\mathrm{Mzr}}, 0)\}$. *Then*

(i) *the Hopfield neural network is contracting with rate* $c > 0$, *i.e., any two trajectories* $x_1(\cdot), x_2(\cdot)$ *of* (5.3) *satisfy*

$$\|x_1(t) - x_2(t)\|_{1,[\eta]} \leq e^{-ct}\|x_1(0) - x_2(0)\|_{1,[\eta]},$$

for all $t \geq 0$, *where* $\eta \in \mathbb{R}^n_{>0}$ *is the dominant left eigenvector of the Metzler matrix* $\lceil A \rceil_{\mathrm{Mzr}}$.

(ii) *the firing-rate neural network is contracting with rate* $c > 0$, *i.e., any two trajectories* $x_1(\cdot), x_2(\cdot)$ *of* (5.4) *satisfy*

$$\|x_1(t) - x_2(t)\|_{\infty,[\xi]^{-1}} \leq e^{-ct}\|x_1(0) - x_2(0)\|_{\infty,[\xi]^{-1}},$$

for all $t \geq 0$ *where* $\xi \in \mathbb{R}^n_{>0}$ *is the dominant right eigenvector of the Metzler matrix* $\lceil A \rceil_{\mathrm{Mzr}}$.

In particular, under Assumption 1 and $\alpha(\lceil A \rceil_{\mathrm{Mzr}}) < 1$ (or equivalently $\alpha(\lceil -I_n + A \rceil_{\mathrm{Mzr}}) < 0$), for each $u \in \mathbb{R}^n$, both the Hopfield and firing-rate neural networks have unique globally exponentially stable equilibria and thus the condition $\alpha(\lceil -I_n + A \rceil_{\mathrm{Mzr}}) <$

0 provides a novel sufficient condition for the existence of a unique globally exponential stable equilibrium along with many additional robustness properties offered by contracting systems such as robustness to uncertainties and entrainment to periodic inputs.

Although in this chapter we primarily study the continuous-time NNs (5.3) and (5.4), we remark that many results apply to classes of discrete-time NNs as well. Specifically, given a continuous-time NN, $\dot{x} = f_{\mathrm{NN}}(x)$, which is contracting, the forward Euler discretization of the continuous-time NN with stepsize $h > 0$ yields a residual neural network

$$x_{k+1} = x_k + h f_{\mathrm{NN}}(x_k), \tag{5.5}$$

which is contracting in the sense of the Banach fixed point theorem for sufficiently small h (see, e.g., [97, Theorem 8]). For recent results on contraction for a different class of discrete-time NNs, we refer to [150].

The condition $\alpha(\lceil -I_n + A \rceil_{\mathrm{Mzr}}) < 0$ is different from the well-known result that Lyapunov diagonal stability (LDS) of $-I_n + A$, i.e., existence of a vector $\eta \in \mathbb{R}^n_{>0}$ satisfying

$$[\eta](-I_n + A) + (-I_n + A)^\top [\eta] \prec 0, \tag{5.6}$$

implies the existence of a unique globally asymptotically stable equilibrium point for the Hopfield neural network [142]. Moreover, the condition $\alpha(\lceil -I_n + A \rceil_{\mathrm{Mzr}}) < 0$ is stronger than LDS of $-I_n + A$, which we prove in Lemma 5.3.2, yet it implies the stronger property of contractivity.

Beyond LDS, an alternative way to establish the stability of the neural networks (5.3) and (5.4) is via absolute stability analysis of Lur'e systems and methods via quadratic Lyapunov functions. These methods are typically based upon linear matrix inequalities (LMIs), see, e.g., [37, 151] and the discussion in [145, Section I.V.]. Compared to these classical approaches, establishing contractivity with respect to diagonally-weighted ℓ_1 or ℓ_∞ norms provides both computational and practical advantages, which we highlight in the following paragraphs.

Computational benefits. In the non-Euclidean contraction analysis of many classes of neural networks, contractivity is checked either via linear programming or, in some simpler instances, the stability of appropriate Metzler matrices. As argued in [152], from a computational point of view, both of these tests are more scalable than LMIs are. Indeed, there exist efficient algorithms for computing Perron eigenvalues and eigenvectors for irreducible Metzler matrices [153].

Practical benefits. Compared to stability with respect to a quadratic Lyapunov function, there are also practical advantages to establishing contractivity with respect to diagonally-weighted ℓ_1 and ℓ_∞ norms. These benefits include (i) the ℓ_1 norm (respectively, the ℓ_∞ norm) is well suited for systems with conserved quantities (respectively, systems with translation invariance), e.g., see the theory of weakly contracting and monotone systems in [154, Chapter 4], (ii) in machine learning, analysis of the adversarial robustness of a NN often needs to be performed in a non-Euclidean norm, because NNs are known to be vulnerable to small disturbances as measured in the ℓ_∞ norm [75], and (iii) contractivity with respect to non-Euclidean norms ensures robustness with respect to edge removals and structural perturbations, e.g., see the notion of connective stability in [155].

To elaborate on point (iii) in the previous paragraph, in continuous-time NNs such as the Hopfield and firing rate neural networks (5.3)-(5.4), the synaptic matrix A defines a graph structure whereby there is an outgoing synapse from neuron j to neuron i provided that $A_{ij} \neq 0$. As we will show in Corollary 5.4.6, if the neural network is contracting with respect to a diagonally-weighted ℓ_1 or ℓ_∞ norm, it is *connectively contracting*. Specifically, the removal of any edge[2] or neuron from the graph[3] yields a new neural network that remains contracting with a rate greater than or equal to the rate of contraction of the original neural network. Note that this property is not enjoyed by stability conditions requiring a matrix to be LDS. Indeed, for the Hopfield neural network (5.3), consider

$$A = \begin{bmatrix} 0 & -1 & 1 \\ 1 & 0 & 15 \\ -1 & -15 & 0 \end{bmatrix}, \quad \tilde{A} = \begin{bmatrix} 0 & -1 & 1 \\ 1 & 0 & 15 \\ -1 & 0 & 0 \end{bmatrix}.$$

Note that A satisfies (5.6) with $\eta = \mathbb{1}_n$ so $-I_n + A$ is LDS and thus the Hopfield neural network is stable. However, zeroing out A_{32} yields $\tilde{A}$ which verifies $\alpha(-I_n + \tilde{A}) > 0$, so the resulting NN is not absolutely stable.

In the following sections, we introduce additional mathematical tools to prove Proposition 5.2.1 under assumptions weaker than those listed in Assumption 1. Specifically, we (i) relax item (i) to C which is diagonal and positive semidefinite, (ii) relax item (ii) to also study $\lceil A \rceil_{\mathrm{Mzr}}$ which may be reducible, and (iii) relax item (iii) to study nonsmooth activation functions which may be nonmonotonic and may have unbounded slope. See Theorems 5.6.1, 5.6.3, and 5.6.5 for these results. Beyond the proof of a more general version of Proposition 5.2.1, we also establish ℓ_∞ contractivity of the Hopfield neural network in Theorem 5.6.2, the ℓ_1 contractivity of the firing-rate neural network in Theorem 5.6.4, and study the contractivity of other classes of neural networks in Section 5.6.4. In the interest of readability, we postpone proofs of most technical results to Appendix 5.8 and include proofs regarding contractivity of classes of NNs in the main body of the text.

[2]Removing an edge corresponds to zeroing a non-diagonal entry of A.
[3]Removing the i-th neuron corresponds to removing the i-th row and column of A.

5.3 Review of relevant matrix analysis

5.3.1 Useful log norms characterizations for diagonally-weighted norms

For diagonally weighted ℓ_1, ℓ_∞, and ℓ_2 norms with $\eta \in \mathbb{R}^n_{>0}$,

$$\begin{aligned}
\mu_{1,[\eta]}(A) &= \max_{i\in\{1,\dots,n\}} A_{ii} + \sum\nolimits_{j=1,j\neq i}^{n} \frac{\eta_j}{\eta_i}|A_{ji}| \\
&= \min\{b \in \mathbb{R} \mid \lceil A\rceil_{\mathrm{Mzr}}^\top \eta \leq b\eta\}, \\
\mu_{\infty,[\eta]^{-1}}(A) &= \max_{i\in\{1,\dots,n\}} A_{ii} + \sum\nolimits_{j=1,j\neq i}^{n} \frac{\eta_j}{\eta_i}|A_{ij}| \\
&= \min\{b \in \mathbb{R} \mid \lceil A\rceil_{\mathrm{Mzr}} \eta \leq b\eta\}, \\
\mu_{2,[\eta]^{1/2}}(A) &= \min\{b \in \mathbb{R} \mid [\eta]A + A^\top[\eta] \preceq 2b[\eta]\}.
\end{aligned}$$

The following result is due to [156] and [61, Lemma 3].

Lemma 5.3.1 (Optimal diagonally-weighted log norms for Metzler matrices)**.** *Given a Metzler matrix $M \in \mathbb{R}^{n\times n}$, $p \in [1,\infty]$, and $\delta > 0$, define $\eta_{M,p,\delta} \in \mathbb{R}^n_{>0}$ by*

$$\eta_{M,p,\delta} = \left(\frac{w_1^{1/p}}{v_1^{1/q}}, \dots, \frac{w_n^{1/p}}{v_n^{1/q}}\right), \tag{5.7}$$

where $q \in [1,\infty]$ is defined by $1/p + 1/q = 1$ (with the convention $1/\infty = 0$) and where v and $w \in \mathbb{R}^n_{>0}$ are the right and left dominant eigenvectors of the irreducible Metzler matrix $M + \delta\mathbb{1}_n\mathbb{1}_n^\top$ (whose existence is guaranteed by the Perron-Frobenius Theorem). Then for each $\varepsilon > 0$ there exists $\delta > 0$ such that

(i) $\alpha(M) \leq \mu_{p,[\eta_{M,p,\delta}]}(M) \leq \alpha(M) + \varepsilon$,

(ii) *if M is irreducible, then* $\alpha(M) = \mu_{p,[\eta_{M,p,0}]}(M)$.

Lemma 5.3.1 also ensures that for Metzler matrices $M \in \mathbb{R}^{n\times n}$, $\inf_{\eta\in\mathbb{R}^n_{>0}} \mu_{p,[\eta]}(M) = \alpha(M)$ for every $p \in [1,\infty]$.

5.3.2 Classes of matrices

We say a matrix $A \in \mathbb{R}^{n\times n}$ is

(i) *Hurwitz stable*, denoted by $A \in \mathcal{H}$, if $\alpha(A) < 0$,

(ii) *totally Hurwitz*, denoted by $A \in \mathcal{TH}$, if all principal submatrices of A are Hurwitz stable,

(iii) *Lyapunov diagonally stable (LDS)*, denoted by $A \in \mathcal{LDS}$, if there exists a $\eta \in \mathbb{R}^n_{>0}$ such that $\mu_{2,[\eta]^{1/2}}(A) < 0$, and

(iv) *M-Hurwitz stable*, denoted by $A \in \mathcal{MH}$, if $\alpha(\lceil A\rceil_{\mathrm{Mzr}}) < 0$.

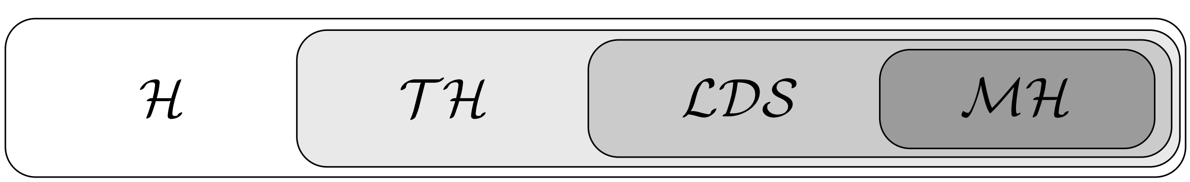
A matrix $A \in \mathbb{R}^{n\times n}$ is *quasidominant* [157] if there exists a vector $\eta \in \mathbb{R}^n_{>0}$ such that

$$\eta_i A_{ii} > \sum\nolimits_{j=1,j\neq i}^{n} \eta_j |A_{ij}|, \quad \text{for all } i \in \{1,\dots,n\}.$$

This is equivalent to $\lceil -A\rceil_{\mathrm{Mzr}}\eta < \mathbb{0}_n$, which, in turn, is equivalent (see, for example, [158, Theorem 15.17]) to the inequality $\alpha(\lceil -A\rceil_{\mathrm{Mzr}}) < 0$, i.e., $-A \in \mathcal{MH}$.

The following results are essentially known in the literature, but not collected in a unified manner.

Lemma 5.3.2 (Inclusions for classes of matrices)**.** *$(A \in \mathcal{MH})$ implies $(A \in \mathcal{LDS})$, $(A \in \mathcal{LDS})$ implies $(A \in \mathcal{TH})$, and $(A \in \mathcal{TH})$ implies $(A \in \mathcal{H})$.*



We show that the counter-implications in Lemma 5.3.2 do not hold.

Example 5. *(i) ($A \in \mathcal{LDS} \not\Rightarrow A \in \mathcal{MH}$) The matrix* $A = \begin{bmatrix} -1 & -1 \\ 2 & -1 \end{bmatrix}$ *satisfies* $\mu_2(A) = -0.5$*, so* $A \in \mathcal{LDS}$*. However,* $\alpha(\lceil A \rceil_{\mathrm{Mzr}}) = \sqrt{2} - 1 > 0$*, so* $A \notin \mathcal{MH}$*.*

(ii) ($A \in \mathcal{TH} \not\Rightarrow A \in \mathcal{LDS}$) is proved in [159, Remark 4].

(iii) ($A \in \mathcal{H} \not\Rightarrow A \in \mathcal{TH}$) The matrix $A = \begin{bmatrix} 1 & 1 \\ -4 & -3 \end{bmatrix}$ *satisfies* $\alpha(A) = -1$*, so* $A \in \mathcal{H}$*. However,* $A \notin \mathcal{TH}$ *since it has a positive diagonal entry.*

In the context of Proposition 5.2.1, the condition $\alpha(\lceil -I_n + A \rceil_{\mathrm{Mzr}}) < 0$, is equivalent to asking $-I_n + A \in \mathcal{MH}$, which, in light of Lemma 5.3.2, implies $-I_n + A \in \mathcal{LDS}$, which was the previously known sufficient condition for asymptotic stability of a unique fixed point of the Hopfield NN.

5.4 Novel log norm results

5.4.1 Optimizing non-Euclidean log norms

First, we provide novel results on optimizing diagonal weights for ℓ_1 and ℓ_∞ log norms and provide computational methods to compute these weights.

Theorem 5.4.1 (Quasiconvexity of μ with respect to diagonal weights)**.** *For fixed* $A \in \mathbb{R}^{n\times n}$*, consider the maps from* $\mathbb{R}^n_{>0}$ *to* $\mathbb{R}$ *defined by*

$$\eta \mapsto \mu_{1,[\eta]}(A), \qquad \eta \mapsto \mu_{\infty,[\eta]^{-1}}(A). \tag{5.8}$$

Then

(i) The maps in (5.8) are continuous, quasiconvex, and their sublevel sets are polytopes.

(ii) Minimizing the maps in (5.8) may be executed via the minimization problems

$$\begin{aligned} \inf_{b\in\mathbb{R},\eta\in\mathbb{R}^n_{>0}} \quad & b \\ s.t. \quad & \lceil A \rceil_{\mathrm{Mzr}}^\top \eta \leq b\eta, \end{aligned} \tag{5.9}$$

for $\mu_{1,[\eta]}(A)$ and

$$\begin{aligned} \inf_{b\in\mathbb{R},\eta\in\mathbb{R}^n_{>0}} \quad & b \\ s.t. \quad & \lceil A \rceil_{\mathrm{Mzr}} \eta \leq b\eta, \end{aligned} \tag{5.10}$$

for $\mu_{\infty,[\eta]^{-1}}(A)$.

Remark 5.4.1. *If $\lceil A \rceil_{\mathrm{Mzr}}$ is irreducible, by Lemma 5.3.1 the optimization problems in (5.9) and (5.10) attain their minima so that the* inf *may be replaced by* min*. Then the problems may be solved by a bisection on $b \in [-\|A\|, \|A\|]$, where each step of the algorithm is a linear program (LP) in η.*

Moreover, the minima in (5.9) and (5.10) exist for many types of reducible matrices, e.g. when $\lceil A \rceil_{\mathrm{Mzr}}$ is a block-diagonal matrix whose diagonal blocks are irreducible.

In the event that the minimum does not exist, let $b^\star$ be the infimum value of either (5.9) or (5.10). Then for any $\epsilon > 0$, one can still apply the bisection algorithm to find a choice of η such that $\mu_{[\eta]}(A) \leq b^\star + \epsilon$, where $\mu_{[\eta]}(\cdot)$ denotes either $\mu_{1,[\eta]}(\cdot)$ or $\mu_{\infty,[\eta]^{-1}}(\cdot)$.

Remark 5.4.2. *Notice that the sets of feasible vectors η in (5.9) and (5.10) are polyhedral cones, that is, if η is feasible, then $\theta\eta$ is also feasible for all $\theta > 0$. Hence, the constraint $\eta \in \mathbb{R}^n_{>0}$ can be replaced by an equivalent constraint $\eta \in [\varepsilon, \infty)^n$, where $\varepsilon > 0$ is an arbitrary constant. This can be useful, because LP solvers usually handle problems with non-strict inequalities.*

Next, we provide closed-form expressions for ℓ_1 and ℓ_∞ log norms over a certain polytopes of matrices. Polytopes of interest are defined by a nominal matrix multiplied by a diagonally-weighted uncertainty and shifted by an additive diagonal matrix. Such matrix polytopes arise in tests verifying the contractivity of Hopfield and firing-rate NNs and will play a critical role in our analysis.

This insert corresponds to Lemma 5.4.3. For $A \in \mathbb{R}^{n\times n}$, $c \in \mathbb{R}^n$, $d_1 \leq d_2 \in \mathbb{R}$, $\overline{d} = \max\{|d_1|, |d_2|\}$, and $\eta \in \mathbb{R}^n_{>0}$,

$$\max_{d\in[d_1,d_2]^n} \mu_{\infty,[\eta]}([c] + [d]A) = \max\big\{\mu_{\infty,[\eta]}([c] + d_1 A), \mu_{\infty,[\eta]}([c] + d_2 A)\big\}, \tag{5.11}$$

$$\max_{d\in[d_1,d_2]^n} \mu_{1,[\eta]}([c] + A[d]) = \max\big\{\mu_{1,[\eta]}([c] + d_1 A), \mu_{1,[\eta]}([c] + d_2 A)\big\}, \tag{5.12}$$

$$\begin{aligned}\max_{d\in[d_1,d_2]^n} \mu_{\infty,[\eta]}([c] + A[d]) = \max\big\{&\mu_{\infty,[\eta]}([c] + \overline{d}A - (\overline{d} - d_1)(I_n \circ A)),\\ &\mu_{\infty,[\eta]}([c] + \overline{d}A - (\overline{d} - d_2)(I_n \circ A))\big\},\end{aligned} \tag{5.13}$$

$$\begin{aligned}\max_{d\in[d_1,d_2]^n} \mu_{1,[\eta]}([c] + [d]A) = \max\big\{&\mu_{1,[\eta]}([c] + \overline{d}A - (\overline{d} - d_1)(I_n \circ A)),\\ &\mu_{1,[\eta]}([c] + \overline{d}A - (\overline{d} - d_2)(I_n \circ A))\big\}.\end{aligned} \tag{5.14}$$

Lemma 5.4.3 (Max value of ℓ_1/ℓ_∞ log norms under multiplicative scalings)**.** *Any $A \in \mathbb{R}^{n\times n}$, $c \in \mathbb{R}^n$, $d_1 \leq d_2 \in \mathbb{R}$, and $\eta \in \mathbb{R}^n_{>0}$ satisfy formulas (5.11)-(5.14) where $\overline{d} = \max\{|d_1|, |d_2|\}$.*

Recall that the log norm is a convex function and that the maximum value of a convex function over a polytope is achieved at one of the vertices of the polytope. In the special case in Lemma 5.4.3, formulas (5.11)-(5.14) ensure that one needs to check only 2 vertices of the polytope, rather than 2^n.

Finally, we show how the optimal diagonal weights that minimize the maximum value of the log norm of a matrix polytope as in Lemma 5.4.3 can be easily computed.

Corollary 5.4.4. *Let A, c, d_1, and d_2 be as in Lemma 5.4.3. Then for $\mu_{[\eta]}(\cdot)$ denoting either $\mu_{1,[\eta]}(\cdot)$ or $\mu_{\infty,[\eta]^{-1}}(\cdot)$ the minimax problems*

$$\inf_{\eta \in \mathbb{R}^n_{>0}} \max_{d \in [d_1, d_2]^n} \mu_{[\eta]}([c] + [d]A),$$

$$\inf_{\eta \in \mathbb{R}^n_{>0}} \max_{d \in [d_1, d_2]^n} \mu_{[\eta]}([c] + A[d]),$$

may each be solved by a bisection algorithm, each step of which is an LP.

Proof. The proof is an immediate consequence of the formulas (5.11)-(5.14) as well as the fact that a max of quasiconvex functions is quasiconvex. Therefore, a bisection algorithm similar to the one in Theorem 5.4.1(ii) may be used to compute the optimal η. $\square$

5.4.2 Monotonicity of diagonally-weighted log norms

Theorem 5.4.2 (Monotonicity of α and μ)**.** *For any $A \in \mathbb{R}^{n \times n}$*

(i) $\alpha(A) \leq \alpha(\lceil A \rceil_{\mathrm{Mzr}})$,

(ii) for all $p \in [1, \infty]$ and $\eta \in \mathbb{R}^n_{>0}$, we have $\mu_{p,[\eta]}(A) \leq \mu_{p,[\eta]}(\lceil A \rceil_{\mathrm{Mzr}})$, with equality holding for $p \in \{1, \infty\}$.

(iii) For $p \in \{1, \infty\}$,

$$\inf_{\eta \in \mathbb{R}^n_{>0}} \mu_{p,[\eta]}(A) = \alpha(\lceil A \rceil_{\mathrm{Mzr}}) \geq \alpha(A).$$

Theorem 5.4.2(iii) demonstrates that using diagonally-weighted ℓ_1 and ℓ_∞ log norms, the best bound one can achieve on $\alpha(A)$ is $\alpha(\lceil A \rceil_{\mathrm{Mzr}})$, which may be conservative. In the following example, we show that the ℓ_2 norm does not have the same conservatism. Despite the conservatism, Theorem 5.4.1 demonstrates that optimizing diagonal weights is computationally efficient, being an LP at every step of the bisection, while optimizing weights for the ℓ_2 norm is a semidefinite program at every step, which is more computationally challenging than an LP of similar dimension.

Example 6. *The matrix* $A_* = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$ *has eigenvalues* $\{1+\mathrm{i}, 1-\mathrm{i}\}$ *whereas* $\lceil A_* \rceil_{\mathrm{Mzr}}$ *has eigenvalues* $\{2, 0\}$*. Therefore,* $\alpha(A_*) = 1 < 2 = \alpha(\lceil A_* \rceil_{\mathrm{Mzr}})$*. Additionally,* $(A_* + A_*^\top)/2 = I_2 \implies \mu_2(A_*) = 1$ *and* $\mu_2(\lceil A_* \rceil_{\mathrm{Mzr}}) = 2$*.*

5.4.3 Log norms of principal submatrices

Given a matrix $A \in \mathbb{R}^n$ and a non-empty index set $\mathcal{I} \subset \{1, \dots, n\}$, let $A_{\mathcal{I}} \in \mathbb{R}^{|\mathcal{I}| \times |\mathcal{I}|}$ denote the *principal submatrix* obtained by removing the rows and columns of A which are not in $\mathcal{I}$. Next, given a non-empty $\mathcal{I} \subset \{1, \dots, n\}$, define the *zero-padding map* $\mathrm{pad}_{\mathcal{I}} \colon \mathbb{R}^{|\mathcal{I}|} \to \mathbb{R}^n$ as follows: $\mathrm{pad}_{\mathcal{I}}(y)$ is obtained by inserting zeros among the entries of y corresponding to the indices in $\{1, \dots, n\} \setminus \mathcal{I}$. For example, with $n = 3$ and $\mathcal{I} = \{1, 3\}$, we define $\mathrm{pad}_{\{1,3\}}(y_1, y_2) = (y_1, 0, y_2)$. Then it is easy to see that given a norm $\|\cdot\|$ on $\mathbb{R}^n$ and non-empty $\mathcal{I} \subset \{1, \dots, n\}$, the map $\|\cdot\|_{\mathcal{I}} \colon \mathbb{R}^{|\mathcal{I}|} \to \mathbb{R}_{\geq 0}$ defined by $\|y\|_{\mathcal{I}} = \|\mathrm{pad}_{\mathcal{I}}(y)\|$ is a norm on $\mathbb{R}^{|\mathcal{I}|}$.

Lemma 5.4.5 (Norm and log norm of principal submatrices)**.** *Assume* $\|\cdot\|$ *is monotonic, let* μ *and* $\mu_{\mathcal{I}}$ *denote the log norms associated to* $\|\cdot\|$ *and* $\|\cdot\|_{\mathcal{I}}$ *respectively. Any matrix* $A \in \mathbb{R}^{n \times n}$ *satisfies*

(i) $\|A_{\mathcal{I}}\|_{\mathcal{I}} \leq \|A\|$,

(ii) $\mu_{\mathcal{I}}(A_{\mathcal{I}}) \leq \mu(A)$,

(iii) if $\mu(A) < 0$, *then* $A \in \mathcal{TH}$.

Corollary 5.4.6. *Suppose* $A \in \mathcal{MH} \subset \mathbb{R}^{n\times n}$. *Then*

(i) $A_{\mathcal{I}} \in \mathcal{MH} \subset \mathbb{R}^{|\mathcal{I}|\times|\mathcal{I}|}$ *for every non-empty* $\mathcal{I} \subset \{1,\dots,n\}$ *and*

(ii) $A - A_{ij}\mathbb{e}_{ij} \in \mathcal{MH}$ *for all* $i,j \in \{1,\dots,n\}, i \neq j$, *where* $\mathbb{e}_{ij}$ *is a matrix with all zeros and unity in its* ij*-th entry.*

In the context of Proposition 5.2.1, since our sufficient condition for the contractivity of the Hopfield and firing-rate NNs is $-I_n + A \in \mathcal{MH}$, Corollary 5.4.6 implies that this sufficient condition implies total and structural contractivity, i.e., the removal of any neuron or edge from the neural network yields a new neural network that remains contracting.

5.5 Nonsmooth contraction theory

In this section we consider locally Lipschitz f and show that in this case, the definition of osL does not depend on the weak pairing and instead depends only on the norm through the log norm.

Theorem 5.5.1 (osL simplification for locally Lipschitz f)**.** *For* $f : U \to \mathbb{R}^n$ *locally Lipschitz on an open convex set,* $U \subseteq \mathbb{R}^n$. *Then for every* $c \in \mathbb{R}$ *the following statements are equivalent:*

(i) $\mathrm{osL}(f) \leq c$,

(ii) $\mu(Df(x)) \leq c$ *for almost every* $x \in U$.

Specifically, $\mathrm{osL}(f) = \operatorname{ess\,sup}_{x\in U} \mu(Df(x))$, *where* ess sup *denotes the essential supremum.*

Recall that $Df(x)$ exists for almost every $x \in U$ by Rademacher's theorem and thus the essential supremum ignores the Lebesgue measure zero set where Df doesn't exist.

Theorem 5.5.1 demonstrates that locally Lipschitz f enjoy a similar simplification in the osL definition as do continuously differentiable functions.

In neural network models, nonsmooth activation functions such as ReLU, LeakyReLU, and nonsmooth saturation functions are prevalent; Theorem 5.5.1 allows us to use standard log norm results to analyze these models. In other words, for a given continuous-time neural network dynamics $\dot{x} = f_{\mathrm{NN}}(x)$, with locally Lipschitz f_{NN}, to establish contractivity, it suffices to verify that $\mu(Df_{\mathrm{NN}}(x)) \leq -c$ for almost every x.

5.6 Contracting neural network dynamics

In this section, we prove Proposition 5.2.1 in greater generality. Specifically, we establish tight estimates for the one-sided Lipschitz constant for both the Hopfield and firing-rate NNs with respect to both diagonally-weighted ℓ_1 and ℓ_∞ norms. In instances where the one-sided Lipschitz constant is negative, we conclude that the neural network is strongly infinitesimally contracting. Beyond the Hopfield and firing-rate NNs, we also study the non-Euclidean contractivity of other classes of NNs.

5.6.1 One-sided Lipschitz characterization of Hopfield NNs

Recall the Hopfield NN dynamics, first introduced in [136]:

$$\dot{x} = -Cx + A\Phi(x) + u =: f_{\mathrm{H}}(x), \tag{5.15}$$

where $C \in \mathbb{R}^{n\times n}$ is a positive semi-definite diagonal matrix, $A \in \mathbb{R}^{n\times n}$ is arbitrary, $u \in \mathbb{R}^n$ is a constant input, and $\Phi\colon \mathbb{R}^n \to \mathbb{R}^n$ is an activation function. We make the following assumption on our activation functions:

Assumption 2 (Activation functions)**.** *Activation functions are locally Lipschitz and diagonal, i.e.,* $\Phi(x) = (\phi_1(x_1), \dots, \phi_n(x_n))$ *where each* $\phi_i \colon \mathbb{R} \to \mathbb{R}$ *satisfies*

$$\begin{aligned} d_1 &:= \inf_{x,y\in\mathbb{R}, x\neq y} \frac{\phi_i(x) - \phi_i(y)}{x - y} > -\infty, \\ d_2 &:= \sup_{x,y\in\mathbb{R}, x\neq y} \frac{\phi_i(x) - \phi_i(y)}{x - y}. \end{aligned} \tag{5.16}$$

When ϕ_i *satisfies* (5.16) *with finite* d_2*, we write* $\phi_i \in \mathrm{slope}[d_1, d_2]$*. If* $d_2 = \infty$*, we write* $\phi_i \in \mathrm{slope}[d_1, \infty]$*.*

Assumption 2 with finite d_2 implies that activation functions are (globally) Lipschitz and that $\phi_i'(x) \in [d_1, d_2]$ for almost every $x \in \mathbb{R}$. Many common activation functions satisfy these assumptions including ReLU, tanh and sigmoids. If, instead, $d_2 = \infty$, we can consider locally Lipschitz activation functions with unbounded slope including rectified polynomials $\phi(x) = \max\{x, 0\}^r$ for $r \in \mathbb{Z}_{\geq 0}$ which have been studied in [160]. Note that compared to Assumption 1, our activation functions do not need to be differentiable and are permitted more arbitrary bounds on their slopes.

The following theorem is the counterpart to Proposition 5.2.1(i) under more general assumptions.

Theorem 5.6.1 (ℓ_1 one-sided Lipschitzness of Hopfield neural network)**.** *Consider the Hopfield neural network* (5.15) *with each* $\phi_i \in \mathrm{slope}[d_1, d_2]$*. Then*

(i) for arbitrary $\eta \in \mathbb{R}^n_{>0}$*,* $\mathrm{osL}_{1,[\eta]}(f_{\mathrm{H}}) = \max\big\{\mu_{1,[\eta]}(-C + d_1 A), \mu_{1,[\eta]}(-C + d_2 A)\big\}$*.*

(ii) the vector η minimizing $\mathrm{osL}_{1,[\eta]}(f_{\mathrm{H}})$ is the solution to

$$\begin{aligned} \inf_{b\in\mathbb{R},\eta\in\mathbb{R}^n_{>0}} \quad & b \\ s.t. \quad & (-C+\lceil d_1 A\rceil_{\mathrm{Mzr}}^\top)\eta \leq b\eta, \\ & (-C+\lceil d_2 A\rceil_{\mathrm{Mzr}}^\top)\eta \leq b\eta, \end{aligned}$$

and if the infimum value is attained at parameter values $b^\star, \eta^\star$, then $\mathrm{osL}_{1,[\eta^\star]}(f_{\mathrm{H}}) = b^\star$.

Further, suppose $\lceil A\rceil_{\mathrm{Mzr}}$ is irreducible. Then

(iii) if $C = cI_n$ and $d_1 \geq 0$, then, with $w_A \in \mathbb{R}^n_{>0}$ being the left dominant eigenvector of $\lceil A\rceil_{\mathrm{Mzr}}$,

$$\inf_{\eta\in\mathbb{R}^n_{>0}} \mathrm{osL}_{1,[\eta]}(f_{\mathrm{H}}) = \mathrm{osL}_{1,[w_A]}(f_{\mathrm{H}}) = -c + \max\{d_1\alpha(\lceil A\rceil_{\mathrm{Mzr}}), d_2\alpha(\lceil A\rceil_{\mathrm{Mzr}})\}. \tag{5.17}$$

(iv) if $d_1 = 0$ and $C \succ 0$, then, with $w_* \in \mathbb{R}^n_{>0}$ being the left dominant eigenvector of $-C + d_2\lceil A\rceil_{\mathrm{Mzr}}$,

$$\inf_{\eta\in\mathbb{R}^n_{>0}} \mathrm{osL}_{1,[\eta]}(f_{\mathrm{H}}) = \mathrm{osL}_{1,[w_*]}(f_{\mathrm{H}}) = \max\big\{\alpha(-C), \alpha(-C + d_2\lceil A\rceil_{\mathrm{Mzr}})\big\}. \tag{5.18}$$

In particular, Theorem 5.6.1 provides *exact values* for the minimal one-sided Lipschitz constant of the Hopfield neural network with respect to diagonally-weighted ℓ_1 norms.

As a consequence of this theorem, suppose the inf in statement (ii) is attained and let $b^\star, \eta^\star$ be the optimal parameters for the LP. If $b^\star < 0$, then the Hopfield neural network (5.15) is strongly infinitesimally contracting with rate $|b^\star|$ with respect to $\|\cdot\|_{1,[\eta^\star]}$. Note, in particular, that if $d_1 = 0, d_2 = 1$, $C = I_n$, and $\alpha(\lceil A\rceil_{\mathrm{Mzr}}) < 1$, statement (iii) is equivalent to the statement in Proposition 5.2.1(i).

Proof of Theorem 5.6.1. Regarding statement (i), for any $\eta \in \mathbb{R}^n_{>0}$,

$$\begin{aligned}\operatorname{osL}_{1,[\eta]}(f_{\mathrm{H}}) &= \sup_{x\in\mathbb{R}^n\setminus\Omega_{f_H}} \mu_{1,[\eta]}(Df_{\mathrm{H}}(x)) = \sup_{x\in\mathbb{R}^n\setminus\Omega_{f_H}} \mu_{1,[\eta]}(-C + A\,D\Phi(x)) \\ &= \max_{d\in[d_1,d_2]^n} \mu_{1,[\eta]}(-C + A[d]) \\ &= \max\big\{\mu_{1,[\eta]}(-C + d_1 A), \mu_{1,[\eta]}(-C + d_2 A)\big\},\end{aligned}$$

where the second-to-last equality holds by Assumption 2 and the last equality holds by Lemma 5.4.3.

Statement (ii) holds by Corollary 5.4.4. Regarding statement (iii), if $C = cI_n$ and $d_1 \geq 0$, then

$$\begin{aligned}\operatorname{osL}_{1,[\eta]}(f_{\mathrm{H}}) &= -c + \max\big\{\mu_{1,[\eta]}(d_1 A), \mu_{1,[\eta]}(d_2 A)\big\} \\ &= -c + \max\big\{d_1\mu_{1,[\eta]}(A), d_2\mu_{1,[\eta]}(A)\big\}.\end{aligned}$$

Additionally, recall that $\eta = w_A$ is the optimal weight from Lemma 5.3.1 for the irreducible Metzler matrix $\lceil A\rceil_{\mathrm{Mzr}}$ with respect to $p = 1$. Therefore,

$$\begin{aligned}\inf_{\eta\in\mathbb{R}^n_{>0}} \operatorname{osL}_{1,[\eta]}(f_{\mathrm{H}}) = \operatorname{osL}_{1,[w_A]}(f_{\mathrm{H}}) &= -c + \max\{d_1\mu_{1,[w_A]}(A), d_2\mu_{1,[w_A]}(A)\} \\ &= -c + \max\{d_1\alpha(\lceil A\rceil_{\mathrm{Mzr}}), d_2\alpha(\lceil A\rceil_{\mathrm{Mzr}})\}\end{aligned}$$

Regarding statement (iv), if $d_1 = 0$ and $C \succ 0$, we compute

$$\operatorname{osL}_{1,[\eta]}(f_{\mathrm{H}}) = \max\big\{\mu_{1,[\eta]}(-C), \mu_{1,[\eta]}(-C + d_2 A)\big\} = \max\big\{\alpha(-C), \mu_{1,[\eta]}(-C + d_2 A)\big\},$$

which holds because $\mu_{1,[\eta]}(-C) = \max_{i\in\{1,\dots,n\}} -c_{ii} = \alpha(-C)$ for every $\eta \in \mathbb{R}^n_{>0}$. Additionally, we have that $\eta = w_*$ is the optimal weight for the irreducible Metzler matrix

$-C + d_2\lceil A \rceil_{\mathrm{Mzr}}$ by Lemma 5.3.1. Thus,

$$\inf_{\eta \in \mathbb{R}^n_{>0}} \mathrm{osL}_{1,[\eta]}(f_H) = \mathrm{osL}_{1,[w_*]}(f_H) = \max\{\alpha(-C), \alpha(-C + d_2\lceil A \rceil_{\mathrm{Mzr}})\},$$

which proves the result. □

Remark 5.6.1 (Comparison to [148, Theorem 1], [154, Exercise 2.22])**.** *For $A \in \mathbb{R}^{n\times n}$, define its nonnegative Metzler majorant $\lceil A \rceil^+_{\mathrm{Mzr}}$ by*

$$(\lceil A \rceil^+_{\mathrm{Mzr}})_{ij} = \begin{cases} \max\{A_{ii}, 0\}, & \text{if } i = j, \\ |A_{ij}|, & \text{if } i \neq j \end{cases}.$$

In [148, Theorem 1], for $d_1 = 0$ and $C = I_n$ it is shown that if $\alpha(-I_n + d_2\lceil A \rceil^+_{\mathrm{Mzr}}) < 0$, then the Hopfield neural network (5.15) is contracting with respect to a diagonally-weighted ℓ_2 norm which is given in Lemma 5.3.1. Compared to the condition $\alpha(-I_n + \lceil A \rceil^+_{\mathrm{Mzr}}) < 0$, the condition in Theorem 5.6.1(iv) replaces $\lceil A \rceil^+_{\mathrm{Mzr}}$ with $\lceil A \rceil_{\mathrm{Mzr}}$ and thus guarantees that a larger class of synaptic matrices still guarantee contractivity of the Hopfield neural network.

Additionally, beyond Proposition 5.2.1(i), we characterize the ℓ_∞ one-sided Lipschitz constant of the Hopfield NN in the following theorem.

Theorem 5.6.2 (ℓ_∞ one-sided Lipschitzness of Hopfield neural network)**.** *Consider the Hopfield neural network (5.15) with each $\phi_i \in \mathrm{slope}[d_1, d_2]$. Let $\overline{d} = \max\{|d_1|, d_2|\}$. Then*

(i) for arbitrary $\eta \in \mathbb{R}^n_{>0}$, $\mathrm{osL}_{\infty,[\eta]^{-1}}(f_{\mathrm{H}}) = \max\{\mu_{\infty,[\eta]^{-1}}(-C + \overline{d}A - (\overline{d} - d_1)(I_n \circ A)),$ $\mu_{\infty,[\eta]^{-1}}(-C + \overline{d}A - (\overline{d} - d_2)(I_n \circ A))\}$.

(ii) the vector η *minimizing* $\mathrm{osL}_{\infty,[\eta]^{-1}}(f_{\mathrm{H}})$ *is the solution to*

$$\begin{aligned} \inf_{b\in\mathbb{R},\eta\in\mathbb{R}^n_{>0}} \quad & b \\ s.t. \quad & (-C + \lceil \overline{d}A - (\overline{d} - d_1)(I_n \circ A)\rceil_{\mathrm{Mzr}})\eta \leq b\eta, \\ & (-C + \lceil \overline{d}A - (\overline{d} - d_2)(I_n \circ A)\rceil_{\mathrm{Mzr}})\eta \leq b\eta, \end{aligned}$$

and if the infimum value is attained at parameter values $b^\star, \eta^\star$*, then* $\mathrm{osL}_{\infty,[\eta^\star]^{-1}}(f_{\mathrm{H}}) = b^\star$.

Proof. Regarding statement (i), in analogy to the proof of Theorem 5.6.1(i), we have

$$\begin{aligned} &\mathrm{osL}_{\infty,[\eta]^{-1}}(f_{\mathrm{H}}) = \max_{d\in[d_1,d_2]^n} \mu_{\infty,[\eta]^{-1}}(-C + A[d]) \\ &= \max\{\mu_{\infty,[\eta]^{-1}}(-C + \overline{d}A - (\overline{d} - d_1)(I_n \circ A)), \mu_{\infty,[\eta]^{-1}}(-C + \overline{d}A - (\overline{d} - d_2)(I_n \circ A))\}, \end{aligned}$$

where the final equality is by Lemma 5.4.3. Statement (ii) is then a consequence of Corollary 5.4.4. □

5.6.2 One-sided Lipschitz characterization of firing-rate NNs

Recall the firing-rate NN dynamics:

$$\dot{x} = -Cx + \Phi(Ax + u) =: f_{\mathrm{FR}}(x). \tag{5.19}$$

The interpretation for this name is that if $\Phi(x)$ is nonnegative for all $x \in \mathbb{R}^n$ (as is ReLU), then the positive orthant is forward-invariant and x is interpreted as a vector of firing-rates, while in the Hopfield neural network, x can be negative and is thus interpreted as a vector of membrane potentials.

The following two theorems are generalizations of Proposition 5.2.1(ii) under more general assumptions. Specifically, Theorem 5.6.3 characterizes one-sided Lipschitzness of the firing-rate NN with respect to diagonally, weighted ℓ_∞ norms, while Theorem 5.6.4 does the same with respect to diagonally-weighted ℓ_1 norms.

Theorem 5.6.3 (ℓ_∞ one-sided Lipschitzness of firing-rate neural network)**.** *Consider the firing-rate neural network (5.19) with each $\phi_i \in \mathrm{slope}[d_1, d_2]$ and invertible A. Then*

(i) for arbitrary $\eta \in \mathbb{R}^n_{>0}$, $\mathrm{osL}_{\infty,[\eta]^{-1}}(f_{\mathrm{FR}}) = \max\{\mu_{\infty,[\eta]^{-1}}(-C + d_1A), \mu_{\infty,[\eta]^{-1}}(-C + d_2A)\}$.

(ii) The choice of η minimizing $\mathrm{osL}_{\infty,[\eta]^{-1}}(f_{\mathrm{FR}})$ is the solution to

$$\begin{aligned}\inf_{b\in\mathbb{R},\eta\in\mathbb{R}^n_{>0}} \quad & b \\ \text{s.t.} \quad & (-C + \lceil d_1A\rceil_{\mathrm{Mzr}})\eta \le b\eta, \\ & (-C + \lceil d_2A\rceil_{\mathrm{Mzr}})\eta \le b\eta,\end{aligned}$$

and if the infimum value is attained at parameter values $b^\star, \eta^\star$, then $\mathrm{osL}_{\infty,[\eta^\star]^{-1}}(f_{\mathrm{FR}}) = b^\star$.

Further, suppose that $\lceil A\rceil_{\mathrm{Mzr}}$ is irreducible. Then

(iii) if $C = cI_n$ and $d_1 \ge 0$, then, with $v_A \in \mathbb{R}^n_{>0}$ being the right dominant eigenvector of $\lceil A\rceil_{\mathrm{Mzr}}$,

$$\inf_{\eta\in\mathbb{R}^n_{>0}} \mathrm{osL}_{\infty,[\eta]}(f_{\mathrm{FR}}) = \mathrm{osL}_{\infty,[v_A]^{-1}}(f_{\mathrm{FR}}) = -c + \max\{d_1\alpha(\lceil A\rceil_{\mathrm{Mzr}}), d_2\alpha(\lceil A\rceil_{\mathrm{Mzr}})\}. \quad (5.20)$$

(iv) if $d_1 = 0$ and $C \succ 0$, then, with $v_* \in \mathbb{R}^n_{>0}$ being the right dominant eigenvector of $-C + d_2\lceil A\rceil_{\mathrm{Mzr}}$,

$$\inf_{\eta\in\mathbb{R}^n_{>0}} \mathrm{osL}_{\infty,[\eta]}(f_{\mathrm{FR}}) = \mathrm{osL}_{\infty,[v_*]^{-1}}(f_{\mathrm{FR}}) = \max\big\{\alpha(-C), \alpha(-C + d_2\lceil A\rceil_{\mathrm{Mzr}})\big\}. \quad (5.21)$$

Proof. Regarding statement (i), for any $\eta \in \mathbb{R}^n_{>0}$ we compute

$$\begin{aligned}\mathrm{osL}_{\infty,[\eta]^{-1}}(f_{\mathrm{FR}}) &= \sup_{x\in\mathbb{R}^n\setminus\Omega_{f_{\mathrm{FR}}}} \mu_{\infty,[\eta]^{-1}}(Df_{\mathrm{FR}}(x)) = \sup_{x\in\mathbb{R}^n\setminus\Omega_{f_{\mathrm{FR}}}} \mu_{\infty,[\eta]^{-1}}(-C + D\Phi(Ax+u)A) \\ &= \max_{d\in[d_1,d_2]^n} \mu_{\infty,[\eta]^{-1}}(-C+[d]A) \\ &= \max\big\{\mu_{\infty,[\eta]^{-1}}(-C+d_1A), \mu_{\infty,[\eta]^{-1}}(-C+d_2A)\big\},\end{aligned}$$

where the second-to-last equality holds by Assumption 2 and because A is invertible. The last equality holds by Lemma 5.4.3.

Statement (ii) is a consequence of Corollary 5.4.4. Regarding statement (iii), if $C = cI_n$ and $d_1 \geq 0$, then

$$\begin{aligned}\mathrm{osL}_{\infty,[\eta]^{-1}}(f_{\mathrm{FR}}) &= -c + \max\big\{\mu_{\infty,[\eta]^{-1}}(d_1A), \mu_{\infty,[\eta]^{-1}}(d_2A)\big\} \\ &= -c + \max\big\{d_1\mu_{\infty,[\eta]^{-1}}(A), d_2\mu_{\infty,[\eta]^{-1}}(A)\big\}.\end{aligned}$$

Additionally, recall that $\eta = v_A$ is the optimal weight from Lemma 5.3.1 for the irreducible Metzler matrix $\lceil A\rceil_{\mathrm{Mzr}}$ with respect to $p=\infty$. Therefore,

$$\begin{aligned}\inf_{\eta\in\mathbb{R}^n_{>0}} \mathrm{osL}_{\infty,[\eta]^{-1}}(f_{\mathrm{FR}}) = \mathrm{osL}_{\infty,[v_A]^{-1}}(f_{\mathrm{FR}}) &= -c + \max\{d_1\mu_{\infty,[v_A]^{-1}}(A), d_2\mu_{\infty,[v_A]^{-1}}(A)\} \\ &= -c + \max\{d_1\alpha(\lceil A\rceil_{\mathrm{Mzr}}), d_2\alpha(\lceil A\rceil_{\mathrm{Mzr}})\}.\end{aligned}$$

Regarding statement (iv), if $d_1 = 0$ and $C \succ 0$, we compute

$$\begin{aligned}\mathrm{osL}_{\infty,[\eta]^{-1}}(f_{\mathrm{FR}}) &= \max\big\{\mu_{\infty,[\eta]^{-1}}(-C), \mu_{\infty,[\eta]^{-1}}(-C+d_2A)\big\} \\ &= \max\big\{\alpha(-C), \mu_{\infty,[\eta]^{-1}}(-C+d_2A)\big\},\end{aligned}$$

which holds because $\mu_{\infty,[\eta]^{-1}}(-C) = \max_{i\in\{1,\dots,n\}} -c_{ii} = \alpha(-C)$ for every $\eta \in \mathbb{R}^n_{>0}$.

Additionally, we have that $\eta = v_*$ is the optimal weight for the irreducible Metzler matrix $-C + d_2\lceil A \rceil_{\mathrm{Mzr}}$ by Lemma 5.3.1. Thus,

$$\inf_{\eta \in \mathbb{R}^n_{>0}} \mathrm{osL}_{\infty,[\eta]^{-1}}(f_{\mathrm{FR}}) = \mathrm{osL}_{\infty,[v_*]^{-1}}(f_{\mathrm{FR}}) = \max\{\alpha(-C), \alpha(-C + d_2\lceil A \rceil_{\mathrm{Mzr}})\},$$

which proves the result. □

Theorem 5.6.4 (ℓ_1 one-sided Lipschitzness of firing-rate neural network)**.** *Consider the firing-rate neural network* (5.19) *with each* $\phi_i \in \mathrm{slope}[d_1, d_2]$*, and invertible* A*. Let* $\overline{d} = \max\{|d_1|, |d_2|\}$*. Then*

(i) for arbitrary $\eta \in \mathbb{R}^n_{>0}$*,* $\mathrm{osL}_{1,[\eta]}(f_{\mathrm{FR}}) = \max\{\mu_{1,[\eta]}(-C + \overline{d}A - (\overline{d} - d_1)(I_n \circ A)),$ $\mu_{1,[\eta]}(-C + \overline{d}A - (\overline{d} - d_2)(I_n \circ A))\}$*.*

(ii) the vector η *minimizing* $\mathrm{osL}_{1,[\eta]}(f_{\mathrm{FR}})$ *is the solution to*

$$\begin{aligned} \inf_{b \in \mathbb{R}, \eta \in \mathbb{R}^n_{>0}} \quad & b \\ s.t. \quad & (-C + \lceil \overline{d}A - (\overline{d} - d_1)(I_n \circ A) \rceil_{\mathrm{Mzr}})^\top \eta \leq b\eta, \\ & (-C + \lceil \overline{d}A - (\overline{d} - d_2)(I_n \circ A) \rceil_{\mathrm{Mzr}})^\top \eta \leq b\eta, \end{aligned}$$

and if the infimum value is attained at parameter values $b^\star, \eta^\star$*, then* $\mathrm{osL}_{1,[\eta^\star]}(f_{\mathrm{FR}}) = b^\star$*.*

Proof. Regarding statement (i), in analogy to the proof of Theorem 5.6.3(i), we have

$$\begin{aligned} \mathrm{osL}_{1,[\eta]}(f_{\mathrm{FR}}) = \max_{d \in [d_1, d_2]^n} \mu_{1,[\eta]}(-C + [d]A) = \max\{&\mu_{1,[\eta]}(-C + \overline{d}A - (\overline{d} - d_1)(I_n \circ A)), \\ &\mu_{1,[\eta]}(-C + \overline{d}A - (\overline{d} - d_2)(I_n \circ A))\}, \end{aligned}$$

where the final equality is by Lemma 5.4.3. Statement (ii) is then a consequence of

Corollary 5.4.4. □

Remark 5.6.2. *For invertible $A \in \mathbb{R}^{n\times n}$, Theorems 5.6.3(i) and 5.6.4(i) provide an* exact value *for the minimal one-sided Lipschitz constant of the firing rate model with respect to a given norm. If A is not invertible, then the closure of the image of the map $x \mapsto D\Phi(Ax+u)$ may not contain all the vertices of the set $[d_1, d_2]^n$. For non-invertible A and arbitrary $\eta \in \mathbb{R}^n_{>0}$, the values presented in these theorems are instead upper bounds on the minimal one-sided Lipschitz constant.*

In Figure 5.1, we plot the phase portrait of a 2-dimensional firing-rate neural network, (5.19), along with level sets of the corresponding Lyapunov function. We highlight the utility of optimizing the weight of the ℓ_∞ norm. Namely, although the firing-rate neural network example is not contracting with respect to the ℓ_∞ norm, it is contracting with respect to a weighted ℓ_∞ norm, where the optimal diagonal weight is $[\eta^\star]^{-1}$, where $\eta^\star$ is the right dominant eigenvector of $\lceil A \rceil_{\mathrm{Mzr}}$.

5.6.3 Contractivity of Hopfield and firing-rate neural networks with unbounded slope

In the spirit of the classic work [142] which studies Hopfield neural networks which have monotone activation functions with unbounded slope, we present the following result on the contractivity of Hopfield and firing-rate neural networks when $\phi_i \in \mathrm{slope}[d_1, \infty]$.

Theorem 5.6.5 (Contractivity under unbounded slope)**.** *Consider the Hopfield neural network (5.15) and firing-rate neural network (5.19) with $\phi_i \in \mathrm{slope}[d_1, \infty]$ and irreducible $\lceil A \rceil_{\mathrm{Mzr}}$ with dominant left and right eigenvectors w_A, v_A, respectively and suppose that*

(A1) $A \in \mathcal{MH}$,

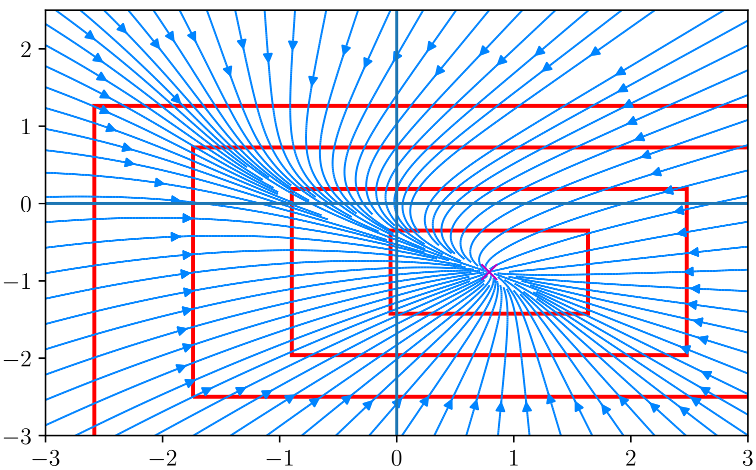


Figure 5.1: The phase portrait for a 2-dimensional firing-rate neural network (5.19) with $C = I_2$, $\Phi = \tanh$, $A = \left[\begin{smallmatrix} -0.1 & -1.3 \\ -0.4 & 0.1 \end{smallmatrix}\right]$, $u = (0, -1)$. The blue curves denote trajectories from varying initial conditions and the purple cross denotes the equilibrium point, $x^\star$. Following Theorem 5.6.2(iv), the right dominant eigenvector of $\lceil A \rceil_{\mathrm{Mzr}}$, $\eta \approx (1.57, 1)$, yields a contraction rate of $1 - \alpha(\lceil A \rceil_{\mathrm{Mzr}}) \approx 0.272$ with respect to $\|\cdot\|_{\infty,[\eta]^{-1}}$. Note that the neural network is not contracting with respect to $\|\cdot\|_\infty$. Level sets of the Lyapunov function $V(x) = \|x - x^\star\|_{\infty,[\eta]^{-1}}$ are shown in red. We remark that this neural network is connectively contracting as described in Section 5.2.

(A2) *$A \in \mathbb{R}^{n\times n}$, $C \succeq 0$, and $d_1 \in \mathbb{R}$ satisfy*

$$-\alpha(-C) + \max\{d_1, 0\}\alpha(\lceil A \rceil_{\mathrm{Mzr}}) > -(|d_1| - d_1) \min_{i\in\{1,\dots,n\}} A_{ii}.$$

Then

(i) The Hopfield neural network (5.15) is strongly infinitesimally contracting with respect to $\|\cdot\|_{1,[w_A]}$ with rate

$$-\alpha(-C) + \max\{d_1, 0\}\alpha(\lceil A \rceil_{\mathrm{Mzr}}) + (|d_1| - d_1) \min_{i\in\{1,\dots,n\}} A_{ii} > 0$$

and

(ii) The firing-rate neural network (5.19) is strongly infinitesimally contracting with

respect to $\|\cdot\|_{\infty,[v_A]^{-1}}$ *with rate*

$$-\alpha(-C) + \max\{d_1, 0\}\alpha(\lceil A\rceil_{\mathrm{Mzr}}) + (|d_1| - d_1) \min_{i\in\{1,\dots,n\}} A_{ii} > 0.$$

Proof. Regarding statement (i), we adopt the shorthand $r_i := A_{ii} + \sum_{j\neq i} |A_{ji}|(w_A)_j/(w_A)_i$. Then we observe that $A \in \mathcal{MH}$ implies that for every $i \in \{1,\dots,n\}$, $r_i \leq \alpha(\lceil A\rceil_{\mathrm{Mzr}})$ with $\alpha(\lceil A\rceil_{\mathrm{Mzr}}) < 0$. Then, for every $x \in \mathbb{R}^n \setminus \Omega_{f_{\mathrm{H}}}$,

$$\begin{aligned}
&\mu_{1,[w_A]}(Df_{\mathrm{H}}(x)) = \mu_{1,[w_A]}(-C + AD\Phi(x)) \\
&= \max_{i\in\{1,\dots,n\}} -c_i + A_{ii}\phi_i'(x_i) + \sum_{j\neq i} |A_{ji}\phi_i'(x_i)| \frac{(w_A)_j}{(w_A)_i} \\
&= \max_{i\in\{1,\dots,n\}} -c_i + A_{ii}\phi_i'(x_i) + |\phi_i'(x_i)| \sum_{j\neq i} |A_{ji}| \frac{(w_A)_j}{(w_A)_i} \\
&= \max_{i\in\{1,\dots,n\}} -c_i + |\phi_i'(x_i)| r_i - (|\phi_i'(x_i)| - \phi_i'(x_i))A_{ii} \\
&\overset{\star}{\leq} \max_{i\in\{1,\dots,n\}} -c_i + \max\{d_1, 0\} r_i - (|d_1| - d_1)A_{ii} \\
&\leq \alpha(-C) + \max\{d_1, 0\}\alpha(\lceil A\rceil_{\mathrm{Mzr}}) - (|d_1| - d_1) \min_{i\in\{1,\dots,n\}} A_{ii}
\end{aligned}$$

where inequality $\overset{\star}{\leq}$ holds because $r_i < 0$ and $A_{ii} < 0$ for all i. Since this inequality holds for all $x \in \mathbb{R}^n \setminus \Omega_{f_{\mathrm{H}}}$, we conclude that $\mathrm{osL}_{1,[w_A]}(f_{\mathrm{H}}) \leq \alpha(-C) + \max\{d_1, 0\}\alpha(\lceil A\rceil_{\mathrm{Mzr}}) - (|d_1| - d_1)\min_{i\in\{1,\dots,n\}} A_{ii}$, which implies the result. The proof of statement (ii) is essentially identical and thus omitted. □

Remark 5.6.3. *Note that in Theorem 5.6.5,*

(i) if $d_1 \geq 0$*, then condition (A1) immediately implies condition (A2). Hence,* $A \in \mathcal{MH}$ *is a sufficient condition for the strong infinitesimal contractivity of* (5.15) *and* (5.19) *with unbounded-slope monotonic activation functions.*

(ii) *alternatively if* $\alpha(\lceil A \rceil_{\mathrm{Mzr}}) = 0$ *and* $C \succ 0$*, then the condition* $-\alpha(-C) > -(|d_1| - d_1) \min_{i \in \{1,\dots,n\}} A_{ii}$ *is a sufficient condition for the strong infinitesimal contractivity of* (5.15) *and* (5.19)*. Note that in this case, we only need* $\lceil A \rceil_{\mathrm{Mzr}}$ *to be marginally stable.*

5.6.4 Contractivity of other continuous-time neural networks

We apply Theorem 5.6.1 and the log norm results in Section 5.4 to the following related neural circuit models, all of which are studied in the classic book [161]. In the following theorems, we assume all Metzler matrices are irreducible.

Theorem 5.6.6 (Contractivity of special Hopfield models). *(i) If* $A \in \mathcal{MH}$*, and* $d_1 > 0$*, the Persidskii-type*[4] *model*

$$\dot{x} = A\Phi(x)$$

with each $\phi_i \in \mathrm{slope}[d_1, d_2]$ *is strongly infinitesimally contracting with respect to norm* $\|\cdot\|_{1,[w_A]}$ *with rate* $d_1|\alpha(\lceil A \rceil_{\mathrm{Mzr}})|$.

(ii) If $-C + d_2 A \in \mathcal{MH}$*, the Hopfield neural network* (5.15) *with* $d_1 = 0$ *and positive diagonal* C *is strongly infinitesimally contracting with respect to* $\|\cdot\|_{1,[w_*]}$ *with rate* $-\max\big\{\alpha(-C), \alpha(-C + d_2\lceil A \rceil_{\mathrm{Mzr}})\big\} > 0$.

Proof. Regarding statement (i), let $f_{\mathrm{P}}(x) := A\Phi(x)$. By Theorem 5.6.1(iii) with $c = 0$, $\mathrm{osL}_{1,[w_A]}(f_{\mathrm{P}}) = \max\{d_1\alpha(\lceil A \rceil_{\mathrm{Mzr}}), d_2\alpha(\lceil A \rceil_{\mathrm{Mzr}})\}$. However, since $A \in \mathcal{MH}$, $\alpha(\lceil A \rceil_{\mathrm{Mzr}}) < 0$, so $\mathrm{osL}_{1,[w_A]}(f_{\mathrm{P}}) = d_1\alpha(\lceil A \rceil_{\mathrm{Mzr}})$. Thus, the Persidskii-type model is strongly infinitesimally contracting with respect to norm $\|\cdot\|_{1,[w_A]}$ with rate $d_1|\alpha(\lceil A \rceil_{\mathrm{Mzr}})|$.

Regarding statement (ii), by Theorem 5.6.1(iv),

$$\mathrm{osL}_{1,[w_*]}(f_{\mathrm{H}}) = \max\big\{\alpha(-C), \alpha(-C + d_2\lceil A \rceil_{\mathrm{Mzr}})\big\}.$$

[4] See [161, Definition 3.2.1]

In particular, since $-C + d_2 A \in \mathcal{MH}$, $\alpha(-C + d_2\lceil A\rceil_{\mathrm{Mzr}}) < 0$ and since C is positive diagonal, we have $\mathrm{osL}_{1,[w_*]}(f_{\mathrm{H}}) < 0$ so that the Hopfield neural network is strongly infinitesimally contracting with respect to $\|\cdot\|_{1,[w_*]}$ with rate $-\max\big\{\alpha(-C), \alpha(-C + d_2\lceil A\rceil_{\mathrm{Mzr}})\big\} > 0$. $\square$

Theorem 5.6.7. *From [161, Theorem 3.2.4], consider*

$$\dot{x} = Ax - C\Phi(x),$$

with diagonal $C \succeq 0$ and each $\phi_i \in \mathrm{slope}[d_1, d_2]$. If $A - d_1 C \in \mathcal{MH}$ with corresponding dominant left eigenvector w_{**}, then this model is strongly infinitesimally contracting with respect to $\|\cdot\|_{1,[w_{**}]}$ with rate $-\alpha(\lceil A\rceil_{\mathrm{Mzr}} - d_1 C) > 0$.

Proof. We compute the one-sided Lipschitz constant of $f(x) := Ax - C\Phi(x)$ with respect to norm $\|\cdot\|_{1,[w_{**}]}$.

$$\begin{aligned}
&\mathrm{osL}_{1,[w_{**}]}(f) = \sup_{x\in\mathbb{R}^n\setminus\Omega_f} \mu_{1,[w_{**}]}(Df(x)) \\
&= \sup_{x\in\mathbb{R}^n\setminus\Omega_f} \mu_{1,[w_{**}]}(A - CD\Phi(x)) \\
&\overset{\bigstar}{=} \max_{d\in[d_1,d_2]^n} \mu_{1,[w_{**}]}(A - C[d]) \overset{\spadesuit}{=} \mu_{1,[w_{**}]}(A - d_1 C) \\
&\overset{\clubsuit}{=} \mu_{1,[w_{**}]}(\lceil A - d_1 C\rceil_{\mathrm{Mzr}}) \overset{\blacklozenge}{=} \alpha(\lceil A\rceil_{\mathrm{Mzr}} - d_1 C).
\end{aligned}$$

where the equality $\overset{\bigstar}{=}$ is due to Assumption 2, equality $\overset{\spadesuit}{=}$ is because $C \succeq 0$, equality $\overset{\clubsuit}{=}$ is by Theorem 5.4.2(ii), and $\overset{\blacklozenge}{=}$ is by Lemma 5.3.1. Moreover, since $A - d_1 C \in \mathcal{MH}$, $\alpha(\lceil A\rceil_{\mathrm{Mzr}} - d_1 C) < 0$ so f is strongly infinitesimally contracting with respect to $\|\cdot\|_{1,[w_{**}]}$ with rate $-\alpha(\lceil A\rceil_{\mathrm{Mzr}} - d_1 C) > 0$. $\square$

Theorem 5.6.8. *From [161, Theorem 3.2.10], consider*

$$\dot{x}_i = \sum_{j=1}^{n} A_{ij}\phi_{ij}(x_j)$$

for each $i \in \{1,\dots,n\}$ *and with each* $\phi_{ij} \in \mathrm{slope}[d_1, d_2]$*. If* $d_1 > 0$ *and*

$$B := d_2 A - (d_2 - d_1)(I_n \circ A) \in \mathcal{MH},$$

with corresponding dominant left and right eigenvectors w_B, v_B*, respectively, then this model is strongly infinitesimally contracting with rate* $-\alpha(\lceil B \rceil_{\mathrm{Mzr}}) > 0$ *with respect to both* $\|\cdot\|_{1,[w_B]}$ *and* $\|\cdot\|_{\infty,[v_B]^{-1}}$*.*

Proof. First note that the assumption $B \in \mathcal{MH}$ implies that $A_{ii} < 0$ for every $i \in \{1,\dots,n\}$ since the diagonal elements of B are $d_1 A_{ii}$ and a necessary condition for $B \in \mathcal{MH}$ is $B_{ii} < 0$ since $\mathcal{MH} \subset \mathcal{TH}$. Let f denote the vector field given by $f_i(x) = \sum_{j=1}^{n} A_{ij}\phi_{ij}(x_j)$. We compute $(Df(x))_{ij} = \frac{\partial}{\partial x_j}\sum_{j=1}^{n} A_{ij}\phi_{ij}(x_j) = A_{ij}\phi'_{ij}(x_j)$ for almost every $x \in \mathbb{R}^n$. In other words, $Df(x) = A \circ D\Phi(x)$, for almost every $x \in \mathbb{R}^n$, where $(D\Phi(x))_{ij} = \phi'_{ij}(x_j)$. We now proceed to elementwise upper bound $\lceil Df(x) \rceil_{\mathrm{Mzr}}$. Observe that for every $i \neq j \in \{1,\dots,n\}$,

$$(\lceil Df(x) \rceil_{\mathrm{Mzr}})_{ij} = |A_{ij}\phi'_{ij}(x_j)| \leq d_2|A_{ij}| = (\lceil B \rceil_{\mathrm{Mzr}})_{ij},$$
$$(\lceil Df(x) \rceil_{\mathrm{Mzr}})_{ii} = A_{ii}\phi'_{ii}(x_i) \leq d_1 A_{ii} = (\lceil B \rceil_{\mathrm{Mzr}})_{ii},$$

where the second inequality holds because $A_{ii} < 0$ for every $i \in \{1,\dots,n\}$. Now observe that for any matrix $A \in \mathbb{R}^{n\times n}$, if $\lceil A \rceil_{\mathrm{Mzr}} \leq A'$ elementwise, then both $\mu_{1,[\eta]}(A) \leq \mu_{1,[\eta]}(A')$

and $\mu_{\infty,[\eta]^{-1}}(A) \leq \mu_{\infty,[\eta]^{-1}}(A')$ hold for any $\eta \in \mathbb{R}^n_{>0}$. Then we can observe that

$$\begin{aligned}
&\mathrm{osL}_{1,[w_B]}(f) = \sup_{x\in\mathbb{R}^n\setminus\Omega_f} \mu_{1,[w_B]}(Df(x)) \\
&= \sup_{x\in\mathbb{R}^n\setminus\Omega_f} \mu_{1,[w_B]}(\lceil Df(x)\rceil_{\mathrm{Mzr}}) \leq \sup_{x\in\mathbb{R}^n\setminus\Omega_f} \mu_{1,[w_B]}(\lceil B\rceil_{\mathrm{Mzr}}) \\
&= \mu_{1,[w_B]}(\lceil B\rceil_{\mathrm{Mzr}}) = \alpha(\lceil B\rceil_{\mathrm{Mzr}}),
\end{aligned}$$

where the final equality holds by Lemma 5.3.1. An analogous computation shows that $\mathrm{osL}_{\infty,[v_B]^{-1}}(f) \leq \alpha(\lceil B\rceil_{\mathrm{Mzr}})$. As a consequence, since $B \in \mathcal{MH}$, this model is strongly infinitesimally contracting with respect to both $\|\cdot\|_{1,[w_B]}$ and $\|\cdot\|_{\infty,[v_B]^{-1}}$ with rate $-\alpha(\lceil B\rceil_{\mathrm{Mzr}}) > 0$. □

The next two theorems serve as non-Euclidean versions of early results on contractivity of Lur'e systems (in application to the entrainment problem) established first in [162].

Theorem 5.6.9 (Contractivity of Lur'e system)**.** *From [161, Theorem 3.2.7], consider the Lur'e system*

$$\begin{aligned}
\dot{x} &= Ax + v\phi(y),\\
y &= w^\top x,
\end{aligned}$$

where $A \in \mathbb{R}^{n\times n}$, $v, w \in \mathbb{R}^n$ and $\phi \in \mathrm{slope}[d_1, d_2]$. Consider the following two infimization problems:

$$\begin{aligned}
\inf_{b\in\mathbb{R},\eta\in\mathbb{R}^n_{>0}} \quad & b \\
s.t. \quad & \lceil A + d_1 vw^\top\rceil_{\mathrm{Mzr}}^\top \eta \leq b\eta, \\
& \lceil A + d_2 vw^\top\rceil_{\mathrm{Mzr}}^\top \eta \leq b\eta,
\end{aligned} \tag{5.22}$$

and

$$\begin{aligned}
\inf_{c\in\mathbb{R},\xi\in\mathbb{R}^n_{>0}} \quad & c \\
s.t. \quad & \lceil A + d_1 v w^\top \rceil_{\mathrm{Mzr}}\, \xi \le c\xi, \\
& \lceil A + d_2 v w^\top \rceil_{\mathrm{Mzr}}\, \xi \le c\xi.
\end{aligned} \tag{5.23}$$

Let $b^\star, c^\star$ be infimum values for (5.22), (5.23), respectively. Then

(i) if $b^\star < 0$, then for every $\varepsilon \in \,]0, |b^\star|[$, there exists $\eta \in \mathbb{R}^n_{>0}$ such that the closed-loop dynamics are strongly infinitesimally contracting with rate $|b^\star| - \varepsilon > 0$ with respect to $\|\cdot\|_{1,[\eta]}$.

(ii) if $c^\star < 0$, then for every $\varepsilon \in \,]0, |c^\star|[$, there exists $\xi \in \mathbb{R}^n_{>0}$ such that the closed-loop dynamics are strongly infinitesimally contracting with rate $|c^\star| - \varepsilon > 0$ with respect to $\|\cdot\|_{\infty,[\xi]^{-1}}$.

Proof. Let $f_{\mathrm{L}}(x) := Ax + v\phi(w^\top x)$. Regarding statement (i), computing the one-sided Lipschitz constant of f_{L} with respect to $\|\cdot\|_{1,[\eta]}$ for arbitrary $\eta \in \mathbb{R}^n_{>0}$ yields

$$\begin{aligned}
\mathrm{osL}_{1,[\eta]}(f_{\mathrm{L}}) &= \sup_{x\in\mathbb{R}^n\setminus\Omega_{f_{\mathrm{L}}}} \mu_{1,[\eta]}(Df_{\mathrm{L}}(x)) \\
&= \sup_{x\in\mathbb{R}^n\setminus\Omega_{f_{\mathrm{L}}}} \mu_{1,[\eta]}(A + v\phi'(w^\top x)w^\top) \\
&\overset{\bigstar}{=} \max_{d\in[d_1,d_2]} \mu_{1,[\eta]}(A + d\, vw^\top) \\
&\overset{\spadesuit}{=} \max\{\mu_{1,[\eta]}(A + d_1 vw^\top), \mu_{1,[\eta]}(A + d_2 vw^\top)\},
\end{aligned}$$

where $\overset{\bigstar}{=}$ holds by Assumption 2 on ϕ and $\overset{\spadesuit}{=}$ holds because the maximum of a convex function (μ in this case) over a compact interval occurs at one of the endpoints of the interval. As a consequence, $\mathrm{osL}_{1,[\eta]}(f_{\mathrm{L}}) < 0$ if and only if

$$\inf_{\eta \in \mathbb{R}^n_{>0}} \max\{\mu_{1,[\eta]}(A + d_1 v w^\top), \mu_{1,[\eta]}(A + d_2 v w^\top)\} < 0.$$

Therefore, if $b^\star < 0$ for problem (5.22), then, by a continuity argument, for every $\varepsilon \in]0, |b^\star|[$, there exists $\eta \in \mathbb{R}^n_{>0}$ such that $\mu_{1,[\eta]}(A + d_1 v w^\top) \leq b^\star + \varepsilon$ and $\mu_{1,[\eta]}(A + d_2 v w^\top) \leq b^\star + \varepsilon$. Therefore, if $b^\star < 0$, then we conclude that the Lur'e system is strongly infinitesimally contracting with respect to $\|\cdot\|_{1,[\eta]}$ with rate $|b^\star| - \varepsilon$. The proof of statement (ii) is essentially identical, replacing $\|\cdot\|_{1,[\eta]}$ with $\|\cdot\|_{\infty,[\xi]^{-1}}$. $\square$

Theorem 5.6.10 (Multivariable Lur'e system). *Consider the multivariable Lur'e system*

$$\begin{aligned} \dot{x} &= Ax + B\Phi(y), \\ y &= Cx, \end{aligned} \tag{5.24}$$

where $A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{n \times m}, C \in \mathbb{R}^{m \times n}$, $\phi_i \in \mathrm{slope}[d_1, d_2]$ *with* $d_1 \geq 0$ *for all* $i \in \{1, \dots, m\}$. *Define* $(\cdot)_+$ *and* $(\cdot)_-$ *by* $(x)_+ = \max\{x, 0\}$ *and* $(x)_- = \min\{x, 0\}$. *Define* $F \in \mathbb{R}^{n \times n}$ *componentwise by*

$$\begin{aligned} F_{ii} &= A_{ii} + d_2 \sum_{j=1}^m (B_{ij} C_{ji})_+ + d_1 \sum_{j=1}^m (B_{ij} C_{ji})_-, \\ F_{ij} &= |A_{ij}| + \max \Bigg\{ d_2 \sum_{k=1}^m (B_{ik} C_{kj})_+ + d_1 \sum_{k=1}^m (B_{ik} C_{kj})_-, \\ &\qquad -d_1 \sum_{k=1}^m (B_{ik} C_{kj})_+ - d_2 \sum_{k=1}^m (B_{ik} C_{kj})_- \Bigg\}, \end{aligned}$$

for $i \neq j$. *Then, if* $F \in \mathcal{MH}$ *with corresponding dominant left and right eigenvectors* w_F, v_F, *the closed-loop dynamics are strongly infinitesimally contracting with rate* $-\alpha(\lceil F \rceil_{\mathrm{Mzr}}) > 0$ *with respect to both* $\|\cdot\|_{1,[w_F]}$ *and* $\|\cdot\|_{\infty,[v_F]^{-1}}$.

Proof. Let $f_{\mathrm{ML}}(x) = Ax + B\Phi(Cx)$ and note $Df_{\mathrm{ML}}(x) = A + B[d]C$ for some $d \in [d_1, d_2]^n$. Also note $(B[d]C)_{ij} = \sum_{k=1}^{m} B_{ik}d_k C_{kj}$. The proof follows from noting that the matrix F is an entry-wise upper bound on $\lceil Df_{\mathrm{ML}}(x)\rceil_{\mathrm{Mzr}}$, for all x, in analogy with the proof of Theorem 5.6.8. □

Finally, we present a sharper condition for the non-Euclidean contractivity of the multivariable Lur'e system with d_1 that can be negative. For $\eta \in \mathbb{R}^n_{>0}$, $\overline{d} = \max\{|d_1|, |d_2|\}$, $M := |A| + \overline{d}|B||C|$, and $g = \|A\|_{\infty,[\eta]^{-1}} + \overline{d}\|BC\|_{\infty,[\eta]^{-1}}$, consider the following mixed-integer linear program (MILP):

$$
\begin{aligned}
\max_{y\in\mathbb{R}, Z\in\mathbb{R}^{n\times n}, d\in[d_1,d_2]^m, W\in\{0,1\}^{n\times n}} \quad & y, \\
\text{subject to} \quad & \\
Z \leq A + B[d]C + 2M \circ (W - (I_n \circ W)), & \\
Z \leq -A - B[d]C + 2M \circ (\mathbb{1}_n\mathbb{1}_n^\top - (W - (I_n \circ W))), & \\
y \leq (A + B[d]C)_{ii} + \sum_{j\neq i} Z_{ij}\frac{\eta_j}{\eta_i} + 2gW_{ii}, \quad \forall i \in \{1, \dots, n\}, & \\
\mathrm{Trace}(W) = n - 1. &
\end{aligned}
\tag{5.25}
$$

Theorem 5.6.11 (One-sided Lipschitzness of multi-variable Lur'e system)**.** *Consider the multi-variable Lur'e system* (5.24)*, let $f_{\mathrm{ML}}(x) = Ax + B\Phi(Cx)$ be the closed-loop dynamics with each $\phi_i \in \mathrm{slope}[d_1, d_2]$ and let $y^\star$ be the optimal value for the MILP* (5.25)*. Then the following statements hold*

(i) $\mathrm{osL}_{\infty,[\eta]^{-1}}(f_{\mathrm{ML}}) \leq y^\star$.

(ii) If C is full row rank, then $\mathrm{osL}_{\infty,[\eta]^{-1}}(f_{\mathrm{ML}}) = y^\star$.

Proof. Note that

$$\mathrm{osL}_{\infty,[\eta]^{-1}}(f_{\mathrm{ML}}) \leq \max_{d\in[d_1,d_2]^m} \mu_{\infty,[\eta]^{-1}}(A + B[d]C),$$

with equality holding if C is full row rank. Therefore, all that remains is to show that $\max_{d\in[d_1,d_2]^m} \mu_{\infty,[\eta]^{-1}}(A + B[d]C) = y^\star$. The proof of this result is a consequence of the formula for $\mu_{\infty,[\eta]^{-1}}$, using a so-called "big-M" formulation (see, e.g., [163, Section III.C.]) with $Z_{ij} \leq |(A + B[d]C)_{ij}|$ for $i \neq j$ and $y \leq \max_{i\in\{1,\dots,n\}}(A + B[d]C)_{ii} + \sum_{j\neq i} Z_{ij}$. □

The challenge of additionally optimizing $\eta \in \mathbb{R}^n_{>0}$ so that $\mathrm{osL}_{\infty,[\eta]^{-1}}(f_{\mathrm{ML}})$ is minimized remains an open problem.

5.7 Discussion

In this chapter, we present novel non-Euclidean log norm results and a non-smooth contraction theory simplification and we apply these results to study the contractivity of continuous-time NN models, primarily focusing on the Hopfield and firing-rate models. We provide efficient algorithms for computing optimal non-Euclidean contraction rates and corresponding norms. Our approach is robust with respect to activation function and additional unmodeled dynamics and, more generally, establishes the strong contractivity property which, in turn, implies strong robustness properties.

As a first direction of future research, we plan to investigate the multistability of continuous-time neural networks via generalizations of contraction theory. Contraction theory ensures the uniqueness of a globally exponentially stable equilibrium, but several classes of neural networks exhibit multiple equilibria [164]. As a second direction, we plan to investigate the role of non-Euclidean contractivity in neural networks for controller design and system identification in the spirit of the works [165, 166, 167]. As a third line of research, we aim to implement non-Euclidean contracting neural networks in machine learning problems akin to methods from [148]. More broadly, we believe that our non-Euclidean contraction framework for continuous-time NNs serves as a first step to analyzing robustness and convergence properties of other classes of neural circuits and other machine learning architectures.

5.8 Proofs

5.8.1 Proof of Theorem 5.4.1

Proof. Regarding statement (i), we provide the proof for $p = 1$ since $p = \infty$ is essentially identical. Continuity is a straightforward consequence of the formula for $\mu_{1,[\eta]}$. Regarding quasiconvexity, we will show that sublevel sets of the map $\eta \mapsto \mu_{1,[\eta]}(A)$ are convex. For fixed $b \in \mathbb{R}$, the set $\{\eta \in \mathbb{R}^n_{>0} \mid \mu_{1,[\eta]}(A) \leq b\}$ is characterized by η satisfying

$$\eta_i A_{ii} + \sum\nolimits_{j=1, j \neq i}^{n} \eta_j |A_{ji}| \leq \eta_i b, \quad \text{for all } i \in \{1, \dots, n\}.$$

Since each of these inequalities is linear in η, for fixed b, the above set is a polytope, proving quasiconvexity. Statement (ii) follows from the definitions of $\mu_{1,[\eta]}(A)$ and $\mu_{\infty,[\eta]^{-1}}(A)$. □

5.8.2 Proof of Lemma 5.4.3

To prove Lemma 5.4.3, we first need a technical result.

Lemma 5.8.1. *For any $\gamma \in \mathbb{R}, A \in \mathbb{R}^{n \times n}$, the following holds:*

$$\lceil \gamma A \rceil_{\mathrm{Mzr}} = \lceil |\gamma| A - (|\gamma| - \gamma)(I_n \circ A) \rceil_{\mathrm{Mzr}}.$$

Proof. The proof follows by checking that the corresponding entries of each matrix are equal. $\square$

Proof of Lemma 5.4.3. First we show (5.11). We use the short-hand

$$r_i := A_{ii} + \sum_{j \neq i} |A_{ij}| \eta_i / \eta_j$$

and $\mathcal{D} := \{d_1, d_2\}$. Then

$$\begin{aligned}
&\max_{d \in [d_1, d_2]^n} \mu_{\infty,[\eta]}([c] + [d]A) \\
&= \max_{d \in [d_1,d_2]^n} \max_{i \in \{1,\dots,n\}} c_i + d_i A_{ii} + \sum_{j \neq i} |d_i A_{ij}| \frac{\eta_i}{\eta_j} \\
&= \max_{i \in \{1,\dots,n\}} \max_{d \in [d_1,d_2]^n} c_i + d_i A_{ii} + \sum_{j \neq i} |d_i A_{ij}| \frac{\eta_i}{\eta_j} \\
&\overset{\bigstar}{=} \max_{i \in \{1,\dots,n\}} \max\{c_i + \gamma A_{ii} + |\gamma| \sum_{j \neq i} |A_{ij}| \frac{\eta_i}{\eta_j} \mid \gamma \in \mathcal{D}\},
\end{aligned}$$

where the equality $\overset{\bigstar}{=}$ holds because the function $d_i \mapsto c_i + d_i A_{ii} + \sum_{j \neq i} |d_i A_{ij}| \eta_i / \eta_j$ is convex. Since the maximum value of a convex function over an interval $d_i \in [d_1, d_2]$ occurs at one of the endpoints, the equality $\overset{\bigstar}{=}$ is justified.

Additionally note that for any $\gamma \in \mathbb{R}$,

$$\gamma A_{ii} + |\gamma| \sum_{j \neq i} |A_{ij}| \frac{\eta_i}{\eta_j} = |\gamma| r_i - (|\gamma| - \gamma) A_{ii}.$$

Therefore,

$$
\begin{aligned}
&\max_{d\in[d_1,d_2]^n} \mu_{\infty,[\eta]}([c]+[d]A) \\
&= \max_{i\in\{1,\dots,n\}} \max\{c_i + |\gamma| r_i - (|\gamma|-\gamma)A_{ii} \mid \gamma\in\mathcal{D}\} \\
&= \max\{\mu_{\infty,[\eta]}([c] + |\gamma|A - (|\gamma|-\gamma)(I_n\circ A)) \mid \gamma\in\mathcal{D}\}, \\
&\overset{\spadesuit}{=} \max\{\mu_{\infty,[\eta]}([c] + \lceil |\gamma|A - (|\gamma|-\gamma)(I_n\circ A)\rceil_{\mathrm{Mzr}}) \mid \gamma\in\mathcal{D}\}, \\
&\overset{\clubsuit}{=} \max\{\mu_{\infty,[\eta]}([c] + \lceil \gamma A\rceil_{\mathrm{Mzr}}) \mid \gamma\in\mathcal{D}\} \\
&\overset{\spadesuit}{=} \max\{\mu_{\infty,[\eta]}([c]+d_1A), \mu_{\infty,[\eta]}([c]+d_2A)\},
\end{aligned}
$$

where equalities $\overset{\spadesuit}{=}$ hold by Theorem 5.4.2(ii) and the equality $\overset{\clubsuit}{=}$ holds by Lemma 5.8.1. Thus, formula (5.11) is proved.

Regarding formula (5.13), we compute

$$
\begin{aligned}
&\max_{d\in[d_1,d_2]^n} \mu_{\infty,[\eta]}([c]+A[d]) \\
&= \max_{d\in[d_1,d_2]^n} \max_{i\in\{1,\dots,n\}} c_i + d_iA_{ii} + \sum_{j\neq i} |d_jA_{ij}|\frac{\eta_i}{\eta_j} \\
&\leq \max_{i\in\{1,\dots,n\}} \max_{d\in[d_1,d_2]^n} c_i + d_iA_{ii} + \overline{d}\sum_{j\neq i} |A_{ij}|\frac{\eta_i}{\eta_j} \\
&\overset{\bigstar}{=} \max_{i\in\{1,\dots,n\}} \max\{c_i + \gamma A_{ii} + \overline{d}\sum_{j\neq i}|A_{ij}|\frac{\eta_i}{\eta_j} \mid \gamma\in\mathcal{D}\},
\end{aligned}
$$

where the equality $\overset{\bigstar}{=}$ holds because the function $d_i \mapsto d_iA_{ii} + \overline{d}\sum_{j\neq i}|A_{ij}|\eta_i/\eta_j$ is convex. Since the maximum value of a convex function over an interval $d_i\in[d_1,d_2]$ occurs at one of the endpoints, the equality $\overset{\bigstar}{=}$ holds.

Additionally, note that for any $\gamma \in \mathbb{R}$,

$$\gamma A_{ii} + \overline{d} \sum_{j \neq i} |A_{ij}| \frac{\eta_i}{\eta_j} = \overline{d} r_i - (\overline{d} - \gamma) A_{ii}.$$

Therefore,

$$\begin{aligned}
&\max_{d \in [d_1, d_2]^n} \mu_{\infty,[\eta]}([c] + A[d]) \\
&\leq \max_{i \in \{1,\dots,n\}} \max\{c_i + \overline{d} r_i - (\overline{d} - \gamma) A_{ii} \mid \gamma \in \mathcal{D}\} \\
&= \max\{\mu_{\infty,[\eta]}([c] + \overline{d} A - (\overline{d} - \gamma)(I_n \circ A)) \mid \gamma \in \mathcal{D}\}.
\end{aligned}$$

To see that this inequality is tight, suppose $\gamma \in \mathcal{D}$ satisfies $|\gamma| = \overline{d}$. Then let:

$$\begin{aligned}
k &\in \operatorname*{arg\,max}_{i \in \{1,\dots,n\}} c_i + d_1 A_{ii} + \sum_{j \neq i} |\gamma A_{ij}| \frac{\eta_i}{\eta_j}, \\
m &\in \operatorname*{arg\,max}_{i \in \{1,\dots,n\}} c_i + d_2 A_{ii} + \sum_{j \neq i} |\gamma A_{ij}| \frac{\eta_i}{\eta_j}.
\end{aligned}$$

Let $\mathbb{e}_k$ and $\mathbb{e}_m$ be unit vectors with 1 in their k-th and m-th entry, respectively, and define

$$d_k = \gamma \mathbb{1}_n - (\gamma - d_1)\mathbb{e}_k, \quad d_m = \gamma \mathbb{1}_n - (\gamma - d_2)\mathbb{e}_m.$$

Then by construction,

$$\begin{aligned}
\mu_{\infty,[\eta]}([c] + A[d_k]) &= c_k + d_1 A_{kk} + \sum_{j \neq k} |\gamma A_{kj}| \frac{\eta_k}{\eta_j} \\
&= c_k + (d_1 - \overline{d}) A_{kk} + \overline{d} A_{kk} + \overline{d} \sum_{j \neq k} |A_{kj}| \frac{\eta_k}{\eta_j} \\
&= c_k + \overline{d} r_k - (\overline{d} - d_1) A_{kk} \\
&= \mu_{\infty,[\eta]}([c] + \overline{d} A - (\overline{d} - d_1)(I_n \circ A)).
\end{aligned} \tag{5.26}$$

Analogously, we have that $\mu_{\infty,[\eta]}([c] + A[d_m]) = \mu_{\infty,[\eta]}([c] + \overline{d}A - (\overline{d} - d_2)(I_n \circ A))$.

Additionally, we see

$$\begin{aligned}
&\max_{d\in[d_1,d_2]^n} \mu_{\infty,[\eta]}([c] + A[d]) \\
&\geq \max\{\mu_{\infty,[\eta]}([c] + A[d_k]), \mu_{\infty,[\eta]}([c] + A[d_m])\} \\
&\overset{(5.26)}{=} \max\{\mu_{\infty,[\eta]}([c] + \overline{d}A - (\overline{d} - d_1)(I_n \circ A)), \\
&\qquad\qquad \mu_{\infty,[\eta]}([c] + \overline{d}A - (\overline{d} - d_2)(I_n \circ A))\}.
\end{aligned}$$

The proofs for (5.12) and (5.14) are straightforward applications of the fact that $\mu_{1,[\eta]}(B) = \mu_{\infty,[\eta]^{-1}}(B^\top)$ and by applying (5.11) and (5.13), respectively. □

Corollary 5.8.2 (Some simplifications)**.** *Using the same notation as in Lemma 5.4.3, suppose*

(i) $\overline{d} = d_2$ *(note that this implies* $d_2 \geq 0$*). Then*

$$\max_{d\in[d_1,d_2]^n} \mu_{\infty,[\eta]}([c] + A[d]) = \max\{\mu_{\infty,[\eta]}([c] + d_2A), \\ \mu_{\infty,[\eta]}([c] + d_2A - (d_2 - d_1)(I_n \circ A))\}.$$

(ii) $\overline{d} = -d_1$ *(note that this implies* $d_1 \leq 0$*). Then*

$$\max_{d\in[d_1,d_2]^n} \mu_{\infty,[\eta]}([c] + A[d]) = \max\{\mu_{\infty,[\eta]}([c] + d_1A), \\ \mu_{\infty,[\eta]}([c] + d_1A - (d_1 - d_2)(I_n \circ A))\}.$$

5.8.3 Proof of Theorem 5.4.2

Proof. From [168, Theorem 8.1.18], for all $A \in \mathbb{R}^{n\times n}$, we have

$$\rho(A) \leq \rho(|A|), \tag{5.27}$$

where ρ denotes the spectral radius of a matrix. Regarding statement (i), pick $\gamma > \max_i |A_{ii}|$ and define $\bar{A} = A + \gamma I_n$ so that $|\bar{A}| = \lceil A \rceil_{\mathrm{Mzr}} + \gamma I_n$. We note that $\alpha(\bar{A}) \leq \rho(\bar{A})$ (which is true for any matrix) and, from inequality (5.27), we know

$$\begin{aligned} \alpha(A) + \gamma = \alpha(\bar{A}) \leq \rho(\bar{A}) \leq \rho(|\bar{A}|) = \alpha(|\bar{A}|) \\ = \alpha(\lceil A \rceil_{\mathrm{Mzr}} + \gamma I_n) = \alpha(\lceil A \rceil_{\mathrm{Mzr}}) + \gamma. \end{aligned} \tag{5.28}$$

Here $\rho(|\bar{A}|) = \alpha(|\bar{A}|)$ follows from the Perron-Frobenius Theorem for non-negative matrices. This proves statement (i).

Regarding statement (ii), note that the norm $\|\cdot\|_{p,[\eta]}$ is monotonic, it is easy to see that, for all matrices B, we have $\|B\|_{p,[\eta]} \leq \||B|\|_{p,[\eta]}$. For small $h > 0$, we note $|I_n + hA| = I_n + h\lceil A \rceil_{\mathrm{Mzr}}$ so that

$$\|I_n + hA\|_{p,[\eta]} \leq \||I_n + hA|\|_{p,[\eta]} = \|I_n + h\lceil A \rceil_{\mathrm{Mzr}}\|_{p,[\eta]}.$$

Therefore, for small enough $h > 0$,

$$\frac{\|I_n + hA\|_{p,[\eta]} - 1}{h} \leq \frac{\|I_n + h\lceil A \rceil_{\mathrm{Mzr}}\|_{p,[\eta]} - 1}{h}.$$

Thus, statement (ii) follows from the definition of the log norm in the limit as $h \to 0^+$. For $p \in \{1, \infty\}$, statement (ii) holds by the formulas for $\mu_{1,[\eta]}$ and $\mu_{\infty,[\eta]}$.

Finally, regarding statement (iii), for $p \in \{1, \infty\}$, by statement (ii) we have

$$\inf_{\eta \in \mathbb{R}^n_{>0}} \mu_{p,[\eta]}(A) = \inf_{\eta \in \mathbb{R}^n_{>0}} \mu_{p,[\eta]}(\lceil A \rceil_{\mathrm{Mzr}}).$$

Moreover, since $\lceil A \rceil_{\mathrm{Mzr}}$ is Metzler, by Lemma 5.3.1, $\inf_{\eta \in \mathbb{R}^n_{>0}} \mu_{p,[\eta]}(\lceil A \rceil_{\mathrm{Mzr}}) = \alpha(\lceil A \rceil_{\mathrm{Mzr}})$.

□

5.8.4 Proofs of Lemma 5.4.5 and Corollary 5.4.6

Proof of Lemma 5.4.5. Regarding statement (i), let $D_{\mathcal{I}}$ denote the diagonal matrix with entries $(D_{\mathcal{I}})_{ii} = 1$ if $i \in \mathcal{I}$ and $(D_{\mathcal{I}})_{ii} = 0$ if $i \notin \mathcal{I}$.

With this notation, we are ready to compute

$$\begin{aligned}
\|A_{\mathcal{I}}\|_{\mathcal{I}} &= \max_{y \in \mathbb{R}^{|\mathcal{I}|}, \|y\|_{\mathcal{I}}=1} \|A_{\mathcal{I}}\, y\|_{\mathcal{I}} && (5.29)\\
&= \max_{y \in \mathbb{R}^{|\mathcal{I}|}, \|y\|_{\mathcal{I}}=1} \|\operatorname{pad}_{\mathcal{I}}(A_{\mathcal{I}} y)\| && (5.30)\\
&= \max_{y \in \mathbb{R}^{|\mathcal{I}|}, \|\operatorname{pad}_{\mathcal{I}}(y)\|=1} \|(D_{\mathcal{I}} A D_{\mathcal{I}})\, \operatorname{pad}_{\mathcal{I}}(y)\| && (5.31)\\
&\leq \max_{x \in \mathbb{R}^n, \|x\|=1} \|(D_{\mathcal{I}} A D_{\mathcal{I}})\, x\| && (5.32)\\
&= \|D_{\mathcal{I}} A D_{\mathcal{I}}\| \leq \|D_{\mathcal{I}}\| \|A\| \|D_{\mathcal{I}}\| = \|A\|. && (5.33)
\end{aligned}$$

The last equality holds because the monotonicity of $\|\cdot\|$ implies $\|D_{\mathcal{I}}\| = 1$. This concludes the proof of (i).

Statement (ii) follows from the definition of log norm and applying statement (i) to the matrix $I_{|\mathcal{I}|} + hA_{\mathcal{I}}$ as a principal submatrix of $I_n + hA$:

$$\mu_{\mathcal{I}}(A_{\mathcal{I}}) := \lim_{h \to 0^+} \frac{\|I_{|\mathcal{I}|} + hA_{\mathcal{I}}\|_{\mathcal{I}} - 1}{h} \leq \lim_{h \to 0^+} \frac{\|I_n + hA\| - 1}{h} = \mu(A).$$

Finally, statement (iii) is an immediate consequence of (ii). ☐

Proof of Corollary 5.4.6. Regarding item (i), since $A \in \mathcal{MH}$, $\alpha(\lceil A \rceil_{\mathrm{Mzr}}) < 0$. By Lemma 5.3.1 and Theorem 5.4.2(ii), for sufficiently small $\epsilon > 0$, there exists $\eta \in \mathbb{R}^n_{>0}$ such that $\mu_A(1,[\eta]) = \mu_{\lceil A \rceil_{\mathrm{Mzr}}}(1,[\eta]) \leq \alpha(\lceil A \rceil_{\mathrm{Mzr}}) + \epsilon < 0$. Then by Lemma 5.4.5(ii), for non-empty $\mathcal{I} \subset \{1,\dots,n\}$, $\mu_{\mathcal{I},1,[\eta]}(A_{\mathcal{I}}) \leq \mu_A(1,[\eta]) < 0$. Moreover, by Theorem 5.4.2(ii), $\alpha(\lceil A_{\mathcal{I}} \rceil_{\mathrm{Mzr}}) \leq \mu_{\mathcal{I},1,[\eta]}(\lceil A_{\mathcal{I}} \rceil_{\mathrm{Mzr}}) = \mu_{\mathcal{I},1,[\eta]}(A_{\mathcal{I}}) < 0$. We conclude that $A_{\mathcal{I}} \in \mathcal{MH}$. Regarding item (ii), note that $A - A_{ij}\mathbb{e}_{ij}$ is the matrix A with its ij-th entry zeroed out. Then since $A \in \mathcal{MH}$, for sufficiently small $\epsilon > 0$, there exists $\eta \in \mathbb{R}^n_{>0}$ such that $\mu_{1,[\eta]}(A) < 0$. The result is then a consequence of the fact that $\alpha(\lceil A - A_{ij}\mathbb{e}_{ij} \rceil_{\mathrm{Mzr}}) \leq \mu_{1,[\eta]}(A - A_{ij}\mathbb{e}_{ij}) \leq \mu_{1,[\eta]}(A) < 0$. ☐

5.8.5 Proof of Theorem 5.5.1

To prove Theorem 5.5.1, we first recall Clarke's generalized Jacobian from nonsmooth analysis.

Definition 13 ([169, Definition 2.6.1])**.** *Let $f\colon U \to \mathbb{R}^m$ be locally Lipschitz on an open set $U \subseteq \mathbb{R}^n$ and let $\Omega_f \subset U$ be the set of points where f is not differentiable. Then* Clarke's generalized Jacobian *at x is*

$$\partial f(x) = \mathrm{conv}\{\lim_{i\to\infty} Df(x_i) \mid x_i \to x \textit{ and } x_i \notin \Omega_f\}. \tag{5.34}$$

The mean-value theorem has the following generalization for locally Lipschitz functions. For any two points $x, y \in \mathbb{R}^n$, denote $[x,y] := \{tx + (1-t)y \mid t \in [0,1]\}$.

Lemma 5.8.3. *For $f\colon U \to \mathbb{R}^m$ locally Lipschitz on an open convex set $U \subseteq \mathbb{R}^n$, let*

$[x, y] \subset U$. Then there exists matrix $A \in \mathbb{R}^{m\times n}$ such that

$$f(x) - f(y) = A(x - y) \text{ and } A \in \operatorname{conv} \bigcup_{u\in[x,y]} \partial f(u). \tag{5.35}$$

Proof. Since $[x, y]$ is a compact subset of $\mathbb{R}^n$, it can be easily seen that there exists a convex open set U_0 such that $[x, y] \subset U_0 \subset \bar{U}_0 \subset U$ (where $\bar{U}_0$ is the closure of U_0). The statement now follows from [169, Proposition 2.6.5]. □

Proof of Theorem 5.5.1. Regarding (ii) $\implies$ (i), Let $x, y \in U$. Since U is convex, $[x, y] \subset U$. Then by Lemma 5.8.3, there exists A satisfying the conditions (5.35). Condition (ii) in implies that $\mu(\cdot)$ does not exceed c on each set $\partial f(u)$ by continuity and convexity of μ, entailing that $\mu(A) \leq c$. Therefore,

$$[\![f(x) - f(y), x - y]\!] = [\![A(x - y), x - y]\!] \leq \mu(A)\|x - y\|^2 \leq c\|x - y\|^2,$$

where we have used Lumer's equality, Lemma 2.2.2. Regarding (i) $\implies$ (ii), let $x \in U$ such that $Df(x)$ exists and let $v \in \mathbb{R}^n$ and $h > 0$. Then by assumption,

$$[\![f(x + hv) - f(x), hv]\!] \leq c\|hv\|^2 \qquad \implies \qquad h\,[\![f(x + hv) - f(x), v]\!] \leq ch^2\|v\|^2,$$

which holds by the weak homogeneity of the weak pairing. Dividing by $h^2 > 0$ and taking the limit as $h \to 0^+$ implies

$$\begin{aligned} \lim_{h\to 0^+} \left[\!\!\left[\frac{f(x + hv) - f(x)}{h}, v\right]\!\!\right] &\leq c\|v\|^2 \\ \implies [\![Df(x)v, v]\!] &\leq c\|v\|^2 \\ \implies \mu(Df(x)) &\leq c, \end{aligned}$$

where the final implication holds by taking the supremum over all $v \in \mathbb{R}^n$ with $\|v\| = 1$ together with Lumer's equality, Lemma 2.2.2. Therefore, statement (ii) holds. $\square$

Chapter 6

Time-Varying Convex Optimization: A Contraction and Equilibrium Tracking Approach

This chapter is accepted and scheduled to appear in the November 2025 edition of the IEEE Transactions on Automatic Control [170].[1]

6.1 Introduction

Problem description and motivation: Mathematical optimization is a fundamental tool in science and engineering research and has pervaded countless application areas. The classical perspective on mathematical optimization is numerical and is motivated through implementation of iterative algorithms on digital devices. An alternative perspective is to view optimization algorithms as dynamical systems and to understand the performance of these algorithms via their dynamical systems properties, e.g., stability and robustness.

Studying optimization algorithms as continuous-time dynamical systems has been an active area of research since the seminal work of Arrow, Hurwicz, and Uzawa [171]. Notable examples include Hopfield and Tank in dynamical neuroscience [172], Kennedy and Chua in analog circuit design [173], and Brockett in systems and control [174]. Recently, the interest in continuous-time dynamics for optimization and computation has been renewed due to the advent of (i) online and dynamic feedback optimization [175], (ii) reservoir computing [176], and (iii) neuromorphic computing [177].

Motivated by these recent developments, we are interested in time-varying convex optimization problems and continuous-time dynamical systems that track their optimal solutions. In many applications of interest, the optimization algorithm must be run in real-time on problems that are time-varying. Such examples include tracking a moving target, estimating the path of a stochastic process, and online learning. In such application areas, we would like our dynamical system to converge to the unique optimal solution when the problem is time-invariant and converge to an explicitly-computable neighborhood of the optimal solution trajectory when the problem is time-varying.

Beyond tracking optimal trajectories, a key desirable feature of optimization algorithms is robustness in the face of uncertainty. In many real-world scenarios, we are not provided the exact value of our cost function but instead noisy estimates and possibly even a time-delayed version of it. Thus, for practical usage of the optimization algorithm, it is essential to ensure that the algorithm has these robustness features *built-in*.

Remarkably, all of these desirable properties, namely tracking for time-varying systems, convergence for time-invariant systems, and robustness to noise and time-delays can be established by ensuring that the dynamical system is *strongly infinitesimally contracting* [23]. To be specific, (i) the effect of the initial condition is exponentially forgotten and the distance between any two trajectories decays exponentially quickly [23], (ii) if a contracting system is time-invariant, it has a unique globally exponentially stable equilibrium point [23], (iii) contracting systems are incrementally input-to-state stable and thus robust to disturbances [178] and delays [179].

Literature review: A recent survey on studying optimization algorithms from a feedback control perspective is available in [180]. Asymptotic and exponential stability of dynamical systems solving convex optimization problems is a classical problem and has been studied in papers including [181, 182, 183, 184, 185] among many others. Compared to papers studying asymptotic and exponential stability, there are far fewer works studying the contractivity of dynamical systems solving optimization problems. A few exceptions include [186, 187] which analyze primal-dual dynamics and [188, 27] which study gradient flows on Riemannian manifolds.

In the context of time-varying convex optimization, algorithms to track the optimal solution are designed based on Newton's method in (i) discrete-time in [189, 190] and in (ii) continuous-time in [191]. See both [192] and [193] and the references therein for reviews of these results and theoretical extensions. These results have been leveraged to study the feedback interconnection of a LTI system and a dynamical system solving an optimization problem in [194]. From a contraction theory perspective, both [186] and [187] provide tracking error bounds for continuous-time time-varying primal-dual dynamics.

Contributions: This chapter makes four main contributions. First, we prove a general theorem regarding parameter-dependent strongly infinitesimally contracting dynamics. Specifically, in Theorem 6.2.1, we prove that both the tracking error, defined as the error between any solution trajectory and the equilibrium trajectory (defined instantaneously in time), and the norm of the vector field are uniformly upper bounded. Moreover, we prove that the tracking error is asymptotically proportional to the rate of change of the parameter with proportionality constant upper bounded by ℓ_θ/c^2 where ℓ_θ is the Lipschitz constant in which the parameter appears and c is the contraction rate of the dynamics. A related result was proved in [186, Lemma 2], but the tracking error bound depends on the knowledge of the rate of change of the equilibrium trajectory, which is unknown, in general. In contrast, Theorem 6.2.1 provides an additional bound on the norm of the vector field and a tracking error bound which depends purely on the rate of change of the parameter, which may be more directly applicable.

Second, in Theorem 6.2.2, we propose an alternative dynamical system which augments the contracting dynamics in Theorem 6.2.1 with a feedforward term. This augmentation ensures that the tracking error is exponentially decaying to zero and does not require any Lipschitz condition on how the parameter appears in the dynamics. A related, continuous-time, treatment is proposed in [191], (see also the early reference [195]) where the authors study a continuous-time Newton method and show how to add a feedforward term to ensure zero tracking error in the Euclidean norm. In discrete-time, the authors of [189, 190] use predictor-corrector methods based on Newton's method with or without projections. Compared to these references, Theorem 6.2.2 is applicable to any contracting dynamics with respect to any norm and need not be limited to the solution of a time-varying optimization problem.

Third, we consider natural transcriptions into contracting dynamics for three canonical strongly convex optimization problems (namely, (i) monotone inclusions, (ii) linear equality-constrained problems, and (iii) composite minimization), and we make specific contributions for each transcription.

(i) For monotone inclusion problems, we consider the *forward-backward splitting dynamics* which were first studied in [196]. These dynamics are a generalization of the projected dynamics studied in [197] and the proximal gradient dynamics studied in [198]. In Theorem 6.3.1, we show that the forward-backward splitting dynamics are contracting, a stronger property than exponential stability as was shown in [197, Theorem 4] and [198, Theorem 2] and show improved rates of exponential convergence in some special cases.

(ii) For linear-equality constrained problems, we study the primal-dual dynamics and prove their strong infinitesimal contractivity in Theorem 6.3.2. Compared to [186], we provide an explicit estimate on the rate of contraction and we show improved rates compared to both [183, 187].

(iii) For composite minimization, we adopt the proximal augmented Lagrangian approach from [185], first introduced in [199], and show that the primal-dual dynamics on the proximal augmented Lagrangian are contracting in Theorem 6.3.3. This result improves on the exponential convergence result from [185, Theorem 3] by allowing for a larger range of parameters. A related result is [183, Theorem 2], which focuses specifically on inequality-constrained minimization problems. Theorem 6.3.3 is a generalization of [183, Theorem 2] to more general composite minimization problems and provides a nonlinear program to estimate the contraction rate. Moreover, to approximate the optimal value of the nonlinear program, we provide a strategy based on a bisection algorithm.

Finally, we apply our general result on tracking error bounds for contracting dynamics from Theorems 6.2.1 and 6.2.2 to each of the aforementioned optimization problems to provide tracking error estimates in time-varying convex optimization problems. To validate our theory, we present numerical and hardware experiments. In Sections 6.4.1 and 6.4.2 we showcase tracking error bounds for time-varying equality and inequality-constrained minimization problems, respectively. In Section 6.4.3, inspired by [200], we present a modern application to online control barrier functions, [201], where we show

how we can leverage our tracking error results for contracting dynamics to ensure safety in a multi-robot collision avoidance scenario without needing to solve a quadratic program at every instance in time.

6.2 Equilibrium tracking for parameter-varying contracting dynamical systems

We begin by considering a dynamical system which is a function of a time-varying parameter, θ. Namely, for a vector field $F\colon \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^n$, consider the system

$$\dot{x}(t) = F\big(x(t), \theta(t)\big), \quad x(0) = x_0 \in \mathbb{R}^n, \tag{6.1}$$

where for all $t \geq 0$, $x(t)$ and $\theta(t)$ take value in $\mathcal{X} \subseteq \mathbb{R}^n$ and $\Theta \subseteq \mathbb{R}^d$, respectively.

We make the following assumptions. There exists a norm $\|\cdot\|_{\mathcal{X}}$ on $\mathcal{X}$ and

(A1) there exists $c > 0$ such that for all θ, the map $x \mapsto F(x, \theta)$ is strongly infinitesimally contracting with respect to $\|\cdot\|_{\mathcal{X}}$ with rate c, i.e., $\mathrm{osL}_x(F) \leq -c$,

(A2) there exists a norm $\|\cdot\|_{\Theta}$ on Θ, and $\ell_\theta \geq 0$ such that for all x, the map $\theta \mapsto F(x, \theta)$ is Lipschitz from $(\Theta, \|\cdot\|_{\Theta})$ to $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ with constant ℓ_θ.

Assumption (A1) implies that, for each $\theta \in \Theta$, there exists a unique $x^\star_\theta \in \mathcal{X}$ satisfying $F(x^\star_\theta, \theta) = \mathbb{0}_n$. Then, we can define the map $x^\star\colon \Theta \to \mathcal{X}$ given by $x^\star(\theta) = x^\star_\theta$. Lemma 6.6.1 in Appendix 6.6 shows that $x^\star(\cdot)$ is Lipschitz from $(\Theta, \|\cdot\|_{\Theta})$ to $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ with constant ℓ_θ/c. With this set up in mind, in the following we define the *time-varying equilibrium curve* which is key in our equilibrium tracking results.

Definition 10 (Time-varying equilibrium curve). *Consider a continuously differentiable curve $\theta\colon \mathbb{R}_{\geq 0} \to \Theta$ and the system (6.1) satisfying Assumptions (A1) and (A2). The* time-varying equilibrium curve *is the map* $t \mapsto x^\star(\theta(t))$.

Since $x^\star(\cdot)$ is Lipschitz, the curve $x^\star(\theta(\cdot))$ is locally Lipschitz (see Lemma 6.6.2 in Appendix 6.6 for details). Additionally, this curve satisfies $F(x^\star(\theta(t)), \theta(t)) = \mathbb{0}_n$ for all $t \geq 0$. In the following theorem, we provide tracking error bounds between any trajectory of (6.1) and the time-varying equilibrium curve.

Theorem 6.2.1 (Equilibrium tracking for contracting dynamics)**.** *Let* $\theta\colon \mathbb{R}_{\geq 0} \to \Theta$ *be continuously differentiable and consider the dynamics* (6.1) *satisfying Assumptions* (A1) *and* (A2)*. Let* $x^\star(\theta(\cdot))$ *be the* time-varying equilibrium curve *of* (6.1)*. Then, for any initial conditions* $x(0) \in \mathbb{R}^n$*,* $\theta(0) \in \mathbb{R}^d$ *and for all* $t \geq 0$*:*

(i) the tracking error $\|x(t) - x^\star(\theta(t))\|_{\mathcal{X}}$ *satisfies*

$$\|x(t) - x^\star(\theta(t))\|_{\mathcal{X}} \leq \mathrm{e}^{-ct}\|x(0) - x^\star(\theta(0))\|_{\mathcal{X}} + \frac{\ell_\theta}{c}\int_0^t \mathrm{e}^{-c(t-\tau)}\|\dot{\theta}(\tau)\|_{\Theta} d\tau;$$

(ii) the residual $\|F(x(t), \theta(t))\|_{\mathcal{X}}$ *satisfies*

$$\|F(x(t), \theta(t))\|_{\mathcal{X}} \leq \mathrm{e}^{-ct}\|F(x(0), \theta(0))\|_{\mathcal{X}} + \ell_\theta\int_0^t \mathrm{e}^{-c(t-\tau)}\|\dot{\theta}(\tau)\|_{\Theta} d\tau;$$

(iii) the following asymptotic bounds hold:

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

Proof. To prove item (i), consider the auxiliary dynamics

$$\dot{x}(t) = F(x(t), \theta(t)) + v(t) := T(x(t), \theta(t), v(t)), \tag{6.2}$$

where $T\colon \mathbb{R}^n \times \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^n$ and $v\colon \mathbb{R}_{\geq 0} \to \mathcal{X}$. Note that by Assumption (A1), for fixed θ and v, the map $x \mapsto T(x, \theta, v)$ is strongly infinitesimally contracting with rate $c > 0$. Moreover, at fixed x, θ, the map $v \mapsto T(x, \theta, v)$ is Lipschitz on $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ with constant $\ell_v = 1$. Consider the inputs $v_1(t) = \mathbb{0}_n$ and $v_2(t) = \dot{x}^\star(\theta(t))$ and note that $\dot{x}^\star(\theta(t)) = F(x^\star(\theta(t)), \theta(t)) + \dot{x}^\star(\theta(t))$ so that the curve $x^\star(\theta(\cdot))$ is a solution to the dynamical system (6.2) with input $v_2(t)$ and initial condition $x^\star(\theta(0))$. Additionally, for any initial condition $x(0) \in \mathcal{X}$, the solution $x(t)$ to the dynamics (6.1) is a solution to the system (6.2) with input $v_1(t)$. By an application of the incremental ISS theorem for contracting dynamical systems, Theorem 2.5.1, to the trajectories $x(\cdot), x^\star(\theta(\cdot))$ arising from inputs $v_1(\cdot), v_2(\cdot)$, we have the bound for a.e. t

$$\begin{aligned}
D^+\|x(t) - x^\star(\theta(t))\|_{\mathcal{X}} &\leq -c\|x(t) - x^\star(\theta(t))\|_{\mathcal{X}} + \|\dot{x}^\star(\theta(t))\|_{\mathcal{X}} \\
&\leq -c\|x(t) - x^\star(\theta(t))\|_{\mathcal{X}} + \frac{\ell_\theta}{c}\|\dot{\theta}(t)\|_\Theta,
\end{aligned}$$

where the last inequality follows from Lemma 6.6.2. Item (i) is then a consequence of the Grönwall inequality for Dini derivatives, e.g. [5, Lemma 11]. To prove item (ii), consider a trajectory $x(t)$ of (6.1) and let $V(t) = \|F(x(t), \theta(t))\|_{\mathcal{X}}$. Then, omitting dependencies of x and θ on time, we compute

$$\begin{aligned}
&D^+V(t) \overset{(\star)}{=} \lim_{h \to 0^+} \frac{\|F(x,\theta) + h\frac{d}{dt}F(x,\theta)\|_{\mathcal{X}} - \|F(x,\theta)\|_{\mathcal{X}}}{h} \\
&\overset{(\triangle)}{\leq} \lim_{h \to 0^+} \frac{\|F(x,\theta) + hD_xF(x,\theta)F(x,\theta)\|_{\mathcal{X}} - \|F(x,\theta)\|_{\mathcal{X}}}{h} + \|D_\theta F(x,\theta)\dot{\theta}\|_{\mathcal{X}} \\
&\overset{\text{(A2)}}{\leq} \|F(x,\theta)\|_{\mathcal{X}} \lim_{h \to 0^+} \frac{\|I_n + hD_xF(x,\theta)\|_{\mathcal{X}} - 1}{h} + \ell_\theta\|\dot{\theta}\|_\Theta \\
&\leq \mu_{\mathcal{X}}(D_xF(x,\theta)V(t) + \ell_\theta\|\dot{\theta}\|_\Theta \overset{\text{(A1)}}{\leq} -cV(t) + \ell_\theta\|\dot{\theta}\|_\Theta,
\end{aligned}$$

where $(\star)$ holds by a Taylor expansion of F in t, inequality $(\triangle)$ is a consequence of $\frac{d}{dt}F(x,\theta) = D_xF(x,\theta)\dot{x} + D_\theta F(x,\theta)\dot{\theta}$ and the triangle inequality, inequalities (A2) and (A1) are a consequence of Assumptions (A2) and (A1), respectively. Item (ii) then follows by the Grönwall inequality. Item (iii) is a consequence of items (i) and (ii). $\square$

Theorem 6.2.1 is a general result that establishes that one does not need to know $\dot{x}^\star(\theta(t))$ in order to get an estimate on the tracking error. Indeed, by using Lemma 6.6.1, we know that the map $x^\star$ has Lipschitz bound ℓ_θ/c and this is one of the key steps in establishing the asymptotic bound (iii). This bound gives designers insight on how they may speed up their dynamics to provide lower values of tracking error.

Additionally, if we have knowledge of $\dot{\theta}$, we can augment the contracting dynamics (6.1) with a feedforward term that ensures an exponential decay to zero tracking error. To do so, consider a parameter-dependent vector field $F\colon \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^n$ continuously differentiable in both arguments and satisfying Assumption (A1). Let $\theta\colon \mathbb{R}_{\geq 0} \to \Theta \subseteq \mathbb{R}^d$ be continuously differentiable. We introduce the *time-varying contracting dynamics with feedforward prediction*:

$$\dot{x}(t) = F\big(x(t),\theta(t)\big) - \big(D_xF(x(t),\theta(t))\big)^{-1} D_\theta F(x(t),\theta(t))\dot{\theta}(t). \tag{6.3}$$

Assumption (A1) implies the inequality $\mu(D_xF(x,\theta)) \leq -c$, for all x,θ. From (1.3d), we know that the eigenvalues of $D_xF(x,\theta)$ are in the open left half plane, which implies invertibility of $D_xF(x,\theta)$ and ensures that the dynamics (6.3) are well-posed.

In the following result, we show that considering the dynamics (6.3) we obtain exponential decay to zero tracking error.

Theorem 6.2.2 (Exact tracking with feedforward prediction). *Let $F\colon \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^n$ be a parameter-dependent vector field, and let $\theta\colon \mathbb{R}_{\geq 0} \to \Theta \subseteq \mathbb{R}^d$ be continuously differentiable. Assume F is continuously differentiable in both arguments and satisfies Assumption (A1).*

Consider the dynamics (6.3)*. Then for all* $t \geq 0$*,*

(i) *the residual* $\|F(x(t), \theta(t))\|_{\mathcal{X}}$ *satisfies*

$$\|F(x(t), \theta(t))\|_{\mathcal{X}} \leq \mathrm{e}^{-ct}\|F(x(0), \theta(0))\|_{\mathcal{X}};$$

(ii) *the tracking error* $\|x(t) - x^{\star}(\theta(t))\|_{\mathcal{X}}$ *satisfies*

$$\|x(t) - x^{\star}(\theta(t))\|_{\mathcal{X}} \leq \frac{1}{c}\mathrm{e}^{-ct}\|F(x(0), \theta(0))\|_{\mathcal{X}};$$

Additionally, if F *is Lipschitz in its first argument with constant* ℓ_x *uniformly in* θ*, then*

$$\|x(t) - x^{\star}(\theta(t))\|_{\mathcal{X}} \leq \frac{\ell_x}{c}\mathrm{e}^{-ct}\|x(0) - x^{\star}(\theta(0))\|_{\mathcal{X}}.$$

Proof. To prove item (i), consider a trajectory $x(t)$ of (6.3), and let $V(t) = \|F(x(t), \theta(t))\|_{\mathcal{X}}$. Then, omitting dependencies of x and θ on time, we compute

$$\begin{aligned}
D^{+}V(t) &\overset{(\star)}{=} \lim_{h \to 0^{+}} \frac{\|F(x, \theta) + h\frac{d}{dt}F(x, \theta)\|_{\mathcal{X}} - \|F(x, \theta)\|_{\mathcal{X}}}{h} \\
&\overset{(6.3)}{=} \lim_{h \to 0^{+}} \frac{\|F(x, \theta) + hD_xF(x, \theta)F(x, \theta)\|_{\mathcal{X}} - \|F(x, \theta)\|_{\mathcal{X}}}{h} \\
&\leq \|F(x, \theta)\|_{\mathcal{X}} \lim_{h \to 0^{+}} \frac{\|I_n + hD_xF(x, \theta)\|_{\mathcal{X}} - 1}{h} \leq \mu_{\mathcal{X}}(D_xF(x, \theta))V(t) \leq -cV(t),
\end{aligned}$$

where $(\star)$ is by a Taylor expansion of F in t and the next equality holds since (6.3) implies that $\frac{d}{dt}F(x, \theta) = D_xF(x, \theta)\dot{x} + D_\theta F(x, \theta)\dot{\theta} = D_xF(x, \theta)F(x, \theta)$. The Grönwall inequality for Dini derivatives implies item (i). Item (ii) is a consequence of the fact that $\|F(x, \theta) - F(x^{\star}(\theta), \theta)\|_{\mathcal{X}} \geq c\|x - x^{\star}(\theta)\|_{\mathcal{X}}$ since $F(x^{\star}(\theta), \theta) = \mathbb{0}_n$ and for fixed θ, the map $x \mapsto F(x, \theta)$ is invertible and the inverse map is Lipschitz on $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ with constant $1/c$ [2, Lemma 3.5]. □

Note that compared to Theorem 6.2.1, Theorem 6.2.2 does not require Assumption (A2) but does additionally require differentiability of F. By assuming knowledge of $\dot{\theta}$, we can leverage this information to exponentially achieve zero tracking error for arbitrary contracting dynamics, F. Comparatively, in Theorem 6.2.1, we do not assume knowledge of $\dot{\theta}$ and cannot expect to achieve zero tracking error as a result.

6.3 Contracting dynamics for canonical convex optimization problems

In this section, we provide a transcription from three canonical optimization problems to continuous-time dynamical systems which are strongly infinitesimally contracting. This transcription allows us to apply Theorems 6.2.1 and 6.2.2 to time-varying instances of these problems. Specifically, we analyze (i) monotone inclusions, (ii) linear equality constrained problems, and (iii) composite minimization problems.

6.3.1 Monotone inclusions

We consider the following general problem which has found many applications in convex optimization, see [74].

Problem 1. *Let* $\mathsf{F}\colon \mathbb{R}^n \to \mathbb{R}^n$ *be monotone and* $g\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ *be convex. We are interested in solving the* monotone inclusion problem

$$\text{Find } x^\star \in \mathbb{R}^n \text{ s.t. } \quad \mathbb{0}_n \in (\mathsf{F} + \partial g)(x^\star). \tag{6.4}$$

We make the following assumptions on F and g:

Assumption 3. $\mathsf{F}\colon \mathbb{R}^n \to \mathbb{R}^n$ *is strongly monotone with parameter* m *and Lipschitz on* $(\mathbb{R}^n, \|\cdot\|_2)$ *with constant* ℓ. *The map* $g\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ *is CCP.* △

Under Assumption 3, the monotone inclusion problem (6.4) has a unique solution due to strong monotonicity of F.

Monotone inclusion problems of the form (6.4) are prevalent in convex optimization and data science and we present two canonical problems which can be stated in terms of the monotone inclusion (6.4).

Example 7 (Convex minimization)**.** *First, consider the convex optimization problem*

$$\min_{x\in\mathbb{R}^n} f(x) + g(x), \tag{6.5}$$

where $f\colon \mathbb{R}^n \to \mathbb{R}$ *is strongly convex and continuously differentiable and* $g\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ *is CCP. In this case, the unique point* $x^* \in \mathbb{R}^n$ *that minimizes* (6.5) *also solves the inclusion problem* (6.4) *with* $\mathsf{F} = \nabla f$.

Example 8 (Variational inequalities)**.** *Second, consider the* variational inequality *defined by the continuous monotone mapping* $\mathsf{F}\colon \mathbb{R}^n \to \mathbb{R}^n$ *and nonempty, convex, and closed set* $\mathcal{C}$ *which is the problem*

$$\textit{Find } x^\star \in \mathcal{C} \textit{ s.t.} \quad \mathsf{F}(x^\star)^\top (x - x^\star) \geq 0, \quad \forall x \in \mathcal{C}. \tag{6.6}$$

We denote problem (6.6) *by* $\mathrm{VI}(\mathsf{F}, \mathcal{C})$. *It is known that* $x^* \in \mathcal{C}$ *solves* $\mathrm{VI}(\mathsf{F}, \mathcal{C})$ *if and only if for all* $\gamma > 0$, x^* *is a fixed point of the map* $\mathrm{P}_{\mathcal{C}} \circ (\mathsf{Id} - \gamma\mathsf{F})$, *i.e.,* $x^* = \mathrm{P}_{\mathcal{C}}(x^* - \gamma\mathsf{F}(x^*))$. *In turn, this fixed-point condition is equivalent to asking* x^* *to solve the monotone inclusion problem* (6.4) *with* $g = \iota_{\mathcal{C}}$, *see, e.g., [72, pp. 37]. Variational inequalities of the form* (6.6) *have found applications in computing Nash and Wardrop equilibria in games [202].*

To solve the monotone inclusion problem (6.4), we consider the following dynamics from [196], called the *continuous-time forward-backward splitting dynamics* with parameter $\gamma > 0$:

$$\dot{x} = -x + \mathrm{prox}_{\gamma g}(x - \gamma \mathsf{F}(x)) =: \mathsf{F}_{\mathrm{FB}}^{\gamma}(x). \tag{6.7}$$

The name forward-backward splitting dynamics comes from the classical forward-backward splitting algorithm from monotone operator theory, see, e.g., [4, Section 26.5].

Remark 6.3.1. *When $g = \iota_{\mathcal{C}}$ for some convex and closed set $\mathcal{C}$, then for any $\gamma > 0$, $\mathrm{prox}_{\gamma g} = \mathrm{P}_{\mathcal{C}}$, we are solving* VI($\mathsf{F}, \mathcal{C}$) (6.6) *and the dynamics* (6.7) *are the* projected dynamics

$$\dot{x} = -x + \mathrm{P}_{\mathcal{C}}(x - \gamma \mathsf{F}(x)),$$

which were studied in [197]. Alternatively, when $\mathsf{F} = \nabla f$ for some continuously differentiable convex function f, we are solving the convex optimization problem (6.5) *and the dynamics* (6.7) *correspond to the* proximal gradient dynamics

$$\dot{x} = -x + \mathrm{prox}_{\gamma g}(x - \gamma \nabla f(x)), \tag{6.8}$$

which were studied in [198]. △

First, we establish that equilibrium points of the dynamics (6.7) correspond to solutions of the inclusion problem (6.4).

Proposition 6.3.2 (Equilibria of (6.7))**.** *Suppose Assumption 3 holds. Then for any $\gamma > 0$, $\mathbb{0}_n \in (\mathsf{F} + \partial g)(x^*)$ if and only if $x^* \in \mathbb{R}^n$ is an equilibrium point of the dynamics* (6.7).

Proof. Note that equilibria of (6.7), $x^* \in \mathbb{R}^n$, satisfy the fixed point equation

$$x^* = \mathrm{prox}_{\gamma g}(x^* - \gamma \mathsf{F}(x^*)). \tag{6.9}$$

Moreover, it is known that fixed points of the form (6.9) also solve the monotone inclusion $\mathbb{0}_n \in (\mathsf{F}+\partial g)(x^*)$, see, e.g., [4, Proposition 26.1(iv)(a)], noting that $\mathrm{prox}_{\gamma g}$ is the resolvent of ∂g with parameter γ. $\square$

Remark 6.3.3. *Proposition 6.3.2 continues to hold under the assumption of monotone* F *[4, Propositon 26.1(iv)(a)].* $\triangle$

Next, we establish that the dynamics (6.7) are contracting under assumptions on the parameter γ.

Theorem 6.3.1 (Contractivity of (6.7))**.** *Suppose Assumption 3 holds. Then*

(i) for every $\gamma \in (0, 2m/\ell^2)$, the dynamics (6.7) are strongly infinitesimally contracting with respect to $\|\cdot\|_2$ with rate $1-\sqrt{1-2\gamma m+\gamma^2\ell^2}$. Moreover, the contraction rate is optimized at $\gamma^* = m/\ell^2$.

Additionally,

(ii) if $\mathsf{F} = \nabla f$ for some strongly convex $f\colon \mathbb{R}^n \to \mathbb{R}$, for every $\gamma \in (0, 2/\ell)$, the dynamics (6.7) are strongly infinitesimally contracting with respect to $\|\cdot\|_2$ with rate $1-\max\{|1-\gamma m|, |1-\gamma\ell|\}$. Moreover, the contraction rate is optimized at $\gamma^* = 2/(m+\ell)$;

(iii) if $\mathsf{F}(x) = Ax+b$ for all $x \in \mathbb{R}^n$, with $A = A^\top \succ 0$, then for every $\gamma \in (1/\lambda_{\min}(A), +\infty)$, the dynamics (6.7) are strongly infinitesimally contracting with respect to the norm $\|\cdot\|_{(\gamma A - I_n)}$ with rate 1.

Proof. Regarding item (i) note that since $g\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ is CCP, for every $\gamma > 0$, $\mathrm{prox}_{\gamma g}$ is a nonexpansive map with respect to the $\|\cdot\|_2$ norm [4, Proposition 12.28]. Moreover, for every $\gamma > 0$ the map $\mathsf{Id} - \gamma\mathsf{F}$ has Lipschitz constant upper bounded by $\sqrt{1-2\gamma m+\gamma^2\ell^2}$

with respect to the $\|\cdot\|_2$ norm [72, pp. 16]. Since the Lipschitz constant of the composition of the two maps is upper bounded by the product of the Lipschitz constants, in light of Lemma 6.6.3 in Appendix 6.6, we conclude that for $\gamma \in (0, 2m/\ell^2)$, $\mathrm{osL}(\mathsf{F}_{\mathrm{FB}}^{\gamma}) \leq -1 + \mathsf{Lip}(\mathrm{prox}_{\gamma g} \circ (\mathsf{Id} - \gamma \mathsf{F})) \leq -1 + \sqrt{1 - 2\gamma m + \gamma^2 \ell^2} < 0$. Moreover, minimizing $\mathrm{osL}(\mathsf{F}_{\mathrm{FB}}^{\gamma})$ corresponds to minimizing $1 - 2\gamma m + \gamma^2 \ell^2$ as a function of $\gamma \in (0, 2m/\ell^2)$. This minimization occurs at $\gamma^* = m/\ell^2$ and yields a one-sided Lipschitz estimate of $\mathrm{osL}(\mathsf{F}_{\mathrm{FB}}^{\gamma^*}) \leq -1 + \sqrt{1 - m^2/\ell^2} < 0$.

Item (ii) follows the same argument as in item (i) where instead one shows that for all $\gamma > 0$, $\mathsf{Lip}(\mathsf{Id} - \gamma \nabla f) \leq \max\{|1 - \gamma m|, |1 - \gamma \ell|\}$ as in, e.g., [72, pp. 15]. Then for all $\gamma \in (0, 2/\ell)$, $\mathrm{osL}(\mathsf{F}_{\mathrm{FB}}^{\gamma}) \leq -1 + \max\{|1 - \gamma m|, |1 - \gamma \ell|\} < 0$. Moreover, the optimal choice of γ is $\gamma^* = 2/(m + \ell)$ and the corresponding bound on $\mathrm{osL}(\mathsf{F}_{\mathrm{FB}}^{\gamma^*})$ is $-1 + (\kappa - 1)/(\kappa + 1)$, where $\kappa := \ell/m \geq 1$ [72, pp. 15].

Regarding item (iii), we compute the Jacobian of $\mathsf{F}_{\mathrm{FB}}^{\gamma}$ for all $x \in \mathbb{R}^n$ for which it exists, i.e., $D\mathsf{F}_{\mathrm{FB}}^{\gamma}(x) = -I_n + D\mathrm{prox}_{\gamma g}(I_n - \gamma(Ax + b))(I_n - \gamma A)$. Note that for all $x \in \mathbb{R}^n$ for which the Jacobian exists, there exists $G = G^\top \in \mathbb{R}^{n \times n}$ with $0 \preceq G \preceq I_n$ satisfying $D\mathrm{prox}_{\gamma g}(I_n - \gamma(Ax + b)) = G$, see Lemma 6.6.4 in Appendix 6.6. Additionally, we recall that, for any matrix A, the log-norm translation property holds. That is, $\mu(A + cI_n) = \mu(A) + c$, for all $c \in \mathbb{R}$. Then for any norm,

$$\sup_x \mu(D\mathsf{F}_{\mathrm{FB}}^{\gamma}(x)) \leq -1 + \max_{0 \preceq G \preceq I_n} \mu(G(I_n - \gamma A)), \tag{6.10}$$

where the sup is over all x for which $D\mathsf{F}_{\mathrm{FB}}^{\gamma}(x)$ exists. Moreover, for $\gamma > 1/\lambda_{\min}(A)$, $\gamma A - I_n$ is positive definite and $G(I_n - \gamma A) = (-G)(\gamma A - I_n)$ is the product of two symmetric matrices. An application of Sylvester's law of inertia implies that $G(I_n - \gamma A)$ has all real eigenvalues and that it has the same number of positive, zero, and negative eigenvalues as $-G$ does, i.e., all eigenvalues are nonpositive. Then from [203, Lemma 2], with the choice of norm $\|\cdot\|_{(\gamma A - I_n)}$, we find

$$\begin{aligned}\mu_{(\gamma A - I_n)}(G(I_n - \gamma A)) &= \mu_{(\gamma A - I_n)}((-G)(\gamma A - I_n)) \\ &= \max\{\mathrm{Re}(\lambda) \mid \lambda \text{ is an eigenvalue of } G(I_n - \gamma A)\} \leq 0,\end{aligned}$$

where in the last equality we used the definition of ℓ_2 log-norm. Since this equality holds for all symmetric G satisfying $0 \preceq G \preceq I_n$, by applying inequality (6.10) we have

$$\sup_{x \in \mathbb{R}^n} \mu_{(\gamma A - I_n)}(D\mathsf{F}^{\gamma}_{\mathrm{FB}}(x)) \leq -1.$$

This inequality proves the result. □

Remark 6.3.4. *The rates of contraction in Theorem 6.3.1(i) and (ii) are essentially consequences of the standard contraction rates of the forward-backward splitting algorithm in monotone operator theory, see [72, pp. 25] and [4, Proposition 26.16]. In contrast, Theorem 6.3.1(iii) provides an improved and sharp rate of contraction in the case of affine* F *for an increased range of* γ*. Note that this rate cannot be improved. To see this fact, consider* $g = \iota_{\{b\}}$ *so that the dynamics are* $\dot{x} = -x + b$*, which are contracting with rate equal to* 1*. It is an open question whether the contraction rates in Theorem 6.3.1(i) and (ii) can be improved to* 1 *with different choice of* γ *and norm.* △

Parameter-varying case Consider the *parameter-varying inclusion problem*

$$\text{For } \theta \in \Theta, \text{ find } x^\star(\theta) \in \mathbb{R}^n \text{ s.t. } \mathbb{0}_n \in (\mathsf{F}_\theta + \partial g_\theta)(x^\star(\theta)), \tag{6.11}$$

where for each $\theta \in \Theta$, the map $\mathsf{F}_\theta \colon \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz and strongly monotone and the map $g_\theta \colon \mathbb{R}^n \to \overline{\mathbb{R}}$ is CCP. Then, for suitable $\gamma > 0$, the corresponding parameter-varying forward-backward splitting dynamics are given by

$$\dot{x} = -x + \mathrm{prox}_{\gamma g_\theta}(x - \gamma \mathsf{F}_\theta(x)) =: \mathsf{F}^{\gamma}_{\mathrm{FB}}(x, \theta). \tag{6.12}$$

For each $\theta \in \Theta$, these dynamics are strongly contracting in view of Theorem 6.3.1 and thus the problem (6.11) has a unique solution $x^\star(\theta)$. Moreover, when $\theta\colon \mathbb{R}_{\geq 0} \to \Theta$ is a continuously differentiable curve, under a Lipschitz condition on the map $\theta \mapsto \mathsf{F}^{\gamma}_{\mathrm{FB}}(x, \theta)$, Theorem 6.2.1 ensures that trajectories of (6.12) track $x^\star(\theta(t))$ with a tracking error proportional to $\|\dot{\theta}(t)\|_\Theta$ after a transient. Such a Lipschitz condition holds if, e.g., the maps $\theta \mapsto \mathrm{prox}_{\gamma g_\theta}(x)$ and $\theta \mapsto \mathsf{F}(x, \theta)$ are Lipschitz uniformly in x[2]. If, additionally, $\mathsf{F}^{\gamma}_{\mathrm{FB}}$ is differentiable in both arguments, one can design a feedforward term to attain zero tracking error leveraging Theorem 6.2.2. In Section 6.4.3, we consider an application that takes this approach.

6.3.2 Linear equality constrained optimization

We study another canonical problem in convex optimization.

Problem 2. *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be convex, $A \in \mathbb{R}^{m\times n}$, and $b \in \mathbb{R}^m$. Consider the equality-constrained problem*

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x), \\ s.t. \quad & Ax = b. \end{aligned} \tag{6.13}$$

We make the following assumptions on f and A:

Assumption 4. *$f\colon \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable, strongly convex and strongly smooth with parameters ρ and ℓ, respectively. The matrix $A \in \mathbb{R}^{m\times n}$ is full row-rank and*

[2] To determine whether the map $\theta \mapsto \mathrm{prox}_{\gamma g_\theta}(x)$ is Lipschitz, one can employ sensitivity analysis of parametric programs. We refer the interested reader to [204] for sufficient conditions.

satisfies $a_{\min} I_m \preceq AA^\top \preceq a_{\max} I_m$ *for* $a_{\min}, a_{\max} \in \mathbb{R}_{>0}$. $\triangle$

Note that Problem 2 is a special case of Problem 1 with $\mathsf{F} = \nabla f$ and $g = \iota_{\mathcal{C}}$ where $\mathcal{C} = \{z \in \mathbb{R}^n \mid Az = b\}$. In this case we have that $\operatorname{prox}_{\alpha g} = \mathrm{P}_{\mathcal{C}}$ and $\mathrm{P}_{\mathcal{C}}(z) = z - A^\dagger(Az - b) = (I_n - A^\dagger A)z + A^\dagger b$, where $A^\dagger$ denotes the pseudoinverse of A. In the context of the forward-backward splitting dynamics (6.7), the dynamics read

$$\dot{x} = -x + (I_n - A^\dagger A)(x - \gamma \nabla f(x)) + A^\dagger b. \tag{6.14}$$

In light of Theorem 6.3.1(ii), for $\gamma \in (0, 2/\ell)$, the dynamics (6.14) are strongly infinitesimally contracting with respect to $\|\cdot\|_2$ with rate $1 - \max\{|1-\gamma\rho|, |1-\gamma\ell|\}$.

The downside to using the dynamics (6.14) is the cost of computing $A^\dagger$. To remedy this issue, a common approach is to leverage duality and jointly solve primal and dual problems. In what follows, we take this approach and study contractivity of the corresponding primal-dual dynamics.

The Lagrangian associated to the problem (6.13) is the map $L \colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ given by $L(x, \lambda) = f(x) + \lambda^\top (Ax - b)$. Computing the gradient descent of L in x and gradient ascent of L in λ, the *continuous-time primal-dual dynamics* (also called Arrow-Hurwicz-Uzawa flow [171]) are

$$\begin{aligned} \dot{x} &= -\nabla_x L(x, \lambda) = -\nabla f(x) - A^\top \lambda, \\ \dot{\lambda} &= \nabla_\lambda L(x, \lambda) = Ax - b. \end{aligned} \tag{6.15}$$

Theorem 6.3.2 (Contractivity of primal-dual dynamics)**.** *Suppose Assumption 4 holds. Then the continuous-time primal-dual dynamics* (6.15) *are strongly infinitesimally con-*

tracting with respect to $\|\cdot\|_P$ *with rate* $c > 0$ *where*

$$P = \begin{bmatrix} I_n & \alpha A^\top \\ \alpha A & I_m \end{bmatrix} \succ 0, \;\; \alpha = \frac{1}{2}\min\left\{\frac{1}{\ell}, \frac{\rho}{a_{\max}}\right\}, \;\; \textit{and} \tag{6.16}$$

$$c = \frac{1}{2}\alpha a_{\min} = \frac{1}{4}\min\left\{\frac{a_{\min}}{\ell}, \frac{a_{\min}}{a_{\max}}\rho\right\}. \tag{6.17}$$

Proof. Since f is continuously differentiable, convex, and strongly smooth, it is almost everywhere twice differentiable so the Jacobian of the dynamics (6.15) exists almost everywhere and is given by $J_{\mathrm{PD}}(z) := \begin{bmatrix} -\nabla^2 f(x) & -A^\top \\ A & 0 \end{bmatrix}$, where $z = (x, \lambda) \in \mathbb{R}^{n+m}$. To prove strong infinitesimal contraction it suffices to show that for all z for which $J_{\mathrm{PD}}(z)$ exists, the bound $\mu_P(J_{\mathrm{PD}}(z)) \leq -c$ holds for P, c given in (6.16) and (6.17), respectively. The assumption of strong convexity and strong smoothness of f further imply that $\rho I_n \preceq \nabla^2 f(x) \preceq \ell I_n$ for all x for which the Hessian exists. Moreover, it holds:

$$\sup_z \mu_P(J_{\mathrm{PD}}(z)) \leq \max_{\rho I_n \preceq B \preceq \ell I_n} \mu_P\left(\begin{bmatrix} -B & -A^\top \\ A & 0 \end{bmatrix}\right),$$

where the sup is over all points for which $J_{\mathrm{PD}}(z)$ exists. The result is then a consequence of Lemma 6.7.1 in Appendix 6.7. □

Remark 6.3.5. *Our method of proof in Lemma 6.7.1 follows the same method as was presented in [183, Lemma 2], but uses a sharper upper bounding to yield a sharper contraction rate of* $\frac{1}{4}\min\left\{\frac{a_{\min}}{\ell}, \frac{a_{\min}}{a_{\max}}\rho\right\}$ *compared to the estimate* $\frac{1}{8}\min\left\{\frac{a_{\min}}{\ell}, \frac{a_{\min}}{a_{\max}}\rho\right\}$ *in [183, Lemma 2]. The sharper upper bounding is a consequence of an appropriate matrix factorization and a less conservative bounding of the square of a difference of matrices.* △

Parameter-varying case Consider the *parameter-dependent equality-constrained minimization problem*:

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & f_\theta(x), \\ \text{s.t.} \quad & Ax = b_\theta, \end{aligned} \tag{6.18}$$

where for each $\theta \in \Theta$, $b_\theta \in \mathbb{R}^m$ and the map f_θ is continuously differentiable, strongly convex, and strongly smooth with parameters ρ and ℓ, respectively. We also assume that A is full row rank.

The *parameter-varying primal-dual dynamics* are

$$\begin{aligned} \dot{x} &= -\nabla f_\theta(x) - A^\top \lambda, \\ \dot{\lambda} &= Ax - b_\theta, \end{aligned} \tag{6.19}$$

and by Theorem 6.3.2, for fixed θ, (6.19) is guaranteed to converge to the unique primal-dual pair solving (6.18). Moreover, when $\theta\colon \mathbb{R}_{\geq 0} \to \Theta$ is a differentiable curve, under the assumptions that $\theta \mapsto \nabla f_\theta(x)$ and $\theta \mapsto b_\theta$ are Lipschitz, the dynamics (6.19) are guaranteed to track $x^\star(\theta(t)), \lambda^\star(\theta(t))$ with a tracking error proportional to $\|\dot{\theta}(t)\|$ after a transient. Further, if $\nabla_\theta f$ is differentiable both in x and θ and b_θ is differentiable in θ, then we can design a feedforward term involving $\dot{\theta}$ leveraging Theorem 6.2.2 to attain zero tracking error.

It is important to note that we have not let A depend on the parameter θ since the norm with respect to which the dynamics (6.15) are contracting depends on A. If A depends on θ, then the norm with respect to which the dynamics are contracting is also parameter-dependent and the results from Theorems 6.2.1 and 6.2.2 do not directly apply.

6.3.3 Composite minimization

Finally, we study a composite minimization problem.

Problem 3. *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ and $g\colon \mathbb{R}^m \to \overline{\mathbb{R}}$ be convex, and $A \in \mathbb{R}^{m\times n}$. We consider the problem*

$$\min_{x\in\mathbb{R}^n} f(x) + g(Ax). \tag{6.20}$$

We make the following assumptions on f, g, and A:

Assumption 5. *$f\colon \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable, strongly convex with parameter ρ, and strongly smooth with parameter ℓ. The map $g\colon \mathbb{R}^m \to \overline{\mathbb{R}}$ is CCP. Finally, $A \in \mathbb{R}^{m\times n}$ satisfies $a_{\min} I_m \preceq AA^\top \preceq a_{\max} I_m$ for $a_{\min}, a_{\max} \in \mathbb{R}_{>0}$.* △

While the optimization problem (6.20) may appear to be a special case of (6.5), it may be computationally challenging to compute the proximal operator of $g \circ A$ even if the proximal operator of g may have a closed-form expression. Thus, we treat this problem separately.

The optimization problem (6.20) is equivalent to

$$\begin{aligned} \min_{x\in\mathbb{R}^n, y\in\mathbb{R}^m} & f(x) + g(y), \\ \text{s.t.} \quad & Ax - y = \mathbb{0}_m. \end{aligned} \tag{6.21}$$

We leverage the proximal augmented Lagrangian approach proposed in [185]. For $\gamma > 0$, define the *augmented Lagrangian* associated to (6.21) $L_\gamma\colon \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ by

$$L_\gamma(x, y, \lambda) = f(x) + g(y) + \lambda^\top (Ax - y) + \frac{1}{2\gamma}\|Ax - y\|_2^2, \tag{6.22}$$

and, by a slight abuse of notation, the *proximal augmented Lagrangian* $L_\gamma\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$

by

$$L_\gamma(x,\lambda) = f(x) + M_{\gamma g}(Ax+\gamma\lambda) - \frac{\gamma}{2}\|\lambda\|_2^2. \tag{6.23}$$

The proximal augmented Lagrangian equals the augmented Lagrangian where the minimization over y has already been performed and the optimal value for y has been substituted; see details in [185, Theorem 1]. Moreover, minimizing (6.20) corresponds to finding saddle points of (6.23). To this end, the proximal augmented Lagrangian primal-dual dynamics are

$$\begin{aligned}\dot{x} &= -\nabla_x L_\gamma(x,\lambda) = -\nabla f(x) - A^\top \nabla M_{\gamma g}(Ax+\gamma\lambda),\\ \dot{\lambda} &= \nabla_\lambda L_\gamma(x,\lambda) = \gamma(-\lambda + \nabla M_{\gamma g}(Ax+\gamma\lambda)).\end{aligned} \tag{6.24}$$

Before providing contraction estimates, we showcase the specific form of the dynamics (6.24) in the case of inequality-constrained minimization problems of the form $\min\{f(x) \mid Ax \leq b\}$. In this case, $g = \iota_{\mathcal{C}}$, where $\mathcal{C} = \{z \in \mathbb{R}^m \mid z \leq b\}$. Then $\mathrm{prox}_{\gamma g}(z) = \mathrm{P}_{\mathcal{C}}(z) = \min\{z, b\}$ and the corresponding gradient of the Moreau envelope is $\nabla M_{\gamma g}(z) = \frac{1}{\gamma}\,\mathrm{ReLU}(z-b)$, where the min and ReLU are applied entrywise. Finally, the dynamics (6.24) take the form

$$\begin{aligned}\dot{x} &= -\nabla f(x) - \frac{1}{\gamma}A^\top \,\mathrm{ReLU}(Ax+\gamma\lambda-b),\\ \dot{\lambda} &= -\gamma\lambda + \mathrm{ReLU}(Ax+\gamma\lambda-b),\end{aligned} \tag{6.25}$$

which we refer to as the *proximal inequality-constrained primal-dual dynamics*.

Next, we turn to contraction analysis of the dynamics (6.24). To provide estimates on the contraction rate and the norm with respect to which the dynamics (6.24) are strongly infinitesimally contracting, we need to introduce a useful nonlinear program.

For $\varepsilon \in (0, 1/\sqrt{a_{\max}})$, consider the nonlinear program

$$\max_{c\geq 0, \alpha\geq 0, \varkappa\geq 0} \quad c \tag{6.26a}$$

$$\text{s.t.} \quad \alpha \leq \min\Big\{\frac{1}{\sqrt{a_{\max}}} - \varepsilon, \frac{\gamma}{a_{\max}}\Big\}, \tag{6.26b}$$

$$\varkappa \geq \frac{2}{3}, \tag{6.26c}$$

$$c \leq \Big(\frac{3}{4} - \frac{1}{2\varkappa}\Big)\alpha a_{\min}, \tag{6.26d}$$

$$h(c, \alpha, \varkappa) \geq 0, \tag{6.26e}$$

with $h\colon \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}$ given by

$$h(c,\alpha,\varkappa) = 2\rho - \mathrm{ReLU}\left(2\alpha - \frac{2}{\gamma}\right)a_{\max} - 2c - \alpha\varkappa\frac{a_{\max}}{a_{\min}}\Big(\gamma^2\frac{a_{\max}}{a_{\min}} + (\ell + \frac{a_{\max}}{\gamma} + 2c)^2 + 2\gamma\frac{a_{\max}}{a_{\min}}(\ell + \frac{a_{\max}}{\gamma} + 2c)\Big).$$

We prove in Lemma 6.7.4 in Appendix 6.7 that there exist finite values $c > 0, \alpha > 0, \varkappa > 0$ that solve the problem (6.26).

Theorem 6.3.3 (Contractivity of the dynamics (6.24))**.** *Suppose Assumption 5 holds and let $\gamma > 0$ be arbitrary. Then the primal-dual dynamics (6.24) are strongly infinitesimally contracting with respect to $\|\cdot\|_P$ with rate $c^\star > 0$ where*

$$P = \begin{bmatrix} I_n & \alpha^\star A^\top \\ \alpha^\star A & I_m \end{bmatrix}, \tag{6.27}$$

and $\alpha^\star > 0, c^\star > 0$ are the arguments solving problem (6.26).

Proof. Let $z = (x, \lambda) \in \mathbb{R}^{n+m}$ and let $\mathsf{F}\colon \mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$ corresponds to the vector field (6.24) for $\dot{z} = \mathsf{F}(z)$. Let $y := Ax + \gamma\lambda$ and define $G(y) := \gamma\nabla^2 M_{\gamma g}(y)$ where it exists. The Jacobian of F is then

$$D\mathsf{F}(z) = \begin{bmatrix} -\nabla^2 f(x) - \frac{1}{\gamma} A^\top G(y) A & -A^\top G(y) \\ G(y)A & -\gamma(I_m - G(y)) \end{bmatrix},$$

which exists for a.e. z. We then aim to show that $\mu_P(D\mathsf{F}(z)) \leq -c$ for all z for which $D\mathsf{F}(z)$ exists. First we note that in light of Lemma 6.6.4 in Appendix 6.6,

$$\sup_z \mu_P(D\mathsf{F}(z)) \leq \max_{\substack{0 \preceq G \preceq I_m \\ \rho I_n \preceq B \preceq \ell I_n}} \mu_P\left(\begin{bmatrix} -B - \frac{1}{\gamma} A^\top G A & -A^\top G \\ GA & \gamma(G - I_m) \end{bmatrix}\right),$$

where the sup is over all z for which $D\mathsf{F}(z)$ exists. The result is then a consequence of Lemma 6.7.3 in Appendix 6.7. □

Remark 6.3.6. *To the best of our knowledge, the solution to the nonlinear program (6.26) provides the most general test for the contractivity of the dynamics (6.24). The original work [185, Theorem 3], proves exponential convergence provided that $\gamma > \ell - \rho$. Instead we prove contraction, a stronger property, for all $\gamma > 0$. We compare contraction and convergence rate estimates in Figure 6.7 in Appendix 6.7.* △

Note that any triple $(c, \alpha, \varkappa) \in \mathbb{R}^3_{\geq 0}$ satisfying the constraints (6.26b)-(6.26e) provides a suboptimal contraction estimate, i.e., the dynamics (6.24) are strongly infinitesimally contracting with rate c (weakly contracting if $c = 0$) with respect to norm $\|\cdot\|_P$, where $P = \begin{bmatrix} I_n & \alpha A^\top \\ \alpha A & I_m \end{bmatrix}$. In what follows, we present computational considerations for estimating the optimal parameters $c^\star, \alpha^\star$ in the nonlinear program (6.26). Let $\varkappa > 2/3$ be fixed (e.g., at a value of 1). Then we have the following bounds on c which we will bisect on:

$$0 \leq c \leq c^{\varkappa}_{\max} := \min\Big\{\rho, \Big(\frac{3}{4} - \frac{1}{2\varkappa}\Big) a_{\min} \alpha_{\max}\Big\},$$

where $\alpha_{\max} = \min\left\{\frac{1}{\sqrt{a_{\max}}} - \varepsilon, \frac{\gamma}{a_{\max}}\right\}$. For any value of $c \in [0, c_{\max}^{\varkappa}]$, we check whether the following linear program (LP) is feasible:

$$\begin{aligned}
\text{Find} \quad & \alpha && (6.28a)\\
\text{s.t.} \quad & 0 \leq \alpha \leq \alpha_{\max}, && (6.28b)\\
& c \leq \Big(\frac{3}{4} - \frac{1}{2\varkappa}\Big)\alpha a_{\min}, && (6.28c)\\
& h(c, \alpha, \varkappa) \geq 0. && (6.28d)
\end{aligned}$$

Although $h(c, \alpha, \varkappa)$ is not linear in α, the problem can be transformed into an equivalent LP since ReLU is piecewise linear. The LP is feasible for $c = 0$ (with $\alpha = 0$). If the LP is feasible for $c = c_{\max}^{\varkappa}$, then $c_{\max}^{\varkappa}$ is the optimal contraction rate for this choice of $\varkappa$. More typically, the LP will not be feasible for $c_{\max}^{\varkappa}$ at which point we bisect on c, checking for feasible α for (6.28) at the prescribed value of c until we find a δ-optimal value of c with corresponding α that is feasible.

Optimizing further over $\varkappa$ can be done numerically either using nonlinear programming solvers or by using a grid search over $\varkappa$ and then the bisecting on c.

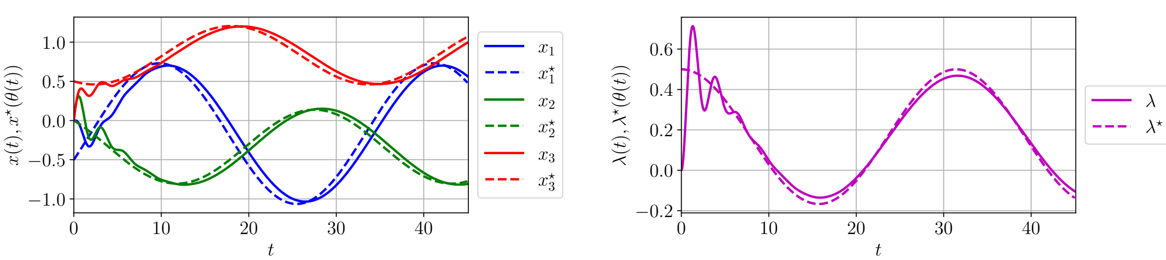


Figure 6.1: Plots of trajectories of the dynamics (6.32) solving the equality-constrained minimization problem (6.31). The left figure shows the trajectories of the primal variables $x(t)$ as solid curves and the trajectories of the instantaneously optimal primal variables $x^\star(\theta(t))$ as dashed curves. The right figure shows the trajectory of the dual variable $\lambda(t)$ as a solid curve and the instantaneously optimal dual variable $\lambda^\star(\theta(t))$ as a dashed curve.

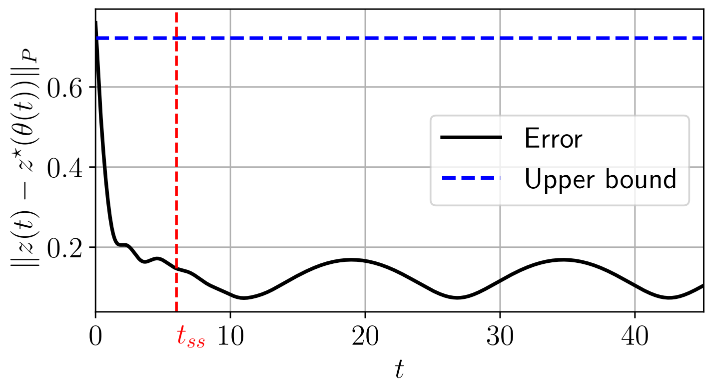


Figure 6.2: Plot of the error $\|z(t) - z^\star(\theta(t))\|_P$ along with the upper bound $\omega\ell/c^2$ where $z(t) = (x(t), \lambda(t))$ and $z^\star(\theta(t)) = (x^\star(\theta(t)), \lambda^\star(\theta(t)))$. We also denote 3 time–constants by t_{ss}, where one time-constant is $1/c = 2$ units of time.

Parameter-varying case Consider the *parameter-dependent composite minimization problem*

$$\min_{x \in \mathbb{R}^n} f_\theta(x) + g_\theta(Ax), \tag{6.29}$$

where for each $\theta \in \Theta$, the function f_θ is continuously differentiable, strongly convex, and strongly smooth, the map g_θ is CCP, and that A is full row rank. Then the proximal augmented Lagrangian primal-dual dynamics are

$$\begin{aligned} \dot{x} &= -\nabla f_\theta(x) - A^\top \nabla M_{\gamma g_\theta}(Ax + \gamma\lambda), \\ \dot{\lambda} &= \gamma(-\lambda + \nabla M_{\gamma g_\theta}(Ax + \gamma\lambda)). \end{aligned} \tag{6.30}$$

For each $\theta \in \Theta$, the minimization problem (6.29) has a unique minimizer $x^\star(\theta)$ and Lagrange multiplier $\lambda^\star(\theta)$ and the dynamics (6.30) converge to them. Let $\mathsf{F}_{\mathrm{PAL}} \colon \mathbb{R}^{n+m} \times \Theta \to \mathbb{R}^{n+m}$ denote the vector field for the dynamics (6.30). When $\theta \colon \mathbb{R}_{\geq 0} \to \Theta$ is a differentiable, time-varying parameter, under the assumption that $\theta \mapsto \mathsf{F}_{\mathrm{PAL}}(z, \theta)$ is Lipschitz uniformly in z, the dynamics (6.30) are guaranteed to track $x^\star(\theta(t)), \lambda^\star(\theta(t))$ with a tracking error proportional to $\|\dot{\theta}(t)\|$ after a transient. The assumption that $\mathsf{F}_{\mathrm{PAL}}$ is Lipschitz in θ is satisfied if, e.g., $\theta \mapsto \nabla f_\theta(x)$ and $\theta \mapsto \mathrm{prox}_{\gamma g_\theta}(x)$ are Lipschitz

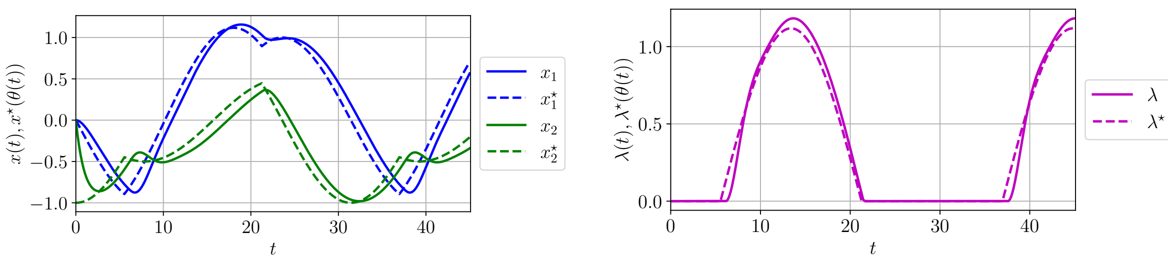


Figure 6.3: Plots of trajectories of the dynamics (6.34) solving the inequality-constrained minimization problem (6.33). The left figure shows the trajectories of the 2 primal variables $x(t)$ as solid curves and the trajectories of the instantaneously optimal primal variables $x^\star(\theta(t))$ as dashed curves. The right figure shows the trajectory of the dual variable $\lambda(t)$ as a solid curve and the instantaneously optimal dual variable $\lambda^\star(\theta(t))$ as a dashed curve.

uniformly in x. Finally, if F_{PAL} is differentiable in both of its arguments, then we can leverage Theorem 6.2.2 to design a feedforward term involving $\dot{\theta}$ to attain zero tracking error.

As in the case of linear equality constrained minimization, we have not let the matrix A depend on the parameter θ. For the same reason as before, the norm with respect to which the dynamics (6.24) are contracting depends on A.

6.4 Numerical and hardware experiments

In this section, we present experiments to showcase the performance of the proposed dynamics.[3] We present an application of Theorem 6.2.1 to a problem with equality constraints, which corresponds to Problem 2 and a case with inequality constraints, which corresponds to Problem 3. Additionally, we consider an application to control barrier function-based design, [201], for collision avoidance in a multi-robot scenario by leveraging the contraction analysis of the proximal gradient dynamics from Problem 1.

[3]Code for our experiments is available at https://github.com/davydovalexander/time-varying-convex.

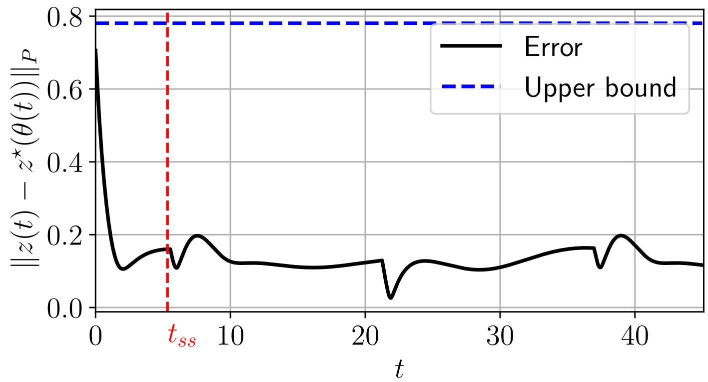


Figure 6.4: Plot of the error $\|z(t) - z^{\star}(\theta(t))\|_P$ along with the upper bound $\omega\ell/c^2$ where $z(t) = (x(t), \lambda(t))$ and $z^{\star}(\theta(t)) = (x^{\star}(\theta(t)), \lambda^{\star}(\theta(t)))$. We also denote 3 time–constants by t_{ss}, where one time-constant is $1/c \approx 1.78$ units of time.

6.4.1 Numerical experiment: Equality constraints

Consider the following time-varying quadratic optimization problem with equality constraints

$$\begin{aligned} \min_{x \in \mathbb{R}^3} \quad & \frac{1}{2}\|x - r(t)\|_2^2, \\ \text{s.t.} \quad & x_1 + 2x_2 + x_3 = \sin(\omega t), \end{aligned} \tag{6.31}$$

where $\omega = 0.2$ and $r(t) = (\sin(\omega t), \cos(\omega t), 1)$. We can see that (6.31) is an instance of (6.18) with $n = 3$ primal variables and $m = 1$ equality constraints by letting $\theta(t) = (\cos(\omega t), \sin(\omega t)) \in \Theta := \{z \in \mathbb{R}^2 \mid \|z\|_2 \leq 1\} \subset \mathbb{R}^2$. Letting $\|\cdot\|_\Theta = \|\cdot\|_2$, the Lipschitz assumptions are verified for $f_{\theta(t)}(x) = \frac{1}{2}\|x - r(t)\|_2^2$ and $b_{\theta(t)} = \sin(\omega t)$. The corresponding primal-dual dynamics for problem (6.31) read

$$\begin{aligned} \dot{x}_1 &= -x_1 + \sin(\omega t) - \lambda, \\ \dot{x}_2 &= -x_2 + \cos(\omega t) - 2\lambda, \\ \dot{x}_3 &= -x_3 + 1 - \lambda, \\ \dot{\lambda} &= x_1 + 2x_2 + x_3 - \sin(\omega t), \end{aligned} \tag{6.32}$$

i.e., the dynamics are that of a linear system.

We simulate the dynamics (6.32) over the time interval $t \in [0, 45]$ with a forward Euler discretization with stepsize $\Delta t = 0.01$ and set the initial conditions $x(0) = \mathbb{0}_3, \lambda(0) = 0$. We plot the trajectories of the dynamics along with the instantaneously optimal values $x^\star(\theta(t)), \lambda^\star(\theta(t))$ in Figure 6.1.

We empirically observe how the trajectories of the dynamics track the instantaneously optimal values $x^\star(\theta(t)), \lambda^\star(\theta(t))$ after a small transient. We then verify that the bound from Theorem 6.2.1 provides valid upper bounds for the tracking error. Finding the norm with respect to which the stable linear system (6.32) is contracting with largest rate corresponds to a bisection algorithm and is detailed in [2, Section 2.5.2]. After executing the bisection algorithm, we find that the primal-dual dynamics for (6.31) are strongly infinitesimally contracting with respect to $\|\cdot\|_P$ with rate $c = 0.5$ for suitably chosen $P = P^\top \succ 0$. Then the corresponding Lipschitz constant for the vector field is computed from $(\Theta, \|\cdot\|_2)$ to $(\mathbb{R}^4, \|\cdot\|_P)$, and is approximately $\ell \approx 0.902$. From Theorem 6.2.1, we know that the asymptotic tracking error as measured in the $\|\cdot\|_P$ norm is upper bounded by $\frac{\ell}{c^2}\omega \approx 0.722$ since $\|\dot{\theta}(t)\|_2 = \omega$ for all $t \geq 0$. In Figure 6.2 we plot $\|z(t) - z^\star(\theta(t))\|_P$, where z is the stacked vector of x and λ, as well as the upper bound $\frac{\ell}{c^2}\omega$ to demonstrate the validity of our bound.

Remark 6.4.1. *Note that in this example we have leveraged the fact that the dynamics (6.32) are linear to get improved rates of contraction. If we had instead simply used the bound on the contraction rate from Theorem 6.3.2, we would instead have $c = \frac{1}{4}$, which would yield looser bounds on the asymptotic error (measured in a different norm, however).* △

6.4.2 Numerical experiment: Inequality constraints

Consider the following time-varying quadratic optimization problem with inequality constraints

$$\begin{aligned} \min_{x \in \mathbb{R}^2} \quad & \frac{1}{2}\|x + r(t)\|_2^2, \\ \text{s.t.} \quad & -x_1 + x_2 \leq \cos(\omega t), \end{aligned} \tag{6.33}$$

where $\omega = 0.2$ and $r(t) = (\sin(\omega t), \cos(\omega t))$. We see that (6.33) is an instance of (6.29) with $g \colon \mathbb{R} \times \Theta \to \overline{\mathbb{R}}$ given by $g(z, \theta) = \iota_{\mathcal{C}_\theta}(z)$ where $\mathcal{C}_\theta = \{z \in \mathbb{R} \mid z \leq \theta_1\}$, the time-varying parameter is $\theta(t) = (\cos(\omega t), \sin(\omega t)) \in \Theta := \{z \in \mathbb{R}^2 \mid \|z\|_2 \leq 1\}$, and $A = [-1, 1]$. Following the expressions (6.25), the correspoding proximal inequality-constrained primal-dual dynamics read

$$\begin{aligned} \dot{x}_1 &= -x_1 - \sin(\omega t) + \frac{1}{\gamma}\operatorname{ReLU}(-x_1 + x_2 + \gamma\lambda - \cos(\omega t)), \\ \dot{x}_2 &= -x_2 - \cos(\omega t) - \frac{1}{\gamma}\operatorname{ReLU}(-x_1 + x_2 + \gamma\lambda - \cos(\omega t)), \\ \dot{\lambda} &= -\gamma\lambda + \operatorname{ReLU}(-x_1 + x_2 + \gamma\lambda - \cos(\omega t)). \end{aligned} \tag{6.34}$$

We simulate the dynamics (6.34) with $\gamma = 10$ over the time interval $t \in [0, 45]$ with a forward Euler discretization with stepsize $\Delta t = 0.01$ and $x(0) = \mathbb{0}_2, \lambda(0) = 0$. We plot the trajectories of the dynamics along with the instantaneously optimal values $x^\star(\theta(t)), \lambda^\star(\theta(t))$ in Figure 6.3.

We empirically observe how the trajectories of the dynamics track the instantaneously optimal values $x^\star(\theta(t)), \lambda^\star(\theta(t))$ after a small transient. We then verify that the bound from Theorem 6.2.1 provides a valid upper bound for the tracking error. Note that the vector field for the dynamics (6.34) is almost everywhere differentiable and its Jacobian has the structure

$$D\mathsf{F}_{\mathrm{PAL}}(z) = \begin{bmatrix} -1 - \frac{1}{\gamma}J & \frac{1}{\gamma}J & J \\ \frac{1}{\gamma}J & -1 - \frac{1}{\gamma}J & -J \\ -J & J & -\gamma + \gamma J \end{bmatrix}, \tag{6.35}$$

where J denotes the derivative of the ReLU evaluated at $-x_1 + x_2 + \gamma\lambda - \cos(\omega t)$ and always takes value in $\{0, 1\}$ when it is defined. In other words, the Jacobian always takes one of two values. Finding the norm which maximizes the contraction rate of the dynamics (6.34) corresponds to the minimization problem

$$\begin{aligned} \min_{c\in\mathbb{R}, P\in\mathbb{R}^{3\times 3}} \quad & c, \\ \text{s.t.} \quad & PD_i + D_i^\top P \preceq 2cP, \quad i \in \{1, 2\}, \\ & P = P^\top \succ 0, \end{aligned} \tag{6.36}$$

where D_1 corresponds to the Jacobian (6.35) with $J = 1$ and D_2 corresponds to the Jacobian (6.35) with $J = 0$. The problem (6.36) can be solved using a bisection algorithm on c as discussed in [2, Section 2.5.2]. After running the bisection algorithm, we find that the dynamics (6.34) are strongly infinitesimally contracting with respect to $\|\cdot\|_P$ with rate $c = 0.5625$ for suitably chosen $P = P^\top \succ 0$. Then the corresponding Lipschitz constant for the vector field is computed from $(\Theta, \|\cdot\|_2)$ to $(\mathbb{R}^3, \|\cdot\|_P)$ and is approximately $\ell \approx 1.235$. From Theorem 6.2.1, we know that the asymptotic tracking error as measured in the $\|\cdot\|_P$ norm is upper bounded by $\frac{\ell}{c^2}\omega \approx 0.781$ since $\|\dot{\theta}(t)\|_2 = \omega$ for all $t \geq 0$. In Figure 6.4 we plot $\|z(t) - z^\star(\theta(t))\|_P$ as well as the upper bound $\frac{\ell}{c^2}\omega$ to demonstrate the validity of our bound.

Remark 6.4.2. *As in the previous example, we leverage the specific structure of the dynamics (6.34) to yield sharper bounds on the contraction rate c instead of using the* *nonlinear program* (6.26). △

6.4.3 Hardware experiment: Collision avoidance with online control barrier functions

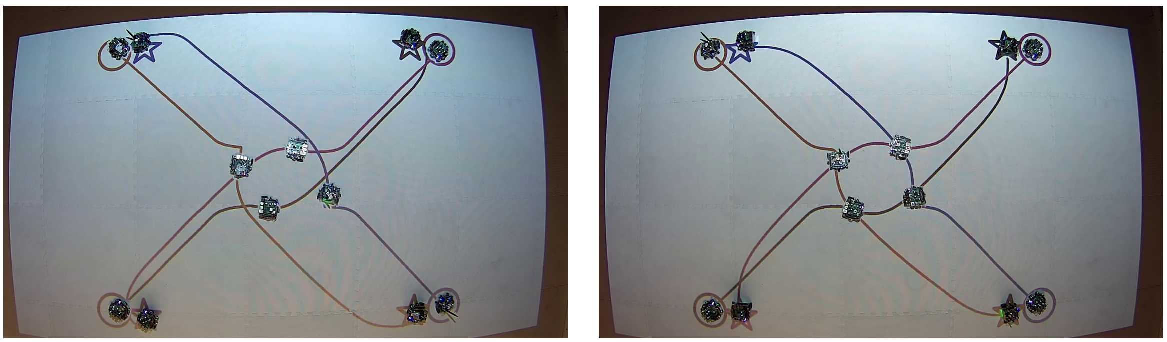


Figure 6.5: Overhead trajectories from the Robotarium experiments shown as photo-composites. The robots' positions are shown at: the start of the maneuver (denoted by circles), the midpoint, and the end of the maneuver (denoted by stars). (Left) Robots executing the dynamics (6.41) and (Right) robots executing the dynamics with feedforward prediction (6.42). Videos are available at `https://bit.ly/TimeVaryingConvex`.

We evaluate our equilibrium tracking framework in the context of control-barrier function (CBF)-based control synthesis [201]. To showcase our framework, we first recall the definition of a CBF.

Suppose we are given a nonlinear control-affine system

$$\dot{x} = F(x) + G(x)u, \tag{6.37}$$

where $F\colon \mathbb{R}^n \to \mathbb{R}^n$ and $G\colon \mathbb{R}^n \to \mathbb{R}^{n\times m}$ are locally Lipschitz, and $u \in \mathbb{R}^m$. We let $h\colon \mathbb{R}^n \to \mathbb{R}$ be sufficiently smooth and $\mathcal{C} := \{x \in \mathbb{R}^n \mid h(x) \geq 0\}$ denote the *safe set* for the system (6.37).

Definition 11 (Control Barrier Function [201, Definition 3]). *The function h is a control barrier function (CBF) for $\mathcal{C}$ if there exists a locally Lipschitz and strictly increasing*

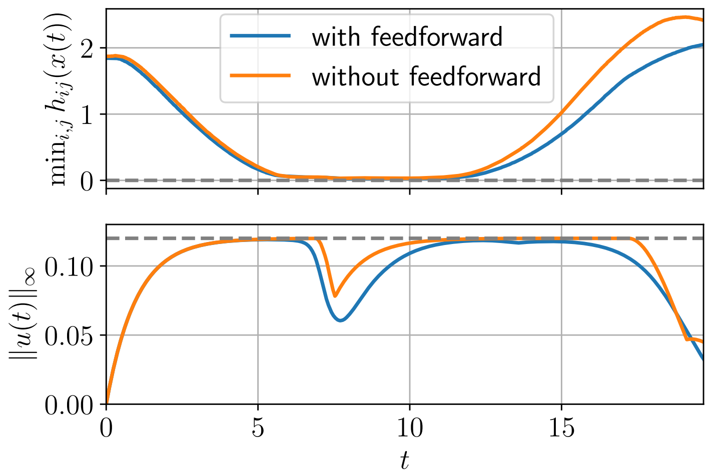


Figure 6.6: The evolution of $\min_{i,j} h_{ij}(x(t))$ and $\|u(t)\|_\infty$ for the experiments in Figure 6.5. We plot $\overline{u} = 0.12$ [m/s] as a dashed line.

function $\alpha\colon \mathbb{R} \to \mathbb{R}$ *with* $\alpha(0) = 0$ *such that for all* $x \in \mathcal{C}$, *there exists* $u \in \mathbb{R}^m$ *with*

$$\nabla h(x)^\top F(x) + \nabla h(x)^\top G(x)u + \alpha(h(x)) \geq 0. \tag{6.38}$$

It is then known that a continuous feedback controller $u\colon \mathbb{R}^n \to \mathbb{R}^m$ which satisfies (6.38) for all $x \in \mathcal{D} \supset \mathcal{C}$, for an open set $\mathcal{D}$, renders $\mathcal{C}$ forward-invariant under the dynamics (6.37) [205].

If $\mathcal{C} = \cap_{i=1}^k \mathcal{C}_i$ and $\mathcal{C}_i := \{x \in \mathbb{R}^n \mid h_i(x) \geq 0\}$, a common approach to synthesize controllers which yield $\mathcal{C}$ forward-invariant is to solve a parametric quadratic program

$$\begin{aligned} \underset{u\in\mathbb{R}^m}{\operatorname{argmin}} \quad & \frac{1}{2}\|u - u_{\mathrm{nom}}(x)\|_2^2, \\ \text{s.t.} \quad & a_i(x)^\top u \leq b_i(x), \quad i \in \{1,\dots,k\} \\ & \|u\|_\infty \leq \overline{u}, \end{aligned} \tag{6.39}$$

where $a_i(x) = -G(x)^\top \nabla h_i(x)$ and $b_i(x) = \nabla h_i(x)^\top F(x) + \alpha(h_i(x))$ for all $i \in \{1,\dots,k\}$, $u_{\mathrm{nom}}\colon \mathbb{R}^n \to \mathbb{R}^m$ is a nominal feedback controller, and $\|u\|_\infty \leq \overline{u}$, where $\overline{u} > 0$, captures actuator constraints.

The complexity of the above approach arises in the computational burden of solving (6.39) at each $x(t)$. To ameliorate this burden, we propose leveraging the equilibrium tracking approach, as was done in [200]. First, we relax the k inequality constraints with log-barrier penalties:

$$\begin{aligned} \underset{u \in \mathbb{R}^m}{\operatorname{argmin}} \quad & \frac{1}{2}\|u - u_{\mathrm{nom}}(x)\|_2^2 - \eta(t) \sum_{i=1}^{k} \log(b_i(x) - a_i(x)^\top u), \\ \text{s.t.} \quad & \|u\|_\infty \le \overline{u}, \end{aligned} \tag{6.40}$$

where $\eta \colon \mathbb{R} \to \mathbb{R}_{\ge 0}$ is a smooth function with $\lim_{t\to\infty} \eta(t) = 0$ which relaxes the log barriers over time. Note that we do not add additional log barrier terms for the actuator constraint since doing so would add $2m$ additional terms to the objective and would result in overly conservative controls initially.

Let $f \colon \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ denote the objective function of (6.40), that is, $f(u, x, \eta) = \frac{1}{2}\|u - u_{\mathrm{nom}}(x)\|_2^2 - \eta \sum_{i=1}^{k} \log(b_i(x) - a_i(x)^\top u)$. Note that both x and η are time-varying parameters in this optimization problem. For $\mathcal{U} = \{u \in \mathbb{R}^m \mid \|u\|_\infty \le \overline{u}\}$, we leverage the proximal gradient dynamics (6.8) to track the solution $u^\star(x(t), \eta(t))$ of (6.40):

$$\dot{u}(t) = -u(t) + \mathrm{P}_{\mathcal{U}}\big(u(t) - \gamma \nabla_u f\big(u(t), x(t), \eta(t)\big)\big). \tag{6.41}$$

To ensure contraction of the dynamics (6.41), we need to make two technical assumptions. First, we assume that there exist $\delta, \gamma_c > 0$ such that for all $\gamma \le \gamma_c$, the set $\{(u, x) \in \mathbb{R}^n \times \mathbb{R}^m \mid a_i(x)^\top u \le b_i(x) - \delta, \text{ for all } i\}$ is forward-invariant for the coupled dynamics (6.37) and (6.41).[4] Secondly, we assume that each a_i is bounded on this set. Under these

[4]This first assumption is reminiscient of a strengthened version of the CBF condition (6.38) as studied, for example, in [205].

technical assumptions, on this forward-invariant set, $f(\cdot, x, \eta)$ is strongly convex and strongly smooth. Then, in light of Theorems 6.2.1 and 6.3.1, we can find suitable $\gamma \in]0, \gamma_c]$ such that the dynamics (6.41) are contracting with respect to $\|\cdot\|_2$ and will track the optimal $u^\star(x(t), \eta(t))$ solving (6.40) with some error depending on $\dot{x}(t)$ and $\dot{\eta}(t)$. This is the approach that is taken in [200], albeit with the primal dual flow on the proximal augmented Lagrangian instead of the proximal gradient flow. In contrast, we will leverage Theorem 6.2.2 to minimize tracking error. To apply Theorem 6.2.2, we first replace $\mathrm{P}_{\mathcal{U}}$ with a smooth approximation, $\Sigma \colon \mathbb{R}^m \to \mathbb{R}^m$ satisfying $\Sigma(z) \in \mathcal{U}$ and $0 \preceq D\Sigma(z) \preceq I_n$ for all z. Then we consider the dynamics with feedforward prediction, omitting the dependencies on time

$$\dot{u} = -u + \Sigma(y) + \big(I_m - D\Sigma(y)(I_m - \gamma \nabla_u^2 f(u, \theta))\big)^{-1} \gamma D\Sigma(y) \frac{\partial \nabla_u f}{\partial \theta}(u, \theta)\dot{\theta}, \tag{6.42}$$

where we have used the shorthands $y = u - \gamma \nabla_u f(u, \theta)$ and $\theta = (x, \eta)$. We can show that all trajectories of (6.42) converge exponentially quickly to a small tube centered at the time-varying equilibrium trajectory of (6.41). To see this rigorously, suppose $\|\Sigma(z) - \mathrm{P}_{\mathcal{U}}(z)\|_2 \leq \varepsilon$ for all z. Intuitively, for any $\varepsilon > 0$, such a Σ exists by "smoothing" $\mathrm{P}_{\mathcal{U}}$ at the points of nonsmoothness. Then, letting $\theta = (x, \eta)$, at fixed θ, we can consider the following two systems:

$$\dot{u}_1 = -u_1 + \mathrm{P}_{\mathcal{U}}(u_1 - \gamma \nabla_u f(u_1, \theta)) \tag{6.43}$$

$$\dot{u}_2 = -u_2 + \mathrm{P}_{\mathcal{U}}(u_2 - \gamma \nabla_u f(u_2, \theta)) + d(u_2, \theta), \tag{6.44}$$

where $d(u_2, \theta) = \Sigma(y) - \mathrm{P}_{\mathcal{U}}(y)$ and $y = u_2 - \gamma \nabla_u f(u_2, \theta)$. In other words, the smoothed dynamics (6.44) are a perturbed version of the nominal, nonsmooth, dynamics (6.43). We know that both of these dynamics are contracting in view of Theorem 2 and, therefore, admit unique equilibria at each θ. Let $u_1^\star(\theta), u_2^\star(\theta)$ denote these parametrized equilibria. From [5, Theorem 38(ii)], if $c - \varepsilon > 0$, we know that the following bound holds:

$$\|u_1^\star(\theta) - u_2^\star(\theta)\|_2 \leq \frac{\|d(u_2^\star, \theta)\|_2}{c - \varepsilon} \leq \frac{\varepsilon}{c - \varepsilon}. \tag{6.45}$$

In other words, as $\varepsilon \to 0$, then $\|u_1^\star(\theta) - u_2^\star(\theta)\|_2 \to 0$.

Extending to the case of time-varying θ, we have that, for all t, $\|u_1^\star(\theta(t)) - u_2^\star(\theta(t))\|_2 \leq \frac{\varepsilon}{c-\varepsilon}$. Namely, the time-varying equilibrium trajectory of (6.44) lies in a small tube around the time-varying equilibrium trajectory of (6.43). Then since (6.44) has smooth dynamics, we may add a feedforward term to (6.44) so that all trajectories converge to $u_2^\star(\theta(t))$ exponentially fast. And since $u_2^\star(\theta(t))$ lies inside an arbitrarily small tube around $u_1^\star(\theta(t))$, we can achieve arbitrarily close tracking depending on how we tune ε. Recall that $\dot{x}$ follows the dynamics (6.37) and that since η is a design choice, $\dot{\eta}$ is known as well, which makes these dynamics well-posed.

We now provide a concrete example. Consider a team of n single-integrator robots, $\dot{x}_i = u_i$, in $\mathbb{R}^2$ attempting to avoid collisions with one another while driving from an initial position to a final one. Let $x = (x_1, \dots, x_n) \in \mathbb{R}^{2n}$ and $u = (u_1, \dots, u_n) \in \mathbb{R}^{2n}$ denote the concatenation of their states and controls, respectively. There are $n(n-1)/2$ CBFs, given by $h_{ij}(x) = \|x_i - x_j\|_2^2 - d^2$, for all $i, j \in \{1, \dots, n\}, i \neq j$ and $d > 0$ is some safety distance between agents. The constraints induced by these CBFs are $-\nabla h_{ij}(x)^\top u \leq \alpha(h_{ij}(x))$.

We implement our proposed controllers in the Robotarium [206], a remotely accessible multi-robot testbed at the Georgia Institute of Technology. We consider $n = 4$ robots and evolve the dynamics (6.41) and (6.42) as the controllers for the robots. We choose $\eta(t) = e^{-0.3t}, \gamma = 0.5, \alpha(r) = 3r, d = 0.182$ [m], and as nominal control, we take $u_{\text{nom}i}(x_i) = x_{\text{des,i}} - x_i$, where $x_{\text{des,i}} \in \mathbb{R}^2$ denotes the desired final position for robot i. To comply with actuation limits in the Robotarium, we take $\overline{u} = 0.12$ [m/s]. For the dynamics (6.42), we take $\Sigma(z) = \overline{u}\tanh(z/\overline{u})$ as a smooth approximation for $\mathrm{P}_{\mathcal{U}}$. The results of the experiment are shown in Figure 6.5 and the evolution of $\min_{i,j} h_{ij}(x)$ and $\|u\|_\infty$ are shown in Figure 6.6.

We can see that both control strategies result in safe execution where each h_{ij} is nonnegative for all t and yield controls that obey the actuation constraint $\|u(t)\|_\infty \leq \overline{u}$ for all t. Note that without the feedforward term, the robots overshoot their target destination before eventually converging to the final location while this does not happen when there is a feedforward correction, see Figure 6.6. This overshoot is due to the lack of knowledge that the log barrier penalty coefficient $\eta(t)$ is decaying to zero. In contrast, the feedforward term allows the robots to account for its exponential decay to zero and thus they avoid overshooting their goal.

6.5 Discussion

In this article, we take a contraction theory approach to the problem of tracking optimal trajectories in time-varying convex optimization problems. We prove in Theorem 6.2.1 that the tracking error between any solution trajectory of a strongly infinitesimally contracting system and its equilibrium trajectory is upper bounded with an explicit estimate on the bound. We additionally prove in Theorem 6.2.2 that any strongly infinitesimally contracting system can be augmented with a feedforward term to ensure that the tracking error converges to zero exponentially quickly. To apply these theorems, we establish the strong infinitesimal contractivity of three dynamical systems solving optimization problems and apply Theorem 6.2.1 to provide explicit tracking error bounds. We validate these bounds in two numerical examples and present a novel application to CBF-based control.

We believe that this work motivates future research in establishing the strong infinitesimal contractivity of dynamical systems solving optimization problems or performing more general computation due to the desirable consequences of contractivity. As future research, we plan to investigate (i) discretization of parameter-varying contracting dynamics and establish similar tracking error bounds for discrete-time contracting systems [207], (ii) contractivity properties of continuous-time stochastic optimization algorithms based on stochastic differential equations [208], and (iii) time-varying nonconvex optimization problems with isolated local minima [209], possibly using the theory of k-contraction [33]. Finally, we believe that a comprehensive comparison to methods based upon incremental quadratic constraints [210] and dissipative systems theory [211] could provide novel design insights.

6.6 Proofs and additional results

We begin with a result on parametrized contractions.

Lemma 6.6.1 (Parametrized contractions). *Consider the system (6.1) satisfying Assumptions (A1) and (A2). Let $x^\star\colon \Theta \to \mathcal{X}$ denote the map given by $x^\star(\theta) = x^\star_\theta$. Then $x^\star(\cdot)$ is Lipschitz from $(\Theta, \|\cdot\|_\Theta)$ to $(\mathcal{X}, \|\cdot\|_\mathcal{X})$ with constant ℓ_θ/c.*

Proof. Consider the dynamics $\dot{x} = F(x,\theta)$ satisfying Assumptions (A1) and (A2), where θ is constant. Given two constant inputs θ_1 and θ_2, the two equilibrium solutions are $x^\star(\theta_1)$ and $x^\star(\theta_2)$. The assumptions of [2, Theorem 3.16] are satisfied with $c = -\operatorname{osL}_x(F)$ and $\ell_\theta = \mathsf{Lip}_\theta(F)$, and the differential inequality [2, Equation 3.39] implies

$$0 \leq -c\|x^\star(\theta_1) - x^\star(\theta_2)\|_\mathcal{X} + \ell_\theta\|\theta_1 - \theta_2\|_\Theta.$$

This concludes the proof. □

Next, we study the Lipschitzness of parametrized time-varying equilibrium trajectories and obtain a bound on their time derivatives.

Lemma 6.6.2 (Lipschitzness of parametrized curves)**.** *Consider $\mathcal{X} \subseteq \mathbb{R}^n$ and $\Theta \subseteq \mathbb{R}^d$ with associated norms $\|\cdot\|_{\mathcal{X}} \colon \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ and $\|\cdot\|_{\Theta} \colon \mathbb{R}^d \to \mathbb{R}_{\geq 0}$, respectively. Let $G \colon \Theta \to \mathcal{X}$ be Lipschitz from $(\Theta, \|\cdot\|_{\Theta})$ to $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ with constant $\mathsf{Lip}(G) \geq 0$. Then for every $a, b \in \mathbb{R}$ with $a < b$ and every continuously differentiable $\theta \colon (a,b) \to \Theta$,*

(i) the curve $x \colon (a,b) \to \mathcal{X}$ given by $x(t) = G(\theta(t))$ is locally Lipschitz;

(ii) $\|\dot{x}(t)\|_{\mathcal{X}} \leq \mathsf{Lip}(G)\|\dot{\theta}(t)\|_{\Theta}$, for a.e. $t \in (a,b)$.

Proof. Item (i) is a consequence of the fact that continuously differentiable mappings are locally Lipschitz and that a composition of Lipschitz mappings is Lipschitz.

To prove item (ii) we first note that item (i) implies that $\dot{x}(t)$ exists almost everywhere by Rademacher's theorem. Next, for all $t \in (a,b)$ for which $\dot{x}(t)$ exists, we have

$$\begin{aligned}\|\dot{x}(t)\|_{\mathcal{X}} &:= \Big\| \lim_{h\to 0^+} \frac{x(t+h)-x(t)}{h} \Big\|_{\mathcal{X}} = \lim_{h\to 0^+} \frac{1}{h}\|x(t+h)-x(t)\|_{\mathcal{X}} \\ &\leq \lim_{h\to 0^+} \frac{\mathsf{Lip}(G)}{h}\|\theta(t+h)-\theta(t)\|_{\Theta} = \mathsf{Lip}(G)\Big\| \lim_{h\to 0^+} \frac{\theta(t+h)-\theta(t)}{h} \Big\|_{\Theta} = \mathsf{Lip}(G)\|\dot{\theta}(t)\|_{\Theta},\end{aligned}$$

where we have used the continuity of the norms $\|\cdot\|_{\mathcal{X}}$ and $\|\cdot\|_{\Theta}$ and the Lipschitzness of the map G. □

Lemma 6.6.3. *Let $\mathsf{T} \colon \mathbb{R}^n \to \mathbb{R}^n$ be Lipschitz with respect to a norm $\|\cdot\|$ with constant $\mathsf{Lip}(\mathsf{T})$. Then the vector field $F \colon \mathbb{R}^n \to \mathbb{R}^n$ defined by the dynamics*

$$\dot{x} = -x + \mathsf{T}(x) =: F(x) \tag{6.46}$$

satisfies $\mathrm{osL}(F) \leq -1 + \mathsf{Lip}(\mathsf{T})$. Moreover, if $\mathsf{Lip}(\mathsf{T}) < 1$, then T has a unique fixed point, x^*, which is the unique equilibrium point of the contracting dynamics (6.46).

Proof. We have $\mathrm{osL}(F) := \mathrm{osL}(-\mathsf{Id} + \mathsf{T}) = -1 + \mathrm{osL}(\mathsf{T}) \leq -1 + \mathsf{Lip}(\mathsf{T})$, where we used the translation property of osL and the upper bound $\mathrm{osL}(\mathsf{T}) \leq \mathsf{Lip}(\mathsf{T})$ [2, Section 3.2.3]. If $\mathsf{Lip}(\mathsf{T}) < 1$, the Banach contraction theorem implies the existence of a unique fixed point and equilibrium points of (6.46) are exactly fixed points of T. Moreover, the dynamics (6.46) are contracting since $\mathrm{osL}(F) < 0$. □

Lemma 6.6.4 (Symmetry and bounds on Jacobians)**.** *Let the map $g\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ be CCP. Then for every $\gamma > 0$, $D\mathrm{prox}_{\gamma g}(x)$ and $\nabla^2 M_{\gamma g}(x)$ exist for a.e. $x \in \mathbb{R}^n$, are symmetric, and satisfy*

$$0 \preceq D\mathrm{prox}_{\gamma g}(x) \preceq I_n, \quad 0 \preceq \nabla^2 M_{\gamma g}(x) \preceq \frac{1}{\gamma} I_n. \tag{6.47}$$

Proof. First note that $D\mathrm{prox}_{\gamma g}(x)$ and $\nabla^2 M_{\gamma g}(x)$ exist for a.e. $x \in \mathbb{R}^n$ by Rademacher's theorem since $\mathrm{prox}_{\gamma g}$ and $\nabla M_{\gamma g}$ are both Lipschitz. Furthermore, $\nabla^2 M_{\gamma g}(x)$ is symmetric for a.e. x by symmetry of second derivatives. Analogously, $D\mathrm{prox}_{\gamma g}(x) = I_n - \gamma \nabla^2 M_{\gamma g}(x)$ by (1.8) so we conclude symmetry of $D\mathrm{prox}_{\gamma g}(x)$ as well. The bounds (6.47) are a consequence of the fact that $\mathrm{prox}_{\gamma g}$ and $\mathsf{Id} - \mathrm{prox}_{\gamma g}$ are both firmly nonexpansive [4, Proposition 12.28]. □

6.7 Logarithmic norm of Hurwitz saddle matrices

In this section, we provide bounds on log norms for the saddle matrices arising from the Jacobians of the dynamics (6.15) and (6.24).

Lemma 6.7.1 (Logarithmic norm of Hurwitz saddle matrices)**.** *Given $B = B^\top \in \mathbb{R}^{n\times n}$ and $A \in \mathbb{R}^{m\times n}$, with $m \leq n$, we consider the* saddle matrix

$$\mathcal{B} = \begin{bmatrix} -B & -A^\top \\ A & 0 \end{bmatrix} \in \mathbb{R}^{(m+n)\times(m+n)}. \tag{6.48}$$

Then, for each matrix pair (B, A) satisfying $\rho I_n \preceq B \preceq \ell I_n$ and $a_{\min} I_m \preceq AA^\top \preceq a_{\max} I_m$, for $\rho, \ell, a_{\min}, a_{\max} \in \mathbb{R}_{>0}$, the following contractivity LMI holds:

$$\mathcal{B}^\top P + P\mathcal{B} \preceq -2cP \quad \iff \quad \mu_{\mathcal{B}}(P) \leq -c, \tag{6.49}$$

where

$$P = \begin{bmatrix} I_n & \alpha A^\top \\ \alpha A & I_m \end{bmatrix} \succ 0, \;\; \alpha = \frac{1}{2}\min\Big\{\frac{1}{\ell}, \frac{\rho}{a_{\max}}\Big\}, \;\; \textit{and} \tag{6.50}$$

$$c = \frac{1}{2}\alpha a_{\min} = \frac{1}{4}\min\Big\{\frac{a_{\min}}{\ell}, \frac{a_{\min}}{a_{\max}}\rho\Big\}. \tag{6.51}$$

Proof. We start by verifying that $P \succ 0$. Using the Schur complement of the $(2,2)$ entry, we need to verify that

$$I_n - \alpha^2 A^\top A \succ 0 \iff 1 - \alpha^2 a_{\max} > 0 \iff \alpha^2 < 1/a_{\max}.$$

The inequality $\alpha^2 < 1/a_{\max}$ follows from the tighter inequality $(2\alpha)^2 \leq \frac{1}{a_{\max}}$ which is proved as follows:

$$\min\Big\{\frac{1}{\ell}, \frac{\rho}{a_{\max}}\Big\}^2 \leq \min\Big\{\frac{1}{\ell}, \frac{\rho}{a_{\max}}\Big\} \cdot \max\Big\{\frac{1}{\ell}, \frac{\rho}{a_{\max}}\Big\} = \frac{1}{\ell} \cdot \frac{\rho}{a_{\max}} \leq \frac{1}{a_{\max}}.$$

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

$$Q = \begin{bmatrix} 2B - 2\alpha A^\top A - 2cI_n & \alpha BA^\top - 2c\alpha A^\top \\ A + \alpha AB - A - 2c\alpha A & 2\alpha AA^\top - 2cI_m \end{bmatrix}.$$

The (2,2) block satisfies the lower bound

$$\begin{aligned} 2\alpha AA^\top - 2cI_m &= 2\big(\tfrac{1}{2}\alpha AA^\top - cI_m\big) + \alpha AA^\top \\ &\succeq 2\big(\tfrac{1}{2}\alpha a_{\min} - c\big)I_m + \alpha AA^\top = \alpha AA^\top \succ 0. \end{aligned}$$

Given this lower bound, we can factorize the resulting matrix as follows:

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

Since $\alpha I_n \succ 0$, it now suffices to show that the Schur complement of the (2,2) block of $n \times n$ matrix is positive semidefinite. We proceed as follows:

$$\begin{aligned} & 2B - 2(\alpha A^\top A + cI_n) - \alpha\big(B - 2cI_n\big)^2 \succeq 0 \\ \iff & 2B - \alpha B^2 + 4\alpha cB \succeq 2(\alpha A^\top A + cI_n) + 4\alpha c^2 I_n \\ \impliedby & \ 2B - \alpha B^2 \succeq 2(\alpha A^\top A + cI_n) \text{ and } 4\alpha cB \succeq 4\alpha c^2 I_n. \end{aligned}$$

To prove $2B - \alpha B^2 \succeq 2(\alpha A^\top A + cI_n)$, we upper bound the right hand side as follows:

$$\begin{aligned} 2(\alpha A^\top A + cI_n) &\overset{(6.50)}{\preceq} \alpha(2a_{\max} + a_{\min})I_n \\ &\overset{\alpha \le \frac{1}{2}\rho/a_{\max}}{\preceq} \frac{1}{2}\frac{\rho}{a_{\max}}(2a_{\max} + a_{\min})I_n \preceq \frac{3}{2}\rho I_n. \end{aligned}$$

Next, since $\alpha \le \frac{1}{2\ell}$, we know $-\alpha\ell \ge -\frac{1}{2}$. We then upper bound the left hand side as follows:

$$2B - \alpha B^2 \succeq 2B - \alpha\ell B \succeq (2 - \tfrac{1}{2})B \succeq \tfrac{3}{2}\rho I_n.$$

Finally, the inequality $4\alpha c B \succeq 4\alpha c^2 I_n$ follows from noting $c \leq \frac{1}{4}\frac{a_{\min}}{a_{\max}}\rho < \rho$. $\square$

The following lemma is presented in [183, Lemma 6]. We include it here for completeness.

Lemma 6.7.2. *Let $X = X^\top \in \mathbb{R}^{m\times m}$ satisfy $0 \preceq X \preceq x_{\max} I_m$ for some $x_{\max} > 0$, $\gamma > 0$, and $\alpha \leq \dfrac{\gamma}{x_{\max}}$. Then for all $d \in [0,1]^m$, the following inequality holds*

$$\alpha(\operatorname{diag}(d)X + X\operatorname{diag}(d)) + 2\gamma(I_m - \operatorname{diag}(d)) \succeq \frac{3}{2}\alpha X. \tag{6.52}$$

Proof. See [183, Lemma 6]. $\square$

The following lemma is a generalization of [183, Lemma 4], where we let the matrix G be dense. The proof method is improved by introducing an auxiliary variable $\varkappa$ which we optimize for in (6.26) whereas in the proof of [183, Lemma 4], $\varkappa = 1$ is chosen.

Lemma 6.7.3 (Generalized saddle matrices)**.** *Given $A \in \mathbb{R}^{m\times n}$, $B = B^\top \in \mathbb{R}^{n\times n}$, and $G = G^\mathsf{T} \in \mathbb{R}^{m\times m}$, with $m \leq n$, and $\gamma > 0$, we consider the* saddle matrix

$$\mathcal{B} = \begin{bmatrix} -B - \frac{1}{\gamma}A^\mathsf{T} G A & -A^\mathsf{T} G \\ GA & -\gamma(I_m - G) \end{bmatrix} \in \mathbb{R}^{(m+n)\times(m+n)}. \tag{6.53}$$

Then, for each matrix triplet (B, A, G) satisfying $\rho I_n \preceq B \preceq \ell I_n$, $a_{\min} I_m \preceq AA^\top \preceq a_{\max} I_m$, and $0 \preceq G \preceq I_m$, for ρ, ℓ, $a_{\min}$, $a_{\max} \in \mathbb{R}_{>0}$, the following contractivity LMI holds:

$$\mathcal{B}^\top P + P\mathcal{B} \preceq -2c^\star P \quad \iff \quad \mu_{\mathcal{B}}(P) \leq -c^\star, \tag{6.54}$$

where

$$P = \begin{bmatrix} I_n & \alpha^\star A^\top \\ \alpha^\star A & I_m \end{bmatrix} \succ 0, \tag{6.55}$$

and $c^\star$ and $\alpha^\star$ are the optimal parameters for the problem (6.26).

Proof. Define the matrix $Q \in \mathbb{R}^{(m+n)\times(m+n)}$ by

$$Q = -\mathcal{B}^\top P - P\mathcal{B} - 2c^\star P = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12}^\top & Q_{22} \end{bmatrix}, \tag{6.56}$$

where $Q_{11} \in \mathbb{R}^{n\times n}$, $Q_{12} \in \mathbb{R}^{n\times m}$, and $Q_{22} \in \mathbb{R}^{m\times m}$. We aim to show that $Q \succeq 0$ for $c^\star$ and $\alpha^\star$ optimal parameters for the problem (6.26). We have

$$\begin{aligned}
Q_{11} &= 2B + \Big(\frac{2}{\gamma} - 2\alpha^\star\Big)A^\top GA - 2c^\star I_n, \\
Q_{12} &= \alpha^\star\gamma A^\top(I_m - G) + \alpha^\star BA^\top + \frac{\alpha^\star}{\gamma}A^\top GAA^\top - 2c^\star\alpha^\star A^\top, \\
Q_{22} &= \alpha^\star(AA^\top G + GAA^\top) + 2\gamma(I_m - G) - 2c^\star I_m.
\end{aligned}$$

To show that $Q \succeq 0$ we use the Schur Complement, which requires to prove that $Q_{22} \succ 0$ and $Q_{11} - Q_{12}Q_{22}^{-1}Q_{12}^\top \succeq 0$. We do this in three steps.

First, we find a lower bound for Q_{22}. Since G is symmetric and satisfies $0 \preceq G \preceq I_m$, there exists an orthogonal matrix $U \in \mathbb{R}^{m\times m}$ and $d \in [0,1]^m$ such that $G = U\mathrm{diag}(d)U^\top$. Substituting this into Q_{22} and multiplying on the left and on the right by U and $U^\top$, respectively, we get

$$U^\top Q_{22}U = \alpha^\star(U^\top AA^\top U\mathrm{diag}(d) + \mathrm{diag}(d)U^\top AA^\top U) + 2\gamma(I_m - \mathrm{diag}(d)) - 2c^\star I_m, \tag{6.57}$$

where we have used the fact that $U \in \mathbb{R}^{m\times m}$ orthogonal, i.e., $UU^\top = U^\top U = I_m$. Moreover, the orthogonality of U implies that $U^{-1} = U^\top$. Thus the eigenvalues of $U^\top AA^\top U$ and $AA^\top$ are equal, and therefore, $a_{\min}I_m \preceq U^\top AA^\top U \preceq a_{\max}I_m$. Next, applying

Lemma 6.7.2 to (6.57), with $X := U^\top AA^\top U$, and, multiplying this on the left and on the right by U and U^T, respectively, we get the following lower bound

$$Q_{22} \succeq \frac{3}{2}\alpha^\star AA^\mathsf{T} - 2c^\star I_m, \tag{6.58}$$

where the inequality holds because $\alpha^\star \leq \frac{\gamma}{a_{\max}}$ – constraint (6.26b). Finally, we note that $Q_{22} \succ 0$ for $\alpha^\star$ and $c^\star$ optimal parameters for the problem (6.26)

Next, we need to prove that $Q_{11} - Q_{12}Q_{22}^{-1}Q_{12}^\mathsf{T} \succeq 0$ for $c^\star$ and $\alpha^\star$ optimal parameters for the problem (6.26). To this purpose, first note that for every $\kappa > \frac{2}{3}\frac{a_{\max}}{a_{\min}}\frac{1}{\alpha^\star}$,

$$Q_{22} \succeq \frac{1}{\kappa}AA^\mathsf{T}. \tag{6.59}$$

Next, we upper bound $Q_{12}Q_{22}^{-1}Q_{12}^\mathsf{T}$. To simplify notation, define $R_1 = B + \frac{1}{\gamma}A^\mathsf{T}GA - 2c^\star I_n \in \mathbb{R}^{n\times n}$, $R_2 = \gamma A^\mathsf{T}(I_m - G) \in \mathbb{R}^{n\times m}$, and note that $Q_{12} = \alpha^\star(R_1A^\mathsf{T} + R_2) \in \mathbb{R}^{n\times m}$. We compute

$$\begin{aligned}
Q_{12}Q_{22}^{-1}Q_{12}^\mathsf{T} &\overset{(6.59)}{\preceq} \kappa Q_{12}(AA^\mathsf{T})^{-1}Q_{12} \\
&\preceq (\alpha^\star)^2\kappa(R_1A^\mathsf{T}(AA^\mathsf{T})^{-1}AR_1^\mathsf{T} + R_1A^\mathsf{T}(AA^\mathsf{T})^{-1}R_2^\mathsf{T} \\
&\quad + R_2(AA^\mathsf{T})^{-1}AR_1^\mathsf{T} + R_2(AA^\mathsf{T})^{-1}R_2^\mathsf{T}) \\
&\preceq (\alpha^\star)^2\kappa(R_1R_1^\mathsf{T} + 2\|R_1A^\mathsf{T}(AA^\mathsf{T})^{-1}R_2^\mathsf{T}\|I_n \\
&\quad + \|R_2(AA^\mathsf{T})^{-1}R_2^\mathsf{T}\|I_n),
\end{aligned} \tag{6.60}$$

where the final inequality holds because $A^\mathsf{T}(AA^\mathsf{T})^{-1}A \preceq I_n$. Note that

$$R_1R_1^\mathsf{T} \preceq \ell^2 I_n + \Big(2\ell + \frac{a_{\max}}{\gamma} + 2c^\star\Big)\Big(\frac{a_{\max}}{\gamma} + 2c^\star\Big)I_n =: h_1(c^\star)I_n,$$

where we have introduced the function $h_1\colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ defined by $h_1(c^\star) = \ell^2 + 2\ell(\frac{a_{\max}}{\gamma} + 2c^\star) + (\frac{a_{\max}}{\gamma} + 2c^\star)^2$. Moreover,

$$\begin{aligned}\|R_2(AA^\top)^{-1}R_2^\top\| &\leq \gamma^2\frac{a_{\max}}{a_{\min}}, \text{ and}\\ 2\|R_1A^\top(AA^\top)^{-1}R_2^\top\| &\leq 2\gamma\frac{a_{\max}}{a_{\min}}\Big(\ell+\frac{a_{\max}}{\gamma}+2c^\star\Big) =: h_2(c^\star),\end{aligned}$$

where, we have introduced the function $h_2\colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ defined by $h_2(c^\star) := 2\gamma\frac{a_{\max}}{a_{\min}}(\ell + \frac{a_{\max}}{\gamma} + 2c^\star)$. Substituting the previous bounds on the LMI (6.60) we get

$$Q_{12}Q_{22}^{-1}Q_{12}^\top \preceq (\alpha^\star)^2\kappa\Big(h_1(c^\star) + h_2(c^\star) + \gamma^2\frac{a_{\max}}{a_{\min}}\Big)I_n.$$

Next, we compute

$$\begin{aligned}Q_{11} &= 2B + \Big(\frac{2}{\gamma} - 2\alpha^\star\Big)A^\top GA - 2c^\star I_n\\ &\succeq \Big(2\rho - \operatorname{ReLU}\Big(2\alpha^\star - \frac{2}{\gamma}\Big)a_{\max} - 2c^\star\Big)I_n.\end{aligned}$$

Finally, we have

$$\begin{aligned}Q_{11}-Q_{12}Q_{22}^{-1}Q_{12}^\top \succeq I_n\Big(2\rho - \operatorname{ReLU}\Big(2\alpha^\star - \frac{2}{\gamma}\Big)a_{\max} - 2c^\star\\ - (\alpha^\star)^2\kappa\Big(h_1(c^\star) + h_2(c^\star) + \gamma^2\frac{a_{\max}}{a_{\min}}\Big)\Big) \succeq 0,\end{aligned}$$

where the last inequality follows from constraint (6.26e). This concludes the proof. □

Lemma 6.7.4. *Consider the nonlinear program (6.26). Then optimal parameters $c^\star, \alpha^\star, \varkappa^\star$ exist, are finite, and are strictly positive.*

Proof. We have the obvious bounds on $c, \alpha, \varkappa$:

$$
\begin{aligned}
&0 \leq c \leq c_{\max} := \min\Big\{\rho, \frac{3}{4}a_{\min}\alpha_{\max}\Big\},\\
&0 \leq \alpha \leq \alpha_{\max} := \min\Big\{\frac{1}{\sqrt{a_{\max}}} - \varepsilon, \frac{\gamma}{a_{\max}}\Big\},\\
&\frac{2}{3} \leq \varkappa.
\end{aligned}
$$

We can see that at fixed $c, \varkappa$, the function h is continuous and decreasing in α and that for fixed $\alpha, \varkappa$, the function h is continuous and decreasing in c. Let $\varkappa > 2/3$ be arbitrary. Pick $\tilde{c} \in \,]0, \rho[$ and note that $h(\tilde{c}, 0, \varkappa) = 2\rho - 2\tilde{c} > 0$. By continuity of h in α, there exists sufficiently small $\alpha > 0$ such that $h(\tilde{c}, \alpha, \varkappa) \geq 0$. Then pick $0 < c < \min\{\tilde{c}, (\frac{3}{4} - \frac{1}{2\varkappa})\alpha a_{\min}\}$. By definition, $c \leq (\frac{3}{4} - \frac{1}{2\varkappa})\alpha a_{\min}$ and $h(c, \alpha, \varkappa) \geq h(\tilde{c}, \alpha, \varkappa) \geq 0$ since h is decreasing in c for fixed $\alpha, \varkappa$. In other words, for every $\varkappa > 2/3$, there exist parameter values $(c, \alpha, \varkappa)$ that are all positive and feasible (at $\varkappa = 2/3$, the optimal value of c is 0). Since c and α are bounded above, we know that optimal values must be finite. We now show that the optimal value of $\varkappa$ is also bounded above.

For $\varkappa \geq 2/3$, define the set

$$
S_\varkappa := \Big\{(c, \alpha) \in \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \mid \alpha \leq \alpha_{\max}, c \leq \Big(\frac{3}{4} - \frac{1}{2\varkappa}\Big)\alpha a_{\min}, h(c, \alpha, \varkappa) \geq 0\Big\}.
$$

Because h is continuous, we can see that for every $\varkappa \geq 2/3$, $S_\varkappa$ is a closed set. Moreover, we can also see that

$$
S_\varkappa \subseteq [0, c_{\max}] \times [0, \alpha_{\max}],
$$

so $S_\varkappa$ is a closed subset of a compact set, so it is also compact. Then for every $\varkappa \geq 2/3$ and $(c, \alpha) \in S_\varkappa$, the mapping $(c, \alpha) \mapsto c$ is continuous on the compact set $S_\varkappa$ and thus attains its maximal value at parameter values $c_\varkappa, \alpha_\varkappa$, both of which are positive for every

$\varkappa > 2/3$ by the above reasoning. Consider specifically $c_1, \alpha_1 > 0$ which are the parameter values which maximize the map over the compact set S_1. It is now straightforward to argue that there exists $M > 0$ sufficiently large such that for all $d \geq M$, $c_d < c_1$ and thus implies that the optimal value for $\varkappa$ lies in the bounded interval $[2/3, M]$. To see this intuitively, if M is sufficiently large, $h(c_M, \alpha_M, M) \geq 0$ can be made to imply that $\alpha_d \leq \frac{4}{3}\frac{c_1}{a_{\min}}$ for every $d \geq M$ (by picking M large enough and since h is decreasing in $\varkappa$) and thus $c_d \leq (\frac{3}{4} - \frac{1}{2d})\alpha_d a_{\min} < \frac{3}{4}\alpha_d a_{\min} \leq c_1$.

Then define the set

$$
\begin{aligned}
S := \Big\{ &(c, \alpha, \varkappa) \in \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \mid \alpha \leq \alpha_{\max}, \\
&c \leq \Big(\frac{3}{4} - \frac{1}{2\varkappa}\Big)\alpha a_{\min}, h(c, \alpha, \varkappa) \geq 0, 2/3 \leq \varkappa \leq M \Big\}.
\end{aligned}
$$

We can easily see that S is closed and is also compact since

$$
S \subseteq [0, c_{\max}] \times [0, \alpha_{\max}] \times [2/3, M].
$$

Moreover, for every $(c, \alpha, \varkappa) \in S$ the mapping $(c, \alpha, \varkappa) \mapsto c$ is continuous on the compact set S and attains its maximal value at parameter values $c^\star, \alpha^\star, \varkappa^\star$ all of which are positive and finite. This concludes the proof. $\square$

Finally, we empirically compare our estimate of the contraction rate for the dynamics (6.24) as a function of γ to existing convergence rates in the literature. We compare against [212, Prop. 3] and [183, Theorem 2]. We elect to compare against [212] rather than [185] since the estimate in [212] depends only on $a_{\min}, a_{\max}$ while the estimate from [185, Theorem 3] depends on A. We note that both [212] and [183] establish exponential convergence rather than contractivity, but the proof methods could be extended to establish contraction. In Figure 6.7, we plot these convergence estimates.

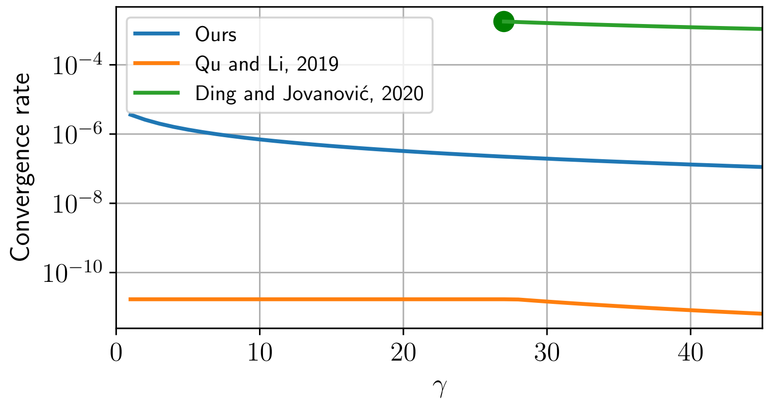


Figure 6.7: Estimates of contraction or exponential convergence rate of the dynamics (6.24) with $(\rho, \ell, a_{\min}, a_{\max}) = (1.03, 27.81, 0.1, 1)$ as a function of γ. We compare the contraction estimates from (6.26) to the exponential convergence rates from [212, Prop. 3] and [183, Theorem 2].

We note that the estimate from [212, Prop. 3] is only valid for $\gamma \geq \ell - \rho$ and this is denoted by the circle in the plot. For $\gamma < \ell - \rho$, our contraction estimate is orders of magnitude better than the estimate in [183]. For $\gamma \geq \ell - \rho$, the estimate in [212] is better. In other words, our estimate appears to be the sharpest-known contraction rate which is valid for all $\gamma > 0$. Our estimates may be well-suited for poorly conditioned problems where $\rho \ll \ell$ which would require very large γ for the analysis in [212] to apply.

Chapter 7

Exponential Stability of Parametric Optimization-Based Controllers via Lur'e Contractivity

This chapter was first published in the IEEE Control Systems Letters [213].[1]

7.1 Introduction

Controllers solving optimization problems are ubiquitous in systems and control. One large class of optimization-based controllers are based upon (i) solving an optimal control problem offline, such as LQR, LQG, or Hamilton-Jacobi PDE and (ii) closing the loop with the resulting controller. Recent interest has focused on a different class of optimization-based controllers, that solve optimization problems at every time-step of the dynamic evolution of the plant. Namely, such controllers are solutions to *parametric optimization problems*, i.e., programs that are functions of the state of the system.

Examples of these controllers include model-predictive control [214], online feedback optimization [175], and control barrier (or Lyapunov) function-based control [201]. While stability and robustness properties of the first class of optimization-based controllers are well understood, fewer studies have focused on stability and robustness properties of parametric optimization-based controllers.

Literature Review: Parametric optimization is a rich subdiscipline of optimization which studies solutions of optimization problems as a function of a parameter; see the textbook [215]. Parametric optimization is ubiquitous in systems and control, especially in model predictive control [214] and CBF-based control [201]. Closed-form solutions for certain classes of parametric programs were studied in [214, Chapter 5]. However, closed-form solutions are not always attainable. Regularity of solutions to parametric programs, namely establishing smoothness properties of their solutions, is a classical problem and has even pervaded systems and control [216, 217]. Compared to regularity results, there are fewer results on the stability of control systems with parametric optimization-based controllers.

One class of systems for which there have been results on stability and safety of systems driven by parametric optimization-based controllers are those coming from CBFs and control Lyapunov functions (CLFs) [201]. In these works, CLF and CBF constraints are jointly enforced in a state-dependent quadratic program (QP). To guarantee feasibility of the QP when the CLF and CBF inequalities cannot be jointly satisfied, the stability is commonly relaxed by introducing a slack variable. This relaxation results in a lack of stability guarantee even for arbitrarily large penalties on the slack variable [218]. Recent work, [219], studied a variant of the CLF-CBF QP controller and demonstrates how to estimate the basin of attraction of the origin.

Contributions: We consider LTI systems equipped with parametric projection-based controllers. As our main contribution, assuming linearity of the nominal controller and various well-posedness conditions, we obtain LMI-based sufficient conditions for exponential stability and the existence of a global Lyapunov function. Our proposed sufficient conditions generalize those presented in [220] which focus only on special classes of parametric QPs, whereas our controllers can incorporate more general convex constraints. Our results can also be seen as using similar ideas to those in [210] regarding sector bounds for projection operators.

Our analysis is based upon the virtual system methods in contraction theory and contractivity of Lur'e systems. For context, contraction theory is a computationally-friendly notion of robust nonlinear stability [154] and the virtual system method, first proposed in [24], is an analysis approach to establish exponential convergence for systems satisfying certain weak contractivity properties. As a tutorial contribution, we provide a novel review of the virtual system method in Section 7.2.1. Specifically, we show that LTI systems with parametric projection-based controllers are in Lur'e form with state-dependent nonlinearity and that an appropriate virtual system can be designed in standard Lur'e form.

As our second main contribution, we establish in Theorem 7.3.1 a novel necessary and sufficient condition for absolute contractivity of Lur'e systems with cocoercive nonlinearities. In contrast, in [221] and [222, Proposition 4], monotone and Lipschitz nonlinearities are considered yet only sufficient conditions are provided. By focusing on cocoercive nonlinearities, we propose a relaxed LMI condition that is necessary and sufficient. See the related discussion in [37, Theorem 4.2] for other sufficient conditions.

As our third main contribution, we study the special LTI case of single integrators. We establish that all trajectories of the closed-loop system converge to the set of equilibria and that all trajectories converging to the origin do so exponentially fast with a known rate. While there are related results in the CBF/CLF literature [223, 224], this convergence result for the general class of parametric projection-based controllers is novel, to the best of our knowledge.

Finally, we study two applications, namely state-dependent saturated control systems and CBF-based control. For state-dependent saturated control systems, the maximal control effort depends on the state of the system and we demonstrate that our sufficient condition can be readily applied to yield a condition for global exponential stability. In CBF-based control, we consider a single integrator avoiding an obstacle and demonstrate that the results hold and provide evidence that the estimated exponential rate of convergence is tight. Specifically, we numerically observe that, in the case of single integrator dynamics, one does not need to enforce any CLF decrease condition to guarantee stability to the origin.

7.2 Prerequisite material

7.2.1 Virtual system method for convergence analysis

The *virtual system* analysis approach is a method to study the asymptotic behavior of a dynamical system that may not enjoy contracting properties. The virtual system approach was first proposed in [24], but we follow the systematic procedure advocated for in [154, Section 5.7]. For completeness sake, we describe this procedure below.

The virtual system analysis approach is as follows. We are given a dynamical system

$$\dot{x} = f(x), \quad x(0) = x_0 \in \mathbb{R}^n \tag{7.1}$$

and we let $\phi_{x_0}(t)$ denote a solution from initial condition $x(0) = x_0$. The analysis proceeds in three steps:

(i) *design* a time-varying dynamical system, called the *virtual system*, of the form

$$\dot{y} = f_{\text{virtual}}(y, \phi_{x_0}(t)), \quad y \in \mathbb{R}^d \tag{7.2}$$

satisfying a strong infinitesimal contractivity property with respect to an appropriate norm, e.g., the existence of a positive definite matrix $P \in \mathbb{R}^{d\times d}$ and a scalar $c > 0$ such that for all $y_1, y_2 \in \mathbb{R}^d$, $z \in \mathbb{R}^n$:

$$(f_{\text{virtual}}(y_1, z) - f_{\text{virtual}}(y_2, z))^\top P(y_1 - y_2) \leq -c\|y_1 - y_2\|_P^2;$$

(The vector field is called virtual since it is different from the nominal vector field, f, and does not correspond to any physically meaningful variation of f.)

(ii) *select* two specific solutions of the virtual system and state their incremental stability property:

$$\|y_1(t) - y_2(t)\|_P \leq \mathrm{e}^{-ct}\|y_1(0) - y_2(0)\|_P; \tag{7.3}$$

(iii) *infer* properties of the trajectory, $\phi_{x_0}(t)$, of the nominal system.

For example, if $d = n$ and $f(x) = f_{\text{virtual}}(x, x)$, then one can see that $\phi_{x_0}(t)$ is a solution for both systems and is often selected as one of the two specific solutions in step (ii). Additionally, if $f_{\text{virtual}}(\mathbb{0}_n, z) = \mathbb{0}_n$ for all $z \in \mathbb{R}^n$, then $\mathbb{0}_n$ is an equilibrium point for the virtual system and can be selected as one of the specific solutions. We refer to [24] for example applications leveraging the virtual system method.

7.3 Absolute contractivity of Lur'e systems

Consider the Lur'e system

$$\dot{x} = Ax + B\varphi(t, Kx), \tag{7.4}$$

where $\varphi\colon \mathbb{R}_{\geq 0} \times \mathbb{R}^m \to \mathbb{R}^m$ is continuous in its first argument and cocoercive second argument. Specifically, φ cocoercive in its second argument means there exists a constant $\rho > 0$ such that for all $y_1, y_2 \in \mathbb{R}^m, t \geq 0$,

$$(\varphi(t, y_1) - \varphi(t, y_2))^\top (y_1 - y_2) \geq \rho \|\varphi(t, y_1) - \varphi(t, y_2)\|_2^2. \tag{7.5}$$

Notably, cocoercivity, (7.5), implies that φ is monotone and Lipschitz continuous with constant ρ^{-1} in its second argument. Many standard nonlinearities satisfy cocoercivity including projections onto convex sets and nonlinearities of the form

$$\varphi(t, y) = (\varphi_1(t, y_1), \dots, \varphi_m(t, y_m))$$

where each φ_i is slope-restricted between 0 and ρ^{-1} in its second argument.

Akin to the classical problem of absolute stability, *absolute contractivity* is the property that the system (7.4) is strongly infinitesimally contracting for any nonlinearity φ obeying the constraint (7.5).

Theorem 7.3.1 (Necessary and sufficient condition for absolute contractivity)**.** *Consider the Lur'e system (7.4) and let $P \in \mathbb{R}^{n \times n}$ be positive definite and let K be full row-rank. The system (7.4) is strongly infinitesimally contracting with respect to $\|\cdot\|_P$ with rate*

$\eta > 0$ for any φ satisfying (7.5) if and only if there exists $\lambda \geq 0$ such that

$$\begin{bmatrix} A^\top P + PA + 2\eta P & PB + \lambda K^\top \\ B^\top P + \lambda K & -2\lambda\rho I_m \end{bmatrix} \preceq 0. \tag{7.6}$$

Proof. Employing the shorthand $\Delta x = x_1 - x_2, \Delta y = y_1 - y_2 = K\Delta x, \Delta u_t = \varphi(t, y_1) - \varphi(t, y_2)$, the contractivity condition for the system (7.4) is equivalently rewritten as

$$\Delta x^\top (PA + A^\top P + 2\eta P)\Delta x + 2\Delta x^\top PB\Delta u_t \leq 0. \tag{7.7}$$

Moreover, by the full row-rankedness of K, the cocoercivity condition (7.5) is equivalent to

$$\Delta u_t^\top (\rho \Delta u_t - K\Delta x) \leq 0. \tag{7.8}$$

Asking when the inequality (7.8) implies (7.7) is equivalent to the inequality (7.6) in light of the necessity and sufficiency of the S-procedure [225]. □

Note that the condition in [221, Theorem 2] corresponds to the inequality (7.6) with $\lambda = 1$. Moreover the matrix in (7.6) has $A^\top P + PA + 2\eta P$ in its $(1, 1)$ block compared to $A^\top P + PA + \eta I_n$ in [221]. This modification ensures that the integral contraction inequality holds rather than a related inequality with $-\frac{\eta}{2}\|x_1 - x_2\|_2^2$ on the right-hand side. Thus, by restricting our focus to cocoercive nonlinearities, we are able to find the sharpest condition for absolute contractivity.

7.4 Parametric projection-based controllers

We are interested in studying a continuous-time LTI system being driven by an parametric optimization-based controller. We say that the optimization problem is parametric since it is a function of the state. Specifically, we look at parametric projection-based controllers. More concretely, for $A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times m}, u^\star\colon \mathbb{R}^n \to \mathbb{R}^m, k\colon \mathbb{R}^n \to \mathbb{R}^m, g\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$, the LTI system and controller are:

$$\begin{aligned} \dot{x} &= Ax + Bu^\star(x), \\ u^\star(x) := \operatorname*{argmin}_{u\in\mathbb{R}^m} \quad & \frac{1}{2}\|u - k(x)\|_2^2 \\ \text{s.t.} \quad & g(x,u) \leq \mathbb{0}_p. \end{aligned} \tag{7.9}$$

In the context of the parametric optimization problem in (7.9), k denotes a nominal feedback controller and g captures constraints on the controller as a function of the state. Such controllers commonly arise in CLF and CBF theory, where the parametric optimization problem in (7.9) is used to enforce that $u^\star$ either causes the closed-loop system to decrease a specified Lyapunov function or keep a certain set forward-invariant, respectively [201].

The question we aim to answer in this chapter is the following: **What are conditions on the LTI system and the parametric optimization problem to ensure exponential stability of (7.9)?** Our main method for establishing sufficient conditions for exponential stability will be via the virtual system method in Section 7.2.1.

7.4.1 Well-posedness and regularity of solutions

In order to study the dynamical system (7.9), we need to ensure that it is well-posed. Several works in the literature have studied sufficient conditions for regularity of $u^\star$, e.g., continuity, Lipschitzness, or differentiability [204, 216, 217]. In this work, we utilize the following proposition from [217] which provides a sufficient condition for $u^\star$ to be continuous.

Proposition 7.4.1 ([217, Proposition 4])**.** *Consider the map $u^\star\colon \mathbb{R}^n \to \mathbb{R}^m$ defined via the solution to the parametric optimization problem*

$$\begin{aligned} u^\star(x) := \underset{u\in\mathbb{R}^m}{\operatorname{argmin}} \quad & f(x,u) \\ \text{s.t.} \quad & g(x,u) \leq \mathbb{0}_p. \end{aligned} \tag{7.10}$$

where $f\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $g\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ are each twice continuously differentiable on $\mathbb{R}^n \times \mathbb{R}^m$. Further assume that for some $x_0 \in \mathbb{R}^n$, $f(x_0,\cdot)$ is strongly convex[2] and $g(x_0,\cdot)$ is convex and that there exists $\hat{u} \in \mathbb{R}^m$ such that $g(x_0,\hat{u}) \ll \mathbb{0}_p$[3]. Then there exists a neighborhood of x_0 such that $u^\star$ is continuous at every point in the neighborhood.

By existence theorems, we know that for the system $\dot{x} = Ax + Bu^\star(x)$, for each initial condition x_0 satisfying the assumptions of Proposition 7.4.1, there exists a positive constant $\tau_{\max}(x_0)$ and a continuously differentiable curve $\phi_{x_0}\colon [0,\tau_{\max}(x_0)) \to \mathbb{R}^n$ satisfying $\frac{d\phi_{x_0}}{dt}(t) = A\phi_{x_0}(t) + Bu^\star(\phi_{x_0}(t))$ for all $t \in [0,\tau_{\max}(x_0))$. We say that the solution ϕ_{x_0} is forward-complete if $\tau_{\max}(x_0) = +\infty$. While Proposition 7.4.1 ensures existence of solutions, we refer to [217] for discussions on conditions for uniqueness of solutions.

[2] A map $f\colon \mathbb{R}^n \to \mathbb{R}$ is strongly convex if there exists $\nu > 0$ such that $\nabla^2 f(x) \succeq \nu I_n$ for all x.
[3] For two vectors, $v,w \in \mathbb{R}^n$, $v \ll w$ if $v_i < w_i$ for all $i \in \{1,\dots,n\}$.

7.4.2 Stability analysis for LTI systems

Consider the dynamical system and its corresponding controller defined via a parametric optimization problem (7.9) and define the following sets

$$\Gamma(x) := \{u \in \mathbb{R}^m \mid g(x,u) \leq \mathbb{0}_p\} \quad \text{and} \tag{7.11}$$

$$\mathcal{K} := \{x \in \mathbb{R}^n \mid \exists \hat{u} \text{ s.t. } g(x,\hat{u}) \ll \mathbb{0}_p\}, \tag{7.12}$$

where $\Gamma(x)$ represents the feasible control actions at the state x and $\mathcal{K}$ denotes the points in state space where the feasible set, $\Gamma(x)$, has an interior.

We make the following assumptions on our problem:

(A1) **(Regularity of g)** The map $g\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ is twice continuously differentiable on $\mathbb{R}^n \times \mathbb{R}^m$ and $g(x,\cdot)$ is convex for all $x \in \mathbb{R}^n$,

(A2) **(Existence of equilibrium and feasibility of zero control)** $\mathbb{0}_n \in \mathcal{K}$ and $\mathbb{0}_m \in \Gamma(x)$ for all $x \in \mathcal{K}$,

(A3) **(Linearity of nominal controller)** the map $k\colon \mathbb{R}^n \to \mathbb{R}^m$ is linear, i.e., $k(x) = Kx$ for some $K \in \mathbb{R}^{m\times n}$,

(A4) **(Dynamical feasibility)** for every $x_0 \in \mathcal{K}$, $\phi_{x_0}(t) \in \mathcal{K}$ for all $t \in [0, \tau_{\max}(x_0))$.

We make comments about some of these assumptions. Assumption (A2) ensures that $\mathbb{0}_n$ is an equilibrium point and that $u = \mathbb{0}_m$ is a feasible control action. Assumption (A4) ensures that the controller $u^\star$ does not drive the system outside the set of points where the feasible set of (7.9) has an interior. Outside of this set, the controller may fail to be continuous and solutions of (7.9) may fail to exist. One simple way to verify Assumption (A4) is to ensure that $\mathcal{K} = \mathbb{R}^n$. Note further that $u^\star(x)$ can compactly be written $u^\star(x) = \mathrm{P}_{\Gamma(x)}(Kx)$, where given a nonempty, closed, convex set $\Omega \subseteq \mathbb{R}^m$, $\mathrm{P}_\Omega(z) := \mathrm{argmin}_{v\in\Omega} \|z - v\|_2$.

We are now ready to state our first main theorem establishing the exponential stability of the system (7.9).

Theorem 7.4.1 (Exponential stability for LTI systems with parametric projection-based controllers)**.** *Consider the dynamics* (7.9) *satisfying Assumptions* (A1)-(A4)*. Further suppose that there exist $P = P^\top \succ 0$, $\eta > 0$, and $\lambda \geq 0$ satisfying the inequality*

$$\begin{bmatrix} A^\top P + PA + 2\eta P & PB + \lambda K^\top \\ B^\top P + \lambda K & -2\lambda I_m \end{bmatrix} \preceq 0. \tag{7.13}$$

Then from any $x_0 \in \mathcal{K}$,

(i) solutions to (7.9)*, ϕ_{x_0}, are forward-complete,*

(ii) the origin is globally exponentially stable with bound

$$\|\phi_{x_0}(t)\|_P \leq \mathrm{e}^{-\eta t}\|x_0\|_P, \tag{7.14}$$

(iii) the mapping $V : \mathcal{K} \to \mathbb{R}_{\geq 0}$ given by $V(x) = x^\top P x$ is a global Lyapunov function for the dynamics (7.9)*.*

Proof. We apply the virtual system method. Let $x_0 \in \mathcal{K}$ be arbitrary and consider the virtual system

$$\dot{y} = Ay + B\,\mathrm{P}_{\Gamma(\phi_{x_0}(t))}(Ky). \tag{7.15}$$

Note that for all $t \in [0, \tau_{\max}(x_0))$, $\mathrm{P}_{\Gamma(\phi_{x_0}(t))}$ obeys the inequality (7.5) with $\rho = 1$ due to cocoercivity of projections, see, e.g., [72, Eq. (2)]. Therefore the virtual system is a Lur'e system of the form (7.4). Theorem 7.3.1 implies that this virtual system is contracting with respect to $\|\cdot\|_P$ with rate $\eta > 0$. In other words, any two trajectories $y_1(\cdot), y_2(\cdot)$ for the virtual system satisfy for all $t \in [0, \tau_{\max}(x_0))$,

$$\|y_1(t) - y_2(t)\|_P \leq \mathrm{e}^{-\eta t}\|y_1(0) - y_2(0)\|_P. \tag{7.16}$$

First note that ϕ_{x_0} is a valid trajectory for the virtual system so we set $y_1(t) = \phi_{x_0}(t)$. Additionally note that $y_2(t) = \mathbb{0}_n$ is a valid trajectory for the virtual system since $\mathbb{0}_m \in \Gamma(x)$ for all $x \in \mathcal{K}$. Substituting these two trajectories implies (7.14) for $t \in [0, \tau_{\max}(x_0))$. We now establish that $\tau_{\max}(x_0) = +\infty$. To this end, note that the bound (7.14) implies that the trajectory ϕ_{x_0} remains in the compact set $\{x \in \mathbb{R}^n \mid \|x\|_P \leq \|x_0\|_P\}$ for $t \in [0, \tau_{\max}(x_0))$ for which it is defined. Since this set is compact, the trajectory cannot escape in a finite amount of time, meaning that the trajectory is forward complete. This reasoning proves statements (i) and (ii). Statement (iii) is a consequence of (ii). To prove statement (iii) note that $V(x) = \|x\|_P^2$, let $x_0 \in \mathcal{K}$ and for $h > 0$ note that the inequality (7.14) implies

$$\|\phi_{x_0}(h)\|_P^2 \leq \mathrm{e}^{-2\eta h}\|x_0\|_P^2.$$

Now subtract $\|x_0\|_P^2$ from both sides, divide by $h > 0$ and take the limit as $h \to 0^+$ to conclude

$$\lim_{h \to 0^+} \frac{\|\phi_{x_0}(h)\|_P^2 - \|x_0\|_P^2}{h} \leq \|x_0\|_P^2 \lim_{h \to 0^+} \frac{\mathrm{e}^{-2\eta h} - 1}{h}.$$

We see that the left-hand side of the above inequality is the definition of the Lie derivative of V along trajectories of (7.9) with $x_0 \in \mathcal{K}$. Moreover, the right-hand side evaluates to $-2\eta\|x_0\|_P^2$. In other words, $\dot{V}(x_0) \leq -2\eta V(x_0)$ for all $x_0 \in \mathcal{K}$, which implies statement (iii). $\square$

The key insight in Theorem 7.4.1 is that LTI systems with parametric projection controllers are a type of state-dependent Lur'e system $\dot{x} = Ax + B\,\mathrm{P}_{\Gamma(x)}(Kx)$ and that the virtual system method transforms them into standard time-varying ones with cocoercive nonlinearities, where Theorem 7.3.1 applies. Specifically, this analysis method allows us to treat general convex constraints in a unifying manner. We remark that one could also use classical results such as the circle criterion to establish stability of (7.9) but we elect to use the virtual system method to highlight its utility.

In the case that Assumptions (A1)-(A4) hold, $K = 0$, and the inequality (7.13) holds, the dynamics simply become $\dot{x} = Ax$ since $\mathbb{0}_n \in \Gamma(x)$ for all $x \in \mathcal{K}$. In other words, $K = 0$ always ensures global exponential stability under the stated assumptions since (7.13) implies that A is Hurwitz. Theorem 7.4.1 provides a novel sufficient condition for the system (7.9) to be exponential stable with $K \neq 0$ which has been designed to make $A + BK$ stable, e.g., using LQR.

7.4.3 Stability analysis for single-integrators

A special class of LTI systems for which we can ensure different results is the single integrator $\dot{x} = u^\star(x)$, namely the system (7.9) with $A = 0$, $B = I_n$. For this system, (7.13) will never hold with $\eta > 0$ since A is not Hurwitz.

Theorem 7.4.2 (Exponential stability for single integrators with parametric projection-based controllers)**.** *Consider the dynamics* (7.9) *with* $A = 0$ *and* $B = I_n$ *and suppose Assumptions* (A1)-(A4) *hold. Further suppose* $K = K^\top \preceq -\eta I_n, \eta > 0$. *Then from any* $x_0 \in \mathcal{K}$,

(i) solutions to (7.9), ϕ_{x_0}, *are forward-complete and*

(ii) solutions asymptotically converge to the set $\mathcal{X}_{\text{eq}} := \{x \in \mathbb{R}^n \mid \mathrm{P}_{\Gamma(x)}(Kx) = \mathbb{0}_n\}$.

Moreover, under the additional assumption that $\mathbb{0}_n \in \operatorname{int}(\Gamma(\mathbb{0}_n))$, *the following statement holds:*

(iii) if $\phi_{x_0}(t) \to \mathbb{0}_n$ as $t \to \infty$, then there exists $M(x_0) > 0$ such that

$$\|\phi_{x_0}(t)\|_2 \leq M(x_0)\mathrm{e}^{-\eta t}\|x_0\|_2. \tag{7.17}$$

Proof. Consider as a Lyapunov function candidate $V(x) = -\frac{1}{2}x^\top Kx$. The Lie derivative of V along trajectories of the dynamical system (7.9) is

$$\begin{aligned}\dot{V}(x) &= -x^\top K\,\mathrm{P}_{\Gamma(x)}(Kx) = -(Kx - \mathbb{0}_n)^\top(\mathrm{P}_{\Gamma(x)}(Kx) - \mathrm{P}_{\Gamma(x)}(\mathbb{0}))\\ &\leq -\|\,\mathrm{P}_{\Gamma(x)}(Kx) - \mathrm{P}_{\Gamma(x)}(\mathbb{0}_n)\|_2^2 \leq 0,\end{aligned}$$

where we have used that $\mathbb{0}_n \in \Gamma(x)$ for all x and cocoercivity of $\mathrm{P}_{\Gamma(x)}$, see, e.g., [72, Eq. (2)]. Since $\dot{V}(x) \leq 0$, we preclude finite escape time and thus conclude statement (i). To establish asymptotic convergence to $\mathcal{X}_{\mathrm{eq}}$, we invoke LaSalle's invariance principle and see that trajectories converge to the largest forward-invariant set in $\{x \in \mathbb{R}^n \mid \dot{V}(x) = 0\}$. However, since $\dot{V}(x) \leq -\|\,\mathrm{P}_{\Gamma(x)}(Kx)\|_2^2$, $\dot{V}(x) = 0$ if and only if $\mathrm{P}_{\Gamma(x)}(Kx) = \mathbb{0}_n$, i.e., $x \in \mathcal{X}_{\mathrm{eq}}$. This argument establishes statement (ii). To establish statement (iii), note that $\mathbb{0}_n \in \mathrm{int}(\Gamma(\mathbb{0}_n))$ implies, by continuity of g, that there exist an open neighborhood, $\mathcal{O}_x$ containing the origin such that $g(x, Kx) \ll \mathbb{0}_p$ for all $x \in \mathcal{O}_x$. In other words, inside this neighborhood, $u^\star(x) = Kx$. Thus, the dynamics (7.9) are locally exponentially stable inside this neighborhood. Since the trajectory is asymptotically converging to the origin and locally exponentially stable inside $\mathcal{O}_x$, we conclude (7.17). □

To prove Theorem 7.4.2, one could alternatively use the virtual system method to establish that $\|\phi_{x_0}(t)\|_{-K} \leq \|x_0\|_{-K}$ and then invoke LaSalle's invariance principle. We opt to give a more direct Lyapunov proof for simplicity.

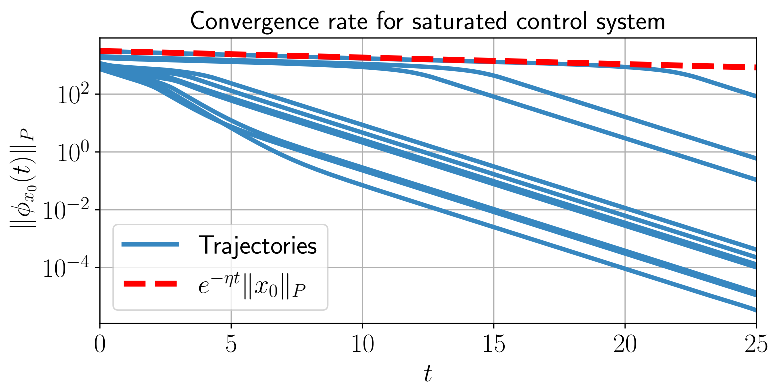


Figure 7.1: The evolution of $\|\phi_{x_0}(t)\|_P$ for 10 different x_0 and P is chosen to maximize η in (7.13). We also plot $\mathrm{e}^{-\eta t}\|x_0\|_P$, where $\|x_0\|_P$ denotes the largest value of $\|x_0\|_P$ over all randomly generated initial conditions.

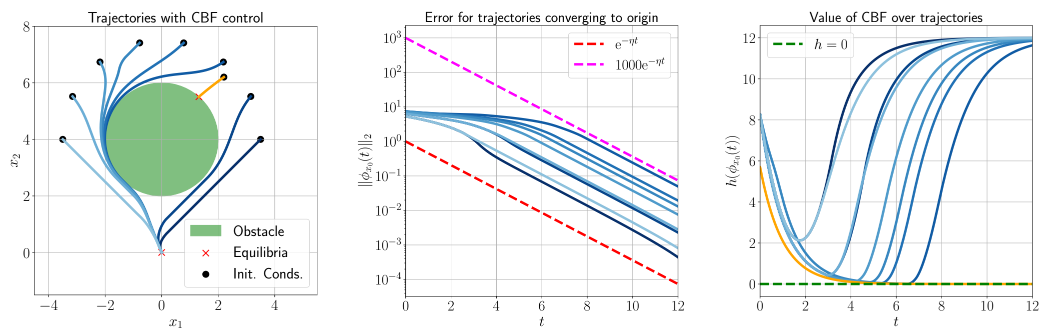


Figure 7.2: The left figure shows plots of trajectories of (7.21). We can see that most trajectories, indicated by shades of blue, converge to the origin, while one converges to a point on the boundary of the safe set, shown in orange. The center figure plots the convergence rate of trajectories that converge to the origin. It also plots $\mathrm{e}^{-\eta t}$ and $1000\mathrm{e}^{-\eta t}$ and demonstrates that the exponential convergence rate in (7.22) cannot be improved in this instance and that $1000 > M(x_0)\|x_0\|_2$ for these initial conditions. The right figure plots the evolution of the CBF, h, along trajectories.

7.5 Applications

7.5.1 State-dependent saturation control

For $v \in \mathbb{R}^n_{>0}$, define the saturation function $\mathrm{sat_v}\colon \mathbb{R}^n \to \mathbb{R}^n$ by

$$\mathrm{sat}_v(x) := \max\{-v, \min\{v, x\}\},$$

where the max and min are applied entrywise. An alternative characterization of the saturation function is via the minimization problem

$$\mathrm{sat}_v(x) = \operatorname*{argmin}_{u \in \mathbb{R}^n} \{\|u - x\|_2^2 \mid -v \leq u \leq v\}.$$

We consider the state-dependent saturated control system

$$\dot{x} = Ax + B\,\mathrm{sat}_{v(x)}(Kx), \tag{7.18}$$

where $v\colon \mathbb{R}^n \to \mathbb{R}^n_{>0}$ is a twice continuously differentiable map dictating actuation constraints as a function of the state.

When v is constant, one can use results from saturated control systems to assess the stability of (7.18). As v is state-dependent, one cannot apply these techniques here. It is straightforward to see that the system (7.18) is of the form (7.9) with $g(x, u) = (u - v(x), -u - v(x))$. Moreover, it is routine to establish that Assumptions (A1)-(A4) hold. Therefore, Theorem 7.4.1 may be applied to provide a sufficient condition for the global exponential stability of the system (7.18).

Example 9. *We consider the system* (7.18)*, where* $n = 3, m = 2$*,* $A = -I_3 + N$*,* $B = \left[\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{smallmatrix}\right]^\top$*, and* $N \in \mathbb{R}^{3\times 3}$ *is a random matrix with entries drawn from the standard normal distribution. We assume that* $K \in \mathbb{R}^{2\times 3}$ *is selected so that* $u = Kx$ *minimizes the objective* $\int_0^\infty (x(t)^\top x(t) + u(t)^\top u(t))dt$ *for* $\dot{x} = Ax + Bu$*. We let* $v(x) = \mathrm{e}^{-\|x\|_2^2/2}\mathbb{1}_m$*, where* $\mathbb{1}_m$ *is the all-ones vector. We find* $\eta > 0, \lambda \geq 0, P \in \mathbb{R}^{3\times 3}$ *satisfying* (7.13) *such that* η *is maximized and plot values of* $\|\phi_{x_0}(t)\|_P$ *for* 10 *different samples of* x_0 *from the multivariate normal distribution* $\mathcal{N}(\mathbb{0}_3, 4I_3)$ *in Figure 7.1.*

We see that all trajectories converge exponentially to the origin and that the estimated exponential convergence rate from Theorem 7.4.1 is $\eta = 0.0525$*. Empirically, we see that* *when trajectories are far from $\mathbb{0}_3$, this rate is tight since $v(x) \approx \mathbb{0}_m$ for x far from $\mathbb{0}_3$ and that the rate is very loose when trajectories are close to $\mathbb{0}_3$ since $v(x) \approx \mathbb{1}_m$ for $x \approx \mathbb{0}_3$.*

Although we have focused on state-dependent saturation control in this section, we would like to note that many structured parametric optimization problems with convex constraints may be handled. Specifically, our methods are not simply restricted to saturations and can be applied to richer classes of examples in a methodological manner.

7.5.2 Stability with control barrier functions

Consider the nonlinear control-affine system

$$\dot{x} = F(x) + G(x)u, \tag{7.19}$$

where $F\colon \mathbb{R}^n \to \mathbb{R}^n, G\colon \mathbb{R}^n \to \mathbb{R}^{n\times m}$ are locally Lipschitz.

Let $\mathcal{C} \subseteq \mathbb{R}^n$ and $h\colon \mathbb{R}^n \to \mathbb{R}$ be a sufficiently smooth function such that $\mathcal{C} = \{x \in \mathbb{R}^n \mid h(x) \geq 0\}$. The set $\mathcal{C}$ is referred to as the "safe set".

Definition 14 (Control Barrier Function [201, Definition 3])**.** *The function h is a control barrier function (CBF) for $\mathcal{C}$ if there exists a locally Lipschitz and strictly increasing function $\alpha\colon \mathbb{R} \to \mathbb{R}$ with $\alpha(0) = 0$ such that for all $x \in \mathcal{C}$, there exists $u \in \mathbb{R}^m$ with*

$$\nabla h(x)^\top F(x) + \nabla h(x)^\top G(x)u + \alpha(h(x)) \geq 0. \tag{7.20}$$

A continuous controller $u\colon \mathbb{R}^n \to \mathbb{R}^m$ which strictly satisfies (7.20) for all $x \in \mathcal{C}$ renders $\mathcal{C}$ forward-invariant under the dynamics (7.19) [205, Theorem 4].

A common way to synthesize controllers that render $\mathcal{C}$ forward invariant is via a parametric QP [201]. To this end, we consider a single-integrator being driven by the

CBF constraint (7.20) and actuator constraints:

$$\begin{aligned} \dot{x} &= u^\star(x), \\ u^\star(x) := \underset{u\in\mathbb{R}^m}{\operatorname{argmin}} \quad & \frac{1}{2}\|u - Kx\|_2^2 \\ \text{s.t.} \quad & -\nabla h(x)^\top u \leq \alpha(h(x)), \\ & -\bar{u}\mathbb{1}_n \leq u \leq \bar{u}\mathbb{1}_n, \end{aligned} \tag{7.21}$$

where $\bar{u} > 0$. While safety of these systems was previously studied in, e.g., [201], we aim to also study their stability properties. To study convergence of (7.21), we can check for conditions under which the hypotheses of Theorem 7.4.2 hold.

Corollary 7.5.1 (Exponential stability for single integrators with CBF-based controllers)**.** *Consider the dynamics (7.21) and suppose (i) h is a CBF for $\mathcal{C}$, (ii) $K = K^\top \preceq -\eta I_n$, (iii) $\mathbb{0}_n \in \operatorname{int}(\mathcal{C})$, and (iv) h and α are thrice and twice continuously differentiable, respectively. Then from any $x_0 \in \operatorname{int}(\mathcal{C})$,*

(i) solutions to (7.21), ϕ_{x_0}, are forward-complete,

(ii) solutions remain in $\mathcal{C}$ for all $t \geq 0$,

(iii) solutions converge to the set of equilibria and

(iv) if $\phi_{x_0}(t) \to \mathbb{0}_n$ as $t \to \infty$, then there exists $M(x_0) > 0$ such that

$$\|\phi_{x_0}(t)\|_2 \leq M(x_0)\mathrm{e}^{-\eta t}\|x_0\|_2. \tag{7.22}$$

Proof. We simply need to verify the assumptions of Theorem 7.4.2, namely Assumptions (A1)-(A4), and that $\mathbb{0}_n \in \operatorname{int}(\Gamma(\mathbb{0}_n))$. We can see Assumption (A1) holds by smoothness of h and α. We can see that Assumption (A2) holds since $\mathbb{0}_n \in \operatorname{int}(\mathcal{C})$ and that $\mathbb{0}_n \in \Gamma(x)$ for $x \in \mathcal{K}$ by the assumption that h is a CBF for $\mathcal{C}$ and that $\alpha(h(x)) > 0$ for $x \in \operatorname{int}(\mathcal{C})$. Assumption (A3) is clear and Assumption (A4) is guaranteed since the feasible set of (7.21) has an interior for all $x \in \operatorname{int}(\mathcal{C})$ since $\alpha(h(x)) > 0$ for $x \in \operatorname{int}(\mathcal{C})$. Finally, we can see that $\mathbb{0}_n \in \operatorname{int}(\Gamma(\mathbb{0}_n))$ since $\mathbb{0}_n \in \operatorname{int}(\mathcal{C})$. Since these assumptions hold, the consequences of Theorem 7.4.2 apply. □

In general, we cannot conclude uniqueness of equilibria. As we will see in the following example, the dynamics may have multiple equilibria, even in the case of simple CBFs. These results agree with the theory presented in [223, 224].

Example 10. *Consider a single integrator in $\mathbb{R}^2$ avoiding a disk-shaped obstacle centered at $(0, 4)$ with radius 2. The corresponding CBF is $h(x_1, x_2) = x_1^2 + (x_2 - 4)^2 - 4$ with $\alpha(r) = r$. We take $K = \left[\begin{smallmatrix} -2 & -0.5 \\ -0.5 & -1 \end{smallmatrix}\right]$ and $\bar{u} = 1$. We plot numerical simulations of (7.21) with these parameters along with the corresponding convergence rate and value of CBF along trajectories in Figure 7.2.*

We observe that trajectories converge to the set of equilibria and the majority of them converge to the origin with exponential convergence rate predicted in (7.22). While the equilibrium point on the boundary of the safe set is unavoidable in this example, we can numerically observe that there are no equilibria inside $\operatorname{int}(\mathcal{C})$ other than the origin. This result is in contrast with the CLF-CBF controllers studied in [223, 224], where there may exist additional equilibria in $\operatorname{int}(\mathcal{C})$. This numerical example provides evidence that, for exponential stability of the origin, one does not need to rely upon any CLF decrease condition, except on a measure zero set.

7.6 Discussion and future work

In this chapter we study LTI systems with controllers solving a special class of parametric programs, namely parametric projections. Using the virtual system method and a novel contractivity result for Lur'e systems, we provide sufficient conditions for the exponential stability of these systems. Separately, for single integrators, we prove convergence to the set of equilibria and an exponential convergence rate for trajectories converging to the origin. As applications, we consider state-dependent saturated control systems and CBF-based control.

We believe that there are many avenues for future research. First, it would be useful to explore relaxing Assumptions (A1)-(A4) to allow for a larger class of LTI systems, possibly using the tools in [226]. Second, it would be useful to characterize the set of trajectories converging to the origin in Theorem 7.4.2. Finally, it is important to study systems whose controllers only approximately solve the parametric program.

7.6.1 Results in set-valued analysis

Standard notions from set-valued analysis are available in [227, Section 2.1.3] (i.e., domain, continuity, limits).

Proposition 7.6.1 ([227, Lemma 2.8.2])**.** *Let $K : \mathbb{R}^p \rightrightarrows \mathbb{R}^n$ be a closed-valued and convex-valued point-to-set map. Let $\bar{x} \in \mathrm{Dom}(K)$ be given. The (single-valued) map $\Phi(x, y) := \mathrm{P}_{K(x)}(y)$ is continuous at $(\bar{x}, y)$ for all $y \in \mathbb{R}^n$ if and only if*

$$\lim_{x \to \bar{x}} K(x) = K(\bar{x}). \tag{7.23}$$

Proposition 7.6.2 ([227, Exercise 2.9.34])**.** *Let $g \colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be continuous. Let $\bar{y} \in \mathbb{R}^m$, such that $g(\cdot, \bar{y})$ is a convex function and there exists $\bar{x} \in \mathbb{R}^n$ such that $g(\bar{x}, \bar{y}) < 0$.*

Then the set-valued map $K : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ *defined by*

$$K(y) := \{x \in \mathbb{R}^n \mid g(x, y) \leq 0\}, \tag{7.24}$$

is continuous at $\bar{y}$.

Proposition 7.6.3 ([227, Proposition 2.1.17(c)])**.** *Suppose* $K : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *is a closed-valued point-to-set map. If* K *is continuous at* x*, then*

$$\lim_{y \to x} K(y) = K(x). \tag{7.25}$$

Corollary 7.6.4. *Let* $g \colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ *be continuous. Let* $\bar{x} \in \mathbb{R}^n$*, such that* $g(\bar{x}, \cdot)$ *is a convex function and there exists* $\bar{y} \in \mathbb{R}^m$ *such that* $g(\bar{x}, \bar{y}) \ll \mathbb{0}_p$*. Define the set-valued map* $K : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ *given by*

$$K(y) := \{x \in \mathbb{R}^n \mid g(x, y) \leq \mathbb{0}_p\}. \tag{7.26}$$

Then the map $\Phi(x, y) := \mathrm{P}_{K(x)}(y)$ *is continuous at* $(\bar{x}, y)$ *for all* $y \in \mathbb{R}^m$.

Proof. We simply need to extend Proposition 7.6.2 to the case that g is vector-valued. For $i \in \{1, \dots, p\}$, define the set-valued maps $K_i(y) := \{x \in \mathbb{R}^n \mid g_i(x, y) \leq 0\}$ and note that $K(y) = \bigcap_{i=1}^p K_i(y)$. We simply need to show upper and lower semicontinuity of K to conclude continuity. Regarding lower-semicontinuity, let $\bar{x} \in \mathrm{Dom}(K)$ be such that there exists $\bar{y}$ with $g(\bar{x}, \bar{y}) \ll \mathbb{0}_p$ and let $\mathcal{U} \subseteq \mathbb{R}^n$ be open and satisfy $K(\bar{x}) \cap \mathcal{U} \neq \emptyset$. By continuity of each of the K_i at $\bar{x}$ (due to Proposition 7.6.2), there exists an open set $\mathcal{N}_i$ containing $\bar{x}$ such that for each $x^i \in \mathcal{N}_i$, $K_i(x^i) \cap \mathcal{U} \neq \emptyset$ for all $i \in \{1, \dots, p\}$. Since a finite intersection of open sets is open, let $\mathcal{N} := \bigcap_{i=1}^p \mathcal{N}_i$. Then it is routine to check that for each $x \in \mathcal{N}$, $K(x) \cap \mathcal{U} \neq \emptyset$. This establishes lower semicontinuity.

To establish upper semicontinuity, let $\bar{x}$ satisfy the stated assumptions and let $\mathcal{V}$ be an open set with $K(\bar{x}) \subseteq \mathcal{V}$. By continuity of each of the K_i at $\bar{x}$, there exists an open neighborhood $\mathcal{N}_i$ of $\bar{x}$ such that for each $x^i \in \mathcal{N}_i$, $K_i(x) \subseteq \mathcal{V}$. Once again taking $\mathcal{N} = \bigcap_{i=1}^p \mathcal{N}$ ensures that for all $x \in \mathcal{N}$, we have that $K(x) \subseteq \mathcal{V}$. This establishes upper semicontinuity and thus continuity of K at $\bar{x}$.

The result is then a consequence of the three propositions above. $\square$

7.6.2 A continuity result from fixed point theory

To establish continuity and Lipschitzness of solutions to parametric optimization problems, it is oftentimes useful to first express them as fixed point problems. From there, we can leverage results from a parametrized version of the Banach fixed point theorem. We reproduce this result in the following proposition.

Proposition 7.6.5 ([228, Section I.6.A. Item (A.4) and Section I.3. Theorem 3.2])**.** *Let $(\mathcal{X}, d)$ be a complete metric space, $(\mathcal{A}, \varrho)$ be a metric space, and $f\colon \mathcal{X} \times \mathcal{A} \to \mathcal{X}$ satisfy the following properties:*

(i) There exists $\ell < 1$ such that for each $a \in \mathcal{A}$, the map $f(\cdot, a)$ is contractive with constant ℓ, i.e., for all $x_1, x_2 \in \mathcal{X}$,

$$d(f(x_1, a), f(x_2, a)) \leq \ell d(x_1, x_2). \tag{7.27}$$

(ii) For each $x \in \mathcal{X}$, the map $a \mapsto f(x, a)$ is continuous.

Then for each $a \in \mathcal{A}$, there exists a unique $x^\star(a)$ satisfying the fixed point equation $x^\star(a) = f(x^\star(a), a)$ and the mapping $a \mapsto x^\star(a)$ is continuous.

Moreover, if there exists $L > 0$ such that for each $x \in \mathbb{R}^n$, the map $a \mapsto f(x, a)$ is Lipschitz with constant L, then the mapping $a \mapsto x^\star(a)$ is Lipschitz with estimate

$L/(1-\ell)$.

7.6.3 Well-posedness of virtual system

Before establishing well-posedness of the virtual system, we require a continuity result for parametric projections.

Lemma 7.6.6. *Let $g\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ be continuous. Let $\bar{x} \in \mathbb{R}^n$, such that $g(\bar{x}, \cdot)$ is a convex function and there exists $\bar{u} \in \mathbb{R}^m$ such that $g(\bar{x}, \bar{u}) \ll \mathbb{0}_p$. Define the set-valued map $\Gamma(x) := \{u \in \mathbb{R}^n \mid g(x,u) \leq \mathbb{0}_p\}$. Then the map $\Phi(x,u) := \mathrm{P}_{\Gamma(x)}(u)$ is continuous at $(\bar{x}, u)$ for all $u \in \mathbb{R}^m$.*

Proof. The proof is a consequence of Exercise 2.9.34, Proposition 2.1.17(c), and Lemma 2.8.2 in [227]. □

Given $x_0 \in \mathcal{K}$, we're interested in the properties of the map $\bar{u}\colon \mathbb{R}^m \times [0, \tau_{\max}(x_0)) \to \mathbb{R}^m$ given by

$$\begin{aligned} \bar{u}(z,t) := \operatorname*{argmin}_{u \in \mathbb{R}^m} \quad & f(z,u) \\ \text{s.t.} \quad & g(\phi_{x_0}(t), u) \leq \mathbb{0}_p. \end{aligned} \tag{7.28}$$

This is exactly the map that appears in the virtual system analysis, i.e., for $f(z,u) = \frac{1}{2}\|u-z\|_2^2$, the virtual system in the proof of Theorem 7.4.1 is $\dot{y} = Ay + B\bar{u}(Ky, t)$.

Lemma 7.6.7. *Suppose that for all $z \in \mathbb{R}^n$, $f(z,\cdot)$ is strongly convex and $g(z,\cdot)$ is convex. Consider the ODE $\dot{x} = Ax + Bu^\star(x)$, where $u^\star(x)$ is the solution to (7.10). For $x_0 \in \mathcal{K}$, let ϕ_{x_0} denote a solution of the ODE from initial condition x_0 on the interval $[0, \tau_{\max}(x_0))$. Assume that for all $t \in [0, \tau_{\max}(x_0))$ that $\phi_{x_0}(t) \in \mathcal{K}$ and that $\nabla_u f$ is Lipschitz in both of its arguments. Then the map $\bar{u}$ in (7.28) is Lipschitz in its first argument and continuous in its second.*

Proof. We first make an important observation. The value of $\bar{u}$ at the point (z,t) solves the fixed point equation

$$\bar{u}(z,t) = \mathrm{P}_{\Gamma(\phi_{x_0}(t))}(\bar{u}(z,t) - \gamma\nabla_u f(z,\bar{u}(z,t))), \tag{7.29}$$

where $\gamma > 0$ is a constant to be tuned to make the map contractive. Let $\bar{f}\colon \mathbb{R}^m \times \mathbb{R}^n \times [0,\tau_{\max}(x_0)) \to \mathbb{R}^m$ denote this fixed-point map, i.e., $\bar{f}(u,z,t) := \mathrm{P}_{\Gamma(\phi_{x_0}(t))}(u - \gamma\nabla_u f(z,u))$. Under the assumption that $\nabla_u f$ is Lipschitz, from standard results on projected gradient algorithms, e.g., [72, pp. 31], we know that for fixed z,t, the map $u \mapsto \bar{f}(u,z,t)$ is a contraction for $0 < \gamma < 2/L$, where L is the Lipschitz constant of $\nabla_u f(z,\cdot)$. Henceforth, we let γ be in this range. To establish Lipschitzness in z, we note that at fixed u,t, the mapping $z \mapsto \bar{f}(u,z,t)$ is Lipschitz under the assumption that the mapping $\nabla_u f(\cdot,u)$ is Lipschitz. Thus, at fixed t, we can apply Proposition 7.6.5 to conclude that the mapping $z \mapsto \bar{u}(z,t)$ is Lipschitz and thus continuous in its first argument.

To establish continuity in t, we first study the point-to-set mapping $t \rightrightarrows \Gamma(\phi_{x_0}(t))$. By assumption that for all t, we have that $\phi_{x_0}(t) \in \mathcal{K}$, we conclude by Corollary 7.6.6 that the mapping $\Phi\colon \mathbb{R}^m \times [0,\tau_{\max}(x_0)) \to \mathbb{R}^m$ given by $\Phi(u,t) := \mathrm{P}_{\Gamma(\phi_{x_0}(t))}(u)$ is continuous on its domain. Since $\bar{f}(u,z,t) = \Phi(u - \gamma\nabla_u f(z,u),t)$, we have that $\bar{f}(u,z,\cdot)$ is continuous as well. And since for all z,t, we know that $u \mapsto \bar{f}(u,z,t)$ is a contraction, by Proposition 7.6.5, we know that for all $z \in \mathbb{R}^m$, the map $t \mapsto \bar{u}(z,t)$ is continuous. $\square$

Chapter 8

Conclusions and Future Work

> There is no beginning, no middle, no end, no suspense, no moral, no causes, no effects. What we love in our books are the depths of many marvelous moments seen all at one time.
>
> *Kurt Vonnegut Jr.,*
> *Slaughterhouse-Five*

8.1 Summary

In this thesis, we have studied contraction theory and presented some application in neural networks, optimization, and control. Specifically, we focused on the special case when dynamical systems are contracting with respect to a metric induced by a norm and presented a comprehensive theory including necessary and sufficient conditions for contraction.

To be specific, in Chapter 2 we introduced weak pairings as a novel tool for contraction analysis and have demonstrated how we can use them to easily establish incremental input to state stability of contracting dynamics and a small-gain type sufficient condition for the interconnection of contracting dynamics to remain contracting.

In Chapter 3, we introduced a non-Euclidean monotone operator theory akin to the standard theory on Hilbert spaces. By adopting the language of weak pairings, we retain many of the desirable consequences of monotone operators and are able to design novel fixed point algorithms for computing zeros of them.

In Chapter 4, we leverage the non-Euclidean contraction theory framework to design a novel implicit neural network architecture and establish its well-posedness and robustness to ℓ_∞ norm-bounded adversarial attacks.

In Chapter 5 we study the contraction of a large class of continuous-time neural networks with respect to diagonally-weighted ℓ_1 and ℓ_∞ norms. In many cases, contraction is guaranteed by checking that a certain Metzler matrix is Hurwitz or by solving a linear program. We highlight key advantages of a non-Euclidean contraction analysis compared to a Euclidean one.

In Chapter 6, we study time-varying convex optimization problems from a contraction theory. First, we establish that many canonical dynamical systems for solving convex optimization problems are in fact contracting with respect to a suitably-defined Euclidean norm. Second, we establish that time-varying contracting dynamics enjoy an equilibrium tracking property whereby the tracking error between any trajectory and the time-varying equilibrium curve is upper bounded. Third, we design a feedforward prediction term for time-varying contracting dynamics that ensures exponential exact tracking of the time-varying equilibrium curve. We apply these results in both numerical and hardware experiments.

Finally, in Chapter 7, we study the stability of an interconnection of a linear control system with a controller which is the solution of a parametric optimization problem. By leveraging the virtual system method of analysis, we present a novel sufficient condition for closed-loop stability based upon the satisfaction of a particular LMI.

8.2 Future work

This thesis opens several avenues for future work. I will highlight possible extensions along with some related literature. Other open problems are presented in our opinion paper [229].

8.2.1 Neural networks

In the direction of neural networks, there has been increased interest in using contraction theory for establishing their stability or robustness. In our own work, we have published [203] which establishes sharp conditions for the Euclidean contraction of Hopfield and firing rate neural networks when their synaptic matrices are symmetric. These results were then used in [230] to study locally competitive networks as firing rate neural networks. Tangentially related, we also studied a generalization of Hopfield neural networks to those that can retrieve more than one memory in [231] although there was no contraction analysis in this work. Possible extensions of these results include (i) sharp conditions for the contractivity of neural networks under synaptic matrices which are Lyapunov diagonally stable, (ii) contraction analysis of more general classes of neural networks inspired by machine learning including LSTMs and simple models of transformers, and (iii) extensions to higher-fidelity biologically inspired neural networks including spiking models and contraction to limit cycles.

On the side of machine learning, in this thesis, we presented robustness of implicit neural networks using Lipschitz bounds. In follow-up work, we additionally studied robustness using interval reachability [232, 233, 234]. Methods based upon interval reachability appear to yield sharper robustness bounds at the expense of more costly training. Beyond these two metrics for robustness, we can consider alternative approaches including the convex outer adversarial polytope from [127] or alternative training approaches such as that taken in [235]. Finally, let us comment that ℓ_∞ norm-bounded adversaries are not the only interesting type of adversarial robustness. In many cases, it is unclear what the correct notion of robustness should be. This is increasingly the case in the age of large language models (LLMs) when "robustness" means that the LLM should not reply to any message that can yield a "dangerous" output.

8.2.2 Optimization and control

In the direction of optimization, we have extended results from [170]. In [236], we revisited the linear equality-constrained minimization problem without the assumption of full row-rankedness and showed semicontraction of the dynamics, i.e., contraction in an invariant subspace and showed the sharpest-known convergence rate for these dynamics. In [237], we studied the proximal gradient dynamics and showed monotonicity of the objective function without convexity and smoothness and demonstrated exponential convergence under a modified PL condition. In [238], we studied dynamical systems for convex but not strongly convex optimization. Under the assumption of local strong contractivity but global nonexpansiveness of the dynamics, we establish a type of semi-global exponential stability referred to as linear-exponential convergence and provide an explicit estimate for the convergence rate. I believe that a sharp local analysis for dynamical systems corresponding to minimization of certain nonconvex optimization problems remains missing. I also believe that an analysis of stochastic optimization problems and convergence in the space of measures akin to the early results of [239] could be of interest.

Finally, in the direction of contraction for control, let me mention the work [240] on learning contracting dynamics from data. The trajectories of these contracting dynamics can be used as dynamical systems-based motion planners in robotics. In this direction, the pioneering work [62] established control contraction metrics and necessary and sufficient conditions for the existence of a controller for which the closed-loop dynamics are contracting. The conditions typically reduce to PDEs and are intractable for moderately-sized control systems. Therefore, follow-up work has studied methods for learning controllers that promote closed-loop contraction, see, e.g., [241]. While these controllers can work well in practice, theoretical guarantees are largely lacking. I believe that novel methods for certifying closed-loop contraction will be critical in learning controllers from data.

Bibliography

[1] 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** (1972), no. 5 480–486.

[2] F. Bullo, *Contraction Theory for Dynamical Systems*. Kindle Direct Publishing, 1.2 ed., 2024.

[3] R. Abraham, J. E. Marsden, and T. S. Ratiu, *Manifolds, Tensor Analysis, and Applications*, vol. 75 of *Applied Mathematical Sciences*. Springer, 2 ed., 1988.

[4] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, 2 ed., 2017.

[5] A. Davydov, S. Jafarpour, and F. Bullo, *Non-Euclidean contraction theory for robust nonlinear stability*, *IEEE Transactions on Automatic Control* **67** (2022), no. 12 6667–6681.

[6] G. Russo, M. Di Bernardo, and E. D. Sontag, *Global entrainment of transcriptional systems to periodic inputs*, *PLoS Computational Biology* **6** (2010), no. 4 e1000739.

[7] 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** (2001), no. 2 360–370.

[8] M. A. Al-Radhawi and D. Angeli, *New approach to the stability of chemical reaction networks: Piecewise linear in rates Lyapunov functions*, *IEEE Transactions on Automatic Control* **61** (2016), no. 1 76–89.

[9] S. Coogan and M. Arcak, *A compartmental model for traffic networks and its dynamical behavior*, *IEEE Transactions on Automatic Control* **60** (2015), no. 10 2698–2703.

[10] 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** (2015), no. 1 57–67.

[11] S. Coogan, *A contractive approach to separable Lyapunov functions for monotone systems*, *Automatica* **106** (2019) 349–357.

[12] J. Monteil, G. Russo, and R. Shorten, *On $\mathcal{L}_\infty$ string stability of nonlinear bidirectional asymmetric heterogeneous platoon systems*, *Automatica* **105** (2019) 198–205.

[13] 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** (2013), no. 5 1328–1331.

[14] Z. Aminzare and E. D. Sontag, *Synchronization of diffusively-connected nonlinear systems: Results based on contractions with respect to general norms*, *IEEE Transactions on Network Science and Engineering* **1** (2014), no. 2 91–106.

[15] D. C. Lewis, *Metric properties of differential equations*, *American Journal of Mathematics* **71** (1949), no. 2 294–312.

[16] B. P. Demidovič, *Dissipativity of a system of nonlinear differential equations in the large*, *Uspekhi Matematicheskikh Nauk* **16** (1961), no. 3(99) 216.

[17] N. N. Krasovskiĭ, *Stability of Motion. Applications of Lyapunov's Second Method to Differential Systems and Equations with Delay.* Stanford University Press, 1963.

[18] 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.

[19] S. M. Lozinskii, *Error estimate for numerical integration of ordinary differential equations. I*, *Izvestiya Vysshikh Uchebnykh Zavedenii. Matematika* **5** (1958) 52–90. (in Russian).

[20] T. Ström, *On logarithmic norms*, *SIAM Journal on Numerical Analysis* **12** (1975), no. 5 741–753.

[21] C. A. Desoer and M. Vidyasagar, *Feedback Systems: Input-Output Properties.* Academic Press, 1975.

[22] M. Vidyasagar, *Nonlinear Systems Analysis.* SIAM, 2002.

[23] W. Lohmiller and J.-J. E. Slotine, *On contraction analysis for non-linear systems*, *Automatica* **34** (1998), no. 6 683–696.

[24] W. Wang and J. J. Slotine, *On partial contraction analysis for coupled nonlinear oscillators*, *Biological Cybernetics* **92** (2005), no. 1 38–53.

[25] Q. C. Pham, N. Tabareau, and J.-J. E. Slotine, *A contraction theory approach to stochastic incremental stability*, *IEEE Transactions on Automatic Control* **54** (2009), no. 4 816–820.

[26] I. Higueras and G. Söderlind, *Logarithmic norms and nonlinear DAE stability*, *BIT Numerical Mathematics* **42** (2002), no. 4 823–841.

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

[28] F. Forni and R. Sepulchre, *A differential Lyapunov framework for contraction analysis*, *IEEE Transactions on Automatic Control* **59** (2014), no. 3 614–628.

[29] Z. Aminzare, Y. Shafi, M. Arcak, and E. D. Sontag, *Guaranteeing spatial uniformity in reaction-diffusion systems using weighted L_2 norm contractions*, in *A Systems Theoretic Approach to Systems and Synthetic Biology I: Models and System Characterizations*, ch. 3, pp. 73–101. Springer, 2014.

[30] I. R. Manchester and J.-J. E. Slotine, *Transverse contraction criteria for existence, stability, and robustness of a limit cycle*, *Systems & Control Letters* **63** (2014) 32–38.

[31] M. Margaliot, E. D. Sontag, and T. Tuller, *Contraction after small transients*, *Automatica* **67** (2016) 178–184.

[32] 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** (2022), no. 3 1285–1300.

[33] C. Wu, I. Kanevskiy, and M. Margaliot, *k-contraction: Theory and applications*, *Automatica* **136** (2022) 110048.

[34] E. Hairer, S. P. Nørsett, and G. Wanner, *Solving Ordinary Differential Equations I. Nonstiff Problems*. Springer, 1993.

[35] W. Lu and T. Chen, *New approach to synchronization analysis of linearly coupled ordinary differential systems*, *Physica D: Nonlinear Phenomena* **213** (2006), no. 2 214–230.

[36] 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** (2005), no. 2059 2257–2267.

[37] L. D'Alto and M. Corless, *Incremental quadratic stability*, *Numerical Algebra, Control and Optimization* **3** (2013) 175–201.

[38] G. Söderlind, *The logarithmic norm. History and modern theory*, *BIT Numerical Mathematics* **46** (2006), no. 3 631–652.

[39] Z. Aminzare and E. D. Sontag, *Logarithmic Lipschitz norms and diffusion-induced instability*, *Nonlinear Analysis: Theory, Methods & Applications* **83** (2013) 31–49.

[40] 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** (1976), no. 6 355–379.

[41] A. F. Filippov, *Differential Equations with Discontinuous Righthand Sides*. Kluwer, 1988.

[42] P. DeLellis, M. Di Bernardo, and G. Russo, *On QUAD, Lipschitz, and contracting vector fields for consensus and synchronization of networks*, *IEEE Transactions on Circuits and Systems I: Regular Papers* **58** (2011), no. 3 576–583.

[43] H. Maeda, S. Kodama, and Y. Ohta, *Asymptotic behavior of nonlinear compartmental systems: Nonoscillation and stability*, *IEEE Transactions on Circuits and Systems* **25** (1978), no. 6 372–378.

[44] J. A. Jacquez and C. P. Simon, *Qualitative theory of compartmental systems*, *SIAM Review* **35** (1993), no. 1 43–79.

[45] 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*, pp. 3835–3847, Dec., 2014.

[46] Z. Aminzare, *On Synchronous Behavior in Complex Nonlinear Dynamical Systems*. PhD thesis, Rutgers, 2015.

[47] S. S. Dragomir, *Semi-Inner Products and Applications*. Nova Science Publishers, 2004.

[48] G. Lumer, *Semi-inner-product spaces*, *Transactions of the American Mathematical Society* **100** (1961) 29–43.

[49] J. R. Giles, *Classes of semi-inner-product spaces*, *Transactions of the American Mathematical Society* **129** (1967), no. 3 436–446.

[50] A. Hamadeh, E. Sontag, and D. Del Vecchio, *A contraction approach to input tracking via high gain feedback*, in *IEEE Conf. on Decision and Control*, pp. 7689–7694, Dec., 2015.

[51] M. Arcak, C. Meissen, and A. Packard, *Networks of Dissipative Systems: Compositional Certification of Stability, Performance, and Safety*. Springer, 2016.

[52] K. Deimling, *Nonlinear Functional Analysis.* Springer, 1985.

[53] J. M. Danskin, *The theory of max-min, with applications*, *SIAM Journal on Applied Mathematics* **14** (1966), no. 4 641–664.

[54] T. Lorenz, *Mutational Analysis: A Joint Framework For Cauchy Problems In And Beyond Vector Spaces.* Springer, 2010.

[55] H. K. Khalil, *Nonlinear Systems.* Prentice Hall, 3 ed., 2002.

[56] W. A. Coppel, *Stability and Asymptotic Behavior of Differential Equations.* Heath, 1965.

[57] A. Pavlov, A. Pogromsky, N. Van de Wouw, and H. Nijmeijer, *Convergent dynamics, a tribute to Boris Pavlovich Demidovich*, *Systems & Control Letters* **52** (2004), no. 3-4 257–261.

[58] D. Angeli, *A Lyapunov approach to incremental stability properties*, *IEEE Transactions on Automatic Control* **47** (2002), no. 3 410–421.

[59] M. Di Bernardo, D. Liuzza, and G. Russo, *Contraction analysis for a class of nondifferentiable systems with applications to stability and network synchronization*, *SIAM Journal on Control and Optimization* **52** (2014), no. 5 3203–3227.

[60] B. Bamieh and M. Dahleh, *On robust stability with structured time-invariant perturbations*, *Systems & Control Letters* **21** (1993), no. 2 103–108.

[61] O. Pastravanu and M. Voicu, *Generalized matrix diagonal stability and linear dynamical systems*, *Linear Algebra and its Applications* **419** (2006), no. 2 299–310.

[62] I. R. Manchester and J.-J. E. Slotine, *Control contraction metrics: Convex and intrinsic criteria for nonlinear feedback design*, *IEEE Transactions on Automatic Control* **62** (2017), no. 6 3046–3053.

[63] F. Forni and R. Sepulchre, *Differentially positive systems*, *IEEE Transactions on Automatic Control* **61** (2016), no. 2 346–359.

[64] A. Davydov, S. Jafarpour, A. V. Proskurnikov, and F. Bullo, *Non-Euclidean monotone operator theory and applications*, *Journal of Machine Learning Research* **25** (2024), no. 307 1–33.

[65] P. L. Combettes and J.-C. Pesquet, *Deep neural network structures solving variational inequalities*, *Set-Valued and Variational Analysis* **28** (2020), no. 3 491–518.

[66] E. Winston and J. Z. Kolter, *Monotone operator equilibrium networks*, in *Advances in Neural Information Processing Systems*, 2020.

[67] P. L. Combettes and J.-C. Pesquet, *Fixed point strategies in data science*, *IEEE Transactions on Signal Processing* **69** (2021) 3878–3905.

[68] A. Simonetto, *Time-varying convex optimization via time-varying averaged operators*, 2017. ArXiv e-print:1704.07338.

[69] A. Bernstein, E. Dall'Anese, and A. Simonetto, *Online primal-dual methods with measurement feedback for time-varying convex optimization*, *IEEE Transactions on Signal Processing* **67** (2019), no. 8 1978–1991.

[70] L. Pavel, *Distributed GNE seeking under partial-decision information over networks via a doubly-augmented operator splitting approach*, *IEEE Transactions on Automatic Control* **65** (2020), no. 4 1584–1597.

[71] T. Chaffey, F. Forni, and R. Sepulchre, *Scaled relative graphs for system analysis*, in *IEEE Conf. on Decision and Control*, pp. 3166–3172, 2021.

[72] E. K. Ryu and S. Boyd, *Primer on monotone operator methods*, *Applied Computational Mathematics* **15** (2016), no. 1 3–43.

[73] P. L. Combettes, *Monotone operator theory in convex optimization*, *Mathematical Programming* **170** (2018) 177–206.

[74] E. K. Ryu and W. Yin, *Large-Scale Convex Optimization via Monotone Operators*. Cambridge University Press, 2022.

[75] I. J. Goodfellow, J. Shlens, and C. Szegedy, *Explaining and harnessing adversarial examples*, in *International Conference on Learning Representations (ICLR)*, 2015.

[76] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, 1997.

[77] G. J. Minty, *Monotone (nonlinear) operators in Hilbert space*, *Duke Mathematical Journal* **29** (1962), no. 3 341–346.

[78] F. E. Browder, *Existence and perturbation theorems for nonlinear maximal monotone operators in Banach spaces*, *Bulletin of the American Mathematical Society* **73** (1967) 322–327.

[79] G. J. Minty, *On the monotonicity of the gradient of a convex function*, *Pacific Journal of Mathematics* **14** (1964), no. 1 243–247.

[80] R. T. Rockafellar, *Characterization of the subdifferentials of convex functions*, *Pacific Journal of Mathematics* **17** (1966), no. 3 497–510.

[81] E. K. Ryu, R. Hannah, and W. Yin, *Scaled relative graphs: nonexpansive operators via 2D Euclidean geometry*, *Mathematical Programming* **194** (2021), no. 1-2 569–619.

[82] G. Gidel, H. Berard, G. Vignoud, P. Vincent, and S. Lacoste-Julien, *A variational inequality perspective on generative adversarial networks*, in *International Conference on Learning Representations*, 2019.

[83] P. L. Combettes and J.-C. Pesquet, *Lipschitz certificates for layered network structures driven by averaged activation operators*, *SIAM Journal on Mathematics of Data Science* **2** (2020), no. 2 529–557.

[84] C. Pabbaraju, E. Winston, and J. Z. Kolter, *Estimating Lipschitz constants of monotone deep equilibrium models*, in *International Conference on Learning Representations*, 2021.

[85] B. Ahn, C. Kim, Y. Hong, and H. J. Kim, *Invertible monotone operators for normalizing flows*, in *Advances in Neural Information Processing Systems*, 2022.

[86] P. Balamurugan and F. Bach, *Stochastic variance reduction methods for saddle-point problems*, in *Advances in Neural Information Processing Systems*, p. 1416–1424, 2016.

[87] J. Diakonikolas, *Halpern iteration for near-optimal and parameter-free monotone inclusion and strong solutions to variational inequalities*, in *Conference on Learning Theory*, vol. 125 of *Proceedings of Machine Learning Research*, pp. 1428–1451, PMLR, 2020.

[88] X. Cai, C. Song, C. A. Guzmán, and J. Diakonikolas, *Stochastic Halpern iteration with variance reduction for stochastic monotone inclusions*, in *Advances in Neural Information Processing Systems*, 2022.

[89] T. Pethick, P. Latafat, P. Patrinos, O. Fercoq, and V. Cevher, *Escaping limit cycles: Global convergence for constrained nonconvex-nonconcave minimax problems*, in *International Conference on Learning Representations*, 2022.

[90] X. Zhang, W. B. Haskell, and Z. Ye, *A unifying framework for variance-reduced algorithms for findings zeroes of monotone operators*, *Journal of Machine Learning Research* **23** (2022), no. 60 1–44.

[91] A. Alacaoglu, A. Böhm, and Y. Malitsky, *Beyond the golden ratio for variational inequality algorithms*, *Journal of Machine Learning Research* **24** (2023), no. 172 1–33.

[92] M. I. Jordan, T. Lin, and M. Zampetakis, *First-order algorithms for nonlinear generalized Nash equilibrium problems*, *Journal of Machine Learning Research* **24** (2023), no. 38 1–46.

[93] T. Yang, M. I. Jordan, and T. Chavdarova, *Solving constrained variational inequalities via a first-order interior point-based method*, in *International Conference on Learning Representations*, 2023.

[94] C. Chidume, *Geometric Properties of Banach Spaces and Nonlinear Iterations.* Springer, 2009.

[95] H. H. Bauschke, J. M. Borwein, and P. L. Combettes, *Bregman monotone optimization algorithms*, *SIAM Journal on Control and Optimization* **42** (2003), no. 2 596–636.

[96] G. Söderlind, *Bounds on nonlinear operators in finite-dimensional Banach spaces*, *Numerische Mathematik* **50** (1986) 27–44.

[97] 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.

[98] 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.

[99] 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., 2022.

[100] A. Davydov, A. V. Proskurnikov, and F. Bullo, *Non-Euclidean contraction analysis of continuous-time neural networks*, *IEEE Transactions on Automatic Control* **70** (2025), no. 1.

[101] G. Giorgi and S. Komlósi, *Dini derivatives in optimization — Part I*, *Rivista di Matematica Per Le Scienze Economiche e Sociali* **15** (1992), no. 1 3–30.

[102] S. Ishikawa, *Fixed points and iteration of a nonexpansive mapping in a Banach space*, *Proceedings of the American Mathematical Society* **59** (1976), no. 1 65–71.

[103] R. Cominetti, J. A. Soto, and J. Vaisman, *On the rate of convergence of Krasnosel'skii-Mann iterations and their connection with sums of Bernoullis*, *Israel Journal of Mathematics* **199** (2014), no. 2 757–772.

[104] B. Mercier, *Inéquations Variationnelles de la Mécanique.* Publications Mathématiques d'Orsay. Université de Paris-Sud, Département de Mathématique, 1980.

[105] J.-J. Moreau, *Fonctions convexes duales et points proximaux dans un espace hilbertien*, *Comptes rendus hebdomadaires des séances de l'Académie des sciences* **255** (1962) 2897–2899.

[106] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, *SIAM Journal on Control and Optimization* **14** (1976), no. 5 877–898.

[107] J. Li, C. Fang, and Z. Lin, *Lifted proximal operator machines*, in *AAAI Conference on Artificial Intelligence*, pp. 4181–4188, 2019.

[108] N. Parikh and S. Boyd, *Proximal algorithms*, *Foundations and Trends in Optimization* **1** (2014), no. 3 127–239.

[109] S. Bai, J. Z. Kolter, and V. Koltun, *Deep equilibrium models*, in *Advances in Neural Information Processing Systems*, 2019.

[110] L. El Ghaoui, F. Gu, B. Travacca, A. Askari, and A. Tsai, *Implicit deep learning*, *SIAM Journal on Mathematics of Data Science* **3** (2021), no. 3 930–958.

[111] S. Diamond and S. Boyd, *CVXPY: A Python-embedded modeling language for convex optimization*, *Journal of Machine Learning Research* **17** (2016), no. 83 1–5.

[112] M. Fazlyab, A. Robey, H. Hassani, M. Morari, and G. J. Pappas, *Efficient and accurate estimation of Lipschitz constants for deep neural networks*, in *Advances in Neural Information Processing Systems*, 2019.

[113] P. Pauli, A. Koch, J. Berberich, P. Kohler, and F. Allgower, *Training robust neural networks using Lipschitz bounds*, *IEEE Control Systems Letters* (2021).

[114] M. Revay, R. Wang, and I. R. Manchester, *Lipschitz bounded equilibrium networks*, *arXiv preprint arXiv:2010.01732* (2020).

[115] P. Tseng, *A modified forward-backward splitting method for maximal monotone mappings*, *SIAM Journal on Control and Optimization* **38** (2000), no. 2 431–446.

[116] D. Davis and W. Yin, *A three-operator splitting scheme and its optimization applications*, *Set-Valued and Variational Analysis* **25** (2017) 829–858.

[117] J. Lee and E. K. Ryu, *Accelerating value iteration with anchoring*, in *Advances in Neural Information Processing Systems*, 2023.

[118] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, *Intriguing properties of neural networks*, in *International Conference on Learning Representations*, 2014.

[119] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud, *Neural ordinary differential equations*, in *Advances in Neural Information Processing Systems*, 2018.

[120] 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.

[121] A. Agrawal, B. Amos, S. Barratt, S. Boyd, S. Diamond, and J. Z. Kolter, *Differentiable convex optimization layers*, in *Advances in Neural Information Processing Systems*, 2019.

[122] B. Amos, I. Jimenez, J. Sacks, B. Boots, and J. Z. Kolter, *Differentiable MPC for end-to-end planning and control*, in *Advances in Neural Information Processing Systems*, 2018.

[123] K. Kawaguchi, *On the theory of implicit deep learning: Global convergence with implicit layers*, in *International Conference on Learning Representations*, 2021.

[124] N. Papernot, P. McDaniel, X. Wu, S. Jha, and A. Swami, *Distillation as a defense to adversarial perturbations against deep neural networks*, in *IEEE Symposium on Security and Privacy*, pp. 582–597, 2016.

[125] N. Carlini and D. Wagner, *Adversarial examples are not easily detected: Bypassing ten detection methods*, in *ACM Workshop on Artificial Intelligence and Security*, pp. 3–14, 2017.

[126] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, *Towards deep learning models resistant to adversarial attacks*, in *International Conference on Machine Learning*, 2018.

[127] E. Wong and J. Z. Kolter, *Provable defenses against adversarial examples via the convex outer adversarial polytope*, in *International Conference on Machine Learning*, pp. 5286–5295, 2018.

[128] M. Fazlyab, M. Morari, and G. J. Pappas, *Safety verification and robustness analysis of neural networks via quadratic constraints and semidefinite programming*, *IEEE Transactions on Automatic Control* **67** (2022), no. 1 1–15.

[129] V. Krishnan, A. A. A. Makdah, and F. Pasqualetti, *Lipschitz bounds and provably robust training by Laplacian smoothing*, in *Advances in Neural Information Processing Systems*, 2020.

[130] M. Revay, R. Wang, and I. R. Manchester, *A convex parameterization of robust recurrent neural networks*, *IEEE Control Systems Letters* **5** (2021), no. 4 1363–1368.

[131] T. C. Lim, *On fixed point stability for set-valued contractive mappings with applications to generalized differential equations*, *Journal of Mathematical Analysis and Applications* **110** (1985), no. 2 436–441.

[132] H. Hosseini, B. Xiao, M. Jaiswal, and R. Poovendran, *On the limitation of convolutional neural networks in recognizing negative images*, in *IEEE International Conference on Machine Learning and Applications*, pp. 352–358, 2017.

[133] Y. Fang and T. G. Kincaid, *Stability analysis of dynamical neural networks*, *IEEE Transactions on Neural Networks* **7** (1996), no. 4 996–1006.

[134] W. He and J. Cao, *Exponential synchronization of chaotic neural networks: a matrix measure approach*, *Nonlinear Dynamics* **55** (2009) 55–65.

[135] A. Davydov, A. V. Proskurnikov, and F. Bullo, *Non-Euclidean contractivity of recurrent neural networks*, in *American Control Conference*, (Atlanta, USA), pp. 1527–1534, May, 2022.

[136] J. J. Hopfield, *Neurons with graded response have collective computational properties like those of two-state neurons*, *Proceedings of the National Academy of Sciences* **81** (1984), no. 10 3088–3092.

[137] K. D. Miller and F. Fumarola, *Mathematical equivalence of two common forms of firing rate models of neural networks*, *Neural Computation* **24** (2012), no. 1 25–31.

[138] 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** (2021) 135–169.

[139] A. N. Michel, J. A. Farrell, and W. Porod, *Qualitative analysis of neural networks*, *IEEE Transactions on Circuits and Systems* **36** (1989), no. 2 229–243.

[140] E. Kaszkurewicz and A. Bhaya, *On a class of globally stable neural circuits*, *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications* **41** (1994), no. 2 171–174.

[141] M. Forti, S. Manetti, and M. Marini, *Necessary and sufficient condition for absolute stability of neural networks*, *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications* **41** (1994), no. 7 491–494.

[142] M. Forti and A. Tesi, *New conditions for global stability of neural networks with application to linear and quadratic programming problems*, *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications* **42** (1995), no. 7 354–366.

[143] E. Nozari and J. Cortés, *Hierarchical selective recruitment in linear-threshold brain networks-part I: Single-layer dynamics and selective inhibition*, *IEEE Transactions on Automatic Control* **66** (2021), no. 3 949–964.

[144] S. Arik, *A note on the global stability of dynamical neural networks*, *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications* **49** (2002), no. 4 502–504.

[145] H. Zhang, Z. Wang, and D. Liu, *A comprehensive review of stability analysis of continuous-time recurrent neural networks*, *IEEE Transactions on Neural Networks and Learning Systems* **25** (2014), no. 7 1229–1262.

[146] S. Jafarpour, A. Davydov, and F. Bullo, *Non-Euclidean contraction theory for monotone and positive systems*, *IEEE Transactions on Automatic Control* **68** (2023), no. 9 5653–5660.

[147] V. Centorrino, F. Bullo, and G. Russo, *Contraction analysis of Hopfield neural networks with Hebbian learning*, in *IEEE Conf. on Decision and Control*, (Cancún, México), Dec., 2022.

[148] L. Kozachkov, M. Ennis, and J.-J. E. Slotine, *RNNs of RNNs: Recursive construction of stable assemblies of recurrent neural networks*, in *Advances in Neural Information Processing Systems*, Dec., 2022.

[149] M. Morita, *Associative memory with nonmonotone dynamics*, *Neural Networks* **6** (1993), no. 1 115–126.

[150] M. Revay and I. Manchester, *Contracting implicit recurrent neural networks: Stable models with improved trainability*, in *Conference on Learning for Dynamics and Control*, vol. 120, pp. 393–403, 2020.

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

[152] A. Rantzer, *Scalable control of positive systems*, *European Journal of Control* **24** (2015) 72–80.

[153] P. Van At, *Diagonal transformation methods for computing the maximal eigenvalue and eigenvector of a nonnegative irreducible matrix*, *Linear Algebra and its Applications* **148** (1991) 93–123.

[154] F. Bullo, *Contraction Theory for Dynamical Systems.* Kindle Direct Publishing, 1.1 ed., 2023.

[155] D. D. Šiljak, *Large-Scale Dynamic Systems Stability & Structure.* North-Holland, 1978.

[156] J. Stoer and C. Witzgall, *Transformations by diagonal matrices in a normed space*, *Numerische Mathematik* **4** (1962) 158–171.

[157] P. J. Moylan, *Matrices with positive principal minors*, *Linear Algebra and its Applications* **17** (1977), no. 1 53–58.

[158] F. Bullo, *Lectures on Network Systems.* Kindle Direct Publishing, 1.6 ed., Jan., 2022.

[159] G. P. Barker, A. Berman, and R. J. Plemmons, *Positive diagonal solutions to the Lyapunov equations*, *Linear and Multilinear Algebra* **5** (1978), no. 4 249–256.

[160] D. Krotov and J. J. Hopfield, *Dense associative memory for pattern recognition*, in *Advances in Neural Information Processing Systems*, 2016.

[161] E. Kaszkurewicz and A. Bhaya, *Matrix Diagonal Stability in Systems and Computation.* Springer, 2000.

[162] V. A. Yakubovich, *Method of matrix inequalities in theory of nonlinear control systems stability. I. Forced oscillations absolute stability*, *Avtomatika i Telemekhanika* **25** (1964), no. 7 1017–1029. (In Russian).

[163] A. Richards and J. How, *Mixed-integer programming for control*, in *American Control Conference*, pp. 2676–2683, 2005.

[164] C.-Y. Cheng, K.-H. Lin, and C.-W. Shih, *Multistability in recurrent neural networks*, *SIAM Journal on Applied Mathematics* **66** (2006), no. 4 1301–1320.

[165] Y. Wang, *A new concept using LSTM neural networks for dynamic system identification*, in *American Control Conference*, pp. 5324–5329, 2017.

[166] H. Yin, P. Seiler, and M. Arcak, *Stability analysis using quadratic constraints for systems with neural network controllers*, *IEEE Transactions on Automatic Control* **67** (2022), no. 4 1980–1987.

[167] A. Harapanahalli, S. Jafarpour, and S. Coogan, *Contraction-guided adaptive partitioning for reachability analysis of neural network controlled systems*, in *IEEE Conf. on Decision and Control*, Dec., 2023.

[168] R. A. Horn and C. R. Johnson, *Matrix Analysis.* Cambridge University Press, 2nd ed., 2012.

[169] F. H. Clarke, *Optimization and Nonsmooth Analysis.* John Wiley & Sons, 1983.

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

[171] K. J. Arrow, L. Hurwicz, and H. Uzawa, eds., *Studies in Linear and Nonlinear Programming.* Stanford University Press, 1958.

[172] J. J. Hopfield and D. W. Tank, *"Neural" computation of decisions in optimization problems*, *Biological Cybernetics* **52** (1985), no. 3 141–152.

[173] M. P. Kennedy and L. O. Chua, *Neural networks for nonlinear programming*, *IEEE Transactions on Circuits and Systems* **35** (1988), no. 5 554–562.

[174] R. W. Brockett, *Dynamical systems that sort lists, diagonalize matrices, and solve linear programming problems*, *Linear Algebra and its Applications* **146** (1991) 79–91.

[175] 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** (2022), no. 1 474–486.

[176] G. Tanaka, T. Yamane, J. B. Héroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano, and A. Hirose, *Recent advances in physical reservoir computing: A review*, *Neural Networks* **115** (2019) 100–123.

[177] C. D. Schuman, S. R. Kulkarni, M. Parsa, J. P. Mitchell, P. Date, and B. Kay, *Opportunities for neuromorphic computing algorithms and applications*, *Nature Computational Science* **2** (2022) 10–19.

[178] S. Xie, G. Russo, and R. H. Middleton, *Scalability in nonlinear network systems affected by delays and disturbances*, *IEEE Transactions on Control of Network Systems* **8** (2021), no. 3 1128–1138.

[179] P. H. A. Ngoc and H. Trinh, *On contraction of functional differential equations*, *SIAM Journal on Control and Optimization* **56** (2018), no. 3 2377–2397.

[180] A. Hauswirth, Z. He, S. Bolognani, G. Hug, and F. Dörfler, *Optimization algorithms as robust feedback controllers*, *Annual Reviews in Control* **57** (2024) 100941.

[181] J. Wang and N. Elia, *A control perspective for centralized and distributed convex optimization*, in *IEEE Conf. on Decision and Control and European Control Conference*, (Orlando, USA), pp. 3800–3805, 2011.

[182] A. Cherukuri, B. Gharesifard, and J. Cortes, *Saddle-point dynamics: Conditions for asymptotic stability of saddle points*, *SIAM Journal on Control and Optimization* **55** (2017), no. 1 486–511.

[183] G. Qu and N. Li, *On the exponential stability of primal-dual gradient dynamics*, *IEEE Control Systems Letters* **3** (2019), no. 1 43–48.

[184] J. Cortés and S. K. Niederländer, *Distributed coordination for nonsmooth convex optimization via saddle-point dynamics*, *Journal of Nonlinear Science* **29** (2019), no. 4 1247–1272.

[185] N. K. Dhingra, S. Z. Khong, and M. R. Jovanović, *The proximal augmented Lagrangian method for nonsmooth composite optimization*, *IEEE Transactions on Automatic Control* **64** (2019), no. 7 2861–2868.

[186] H. D. Nguyen, T. L. Vu, K. Turitsyn, and J.-J. E. Slotine, *Contraction and robustness of continuous time primal-dual dynamics*, *IEEE Control Systems Letters* **2** (2018), no. 4 755–760.

[187] P. Cisneros-Velarde, S. Jafarpour, and F. Bullo, *Distributed and time-varying primal-dual dynamics via contraction analysis*, *IEEE Transactions on Automatic Control* **67** (2022), no. 7 3560–3566.

[188] P. M. Wensing and J.-J. E. Slotine, *Beyond convexity — Contraction and global convergence of gradient descent*, *PLoS One* **15** (2020), no. 8 1–29.

[189] A. Simonetto, A. Mokhtari, A. Koppel, G. Leus, and A. Ribeiro, *A class of prediction-correction methods for time-varying convex optimization*, *IEEE Transactions on Signal Processing* **64** (2016), no. 17 4576–4591.

[190] A. Simonetto and E. Dall'Anese, *Prediction-correction algorithms for time-varying constrained optimization*, *IEEE Transactions on Signal Processing* **65** (2017), no. 20 5481–5494.

[191] M. Fazlyab, S. Paternain, V. M. Preciado, and A. Ribeiro, *Prediction-correction interior-point method for time-varying convex optimization*, *IEEE Transactions on Automatic Control* **63** (2018), no. 7 1973–1986.

[192] E. Dall'Anese, A. Simonetto, S. Becker, and L. Madden, *Optimization and learning with information streams: Time-varying algorithms and applications*, *IEEE Signal Processing Magazine* **37** (2020), no. 3 71–83.

[193] A. Simonetto, E. Dall'Anese, S. Paternain, G. Leus, and G. B. Giannakis, *Time-varying convex optimization: Time-structured algorithms and applications*, *Proceedings of the IEEE* **108** (2020), no. 11 2032–2048.

[194] M. Colombino, E. Dall'Anese, and A. Bernstein, *Online optimization as a feedback controller: Stability and tracking*, *IEEE Transactions on Control of Network Systems* **7** (2020), no. 1 422–432.

[195] Y. Zhao and M. N. S. Swamy, *A novel technique for tracking time-varying minimum and its applications*, in *IEEE Canadian Conference on Electrical and Computer Engineering*, CCECE-98, (Waterloo, Canada), 1998.

[196] B. Abbas and H. Attouch, *Dynamical systems and forward–backward algorithms associated with the sum of a convex subdifferential and a monotone cocoercive operator*, *Optimization* **64** (2015), no. 10 2223–2252.

[197] X.-B. Gao, *Exponential stability of globally projected dynamic systems*, *IEEE Transactions on Neural Networks* **14** (2003), no. 2 426–431.

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

[199] M. Fortin, *Minimization of some non-differentiable functionals by the augmented Lagrangian method of Hestenes and Powell*, *Applied Mathematics and Optimization* **2** (1975) 236–250.

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

[201] A. D. Ames, X. Xu, J. W. Grizzle, and P. Tabuada, *Control barrier function based quadratic programs for safety critical systems*, *IEEE Transactions on Automatic Control* **62** (2017), no. 8 3861–3876.

[202] D. Paccagnan, B. Gentile, F. Parise, M. Kamgarpour, and J. Lygeros, *Nash and Wardrop equilibria in aggregative games with coupling constraints*, *IEEE Transactions on Automatic Control* **64** (2019), no. 4 1373–1388.

[203] V. Centorrino, A. Gokhale, A. Davydov, G. Russo, and F. Bullo, *Euclidean contractivity of neural networks with symmetric weights*, *IEEE Control Systems Letters* **7** (2023) 1724–1729.

[204] A. V. Fiacco, *Sensitivity analysis for nonlinear programming using penalty methods*, *Mathematical Programming* **10** (1976), no. 1 287–311.

[205] R. Konda, A. D. Ames, and S. Coogan, *Characterizing safety: Minimal control barrier functions from scalar comparison systems*, *IEEE Control Systems Letters* **5** (2021), no. 2 523–528.

[206] S. Wilson, P. Glotfelter, L. Wang, S. Mayya, G. Notomista, M. Mote, and M. Egerstedt, *The Robotarium: Globally impactful opportunities, challenges, and lessons learned in remote-access, distributed control of multirobot systems*, *IEEE Control Systems* **40** (2020), no. 1 26–44.

[207] G. Russo and F. Wirth, *Matrix measures, stability and contraction theory for dynamical systems on time scales*, *Discrete & Continuous Dynamical Systems - B* **27** (2022), no. 6 3345–3374.

[208] Z. Aminzare, *Stochastic logarithmic Lipschitz constants: A tool to analyze contractivity of stochastic differential equations*, *IEEE Control Systems Letters* **6** (2022) 2311–2316.

[209] Y. Ding, J. Lavaei, and M. Arcak, *Time-variation in online nonconvex optimization enables escaping from spurious local minima*, *IEEE Transactions on Automatic Control* **68** (2023), no. 1 156–171.

[210] L. Lessard, B. Recht, and A. Packard, *Analysis and design of optimization algorithms via integral quadratic constraints*, *SIAM Journal on Optimization* **26** (2016), no. 1 57–95.

[211] L. Lessard, *The analysis of optimization algorithms: A dissipativity approach*, *IEEE Control Systems* **42** (2022), no. 3 58–72.

[212] D. Ding and M. R. Jovanović, *Global exponential stability of primal-dual gradient flow dynamics based on the proximal augmented Lagrangian: A Lyapunov-based approach*, in *IEEE Conf. on Decision and Control*, p. 4836–4841, 2020.

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

[214] F. Borrelli, A. Bemporad, and M. Morari, *Predictive Control for Linear and Hybrid Systems*. Cambridge University Press, 2017.

[215] B. Bank, J. Guddat, D. Klatte, B. Kummer, and K. Tammer, *Non-Linear Parametric Optimization*. 1982.

[216] B. J. Morris, M. J. Powell, and A. D. Ames, *Continuity and smoothness properties of nonlinear optimization-based feedback controllers*, in *IEEE Conf. on Decision and Control*, pp. 151–158, 2015.

[217] P. Mestres, A. Allibhoy, and J. Cortés, *Robinson's counterexample and regularity properties of optimization-based controllers*, *arXiv e-print:2311.13167* (2023).

[218] M. Jankovic, *Robust control barrier functions for constrained stabilization of nonlinear systems*, *Automatica* **96** (2018) 359–367.

[219] P. Mestres and J. Cortés, *Optimization-based safe stabilizing feedback with guaranteed region of attraction*, *IEEE Control Systems Letters* **7** (2023) 367–372.

[220] G. Li, B. Lennox, and W. P. Heath, *Concise stability conditions for systems with static nonlinear feedback expressed by a quadratic program*, *IET Control Theory & Applications* **2** (2008), no. 7 554–563.

[221] V. Andrieu and S. Tarbouriech, *LMI conditions for contraction and synchronization*, in *IFAC Symposium on Nonlinear Control Systems*, vol. 52, pp. 616–621, 2019.

[222] M. Giaccagli, V. Andrieu, S. Tarbouriech, and D. Astolfi, *LMI conditions for contraction, integral action, and output feedback stabilization for a class of nonlinear systems*, *Automatica* **154** (2023) 111106.

[223] M. F. Reis, A. P. Aguiar, and P. Tabuada, *Control barrier function-based quadratic programs introduce undesirable asymptotically stable equilibria*, *IEEE Control Systems Letters* **5** (2021), no. 2 731–736.

[224] X. Tan and D. V. Dimarogonas, *On the undesired equilibria induced by control barrier function based quadratic programs*, *Automatica* **159** (2024) 111359.

[225] I. Pólik and T. Terlaky, *A survey of the S-lemma*, *SIAM Review* **49** (2007), no. 3 371–418.

[226] A. V. Proskurnikov, A. Davydov, and F. Bullo, *The Yakubovich S-Lemma revisited: Stability and contractivity in non-Euclidean norms*, *SIAM Journal on Control and Optimization* **61** (2023), no. 4 1955–1978.

[227] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems.* Springer, 2003.

[228] A. Granas and J. Dugundji, *Fixed Point Theory.* Springer, 2003.

[229] A. Davydov and F. Bullo, *Perspectives on contractivity in control, optimization and learning*, *IEEE Control Systems Letters* **8** (2024) 2087–2098.

[230] V. Centorrino, A. Gokhale, A. Davydov, G. Russo, and F. Bullo, *Positive competitive networks for sparse reconstruction*, *Neural Computation* **36** (2024), no. 6 1163–1197.

[231] A. Davydov, S. Jaffe, A. K. Singh, and F. Bullo, *Retrieving k-nearest memories with modern Hopfield networks*, in *2023 NEURIPS Workshop on Associative Memory and Hopfield Networks*, Dec., 2023.

[232] S. Jafarpour, M. Abate, A. Davydov, F. Bullo, and S. Coogan, *Robustness certificates for implicit neural networks: A mixed monotone contractive approach*, in *Learning for Dynamics and Control Conference*, June, 2022.

[233] A. Davydov, S. Jafarpour, M. Abate, F. Bullo, and S. Coogan, *Comparative analysis of interval reachability for robust implicit and feedforward neural networks*, in *IEEE Conf. on Decision and Control*, (Cancún, México), 2022.

[234] S. Jafarpour, A. Davydov, M. Abate, F. Bullo, and S. Coogan, *Robust training and verification of implicit neural networks: A non-Euclidean contractive approach*, in *ICML Workshop on Formal Verification of Machine Learning*, July, 2022.

[235] H. Zhang, Y. Yu, J. Jiao, E. Xing, L. El Ghaoui, and M. Jordan, *Theoretically principled trade-off between robustness and accuracy*, in *International Conference on Machine Learning*, vol. 97, pp. 7472–7482, 2019.

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

[237] A. Gokhale, A. Davydov, and F. Bullo, *Proximal gradient dynamics: Monotonicity, exponential convergence, and applications*, *IEEE Control Systems Letters* **8** (2024) 2853–2858.

[238] V. Centorrino, A. Davydov, A. Gokhale, G. Russo, and F. Bullo, *On weakly contracting dynamics for convex optimization*, *IEEE Control Systems Letters* **8** (2024) 1745–1750.

[239] S. Betteti and F. Bullo, *Contraction and concentration of measures with applications to theoretical neuroscience*, in *IEEE Conf. on Decision and Control*, Mar., 2025. Submitted.

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

[241] D. Sun, S. Jha, and C. Fan, *Learning certified control using contraction metric*, in *Conference on Robot Learning*, vol. 155, pp. 1519–1539, 2021.