Advances in Contraction Theory for Robust Optimization, Control, and Neural Computation

Francesco Bullo[1] Samuel Coogan[2] Emiliano Dall'Anese[3] Ian R. Manchester[4] Giovanni Russo[5]

***Abstract*— This tutorial provides an overview of recent developments in contraction theory, highlighting theoretical advances, practical applications, and emerging extensions. We explore topics including time-varying convex optimization through equilibrium tracking, biologically plausible optimization in neural networks, and the analysis of interconnected and sampled-data systems. Additional focus is given to linear differential inclusions, reachability analysis, and the integration of contraction theory with robust, control-oriented machine learning.**

INTRODUCTION

Contraction theory is a powerful mathematical framework for analyzing convergence, robustness, and modularity in dynamical systems, optimization algorithms, and learning methods. This modern framework offers a unified approach to studying systems with exponential and incremental stability, as well as robustness to disturbances and uncertainty.

The study of contraction theory traces back to the groundbreaking work by Banach [11]. While fixed point theory has been extensively developed (see, e.g., [55], [65], [117]), applications of contraction theory to dynamical systems remain more recent and comparatively less explored. Key contributions include the classic works on stability and convergence by Demidovič [35], Krasovskiĭ [68], Desoer and Haneda [36], the pioneering efforts of Lohmiller and Slotine [71], [72], and subsequent advances such as [32], [45], [75], [87], [94]. Important applications include the analysis of coupled oscillators and network [6], [38], [53], [106]. Recent surveys and theses provide broader context and perspectives [3], [5], [51], [93], [105], [114].

This tutorial introduces the core concepts and tools of contraction theory and surveys recent developments in theory, computation, and applications. We highlight its use in control, online and biologically plausible optimization, verification, reachability analysis, and machine learning, with an emphasis on computational efficiency, modular design, and robustness.

Here is a synopsis of the content and contributions in each of the following sections. First, Section I presents recent results on time-varying convex optimization using contraction theory to analyze and design dynamical systems that track changing optimal solutions. Under assumptions of strong contractivity and Lipschitz dependence on time-varying parameters, the analysis provides explicit bounds on tracking error and residuals. When the parameter derivative is known, a feedforward correction improves performance, achieving exponential convergence. The section also extends contractivity results to standard optimization problems, including monotone inclusions, equality-constrained problems, and composite minimization.

Section II studies how contraction theory supports the analysis and design of biologically plausible neural networks that solve composite convex optimization problems. For concreteness, firing-rate networks are considered, with a focus on a specific application relevant in a broad range of domains – positive sparse reconstruction, modeled through proximal gradient dynamics. These dynamics lead to a positive system called the positive firing rate competitive network (PFCN), whose equilibria solve the original optimization problem. The PFCN is shown to be weakly contracting and locally strongly contracting under certain conditions (e.g., RIP of the dictionary), ensuring convergence to a sparse solution.

Section III studies the stability of systems combining continuous-time dynamics with discrete-time updates through sampling and zero-order hold. Using contraction theory, this section analyzes systems where control inputs are computed via finite-iteration optimization algorithms, a common setting in online and model predictive control. The results show that global exponential stability can be ensured under certain Lipschitz and contractivity conditions, either by verifying contraction of a reduced continuous-time model or by applying a small-gain condition. Explicit bounds are provided to guide the selection of sampling time and iteration count.

Section IV analyzes the contractivity properties of linear differential inclusions (LDIs), showing how contraction can be verified by bounding the Jacobian over a set of states. The section introduces a refined LDI tailored to convergence toward a known trajectory, based on a mixed Jacobian construction that partially fixes the state to obtain tighter bounds. This approach enables more accurate contraction analysis, particularly in reachability and tracking contexts. The method is computationally efficient via interval analysis

FB is supported in part by AFOSR grant FA9550-22-1-0059 and ARO MURI grant W911NF-24-1-0228. SC is supported in part by AFOSR grant FA9550-23-1-0303. ED is supported in part by the NSF award 2444163 and the AFOSR grant FA9550-23-1-0740. IM is supported in part by the Australian Research Council (DP230101014) and Google LLC. GR is supported by the European Union-Next Generation EU Mission 4 Component 1 CUP E53D23014640001.

[1] Center for Control, Dynamical Systems, and Computation, UC Santa Barbara, CA, USA. bullo@ucsb.edu.

[2] School of Electrical and Computer Engineering, Georgia Institute of Technology, GA, USA. scoogan@gatech.edu.

[3] Department of Electrical and Computer Engineering, Boston University, MA, USA. edallane@bu.edu.

[4] Australian Centre for Robotics and School of Aerospace, Mechanical and Mechatronic Engineering, University of Sydney, Australia. ian.manchester@sydney.edu.au

[5] DIEM, University of Salerno, Italy giovarusso@unisa.it.

and is illustrated on a robot arm model, where it yields significantly tighter reachable set overapproximations compared to existing techniques.

Finally, Section V presents contraction-based methods for ensuring robustness and stability in machine learning models used for control. The section outlines how direct parameterizations can enable scalable training of models—such as Lipschitz and bi-Lipschitz neural networks—by embedding stability constraints into unconstrained optimization. These methods apply to both static (feedforward) and dynamic (recurrent) networks, and allow learning models that are certifiably stable, contractive, or invertible. Applications include robust control policies, observer design, neural Lyapunov functions, and training dynamics with guaranteed convergence and robustness properties.

I. Time-Varying Convex Optimization: A Contraction and Equilibrium Tracking Approach

This first topic is a summary of the recent results on time-varying convex optimization and contracting dynamics in [31]. While time-invariant contracting dynamics is extensively studied, the behavior of time-varying systems is much less understood.

A. Introduction

Mathematical optimization is fundamental in various scientific and engineering disciplines, traditionally approached via numerical iterative methods. A complementary viewpoint treats optimization algorithms as continuous-time dynamical systems, analyzing their properties such as stability and robustness. This perspective, originating from early works by Arrow, Hurwicz, and Uzawa [10], has gained renewed interest due to applications in online optimization [14], reservoir computing [102], and neuromorphic computing [95].

The recent paper [31] addresses dynamical systems designed for time-varying convex optimization problems. Such problems require algorithms capable of accurately tracking changing optimal solutions while remaining robust to practical uncertainties such as noise and delays. These critical properties can be simultaneously ensured by establishing that the dynamical system is strongly infinitesimally contracting [18], [71], guaranteeing exponential convergence, robustness, and incremental stability. Previous studies have extensively analyzed optimization algorithms' stability [26], [28], [37], [89], [109], yet few explicitly address contractivity, with exceptions including primal-dual dynamics [27], [84]. For time-varying optimization, Newton-type methods have been explored in discrete-time [97], [99] and continuous-time [42], with recent reviews available in [29], [98].

B. Contributions

Consider the dynamical system

$$\dot{x}(t) = F(x(t), \theta(t)), \quad x(0) = x_0, \tag{1}$$

where the vector field $F : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^n$ depends explicitly on a time-varying parameter trajectory $\theta : \mathbb{R}_{\geq 0} \to \Theta \subseteq \mathbb{R}^d$. The parameter curve $\theta(t)$ is assumed to be continuously differentiable. To analyze this system's behavior, we introduce the notion of a *time-varying equilibrium trajectory* $t \mapsto x^\star(\theta(t))$, defined implicitly by $F(x^\star(\theta(t)), \theta(t)) = \mathbb{0}_n$ for all $t \geq 0$.

We impose two critical assumptions on the vector field F:

Assumption 1 (Contraction property): There exists a norm $\|\cdot\|$ on $\mathbb{R}^n$ and a constant $c > 0$ such that, for every fixed parameter $\theta \in \Theta$, the map $x \mapsto F(x, \theta)$ is strongly infinitesimally contracting with rate c, that is

$$(F(x_1, \theta) - F(x_2, \theta))^\top (x_1 - x_2) \leq -c\|x_1 - x_2\|^2$$

for all $x_1, x_2 \in \mathbb{R}^n$.

Assumption 2 (Lipschitz parameter-dependence): The vector field F is Lipschitz continuous with respect to the parameter θ with constant $\ell_\theta > 0$, uniformly over all states $x \in \mathbb{R}^n$. Specifically, there exists a norm $\|\cdot\|_\Theta$ on the parameter space Θ such that:

$$\|F(x, \theta_1) - F(x, \theta_2)\| \leq \ell_\theta \|\theta_1 - \theta_2\|_\Theta,$$

for all $x \in \mathbb{R}^n, \ \theta_1, \theta_2 \in \Theta$.

These assumptions ensure that (i) for each fixed parameter θ, the system admits a unique equilibrium $x^\star(\theta)$ satisfying $F(x^\star(\theta), \theta) = \mathbb{0}_n$, and (ii) the equilibrium map $\theta \mapsto x^\star(\theta)$ is itself Lipschitz continuous with constant ℓ_θ / c. Hence, the time-varying equilibrium trajectory $x^\star(\theta(\cdot))$ inherits this Lipschitz continuity, providing a natural reference for tracking performance. Under these assumptions, the following theorem provides quantitative bounds for how trajectories of the original system track the time-varying equilibrium trajectory, see also Figure 1.

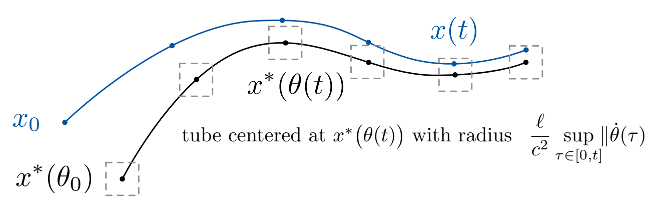


Fig. 1: Each solution $x(t)$ to the contracting dynamics (1) asymptotically approaches or remains inside a tube around the time-varying equilibrium trajectory $x^\star(\theta(t))$.

Theorem 1 (Equilibrium tracking): Let the dynamics $\dot{x}(t) = F(x(t), \theta(t))$ satisfy Assumptions 1 and 2. Then, for all $t \geq 0$, the trajectory tracking error satisfies the integral inequality:

$$\|x(t) - x^\star(\theta(t))\| \leq \mathrm{e}^{-ct}\|x(0) - x^\star(\theta(0))\| + \frac{\ell_\theta}{c}\int_0^t \mathrm{e}^{-c(t-\tau)}\|\dot{\theta}(\tau)\|_\Theta d\tau. \tag{2}$$

Additionally, the residual of the vector field satisfies:

$$\|F(x(t), \theta(t))\| \leq \mathrm{e}^{-ct}\|F(x(0), \theta(0))\| + \ell_\theta \int_0^t \mathrm{e}^{-c(t-\tau)}\|\dot{\theta}(\tau)\|_\Theta d\tau. \tag{3}$$

In the long-term limit, assuming the parameter rate $\|\dot{\theta}(t)\|_\Theta$ remains bounded, the following asymptotic bounds hold:

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

and

$$\limsup_{t\to\infty} \|F(x(t), \theta(t))\| \leq \frac{\ell_\theta}{c} \limsup_{t\to\infty} \|\dot{\theta}(t)\|_\Theta.$$

This theorem provides explicit and practical insights, establishing that the asymptotic tracking error is directly proportional to the maximum rate of change of the parameter trajectory, inversely scaled by the square of the contraction rate c. Thus, increasing the contraction rate improves the system's ability to track the moving equilibrium.

Next, consider the scenario where the dynamics F are continuously differentiable, and the time derivative $\dot{\theta}(t)$ is known or measurable. In this situation, one can design modified dynamics incorporating a *feedforward prediction term* that exploits the knowledge of $\dot{\theta}(t)$ to improve tracking performance:

$$\begin{aligned}\dot{x}(t) &= F(x(t), \theta(t)) \\ &\quad - (D_x F(x(t), \theta(t)))^{-1} D_\theta F(x(t), \theta(t)) \dot{\theta}(t). \end{aligned} \quad (4)$$

Theorem 2 (Exact tracking with feedforward prediction): Assuming that F is continuously differentiable and contracting 1, the contracting dynamics with feedfoward prediction (4) ensure that both the residual and tracking error exponentially decay to zero. Specifically, for all $t \geq 0$:

$$\|F(x(t), \theta(t))\| \leq \mathrm{e}^{-ct} \|F(x(0), \theta(0))\|, \quad (5)$$

and

$$\|x(t) - x^\star(\theta(t))\| \leq \frac{1}{c} \mathrm{e}^{-ct} \|F(x(0), \theta(0))\|. \quad (6)$$

If the vector field F is additionally Lipschitz continuous in the state variable x with constant ℓ_x, uniformly in θ, then the following stronger bound holds:

$$\|x(t) - x^\star(\theta(t))\| \leq \frac{\ell_x}{c} \mathrm{e}^{-ct} \|x(0) - x^\star(\theta(0))\|.$$

The second theorem significantly strengthens the result by guaranteeing exponential convergence to zero tracking error, given accurate knowledge of the parameter's rate of change. This insight is valuable in applications where precise tracking of time-varying equilibria is critical. This theorem generalizes previous Euclidean-norm results [42], [97], [99], [119]. Taken together, these two theorems provide comprehensive guidance for the analysis and design of robust, contraction-based dynamical systems capable of accurately tracking time-varying equilibria, with clear and explicit relationships between key system parameters and performance metrics.

Additionally, the paper [31] studies canonical optimization frameworks: (i) monotone inclusions, establishing strong contractivity for forward-backward splitting dynamics, thus enhancing earlier exponential stability findings [47], [58]; (ii) linear equality-constrained optimization, rigorously analyzing primal-dual dynamics contractivity, improving upon existing studies [27], [84], [89]; and (iii) composite minimization, extending proximal augmented Lagrangian methods' results to demonstrate enhanced contractivity properties, improving upon previous exponential convergence results [37], [89].

Followup works include the application and extension of these ideas to control barrier functions in [77]–[79] and related broad perspectives in [30].

II. Contraction in Neural Networks And Biologically Plausible Optimization

This topic, a summary of recent results from [20]–[23], focuses on network-level aspects of contraction theory. Emphasis is on the role of contraction in the analysis and design of biologically plausible neural networks computing the optimal solution of composite convex optimization problems.

A. Introduction

The design and analysis of neural circuits is central to several scientific and engineering domains, such as machine learning, control, neuroscience, signal processing. In this context, two key design/analysis steps are: *Step (i)* transcribing optimization problems into a neural dynamics of the form

$$\dot{x}_F(t) = -x_F(t) + \Psi\left(W x_F(t) + u_F(t)\right). \quad (7)$$

In this *firing-rate* dynamics (FNN) $x_F \in \mathbb{R}^n$ is the state of n neurons, Ψ is the activation function, W is the synaptic matrix, u_F is an external stimulus [33]; *Step (ii)* characterizing convergence of (7) to an equilibrium that is also the optimal solution of the problem. We are interested in FNNs since these dynamics might hold an advantage over Hopfield models in terms of biological plausibility: if Ψ is non-negative, the FNN is a positive system and the state can be interpreted as a vector of firing rates. Thus, FNNs offer a more natural interpretation of negative (positive) synaptic connections as inhibitory (excitatory). Here, as an example, these two steps are illustrated on positive sparse reconstruction (SR) problems. Given an input $u \in \mathbb{R}^m$ (e.g., an m-pixel picture), the positive SR problem consists in reconstructing u with a linear combination of a sparse vector $y \in \mathbb{R}^n_{\geq 0}$ and a *dictionary* $\Phi \in \mathbb{R}^{m \times n}$ composed of n (unit-norm) vectors. This problem can be formulated as:

$$\min_{y \in \mathbb{R}^n_{\geq 0}} \frac{1}{2} \left\| u - \Phi y \right\|_2^2 + \lambda \|y\|_1. \quad (8)$$

Definition 3 (k-sparse vector, RIP [19]): Let $k < n$ be natural numbers. A vector $x \in \mathbb{R}^n$ is *k-sparse* if it has at most k non-zero entries. A matrix $\Phi \in \mathbb{R}^{n \times m}$ satisfies the *restricted isometry property (RIP)* of order k if there exist a constant $\delta \in [0, 1)$, such that for all k-sparse $x \in \mathbb{R}^n$, $(1-\delta)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1+\delta)\|x\|_2^2$.

B. Contributions

Step (i): consider the composite convex optimization problem

$$\min_{x\in\mathbb{R}^n} f(x,u) + g(x), \tag{9}$$

where: (i) $u \in \mathbb{R}^m$; (ii) $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is a convex, closed and proper (CCP) differentiable function for each u; (iii) $g : \mathbb{R}^n \to \bar{\mathbb{R}} :=]-\infty, +\infty[$ is CCP. Next, we leverage proximal operators [17], [86]. The proximal operator of $g : \mathbb{R}^n \to \bar{R}$ with parameter $\gamma > 0$, $\mathrm{prox}_{\gamma g} : \mathbb{R}^n \to \mathbb{R}^n$, is $\mathrm{prox}_{\gamma g}(x) = \arg\min_{z\in\mathbb{R}^n} g(z) + \frac{1}{2\gamma}\|x-z\|_2^2$, $\forall x \in \mathbb{R}^n$. If $g(x) = \sum_{i=1}^n g_i(x_i)$, with $g_i : \mathbb{R} \to \mathbb{R}$ being CCP, then $\mathrm{prox}_{\gamma g}(x)$ exists, is unique and is given by

$$\big(\mathrm{prox}_{\gamma g}(x)\big)_i = \mathrm{prox}_{\gamma g_i}(x_i), \quad i \in \{1,\dots,n\}.$$

That is, if $g(x)$ is CCP and separable, the i-th component of $\mathrm{prox}_{\gamma g}(x)$ is the proximal operator of $g_i(x_i)$. In this setting [31], [58], the continuous-time proximal gradient dynamics associated to (9) is defined as

$$\dot{x} = -x + \mathrm{prox}_{\gamma g}\left(x - \gamma \nabla f(x,u)\right), \tag{10}$$

with $\gamma > 0$. Consider now the positive SR problem in (8) and note that this can be formulated as

$$\min_{y\in\mathbb{R}^n} \frac{1}{2}\|u - \Phi y\|_2^2 + \lambda\|y\|_1 + \iota_{\mathbb{R}^n_{\geq 0}}(y), \tag{11}$$

where $\iota_{\mathbb{R}^n_{\geq 0}} : \mathbb{R}^n \to [0, +\infty]$ is the *zero-infinity indicator function on* $\mathbb{R}^n_{\geq 0}$, defined as $\iota_{\mathbb{R}^n_{\geq 0}}(x) = 0$ if $x \in \mathbb{R}^n_{\geq 0}$ and $\iota_{\mathbb{R}^n_{\geq 0}}(x) = +\infty$ otherwise. As shown in [23], the function $S_1(y) := \|y\|_1 + \frac{1}{\lambda}\iota_{\mathbb{R}^n_{\geq 0}}(y)$ is CCP and separable; also, its proximal operator is the shifted ReLU. That is, $\mathrm{prox}_{\lambda S_1}(x) = \mathrm{ReLU}(x - \lambda \mathbb{1}_n)$, for all $x \in \mathbb{R}^n$. Therefore, the proximal gradient dynamics associated to (11) is [23]

$$\dot{x}(t) = -x(t) + \mathrm{ReLU}\big((I_n - \Phi^\top\Phi)x(t) + \Phi^\top u(t) - \lambda\mathbb{1}_n\big), \tag{12}$$

with $y(t) = x(t)$ and $\mathbb{1}_n$ the n-dimensional vector of ones. We term (12) as Positive Firing Rate Competitive Network (PFCN) and give the following result.

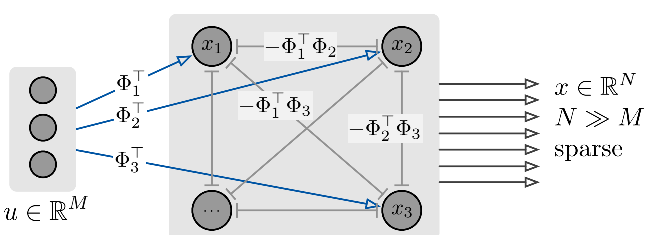


Fig. 2: An illustration of the PFCN network in equation (12).

Lemma 4: [PFCN properties [23]] The PFCN is a positive system and the vector $x^\star \in \mathbb{R}^n$ is an optimal solution of (8) if and only if it is an equilibrium of (12).

Hence, (12) is a positive system and therefore it is a FNN. Moreover, an optimal solution of (8) must also be an equilibrium of (12). We also give the following:

Definition 5 (Active and inactive neuron): Given a neural state $x^\star \in \mathbb{R}^n$, an input $u \in \mathbb{R}^m$, and a parameter $\lambda > 0$, the i-th neuron is *active* if $\mathrm{ReLU}\big(((I_n - \Phi^\top\Phi)x^\star + \Phi^\top u - \lambda\mathbb{1}_n)_i\big) \neq 0$, *inactive* otherwise.

Next, we need to characterize convergence of (12) to a sparse equilibrium vector and contraction is central to this *Step (ii)*. In particular, one can show that: (i) the PFCN is weakly contracting, i.e., non-expansive, in the positive orthant; (ii) an equilibrium of (12), say $x^\star$, with n_a active neurons is locally asymptotically stable and contracting if the dictionary is RIP of order n_a and $\delta \in [0,1)$. These two properties, yield a convergence result informally stated as follows:

Informal statement 1 (PFCN Convergence): The trajectories of (12) are bounded. If the dictionary Φ is RIP, then:

(i) the PFCN converges to an equilibrium point that is also the optimal solution of the positive SR problem (11);
(ii) convergence is linear-exponential, in the sense that the trajectory's distance from the equilibrium point initially decays at worst linearly, and then, after a transient, exponentially.

This result, formally stated in [23], not only gives convergence bounds (further sharpened in [21]) that depend on explicit biologically meaningful parameters, but can also be exploited in other settings, with different regularizers and constraints [22]. A possible extension of the result includes learning the synaptic weights. In this context, as recently shown in [20] contraction can be a valuable tool to study convergence of these neural-synaptic dynamics.

III. Contraction-theoretic Analysis of Sampled-data Systems

In this section, we provide a summary of recent stability results for the interconnection of continuous-time and discrete-time systems through sampling and zero-order hold [24]. The results leverage tools from contraction theory.

A. Introduction

We are interested in sampled-data systems of the form:

$$\dot{x}(t) = F(x(t), z(t)) \tag{13a}$$

$$z_k = \mathsf{G}^n(x(kT), z_{k-1}), \tag{13b}$$

$$z(t) = z_k, \ t \in [kT, (k+1)T) \tag{13c}$$

where $x \in \mathcal{X} \subseteq \mathbb{R}^n$ and $z \in \mathcal{Z} \subseteq \mathbb{R}^d$ are the states of the continuous and discrete-time systems, respectively, $t \in \mathbb{R}_{\geq 0}$ denotes time, $k \in \mathbb{Z}_{\geq 0}$ is the index for the updates of $z(t)$, and $T > 0$ is a given time interval. We define $\mathsf{G}^n(x,z)$ as the n-time composition of a map $z \mapsto \mathsf{G}(x,z)$, for fixed x; that is, $\mathsf{G}^1(x,z) = \mathsf{G}(x,z), \dots, \mathsf{G}^n(x,z) = \mathsf{G}(x, \mathsf{G}^{n-1}(x,z))$, for any $n \in \mathbb{Z}_{>0}$. We assume that $F : \mathcal{X} \times \mathcal{Z} \to \mathcal{X}$ and $\mathsf{G} : \mathcal{X} \times \mathcal{Z} \to \mathcal{Z}$ are continuous in their arguments.

The mathematical model in (13) is motivated by optimization-based control methods—which includes model predictive control (MPC) as a prime example [39], [81]—where the control input is implicitly determined as the solution to an optimization problem that encodes both performance objectives and system constraints. System (13) is

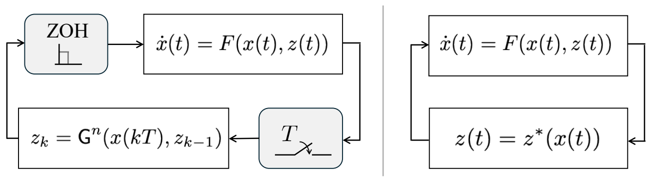


Fig. 3: *(Left)* Interconnected system (13). Here, $\mathsf{G}^n(x,z)$ is defined as $\mathsf{G}^1(x,z) = \mathsf{G}(x,z), \dots, \mathsf{G}^n(x,z) = \mathsf{G}(x, \mathsf{G}^{n-1}(x,z))$. *(Right)* Reduced model, where $z^*(x)$ is defined by $z^*(x) = \mathsf{G}(x, z^*(x))$ for any x.

particularly useful for modeling *online* and *sampled-data* implementations: in this case, $\mathsf{G}(x,z)$ represents the mapping of an algorithm used to solve the optimization problem; here, "online" refers to the scenario where only a finite number n of iterations are performed, rather than solving the optimization problem to full convergence.

For the particular case of online MPC, existing stability analysis focus on cases where the dynamics of $x(t)$ and $z(t)$ are both in continuous-time [116] or in discrete-time [70]. Recently, the discrete-time setting has been extended to more general online control frameworks in [62]. These results leverage dissipativity [116] or small-gain-type arguments [61].

The study of (13) inspired by the foundational works on two-time-scale continuous-time systems using singular perturbation theory [66] and contraction theory [34]. In (13), T and n can be seen as parameters that induce a time-scale separation between the (fast) dynamics of (13a) and the (slower) updates of $z(t)$, executed within intervals of length T and with a limited number of iterations n. When considering the case $n = +\infty$, $T > 0$, (13) encompasses classical sampled-data control systems [59], [82].

B. Contributions

To lay the foundations for our results, we first introduce the working assumptions adopted in this section.

Assumption 3 (Lipschitz dynamics (13a)): The map $x \mapsto F(x,z)$ is Lipschitz with constant $\ell_{F,z} > 0$ uniformly over $\mathcal{Z}$.

Assumption 4 (Contractive dynamics (13b)): The map $z \mapsto \mathsf{G}(x,z)$ is uniformly Lipschitz with constant $c_{\mathsf{G}} \in (0,1)$ with respect to $\|\cdot\|_{\mathcal{Z}}$.

Assumption 5 (Lipschitz interconnection): The map $z \mapsto F(x,z)$ is uniformly Lipschitz (over $x \in \mathcal{X}$) with respect to $\|\cdot\|_{\mathcal{X}}$ and $\|\cdot\|_{\mathcal{Z}}$ with constant $\ell_{F,z} > 0$. Similarly, the map $x \mapsto \mathsf{G}(x,z)$ is uniformly Lipschitz (over $z \in \mathcal{Z}$) with constant $\ell_{\mathsf{G},x} > 0$.

We note that, under Assumption 3, the one-sided Lipschitz constant of F, denoted by $\mathrm{osL}_x(F)$ is finite. Due to the Banach Contraction Theorem, Assumption 4 implies that for any $x \in \mathcal{X}$, there exists a unique fixed point $z^*(x)$ for the map $\mathsf{G}(x,z)$; i.e., $z^*(x) = \mathsf{G}(x, z^*(x))$. Moreover, for any fixed x, one has that $\lim_{n\to+\infty} \mathsf{G}^n(x,z) = z^*(x)$ and, additionally, the map $x \mapsto z^*(x)$ is Lipschitz.

Mirroring two-time-scale continuous-time dynamical models [66], we formalize the notion of a *reduced model* (RM) associated with (13); we defined the RM as:

$$\dot{x}_{\mathsf{r}}(t) = \mathsf{R}(x_{\mathsf{r}}(t)) := F(x_{\mathsf{r}}(t), z^\star(x_{\mathsf{r}}(t))), \quad t \geq 0 \tag{14}$$

with $x_{\mathsf{r}}(0) \in \mathcal{X}$. The RM represents the limiting dynamics under two idealized assumptions: *(i)* $n \to +\infty$, so that $z^*(x)$ is computed for any x; and, *(ii)* $T \to 0^+$, so that $z^*(x)$ is applied continuously in time. Before proceeding, we connect our model to optimization-based methods.

Remark 6: In our model, $z^*(x)$ represents the unique optimal solution of a convex optimization problem defining the control law for the system (13a). The sub-system (13b) accounts for practical settings where the state $x(t)$ may be measured at given intervals $\{kT, k \in \mathbb{Z}_{\geq 0}\}$, and the solution $z^*(x)$ is approximated by executing only a finite number n of algorithmic iterations, including the case where $n = 1$.

We now state the main stability results for (13). Without loss of generality, the results are stated for the case where $F(0,0) = 0$, $\mathsf{G}(0,0) = 0$, and $z^*(0) = 0$. The result leverages the notion of weighted norm $\|x\|_{p,[\eta]}$ in $\mathbb{R}^2$, which is defined as $\|x\|_{p,[\eta]} := (\eta_1|x_1|^p + \eta_2|x_2|^p)^{\frac{1}{p}}$, where $\eta_1, \eta_2 > 0$.

Theorem 7 (Stability from RM contractivity): Consider the interconnected system (13), and let Assumptions 3, 4, and 5 hold with norms $\|\cdot\|_{\mathcal{X}}$ and $\|\cdot\|_{\mathcal{Z}}$. Let $\xi \geq \mathrm{osL}_x(F)$, $\xi \neq 0$. Assume that $\zeta > 0$ is such that $\mathrm{osL}_x(\mathsf{R}) \leq -\zeta$. Then, for any $n \in \mathbb{Z}_{>0}$ there exists a $T(n) > 0$ and a weighted vector norm $\|\cdot\|_{2,[\eta]}$ so that

$$\left\| \begin{bmatrix} \|x(t)\|_{\mathcal{X}} \\ \|z(t)\|_{\mathcal{Z}} \end{bmatrix} \right\|_{2,[\eta]} \leq \kappa \mathrm{e}^{-\alpha t} \left\| \begin{bmatrix} \|x(0)\|_{\mathcal{X}} \\ \|z(0)\|_{\mathcal{Z}} \end{bmatrix} \right\|_{2,[\eta]}, \forall\, t \geq 0, \tag{15}$$

with $\alpha > 0$ and $\kappa \geq 0$, for any $0 < T < T(n)$. In particular, such a $T(n)$ is given by:

$$T(n) = \frac{1}{\xi} \log \left(\frac{\xi(1 - c_{\mathsf{G}}^n)}{C_2(n) + C_1/\zeta} + 1 \right) \tag{16}$$

where $C_2(n) := c_{\mathsf{G}}^n(1 - c_{\mathsf{G}})\ell_{\mathsf{G},x}\ell_{F,z}$ and

$$C_1 := \frac{\ell_{F,z}\ell_{\mathsf{G},x}}{1 - c_{\mathsf{G}}} \left(\ell_{F,x} + \frac{\ell_{F,z}\ell_{\mathsf{G},x}}{1 - c_{\mathsf{G}}} \right).$$

Detailed expressions for α and κ are provided in [24]. When the RM is strongly infinitesimally contracting, Theorem 7 asserts that, for any $n \in \mathbb{Z}_{>0}$, global exponential stability (GES) of the interconnected system (13) can be guaranteed so long as $T < T(n)$. We note that $T(n)$ depends solely on Lipschitz constants of the maps F and G, and on the contractivity rate ζ of the RM.

In the following, we provide an extension of classical small-gain and network contraction theorems.

Theorem 8 (Stability from small gain): Consider the interconnected system (13), and let Assumptions 3, 4, and 5 hold with norms $\|\cdot\|_{\mathcal{X}}$ and $\|\cdot\|_{\mathcal{Z}}$. If

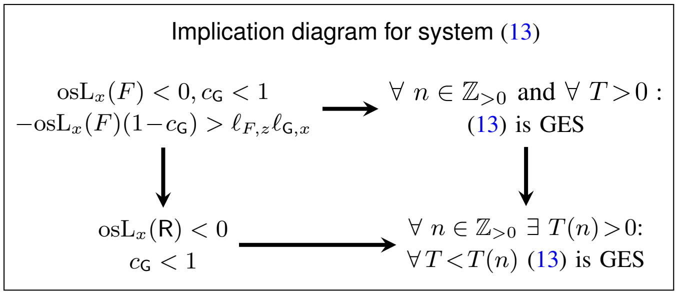
$$-\mathrm{osL}_x(F)(1 - c_{\mathsf{G}}) > \ell_{F,z}\ell_{\mathsf{G},x}, \tag{17}$$

then for any $n \in \mathbb{Z}_+$ and for any $T > 0$, there exists a norm $\|[\|x\|_{\mathcal{X}}, \|z\|_{\mathcal{Z}}]^\top\|_{2,[\eta]}$ on $\mathbb{R}^{n_x+n_z}$, $\eta > 0$, such that the interconnected system (13) renders the origin GES.

Details on the transient bound can be found in [24]. We note that Theorem 8 does not impose any conditions on T and n. However, Theorem 8 requires conditions for stability that are stricter than the ones in Theorem 7, as highlighted in the following.

Theorem 9 (Small gain implies contractivity of the RM): Let Assumption 3, 5 and 4 be satisfied with norms $\|\cdot\|_{\mathcal{X}}$ and $\|\cdot\|_{\mathcal{Z}}$. If (17) holds, then the RM (14) is strongly infinitesimally contracting with respect to the norm $\|\cdot\|_{\mathcal{X}}$.

The results and conditions provided above are summarized in the following implication diagram for the system (13).



When considering optimization-based control methods, Theorem 7 and Theorem 8 suggest two different approaches to designing the controllers. In particular, Theorem 7 offers practical guidance for selecting the sampling time T. The case where $n = 1$ yields a significant contribution in MPC, as it establishes conditions that guarantee stability for single-iteration suboptimal MPC implementations.

IV. A Linear Differential Inclusions Perspective Towards Contraction

In this section, we first recall a connection between contraction theory and a linear differential inclusion (LDI) characterizing the error dynamics between pairs of trajectories. We then summarize results in [56] showing that an alternative, generally tighter, LDI can be constructed when studying the error dynamics to a particular known trajectory, a setting common in, e.g., trajectory tracking and reachability analysis. This new LDI remains amenable to familiar analysis tools from contraction theory.

A. LDIs and Contraction Theory

An LDI is given by [16, p.52]

$$\dot{x} \in \Omega x, \quad x(t_0) = x_0, \tag{18}$$

where $\Omega \subseteq \mathbb{R}^{n\times n}$ is a set of matrices. Any $t \mapsto x(t)$ satisfying (18) is called a *trajectory* of the LDI. A proof of the following proposition based on the mean value theorem is given in [16, Section 4.3.1, p.55].

Proposition 10: Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be differentiable and $X \subseteq \mathbb{R}^n$ be convex. If $\mathcal{J} \subseteq \mathbb{R}^{n\times n}$ satisfies $\frac{\partial F}{\partial x}(X) \subseteq \mathcal{J}$, where $\frac{\partial F}{\partial x}(X) = \left\{\frac{\partial F}{\partial x}(x) : x \in X\right\}$, then

$$F(x) - F(x') \in \overline{\mathrm{co}}(\mathcal{J})(x - x') \tag{19}$$

for every $x, x' \in X$, where $\overline{\mathrm{co}}$ denotes closed convex hull.

Consider now the system $\dot{x} = F(x)$ where $F : \mathbb{R}^n \to \mathbb{R}^n$ is differentiable, and let $X \subseteq \mathbb{R}^n$ be a convex set. If $\frac{\partial F}{\partial x}(X) \subseteq \mathcal{J}$, then Proposition 10 leads to the LDI

$$\dot{\varepsilon}(t) = F(x(t)) - F(x'(t)) \in \overline{\mathrm{co}}(\mathcal{J})\varepsilon(t)$$

for any two trajectories $t \mapsto x(t)$, $t \mapsto x'(t)$ of the system and $\varepsilon(t) = x(t) - x'(t)$. By continuity and convexity of the log norm μ, and by application of Coppel's inequalities (e.g., [18, Theorem 2.3]), we recover the familiar contraction result:

$$\mu(J) \le -c\ \forall J \in \mathcal{J} \implies \|\varepsilon(t)\| \le e^{-ct}\|\varepsilon(0)\|.$$

We highlight two possible advantages of the LDI perspective on contraction analysis. The first is that this perspective makes clear that any computational or analytical approach for studying a general LDI is applicable for contraction analysis. For example, if $\mathcal{J}$ is in one of several common forms such as a polytope, then the semi-definite programming based quadratic stability analysis of LDIs in, e.g., [16, Ch. 4 and 5] offers a computationally efficient way to analyze the contraction properties of the system, where we especially recall the equivalence

$$\mu_{2,P^{1/2}}(M) \le -c \iff M^T P + PM \preceq -2cP$$

where $\mu_{2,P^{1/2}}$ is the log norm associated to the weighted Euclidean norm $\|x\|_{2,P^{1/2}} = \sqrt{x^T P x}$ for positive definite P. Further, it is apparent that we may use any computational method for obtaining $\mathcal{J}$ that satisfies $\frac{\partial F}{\partial x}(X) \subseteq \mathcal{J}$. Of particular note, interval analysis offers an automated way for obtaining $\mathcal{J}$ as an interval set given an interval X and the symbolic form of $\frac{\partial F}{\partial x}$. This observation leads to our second main advantage, developed next.

B. An Alternative LDI for Contraction to a Known Trajectory

A key contribution of [56] is the introduction of a certain *mixed Jacobian* operator inspired by similar ideas in the literature on interval analysis [60]. We consider here a simplified instantiation of this operator applied to interval sets.

Lemma 11: Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be differentiable and $X = X_1 \times \cdots \times X_n \subseteq \mathbb{R}^n$ be an interval with each $X_i \subseteq \mathbb{R}$, and consider some fixed $x' \in X$. If $\mathcal{M} \subseteq \mathbb{R}^{n\times n}$ is an interval set of matrices satisfying

$$\frac{\partial F_i}{\partial x_j}(X_1, \ldots, X_j, x'_{j+1}, \ldots, x'_n) \subseteq \mathcal{M}_{ij}, \tag{20}$$

then $F(x) - F(x') \in \mathcal{M}(x - x')$.

The proof of this lemma applies the mean value theorem n times on a path connecting x to x' composed of n straight lines, each varying only one coordinate at a time. If, in Proposition 10, we restrict to interval sets X and $\mathcal{J}$ and compare to Lemma 11, the key difference is that, when considering the j-th column of the Jacobian $\frac{\partial F}{\partial x}(x)$, the last $j + 1$ through n arguments are fixed to x'_{j+1} through x'_n, rather than the entire intervals X_{j+1} through X_n. Obtaining interval matrices satisfying (20) given differentiable

F is automatic using interval analysis toolboxes such as `immrax` [57], which also includes automatic differentiation capabilities.

Now, consider the time-varying system

$$\dot{x} = F(t, x), \quad x(t_0) = x_0, \tag{21}$$

where $F : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ is continuous with continuous partial derivatives with respect to x.

Proposition 12: Let $\|\cdot\|$ be a norm with induced logarithmic norm μ, and let $X = X_1 \times \cdots \times X_n \subseteq \mathbb{R}^n$ be an interval set. Consider the dynamical system (21). Let $t \mapsto x'(t) \in X$ be a known trajectory defined on $[t_0, \infty)$. For each t, let $\mathcal{M}(t) \subseteq \mathbb{R}^{n\times n}$ be a set of matrices satisfying

$$\frac{\partial F_i}{\partial x_j}(X_1, \dots, X_j, x'_{j+1}(t), \dots, x'_n(t)) \subseteq \mathcal{M}(t)_{ij}.$$

If for some $c \in \mathbb{R}$,

$$\mu(M) \leq -c \text{ for all } M \in \mathcal{M}(t) \text{ and all } t \geq t_0,$$

then, for any trajectory $t \mapsto x(t) \in X$ defined on $[t_0, \infty)$,

$$\|x(t) - x'(t)\| \leq e^{-c(t-t_0)}\|x(t_0) - x'(t_0)\|,$$

for every $t \geq t_0$.

C. Example

A common application of contraction analysis is to compute simulation-guided overapproximating reachable sets by computing a nominal trajectory $x'(t)$, bounding the log norm in a region around $x'(t)$, and expanding/contracting norm balls using the log norm rate [40], [74]. Specifically, [40] bounds the log norm by constructing an interval Jacobian matrix that overapproximates the set of linearizations of the dynamics. Using Lemma 11, since $x'(t)$ is fixed, we replace the interval Jacobian with an interval mixed Jacobian for immediate benefit via Proposition 12.

For example, consider the robot arm model [8]

$$\begin{aligned}
&\dot{q}_1 = z_1, \quad \dot{q}_2 = z_2,\\
&\dot{z}_1 = \tfrac{1}{mq_2^2 + ML^2/3}(-2mq_2 z_1 z_2 - k_{d_1} z_1 + k_{p_1}(u_1 - q_1)),\\
&\dot{z}_2 = q_2 z_1^2 + \tfrac{1}{m}(-k_{d_2} z_2 + k_{p_2}(u_2 - q_2)),
\end{aligned}$$

with $u_1 = 2$, $u_2 = 1$, $m = M = 1$, $L = \sqrt{3}$, $k_{p_1} = 2$, $k_{p_2} = 1$, $k_{d_1} = 2$, $k_{d_2} = 1$. The method from [40] results in Figure 4, left. We repeat the analysis using the tighter LDI from Proposition 12, obtaining significantly tighter approximations of the reachable set, as shown in Figure 4, right.

V. Contraction Approaches to Robust and Control-Oriented Machine Learning

Learning-based control is one domain in which contraction and incremental stability properties have a particularly strong motivation. In classical stability analysis and control methods, the objective is often to characterize or achieve the stability of a particular *known* equilibrium or trajectory. In learning some training data may be known but the objective is to find a model or policy that certifiably generalizes well

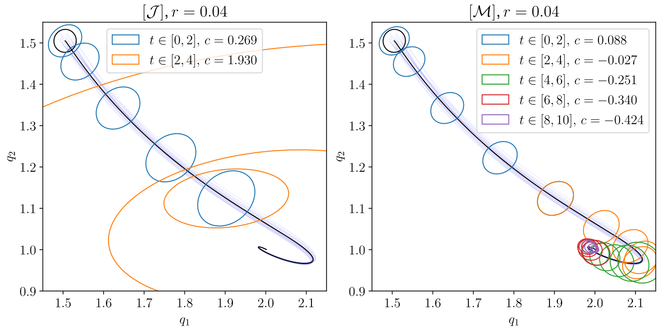


Fig. 4: Ellipsoidal reachable set overapproximations for a robot arm model obtained by uniformly bounding the log norm of the Jacobian linearization of the dynamics in a neighborhood of a nominal trajectory (left) and using the tighter mixed Jacobian LDI from Proposition 12 (right).

to *previously unseen* conditions. In real problems, these may be quite different to the training data.

The remarkable recent achievements applying neural networks to prediction and control—along with their well-known lack of robustness [101]—call for a theory of stability and robustness that is amenable this nonlinear model family. In this section we argue that contraction and incremental robustness conditions can meet this need, in that they are:

- compatible with classical system analysis tools such as small gain, passivity, IQC, and Lyapunov methods, and
- compatible with modern approaches to learning large-scale models and policies such as auto-differentiation and unconstrained first-order optimization.

A. Direct Parameterization and Unconstrained Optimization

We suppose that some data-set available, denoted $\mathcal{D}$, some nonlinear model class parameterized by a vector $\theta \in \mathbb{R}^p$, where models may be static functions or dynamic systems, and some loss function $l(\theta, \mathcal{D})$ which is to be optimized. This formulation can include classical supervised learning and system identification problems, as well as reinforcement learning problems in which the model is a control policy, $\mathcal{D}$ consists of environmental conditions, and the loss incorporates simulation of the closed-loop system.

In all such cases, $\mathcal{D}$ consists of *representative* scenarios, and is not assumed to be exhaustive. We consider the case in which *certified* stability and/or robustness properties of the model related are needed. In many cases it is possible to use robust and nonlinear control theory to construct a subset of parameter space $\Theta \subset \mathbb{R}^p$ on which the desired property is certified, leading to the constrained optimization problem:

$$\min_{\theta\in\Theta} l(\theta, \mathcal{D}). \tag{22}$$

Unfortunately, the constraint $\theta \in \Theta$ often takes the form of a set of parameterized linear matrix inequalities (LMIs). Standard constrained optimization methods require either repeated projection onto Θ or computation of barrier terms, and while LMIs are tractable at moderate scale they are prohibitive for the kinds of large models used in applications.

An alternative approach is to construct a new parameterization consisting of a differentiable mapping $\theta = m(\phi)$, where $\phi \in \mathbb{R}^N$ with $N > p$ in general, such that $m(\phi) \in \Theta$ for all $\phi \in \mathbb{R}^N$. If m is surjective, i.e. the image of $m(\mathbb{R}^N) = \Theta$ then we refer to this as a *direct parameterization*. Then (22) can be replaced with the unconstrained optimization:

$$\min_{\phi \in \mathbb{R}^N} l(m(\phi), \mathcal{D}), \tag{23}$$

to which scalable methods such as stochastic gradient descent and its variants can be directly applied.

a) Lipschitz Neural Networks: The first task we consider is to learn static (feedforward) neural networks

$$y = f(x) := W_L \sigma(W_{L-1}\sigma(\cdots \sigma(W_0 x + b_0)) + b_{L-1}) + b_L$$

with a bound on their Lipschitz constant:

$$\|f(x_1) - f(x_2)\| \leq \gamma \|x_1 - x_2\|, \; \forall x_1, x_2 \in \mathbb{R}^n$$

which in the systems theory context can be considered an incremental ℓ^2 gain condition.

In [43], incremental IQC methods were used to show that f is γ-Lipschitz if $\exists \Lambda_0, \dots \Lambda_{L-1}$ (positive diagonal) s.t.

$$H := \begin{bmatrix} \gamma I & -\hat{W}_0^\top & & & \\ -\hat{W}_0 & 2\Lambda_0 & -\hat{W}_1^\top & & \\ & \ddots & \ddots & \ddots & \\ & & -\hat{W}_{L-1} & 2\Lambda_{L-1} & -\hat{W}_L^\top \\ & & & -\hat{W}_L & \gamma I \end{bmatrix} \succeq 0 \tag{24}$$

where $\hat{W}_k = \Lambda_k W_k$, $\hat{W}_L = W_L$. While offering the tightest-known bounds, verification of this condition is not scalable to models with more than a few thousand neurons, and difficult to apply directly to convolutional models.

In [112] a direct parameterization of condition (24) was constructed. The basic idea is a square-root parameterization: $H = PP^\top$, so H is positive-definite by construction. However, H in (24) has a particular structure: it is block-tridiagonal and the diagonal blocks are themselves diagonal matrices. The breakthrough in [112] was a method to parameterize P so that H has the desired structure, arrived at via a version of the Cayley transform, and such that and the neural network weights W_i can be recovered for inference.

Lipschitz bounded neural networks have numerous applications [9], [54], [112], but in particular we can highlight their utility in parameterizing control policies for reinforcement learning can enforce closed-loop contraction via the incremental small-gain theorem, and robustness to adversarial attacks and measurement errors [92], see e.g. [13] in which Lipschitz convolutional policies were trained for Atari Pong and achieved significantly improved robustness.

B. Contracting and Lipschitz Dynamic Models

In [91] this approach was extended to *dynamic* (recurrent) models. A recurrent equilibrium network (REN) is a feedback interconnection of an LTI G and activation functions σ:

$$\left.\begin{aligned} x_+ &= Ax + B_1 w + B_2 u + b_x \\ v &= C_1 x + D_{11} w + D_{12} u + b_v \\ y &= C_2 x + D_{21} w + D_{22} u + b_y \end{aligned}\right\} = G, \quad w = \sigma(v) \tag{25}$$

Using similar methods, RENs can be parameterized to be contracting and Lipschitz or more generally satisfy incremental dissipativity properties. Indeed, contraction can be thought of as a form of Lipschitzness of the mapping from initial conditions to solutions.

Contracting RENs have been used for physics-informed learning of nonlinear observers, parameterization of stabilizing feedback controllers [46], [91], [110], and data-driven learning of optimization algorithms [76].

C. Bi-Lipschitz Networks and Neural Lyapunov Functions

One interesting extension is the parameterization of bi-Lipschitz models [111]. The starting point is to construct and directly-parameterize neural networks that are both Lipschitz and strongly input-output monotone:

$$\langle f(x_1) - f(x_2), x_1 - x_2 \rangle \geq \alpha \|x_1 - x_2\|^2$$

which implies that f is *invertible* and bi-Lipschitz, i.e.

$$\mu \|x_1 - x_2\| \leq \|f(x_1) - f(x_2)\| \leq \nu \|x_1 - x_2\|$$

for some $\nu \geq \nu > 0$. From this building block, more flexible bi-Lipschitz bijections can be constructed by composition with orthogonal affine layers $\mathcal{O}_i$, which can also be directly-parameterized via the Cayley transform [104]:

$$y = \mathcal{G}(x) := \mathcal{O}_{K+1} \circ f_K \circ \mathcal{O}_K \circ \cdots \circ f_1 \circ \mathcal{O}_1(x)$$

The concept of a bi-Lipschitz invertible model was recently extended to contracting and Lipschitz dynamic models with contracting and Lipschitz inverses [118], providing an incremental notion of robustly-invertible systems generalizing the concept of a minimum-phase linear system.

Bi-Lipschitz and invertible neural networks have many potential applications [111], including learning Koopman embeddings for contracting systems [41], [115]. One control-oriented example is the construction of neural Lyapunov functions [25], i.e. $V(x) = \|\mathcal{G}(x)\|^2$. By construction, such functions are positive-definite and satisfy $V(x^\star) = 0$ at a unique point $x^\star = \mathcal{G}^{-1}(0)$.

This construction is further motivated by the fact that all Lyapunov functions of an asymptotically stable system have level sets that are homeomorphisms of the unit ball [113]. Furthermore, the bi-Lispchitz condition leads to

$$\mu^2 \|x - x^\star\|^2 \leq V(x) \leq \nu^2 \|x - x^\star\|^2, \quad \forall x \in \mathbb{R}^n$$

which is a widely-used condition for verifying exponential stability. Having parameterized a rich family of Lyapunov functions, one can parameterize stable neural dynamics via *smooth descent* directions of V [25]:

$$\dot{x} = [J(x) - R(x)]\nabla V(x),$$

where $J(x) = -J(x)^\top, R(x) = R(x)^\top \succeq \epsilon I$ which is globally exponentially stable at $x^\star$ with converge rate of ϵ.

This formulation can be extended to passive and negative-imaginary systems [25], [96] for learning stable physical interaction policies.

Concluding remarks on other recent and ongoing work

We note that this document provides only a partial overview of the many recent developments in contraction theory. To acknowledge the broader scope of ongoing work, we include below a (necessarily incomplete) list of recent other contributions in contraction theory and its applications: [4], [15], [83] on contractivity of stochastic differential equations, [103] on contractivity of functional differential equations, [44], [51], [52], [69] on numerical computation and verification of contraction metrics, [73] on adaptive control via contraction metrics, [88] on regular pairings for non-quadratic analysis, [2] on structural contraction of biological interaction networks, [67] on Riemannian contraction in supervised learning, [7], [48]–[50], [90] on LMI conditions for stability, contraction, and synchronization of systems in Luré form, [63], [64] on contractivity of virtually positive systems and discrete-time stochastic systems, [12], [85] on k-contractivity, [108] on self-triggered stabilization, [107] on the incremental input-to-state properties of hybrid integrator-gain systems, and [1] on contractive dynamical imitation learning. Finally, a special mention goes to the recent authoritative text [100] on logarithmic norms for matrices, nonlinear maps and linear differential operators.

VI. Concluding remarks on future research

Several promising directions emerge for the development of contraction-theoretic methods. A first priority is to sharpen estimates of contraction rates for important example systems and, in the case of locally contractive systems, to characterize domains of contraction. Such results are particularly relevant to optimization algorithms and to equilibrium-tracking problems in discrete-time, multi-agent, and stochastic settings. Extending these ideas to sampled-data systems, including continuous–discrete interconnections and systems with delays, is a natural and timely research challenge with connections to recent work on model predictive control [24].

A second theme concerns the development of computational and data-driven tools. Robust codebases and standardized benchmarks—along the lines of community efforts such as the OpenAI Gym—would accelerate progress by enabling systematic evaluation of algorithms and by promoting best practices from the machine learning community. This direction includes exploring the trade-off between physically motivated metrics and computationally favorable ones that yield sharper contraction rates, as well as building scalable LMI-based methods (e.g., adopting strategies in [90]) for certifying local contractivity.

Third, there is strong interest in extending contraction theory to broader mathematical settings, while retaining computational tractability. This includes contraction in probability space, with potential implications for imitation learning, as well as computationally-tractable generalizations to classes of manifolds. Such extensions could further connect contraction theory with questions of static and dynamic regret, expanding the treatments in [80], and open new possibilities for the analysis and control of learning and adaptive systems.

References

[1] A. Abyaneh, M. G. Boroujeni, H.-C. Lin, and G. Ferrari-Trecate. Contractive dynamical imitation policies for efficient out-of-sample recovery. *arXiv preprint arXiv:2412.07544*, 2024. doi:10.48550/arXiv.2412.07544.

[2] M. A. Al-Radhawi, D. Angeli, and E. D. Sontag. On structural contraction of biological interaction networks. *arXiv preprint arXiv:2307.13678*, 2023. doi:10.48550/arXiv.2307.13678.

[3] Z. Aminzare. *On Synchronous Behavior in Complex Nonlinear Dynamical Systems*. PhD thesis, Rutgers, 2015.

[4] Z. Aminzare. Stochastic logarithmic Lipschitz constants: A tool to analyze contractivity of stochastic differential equations. *IEEE Control Systems Letters*, 6:2311–2316, 2022. doi:10.1109/LCSYS.2022.3148945.

[5] Z. Aminzare and E. D. Sontag. Contraction methods for nonlinear systems: A brief introduction and some open problems. In *IEEE Conf. on Decision and Control*, pages 3835–3847, December 2014. doi:10.1109/CDC.2014.7039986.

[6] 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(2):91–106, 2014. doi:10.1109/TNSE.2015.2395075.

[7] V. Andrieu and S. Tarbouriech. LMI conditions for contraction and synchronization. In *IFAC Symposium on Nonlinear Control Systems*, volume 52, pages 616–621, 2019. doi:10.1016/j.ifacol.2019.12.030.

[8] D. Angeli, E. D. Sontag, and Y. Wang. A characterization of integral input-to-state stability. *tac*, 45(6):1082–1097, 2000. doi:10.1109/9.863594.

[9] A. Araujo, A. Havens, B. Delattre, A. Allauzen, and B. Hu. A unified algebraic perspective on Lipschitz neural networks. *arXiv preprint arXiv:2303.03169*, 2023. doi:10.48550/arXiv.2303.03169.

[10] K. J. Arrow, L. Hurwicz, and H. Uzawa, editors. *Studies in Linear and Nonlinear Programming*. Stanford University Press, 1958.

[11] S. Banach. Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales. *Fundamenta Mathematicae*, 3(1):133–181, 1922. doi:10.4064/fm-3-1-133-181.

[12] E. Bar-Shalom, O. Dalin, and M. Margaliot. Compound matrices in systems and control theory: a tutorial. *Mathematics of Control, Signals, and Systems*, 35(3):467–521, 2023. doi:10.1007/s00498-023-00351-8.

[13] N. H. Barbara, R. Wang, and I. Manchester. On robust reinforcement learning with Lipschitz-bounded policy networks. 2024. doi:10.48550/arXiv.2405.11432.

[14] G. Bianchin, J. Cortés, J. I. Poveda, and E. Dall'Anese. Time-varying optimization of LTI systems via projected primal-dual gradient flows. *IEEE Transactions on Control of Network Systems*, 9(1):474–486, 2022. doi:10.1109/TCNS.2021.3112762.

[15] N. M. Boffi and J.-J. E. Slotine. A continuous-time analysis of distributed stochastic gradient. *Neural Computation*, 32(1):36–96, 2020. doi:10.1162/neco_a_01248.

[16] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*. SIAM, 1994, ISBN 089871334X.

[17] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–124, 2010. doi:10.1561/2200000016.

[18] F. Bullo. *Contraction Theory for Dynamical Systems*. Kindle Direct Publishing, 1.2 edition, 2024, ISBN 979-8836646806. URL: https://fbullo.github.io/ctds.

[19] E. J. Candès and T. Tao. The Dantzig selector: Statistical estimation when p is much larger than n. *Quality Control and Applied Statistics*, 54(1):83–84, 2009. doi:10.1214/009053606000001523.

[20] V. Centorrino, F. Bullo, and G. Russo. Modelling and contractivity of neural-synaptic networks with Hebbian learning. *Automatica*, 164:111636, 2024. doi:10.1016/j.automatica.2024.111636.

[21] V. Centorrino, A. Davydov, A. Gokhale, G. Russo, and F. Bullo. On weakly contracting dynamics for convex optimization. *IEEE Control Systems Letters*, 8:1745–1750, 2024. doi:10.1109/LCSYS.2024.3414348.

[22] V. Centorrino, A. Gokhale, A. Davydov, G. Russo, and F. Bullo. Euclidean contractivity of neural networks with symmetric weights. *IEEE Control Systems Letters*, 7:1724–1729, 2023. doi:10.1109/LCSYS.2023.3278250.

[23] V. Centorrino, A. Gokhale, A. Davydov, G. Russo, and F. Bullo. Positive competitive networks for sparse reconstruction. *Neural Computation, Theoretical Neuroscience*, 36(6):1163–1197, 2024. doi:10.1162/neco_a_01657.

[24] Y. Chen, F. Bullo, and E. Dall'Anese. Sampled-data systems: Stability, contractivity and single-iteration suboptimal MPC. *arXiv preprint arXiv:2505.18336*, 2025. doi:10.48550/arXiv.2505.18336.

[25] J. Cheng, R. Wang, and I. R. Manchester. Learning stable and passive neural differential equations. In *IEEE Conference on Decision and Control (CDC)*, 2024. doi:10.1109/CDC56724.2024.10886407.

[26] A. Cherukuri, B. Gharesifard, and J. Cortes. Saddle-point dynamics: Conditions for asymptotic stability of saddle points. *SIAM Journal on Control and Optimization*, 55(1):486–511, 2017. doi:10.1137/15M1026924.

[27] P. Cisneros-Velarde, S. Jafarpour, and F. Bullo. A contraction analysis of primal-dual dynamics in distributed and time-varying implementations. *IEEE Transactions on Automatic Control*, 67(7):3560–3566, 2022. doi:10.1109/TAC.2021.3103865.

[28] J. Cortés and S. K. Niederländer. Distributed coordination for nonsmooth convex optimization via saddle-point dynamics. *Journal of Nonlinear Science*, 29(4):1247–1272, 2019. doi:10.1007/s00332-018-9516-4.

[29] 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(3):71–83, 2020. doi:10.1109/MSP.2020.2968813.

[30] A. Davydov and F. Bullo. Perspectives on contractivity in control, optimization and learning. *IEEE Control Systems Letters*, 8:2087–2098, 2024. doi:10.1109/LCSYS.2024.3436127.

[31] A. Davydov, V. Centorrino, A. Gokhale, G. Russo, and F. Bullo. Time-varying convex optimization: A contraction and equilibrium tracking approach. *IEEE Transactions on Automatic Control*, 70(11), 2025. To appear. doi:10.1109/TAC.2025.3576043.

[32] A. Davydov, S. Jafarpour, and F. Bullo. Non-Euclidean contraction theory for robust nonlinear stability. *IEEE Transactions on Automatic Control*, 67(12):6667–6681, 2022. doi:10.1109/TAC.2022.3183966.

[33] P. Dayan and L. F. Abbott. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. MIT Press, 2005.

[34] D. Del Vecchio and J.-J. E. Slotine. A contraction theory approach to singularly perturbed systems. *tac*, 58(3):752–757, 2012. doi:10.1109/TAC.2012.2211444.

[35] B. P. Demidovič. Dissipativity of a system of nonlinear differential equations in the large. *Uspekhi Matematicheskikh Nauk*, 16(3(99)):216, 1961.

[36] C. A. Desoer and H. Haneda. The measure of a matrix as a tool to analyze computer algorithms for circuit analysis. *IEEE Transactions on Circuit Theory*, 19(5):480–486, 1972. doi:10.1109/TCT.1972.1083507.

[37] 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(7):2861–2868, 2019. doi:10.1109/TAC.2018.2867589.

[38] M. Di Bernardo, D. Fiore, G. Russo, and F. Scafuti. Convergence, consensus and synchronization of complex networks via contraction theory. In *Complex Systems and Networks*, pages 313–339. Springer, 2016. doi:10.1007/978-3-662-47824-0_12.

[39] M. Diehl, H. G. Bock, and J. P. Schlöder. A real-time iteration scheme for nonlinear optimization in optimal feedback control. *SIAM Journal on Control and Optimization*, 43(5):1714–1736, 2005. doi:10.1137/S0363012902400713.

[40] C. Fan, J. Kapinski, X. Jin, and S. Mitra. Simulation-driven reachability using matrix measures. *ACM Transactions on Embedded Computing Systems*, 17(1):1–28, 2017. doi:10.1145/3126685.

[41] F. Fan, B. Yi, D. Rye, G. Shi, and I. R. Manchester. Learning stable Koopman embeddings for identification and control. *arXiv preprint arXiv:2401.08153*, 2024. doi:10.48550/arXiv.2401.08153.

[42] 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(7):1973–1986, 2018. doi:10.1109/TAC.2017.2760256.

[43] M. Fazlyab, A. Robey, H. Hassani, M. Morari, and G. J. Pappas. Efficient and accurate estimation of Lipschitz constants for deep neural networks. *arXiv preprint arXiv:1906.04893*, 2019. doi:10.48550/arXiv.1906.04893.

[44] M. Fitzsimmons and J. Liu. Computation and formal verification of neural network contraction metrics. *IEEE Control Systems Letters*, 8:2367–2372, 2024. doi:10.1109/lcsys.2024.3478272.

[45] F. Forni and R. Sepulchre. A differential Lyapunov framework for contraction analysis. *IEEE Transactions on Automatic Control*, 59(3):614–628, 2014. doi:10.1109/TAC.2013.2285771.

[46] C. L. Galimberti, L. Furieri, and G. Ferrari-Trecate. Parametrizations of all stable closed-loop responses: From theory to neural network control design. *arXiv preprint arXiv:2412.19280*, 2024. doi:10.48550/arXiv.2412.19280.

[47] X.-B. Gao. Exponential stability of globally projected dynamic systems. *IEEE Transactions on Neural Networks*, 14(2):426–431, 2003. doi:10.1109/tnn.2003.809409.

[48] M. Giaccagli, V. Andrieu, S. Tarbouriech, and D. Astolfi. Infinite gain margin, contraction and optimality: An LMI-based design. *European Journal of Control*, 68:100685, 2022. doi:10.1016/j.ejcon.2022.100685.

[49] 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*, 154:111106, 2023. doi:10.1016/j.automatica.2023.111106.

[50] M. Giaccagli, S. Zoboli, D. Astolfi, V. Andrieu, and G. Casadei. Synchronization in networks of nonlinear systems: Contraction analysis via Riemannian metrics and deep-learning for feedback estimation. *IEEE Transactions on Automatic Control*, 69(11):8041–8048, 2024. doi:10.1109/tac.2024.3407015.

[51] P. Giesl, S. Hafstein, and C. Kawan. Review on contraction analysis and computation of contraction metrics. *Journal of Computational Dynamics*, 10(1):1–47, 2023. doi:10.3934/jcd.2022018.

[52] P. Giesl, S. Hafstein, and I. Mehrabinezhad. Computation and verification of contraction metrics for exponentially stable equilibria. *Journal of Computational and Applied Mathematics*, 390:113332, 2021. doi:10.1016/j.cam.2020.113332.

[53] B. C. Goodwin. Oscillatory behavior in enzymatic control processes. *Advances in Enzyme Regulation*, 3:425–437, 1965. doi:10.1016/0065-2571(65)90067-1.

[54] H. Gouk, E. Frank, B. Pfahringer, and M. J. Cree. Regularisation of neural networks by enforcing Lipschitz continuity. *Machine Learning*, 110:393–416, 2021. doi:10.1007/s10994-020-05929-w.

[55] A. Granas and J. Dugundji. *Fixed Point Theory*. Springer, 2003, ISBN 978-1-4419-1805-5. doi:10.1007/978-0-387-21593-8.

[56] A. Harapanahalli and S. Coogan. A linear differential inclusion for contraction analysis to known trajectories. *arXiv preprint arXiv:2411.11587*, 2024. doi:10.48550/arXiv.2411.11587.

[57] A. Harapanahalli, S. Jafarpour, and S. Coogan. immrax: A parallelizable and differentiable toolbox for interval analysis and mixed monotone reachability in JAX. In *IFAC Conference on Analysis and Design of Hybrid Systems (ADHS)*, 2024. doi:10.1016/j.ifacol.2024.07.428.

[58] S. Hassan-Moghaddam and M. R. Jovanović. Proximal gradient flow and Douglas-Rachford splitting dynamics: Global exponential stability via integral quadratic constraints. *Automatica*, 123:109311, 2021. doi:10.1016/j.automatica.2020.109311.

[59] W. Heemels and N. Van de Wouw. Stability and stabilization of networked control systems. In *Networked Control Systems*, pages 203–253. Springer, 2010. doi:10.1007/978-0-85729-033-5_7.

[60] L. Jaulin, M. Kieffer, O. Didrit, and É. Walter. *Applied Interval Analysis*. Springer London, 2001.

[61] Z.-P. Jiang, Y. Lin, and Y. Wang. Nonlinear small-gain theorems for discrete-time feedback systems and applications. *Automatica*, 40(12):2129–2136, 2004. doi:10.1016/j.automatica.2004.08.002.

[62] A. Karapetyan, E. C. Balta, A. Iannelli, and J. Lygeros. Closed-loop finite-time analysis of suboptimal online control. *tac*, 2025. doi:10.1109/TAC.2025.3539988.

[63] Y. Kawano and M. Cao. Contraction analysis of virtually positive systems. *Systems & Control Letters*, 168:105358, 2022. doi:10.1016/j.sysconle.2022.105358.

[64] Y. Kawano and Y. Hosoe. Contraction analysis of discrete-time stochastic systems. *IEEE Transactions on Automatic Control*, 69(2):982–997, 2024. doi:10.1109/tac.2023.3283678.

[65] M. A. Khamsi and W. A. Kirk. *An Introduction to Metric Spaces and Fixed Point Theory*. John Wiley & Sons, 2001, ISBN 9781118031322.

[66] P. Kokotovic and P. Sannuti. Singular perturbation method for reducing the model order in optimal control design. *tac*, 13(4):377–384, 1968. doi:10.1109/TAC.1968.1098927.

[67] L. Kozachkov, P. M. Wensing, and J.-J. Slotine. Generalization as dynamical robustness–The role of Riemannian contraction in supervised learning. *Transactions on Machine Learning Research*, 2023. URL: https://openreview.net/forum?id=Sb6p5mcefw.

[68] N. N. Krasovskiĭ. *Stability of Motion. Applications of Lyapunov's Second Method to Differential Systems and Equations with Delay*. Stanford University Press, 1963.

[69] H. Li, X. Zhong, B. Hu, and H. Zhang. Neural contraction metrics with formal guarantees for discrete-time nonlinear dynamical systems. In N. Ozay, L. Balzano, D. Panagou, and A. Abate, editors, *Proceedings of the 7th Annual Learning for Dynamics & Control Conference*, volume 283 of *Proceedings of Machine Learning Research*, pages 1447–1459. PMLR, 2025. URL: https://proceedings.mlr.press/v283/li25e.html.

[70] D. Liao-McPherson, M. M. Nicotra, and I. Kolmanovsky. Time-distributed optimization for real-time model predictive control: Stability, robustness, and constraint satisfaction. *Automatica*, 117:108973, 2020. doi:10.1016/j.automatica.2020.108973.

[71] W. Lohmiller and J.-J. E. Slotine. On contraction analysis for non-linear systems. *Automatica*, 34(6):683–696, 1998. doi:10.1016/S0005-1098(98)00019-3.

[72] W. Lohmiller and J.-J. E. Slotine. Control system design for mechanical systems using contraction theory. *IEEE Transactions on Automatic Control*, 45(5):984–989, 2000. doi:10.1109/9.855568.

[73] B. T. Lopez and J.-J. E. Slotine. Adaptive nonlinear control with contraction metrics. *IEEE Control Systems Letters*, 5(1):205–210, 2021. doi:10.1109/LCSYS.2020.3000190.

[74] J. Maidens and M. Arcak. Reachability analysis of nonlinear systems using matrix measures. *IEEE Transactions on Automatic Control*, 60(1):265–270, 2015. doi:10.1109/TAC.2014.2325635.

[75] 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(6):3046–3053, 2017. doi:10.1109/TAC.2017.2668380.

[76] A. Martin and L. Furieri. Learning to optimize with convergence guarantees using nonlinear system theory. *IEEE Control Systems Letters*, 2024. doi:10.1109/LCSYS.2024.3406967.

[77] Z. Marvi, F. Bullo, and A. G. Alleyne. Control barrier proximal dynamics: A contraction theoretic approach for safety verification. *IEEE Control Systems Letters*, 8:880–885, 2024. doi:10.1109/LCSYS.2024.3402188.

[78] Z. Marvi, F. Bullo, and A. G. Alleyne. Robust and exponential stability in barrier-certified systems via contracting piecewise smooth dynamics. *IEEE Control Systems Letters*, 8:3279–3284, 2024. doi:10.1109/LCSYS.2024.3524369.

[79] Z. Marvi, F. Bullo, and A. G. Alleyne. Air cooled battery pack thermal management via control barrier proximal dynamics. *IEEE Transactions on Control Systems Technology*, June 2025. Submitted.

[80] A. Mokhtari, S. Shahrampour, A. Jadbabaie, and A. Ribeiro. Online optimization in dynamic environments: Improved regret rates for strongly convex problems. In *IEEE Conf. on Decision and Control*, pages 7195–7201, 2016. doi:10.1109/CDC.2016.7799379.

[81] M. Morari, C. E. Garcia, and D. M. Prett. Model predictive control: Theory and practice. *IFAC Proceedings Volumes*, 21(4):1–12, 1988. doi:10.1016/B978-0-08-035735-5.50006-1.

[82] P. Naghshtabrizi, J. P. Hespanha, and A. R. Teel. On the robust stability and stabilization of sampled-data systems: A hybrid system approach. In *IEEE Conference on Decision and Control (CDC)*, pages 4873–4878, 2006. doi:10.1109/CDC.2006.377315.

[83] P. H. A. Ngoc. Contraction of stochastic differential equations. *Communications in Nonlinear Science and Numerical Simulation*, 95:105613, 2021. doi:10.1016/j.cnsns.2020.105613.

[84] 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(4):755–760, 2018. doi:10.1109/LCSYS.2018.2847408.

[85] R. Ofir, A. Ovseevich, and M. Margaliot. Contraction and k-contraction in Lurie systems with applications to networked systems. *Automatica*, 159:111341, 2024. doi:10.1016/j.automatica.2023.111341.

[86] N. Parikh and S. Boyd. Proximal algorithms. *Foundations and Trends in Optimization*, 1(3):127–239, 2014. doi:10.1561/2400000003.

[87] A. Pavlov, A. Pogromsky, N. Van de Wouw, and H. Nijmeijer. Convergent dynamics, a tribute to Boris Pavlovich Demidovich. *Systems & Control Letters*, 52(3-4):257–261, 2004. doi:10.1016/j.sysconle.2004.02.003.

[88] A. V. Proskurnikov and F. Bullo. Regular pairings for non-quadratic Lyapunov functions and contraction analysis. *SIAM Journal on Control and Optimization*, September 2024. Submitted. doi:10.48550/arXiv.2408.17350.

[89] G. Qu and N. Li. On the exponential stability of primal-dual gradient dynamics. *IEEE Control Systems Letters*, 3(1):43–48, 2019. doi:10.1109/LCSYS.2018.2851375.

[90] I. Queinnec, S. Tarbouriech, G. Valmorbida, and L. Zaccarian. Design of saturating state feedback with sign-indefinite quadratic forms. *IEEE Transactions on Automatic Control*, 67(7):3507–3520, 2022. doi:10.1109/tac.2021.3106878.

[91] M. Revay, R. Wang, and I. R. Manchester. Recurrent equilibrium networks: Flexible dynamic models with guaranteed stability and robustness. *IEEE Transactions on Automatic Control*, 69(5):2855–2870, 2023. doi:10.1109/TAC.2023.3294101.

[92] A. Russo and A. Proutiere. Towards optimal attacks on reinforcement learning policies. In *American Control Conference*, pages 4561–4567, 2021. doi:10.23919/acc50511.2021.9483025.

[93] G. Russo. *Analysis, Control and Synchronization of Nonlinear Systems and Networks via Contraction Theory: Theory and Applications*. PhD thesis, Universita degli Studi di Napoli Federico II, 2010.

[94] G. Russo, M. Di Bernardo, and E. D. Sontag. Global entrainment of transcriptional systems to periodic inputs. *PLoS Computational Biology*, 6(4):e1000739, 2010. doi:10.1371/journal.pcbi.1000739.

[95] 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*, 2:10–19, 2022. doi:10.1038/s43588-021-00184-y.

[96] K. Shi, R. Wang, and I. R. Manchester. Negative imaginary neural ODEs: Learning to control mechanical systems with stability guarantees. In *Proceedings of the IFAC Symposium on Nonlinear Control Systems (NOLCOS)*, 2025. to appear.

[97] A. Simonetto and E. Dall'Anese. Prediction-correction algorithms for time-varying constrained optimization. *IEEE Transactions on Signal Processing*, 65(20):5481–5494, 2017. doi:10.1109/TSP.2017.2728498.

[98] 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(11):2032–2048, 2020. doi:10.1109/JPROC.2020.3003156.

[99] 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(17):4576–4591, 2016. doi:10.1109/TSP.2016.2568161.

[100] G. Söderlind. *Logarithmic Norms*. Springer, 2024, ISBN 9783031743795. doi:10.1007/978-3-031-74379-5.

[101] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus. Intriguing properties of neural networks. 2014. doi:10.48550/arXiv.1312.6199.

[102] 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*, 115:100–123, 2019. doi:10.1016/j.neunet.2019.03.005.

[103] K. Q. Tran and P. H. A. Ngoc. On contraction of functional differential equations with markovian switching. *Systems & Control Letters*, 183:105675, January 2024. doi:10.1016/j.sysconle.2023.105675.

[104] A. Trockman and J. Z. Kolter. Orthogonalizing convolutional layers with the Cayley transform. *arXiv preprint arXiv:2104.07167*, 2021. doi:10.48550/arXiv.2104.07167.

[105] H. Tsukamoto, S.-J. Chung, and J.-J. E. Slotine. Contraction theory for nonlinear stability analysis and learning-based control: A tutorial overview. *Annual Reviews in Control*, 52:135–169, 2021. doi:10.1016/j.arcontrol.2021.10.001.

[106] A. M. Turing. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 237(641):37–72, 1952. doi:10.1098/rstb.1952.0012.

[107] S. J. A. M. van den Eijnden, M. F. Heertjes, H. Nijmeijer, and W. P. M. H. Heemels. A small-gain approach to incremental input-to-state stability analysis of hybrid integrator-gain systems. *IEEE Control Systems Letters*, 7:2443–2448, 2023. doi:10.1109/lcsys.2023.3284601.

[108] M. Wakaiki. Self-triggered stabilization of contracting systems under quantization. *IEEE Transactions on Automatic Control*, pages 1–16, 2025. doi:10.1109/tac.2025.3540727.

[109] 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*, pages 3800–3805, Orlando, USA, 2011. doi:10.1109/CDC.2011.6161503.

[110] R. Wang, N. H. Barbara, M. Revay, and I. R. Manchester. Learning over all stabilizing nonlinear controllers for a partially-observed linear system. *IEEE Control Systems Letters*, 7:91–96, 2022. doi:10.1109/LCSYS.2022.3184847.

[111] R. Wang, K. D. Dvijotham, and I. Manchester. Monotone, Bi-Lipschitz, and Polyak-Łojasiewicz networks. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 50379–50399. PMLR, 2024. URL: https://proceedings.mlr.press/v235/wang24p.html.

[112] R. Wang and I. Manchester. Direct parameterization of lipschitz-bounded deep networks. In *International Conference on Machine Learning (ICML)*, pages 36093–36110. PMLR, 2023. URL: https://proceedings.mlr.press/v202/wang23am.html.

[113] F. W. Wilson. The structure of the level surfaces of a Lyapunov function. *Journal of Differential Equations*, 3(3):323–329, 1967. doi:10.1016/0022-0396(67)90035-6.

[114] D. Wu. *Contraction Analysis of Nonlinear Systems on Riemannian Manifolds*. PhD thesis, Thèse de doctorat de l'université Paris-Saclay et de Harbin Institute of Technology, 2022.

[115] B. Yi and I. R. Manchester. On the equivalence of contraction and Koopman approaches for nonlinear stability and control. *IEEE Transactions on Automatic Control*, 69(7):4336–4351, 2023. doi:10.1109/tac.2023.3319051.

[116] K. Yoshida, M. Inoue, and T. Hatanaka. Instant MPC for linear systems and dissipativity-based stability analysis. *lcsys*, 3(4):811–816, 2019. doi:10.1109/LCSYS.2019.2918095.

[117] E. Zeidler. *Nonlinear Functional Analysis and Its Applications. I: Fixed Point Theorems*. Springer, 1986, ISBN 978-0-387-90914-1.

[118] Y. Zhang, R. Wang, and I. R. Manchester. Robustly invertible nonlinear dynamics and the BiLipREN: Contracting neural models with contracting inverses. *arXiv preprint arXiv:2505.03069*, 2025. doi:10.48550/arXiv.2505.03069.

[119] 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. doi:10.1109/ccece.1998.685646.