Contraction theory for nonlinear stability analysis and learning-based control: A tutorial overview

Contraction Theory for Nonlinear Stability Analysis and Learning-based Control: A Tutorial Overview

Hiroyasu Tsukamoto[a], Soon-Jo Chung[a,b], Jean-Jacques Slotine[c]

[a] *Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA, USA*
[b] *NASA Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA*
[c] *Nonlinear Systems Laboratory, Massachusetts Institute of Technology, Cambridge MA, USA*

Abstract

Contraction theory is an analytical tool to study differential dynamics of a non-autonomous (i.e., time-varying) nonlinear system under a contraction metric defined with a uniformly positive definite matrix, the existence of which results in a necessary and sufficient characterization of incremental exponential stability of multiple solution trajectories with respect to each other. By using a squared differential length as a Lyapunov-like function, its nonlinear stability analysis boils down to finding a suitable contraction metric that satisfies a stability condition expressed as a linear matrix inequality, indicating that many parallels can be drawn between well-known linear systems theory and contraction theory for nonlinear systems. Furthermore, contraction theory takes advantage of a superior robustness property of exponential stability used in conjunction with the comparison lemma. This yields much-needed safety and stability guarantees for neural network-based control and estimation schemes, without resorting to a more involved method of using uniform asymptotic stability for input-to-state stability. Such distinctive features permit systematic construction of a contraction metric via convex optimization, thereby obtaining an explicit exponential bound on the distance between a time-varying target trajectory and solution trajectories perturbed externally due to disturbances and learning errors. The objective of this paper is therefore to present a tutorial overview of contraction theory and its advantages in nonlinear stability analysis of deterministic and stochastic systems, with an emphasis on deriving formal robustness and stability guarantees for various learning-based and data-driven automatic control methods. In particular, we provide a detailed review of techniques for finding contraction metrics and associated control and estimation laws using deep neural networks.

1. Introduction

Lyapunov theory is one of the most widely-used approaches to stability analysis of a nonlinear system [1–6], which provides a condition for stability with respect to an equilibrium point, a target trajectory, or an invariant set. Contraction theory [7–11] rewrites suitable Lyapunov stability conditions using a quadratic Lyapunov function of the differential states, defined by a Riemannian contraction metric and its uniformly positive definite matrix, thereby characterizing a necessary and sufficient condition for incremental exponential convergence of the multiple nonlinear system trajectories to one single trajectory. It can be regarded as a generalization of Krasovskii's theorem [5, p. 83] applied to nonlinear incremental stability analysis [7, 12], where the differential formulation permits a pure differential coordinate change with a non-constant metric for simplifying its stability proofs [7].

The differential nature of contraction theory implies we can exploit the Linear Time-Varying (LTV) systems-type techniques for nonlinear stability analysis and control/estimation synthesis [13–17] (see Table 1). We emphasize that some of these methodological simplifications in contraction theory are accomplished by its extensive use of exponential stability along with the comparison lemma [3, pp. 102-103, pp. 350-353], in lieu of Input-to-State Stability (ISS) or uniform asymptotic stability which often renders nonlinear stability analysis more involved [1–6]. Several studies related to the notion of contraction, although not based on direct differential analysis, can be traced back to [18–20].

The objective of this tutorial paper is to elucidate how contraction theory may be utilized as a method of providing provable incremental exponential robustness and stability guarantees of learning-based and data-driven automatic control techniques. In pursuit of this goal, we also provide an overview of the advantages of contraction theory and present a systematic convex optimization formulation to explicitly construct an optimal contraction metric and a differential Lyapunov function for general nonlinear deterministic and stochastic systems.

1.1. Paper Organization

This tutorial paper is organized into the following two groups of sections.

*This work is in part funded by the NASA Jet Propulsion Laboratory, California Institute of Technology.

Email addresses: htsukamoto@caltech.edu (Hiroyasu Tsukamoto), sjchung@caltech.edu (Soon-Jo Chung), jjs@mit.edu (Jean-Jacques Slotine)

Table 1: Differences between Contraction Theory and Lyapunov Theory.

	Contraction theory (which constructs positive definite matrix $M(x,t)$ that defines contraction metric)	Lyapunov direct method (which constructs Lyapunov function $V(x,t)$)
1. Lyapunov function	Always quadratic function of differential state δx ($V = \delta x^\top M(x,t)\delta x$)	Any function of x, including $V(x,t) = x^\top M(x,t)x$
2. Stability condition	Exponential stability of trajectories including points/invariant sets*	Asymptotic or exponential stability of points and invariant sets
3. Incremental stability	Incremental stability of trajectories using differential displacements ($\lim_{t\to\infty} \delta x(t) = 0$)	Incremental stability via stability of points ($\lim_{t\to\infty}(x(t) - x_d(t)) = 0$ for given $x_d(t)$)
4. Non-autonomous system	Same theory as for autonomous systems	Additional conditions required for non-autonomous stability analysis
5. Robustness analysis	Intuitive both for ISS and finite-gain $\mathcal{L}_p$ stability due to extensive use of exponential stability	Same as contraction theory if exponentially stable; more involved if uniformly asymptotically stable
6. Analogy to linear system	LTV-like differential dynamics for global convergence	Indirect methods use linearization for local stability (direct methods use motion integrals)
7. $\mathcal{L}_2$ stability condition	Reduces to LMI conditions in terms of contraction metric defined by positive definite matrix M	Hamilton-Jacobi inequality (PDE) in terms of Lyapunov function V [3, p. 211]
8. Modular stability	Differential analysis handles hierarchical, feedback, and parallel combinations [10]	Passivity is not intuitive for hierarchical combinations

*A semi-contracting system with a negative semi-definite generalized Jacobian matrix can be used to analyze asymptotic stability (see Sec. 8).

Part I: Nonlinear Stability Analysis (Sec. 2–4)

In Sec. 2, we present fundamental results of contraction theory for nonlinear robustness and stability analysis. Sections 3 and 4 consider nonlinear optimal feedback control and estimation problems from the perspective of contraction theory, deriving and delineating a convex optimization-based method for constructing contraction metrics. Section 3 also presents some new results on relating contraction theory to the bounded real lemma [21] and Kalman–Yakubovich–Popov (KYP) lemma [22, p. 218].

Part II: Learning-based Control (Sec. 5–9)

In Sec. 5, we derive several theorems which form the basis of learning-based control using contraction theory. Sections 6 and 7 present frameworks for learning-based control, estimation, and motion planning via contraction theory using deep neural networks for designing contraction metrics, and Section 8 extends these results to parametric uncertain nonlinear systems with adaptive control techniques. In Sec. 9, we propose model-free versions of contraction theory for learning-based and data-driven control.

1.2. Related Work

In the remainder, we give an overview of each section (Sec. 2–9) as well as a survey of related work.

Contraction Theory (Sec. 2)

According to contraction theory, all the solution trajectories of a given nonlinear system converge to one single trajectory incrementally and exponentially, regardless of the initial conditions, if the system has a contraction metric and its associated quadratic Lyapunov function of the differential state [7] (see 1–3 of Table 1). This paper primarily considers this generalized notion of stability, called incremental exponential stability, which enables systematic learning-based and data-driven control synthesis with formal robustness and stability guarantees. The purpose of this section is not for proposing that other notions of stability, such as traditional Lyapunov-based stability or incremental asymptotic stability for semi-contraction systems, should be replaced by incremental exponential stability, but for clarifying its advantages to help determine which of these concepts is the best fit when analyzing nonlinear robustness and stability (see Sec. 2 for the illustrative examples).

In keeping with the use of the comparison lemma [3, pp. 102-103, pp. 350-353], incremental exponential stability naturally holds for non-autonomous nonlinear systems without any additional conditions or modifications unlike Lyapunov techniques (see, e.g., the examples and theorems in [23]). Such aspects of contraction theory, including the extensive use of exponential stability, result in intuitive proofs on ISS and finite-gain $\mathcal{L}_p$ stability both for autonomous and non-autonomous nonlinear systems, without resorting to uniform asymptotic stability which makes stability analysis much more involved than necessary [1–6]. In particular, perturbed systems with a time-varying target trajectory are non-autonomous, and thus contraction theory allows us to easily obtain an explicit exponential bound on its tracking error [7, 13, 16, 17, 24], leveraging incremental stability of the perturbed system trajectories with respect to the target trajectory (see 4 of Table 1). Having such analytical bounds on the tracking error is almost essential for the safe and robust implementation of automatic control schemes in real-world scenarios.

Contraction theory also simplifies input-output stability analysis, such as $\mathcal{L}_p$ gain analysis of nonlinear systems including the $\mathcal{H}_\infty$ nonlinear optimal control problem [25–32]. For example, in Lyapunov theory, the problem of

finding a suitable Lyapunov function with the smallest $\mathcal{L}_2$ gain boils down to solving a Partial Differential Equation (PDE) called the Hamilton-Jacobi inequality [3, p. 211], [25] in terms of its associated Lyapunov function. In essence, since contraction theory utilizes a quadratic Lyapunov function of the differential state for stability analysis, the problem could be solved with a Linear Matrix Inequality (LMI) constraint [21] analogous to the KYP lemma [22, p. 218] in LTV systems theory [13, 16, 17, 33], as shall be shown in Sec. 3–4 (see 5–7 of Table 1). There exist stochastic analogues of these stability results for nonlinear systems with stochastic perturbations [13, 16, 17, 34, 35], as shall be outlined also in this paper.

Another notable feature of contraction theory is modularity, which preserves contraction through parallel, feedback, and hierarchical combinations [9, 10], specific time-delayed feedback communications [36], synchronized coupled oscillations [11, 37], and synchronized networks [10, 24, 24, 37–40], expanding the results obtainable with the passivity formalism of Lyapunov theory [41], [3, p. 227], [5, p. 132] (see 8 of Table 1). Due to all these useful properties, extensions of contraction theory have been considered in many different settings. These include, but are not limited to, stochastic contraction (Gaussian white noise [13, 16, 17, 34], Poisson shot noise and Lévy noise [35]), contraction for discrete and hybrid nonlinear systems [7, 8, 13, 17, 37, 42, 43], partial contraction [11], transverse contraction [44], incremental stability analysis of nonlinear estimation (the Extended Kalman Filter (EKF) [45], nonlinear observers [16, 46], Simultaneous Localization And Mapping (SLAM) [47]), generalized gradient descent based on geodesical convexity [48], contraction on Finsler and Riemannian manifolds [49–51], contraction on Banach and Hilbert spaces for PDEs [52–54], non-Euclidean contraction [55], contracting learning with piecewise-linear basis functions [56], incremental quadratic stability analysis [57], contraction after small transients [58], immersion and invariance stabilizing controller design [59, 60], and Lipschitz-bounded neural networks for robustness and stability guarantees [61–64].

Construction of Contraction Metrics (Sec. 3–4)

The benefits of contraction theory reviewed so far naturally lead to a discussion on how to design a contraction metric and corresponding Lyapunov function. There are some cases in which we can analytically find them using special structures of systems in question [1, 3, 5]. Among these are Lagrangian systems [5, p. 392], where one easy choice of positive definite matrices that define a contraction metric is the inertia matrix, or feedback linearizable systems [65–69], where we could solve the Riccati equation for a contraction metric as in LTV systems. This is also the case in the context of state estimation (e.g., the nonlinear SLAM problem can be reformulated as an LTV estimation problem using virtual synthetic measurements [16, 47]). Once we find a contraction metric and Lyapunov function of a nominal nonlinear system for the sake of stability, they could be used as a Control Lyapunov Function (CLF) to attain stabilizing feedback control [3, 70, 71] or could be augmented with an integral control law called adaptive backstepping to recursively design a Lyapunov function for strict- and output-feedback systems [72–75]. However, deriving an analytical form of contraction metrics for general nonlinear systems is challenging, and thus several search algorithms have been developed for finding them at least numerically using the LMI nature of the contraction condition.

The simplest of these techniques is the method of State-Dependent Riccati Equation (SDRE) [76–78], which uses the State-Dependent Coefficient (SDC) parameterization (also known as extended linearization) of nonlinear systems for feedback control and state estimation synthesis. Motivating optimization-based approaches to design a contraction metric, it is proposed in [13, 16, 17] that the Hamilton-Jacobi inequality for the finite-gain $\mathcal{L}_2$ stability condition can be expressed as an LMI when contraction theory is equipped with the extended linearity of the SDC formulation. Specifically, in [13–15, 17], a convex optimization-based framework for robust feedback control and state estimation, named ConVex optimization-based Steady-state Tracking Error Minimization (CV-STEM), is derived to find a contraction metric that minimizes an upper bound of the steady-state distance between perturbed and unperturbed system trajectories. In this context, we could utilize Control Contraction Metrics (CCMs) [33, 79–83] for extending contraction theory to the systematic design of differential feedback control $\delta u = k(x, \delta x, u, t)$ via convex optimization, achieving greater generality at the expense of computational efficiency in obtaining u. Applications of the CCM to estimation, adaptive control, and motion planning are discussed in [84], [85–87], and [80, 88–91], respectively, using geodesic distances between trajectories [50]. It is also worth noting that the objective function of CV-STEM has the condition number of a positive definite matrix that defines a contraction metric as one of its arguments, rendering it applicable and effective even to machine learning-based automatic control frameworks as shall be seen in Sec. 5–9.

Contraction Theory for Learning-based Control (Sec. 5)

One drawback of these numerical schemes is that they require solving optimization problems or nonlinear systems of equations at each time instant online, which is not necessarily realistic in practice. In Lyapunov theory, approximating functions in a given hypothesis space has therefore been a standard technique [92–102], where examples of its function classes include piecewise quadratic functions [92], linearly parameterized non-quadratic functions [93], a linear combination of radial basis functions [94], Sum-Of-Squares (SOS) functions [95], and neural networks [96, 97, 102]. In [68, 79], the SOS approximation is investigated for the case of contraction theory, showing that the contraction condition can be relaxed to SOS conditions for dynamics with polynomial or rational vector fields. Although

computationally tractable, it still has some limitations in that the problem size grows exponentially with the number of variables and basis functions [103]. Learning-based and data-driven control using contraction theory [14, 15, 104, 105] has been developed to refine these ideas, using the high representational power of DNNs [106–108] and their scalable training realized by stochastic gradient descent [48, 109].

The major advantage of using contraction theory for learning-based and data-driven control is that, by regarding its internal learning error as an external disturbance, we can ensure the distance between the target and learned trajectories to be bounded exponentially with time as in the CV-STEM results [14, 15, 104, 105], with its steady-state upper bound proportional to the learning error. Such robustness and incremental stability guarantees are useful for formally evaluating the performance of machine learning techniques such as reinforcement learning [110–113], imitation learning [114–118], or neural networks [106–108]. This implies contraction theory could be utilized as a central tool in realizing safe and robust operations of learning-based and data-driven control, estimation, and motion planning schemes in real-world scenarios. We especially focus on the following areas of research.

Learning-based Robust Control and Estimation (Sec. 6)

In order to achieve real-time computation of a contraction metric, mathematical models based on a Deep Neural Network (DNN) called a Neural Contraction Metric (NCM) [14] and Neural Stochastic Contraction Metric (NSCM) [15] are derived to compute optimal CV-STEM contraction metrics for nonlinear systems perturbed by deterministic and stochastic disturbances, respectively. It can be proven that the NCM and NSCM still yield robustness and optimality associated with the CV-STEM framework despite having non-zero modeling errors [119]. These metrics could also be synthesized and learned simultaneously with their feedback control laws directly by DNNs [91, 119, 120] at the expense of the convex property in the CV-STEM formulation.

Learning-based Robust Motion Planning (Sec. 7)

In [80, 121], contraction theory is leveraged to develop a tracking feedback controller with an optimized control invariant tube, solving the problem of robust motion planning under bounded external disturbances. This problem is also considered for systems with changing operating conditions [88] and parametric uncertainty [89–91] for its broader use in practice. As these methods still require online computation of a target trajectory, Learning-based Autonomous Guidance with RObustness and Stability (LAG-ROS) [105] is developed to model such robust control laws including the CV-STEM by a DNN, without explicitly requiring the target or desired trajectory as its input. While this considers motion planning algorithms only implicitly to avoid solving them in real-time, it is shown that contraction theory still allows us to assure a property of robustness against deterministic and stochastic disturbances following the same argument as in the NCM and NSCM work [105]. Note that LAG-ROS using contraction theory is not intended to derive new learning-based motion planning, but rather to augment any existing motion planner with a real-time method of guaranteeing formal incremental robustness and stability, and thus still applicable to other methods such as tube-based robust Model Predictive Control (MPC) [80, 121–125], its dynamic and adaptive counterparts [88–91], and CCM-based learning certified control [120].

Learning-based Adaptive Control (Sec. 8)

Adaptive control using contraction theory is studied in [75] for parametric strict-feedback nonlinear systems, and recently generalized to deal with systems with unmatched parametric uncertainty by means of parameter-dependent CCM feedback control [85, 86]. This method is further explored to develop an adaptive Neural Contraction Metric (aNCM) [104], a parameter-dependent DNN model of the adaptive CV-STEM contraction metric. As the name suggests, the aNCM control makes adaptive control of [85, 86] implementable in real-time for asymptotic stabilization, while maintaining the learning-based robustness and CV-STEM-type optimality of the NCM. Although it is designed to avoid the computation for evaluating integrals involving geodesics unlike [85, 86], these differential state feedback schemes could still be considered, trading off added computational cost for generality. It is demonstrated in [104] that the aNCM is applicable to many types of systems such as robotics systems [5, p. 392], spacecraft high-fidelity dynamics [126, 127], and systems modeled by basis function approximation and DNNs [128, 129]. Discrete changes could be incorporated in this framework using [87, 130].

Contraction Theory for Learned Models (Sec. 9)

Recent applications of machine learning often consider challenging scenarios in the field of systems and control theory, where we only have access to system trajectory data generated by unknown underlying dynamics, and the assumptions in the aforementioned adaptive control techniques are no longer valid. For situations where the data is used for system identification of full/residual dynamics [61, 64, 81, 131] by a spectrally-normalized DNN [132], we can show by contraction theory that the model-based approaches (e.g. CV-STEM and NCM) are still utilizable to guarantee robustness against dynamics modeling errors and external disturbances. It is also proposed in [133] that we could directly learn certificate functions such as contraction metrics and their associated Lyapunov functions using trajectory data. Note that some of the theoretical results on gradient descent algorithms, essential in the field of data-driven machine learning, can be replaced by more general ones based on contraction and geodesical convexity [48].

Table 2: Notations used in this paper.

$\|x\|$	Euclidean norm of $x \in \mathbb{R}^n$
δx	Differential displacement of $x \in \mathbb{R}^n$
$\|A\|$	Induced 2-norm of $A \in \mathbb{R}^{n\times m}$
$\|A\|_F$	Frobenius norm of $A \in \mathbb{R}^{n\times m}$
$\mathrm{sym}(A)$	Symmetric part of $A \in \mathbb{R}^{n\times n}$, i.e., $(A + A^\top)/2$
$\lambda_{\min}(A)$	Minimum eigenvalue of $A \in \mathbb{R}^{n\times n}$
$\lambda_{\max}(A)$	Maximum eigenvalue of $A \in \mathbb{R}^{n\times n}$
I	Identity matrix of appropriate dimensions
$\mathbb{E}$	Expected value operator
$\mathbb{P}$	Probability measure
$\mathbb{R}_{>0}$	Set of positive reals, i.e., $\{a \in \mathbb{R} \vert a \in (0, \infty)\}$
$\mathbb{R}_{\geq 0}$	Set of non-negative reals, i.e., $\{a \in \mathbb{R} \vert a \in [0, \infty)\}$

1.3. Notation

For a square matrix $A^{n\times n}$, we use the notation $A \succ 0$, $A \succeq 0$, $A \prec 0$, and $A \preceq 0$ for the positive definite, positive semi-definite, negative definite, negative semi-definite matrices, respectively. The $\mathcal{L}_p$ norm in the extended space $\mathcal{L}_{pe}$ [3, pp. 196-197], $p \in [1, \infty]$, is defined as $\|(y)_\tau\|_{\mathcal{L}_p} = \left(\int_0^\tau \|y(t)\|^p\right)^{1/p} < \infty$ for $p \in [1, \infty)$ and $\|(y)_\tau\|_{\mathcal{L}_\infty} = \sup_{t\geq 0} \|(y(t))_\tau\| < \infty$ for $p = \infty$, where $(y(t))_\tau$ is a truncation of $y(t)$, i.e., $(y(t))_\tau = 0$ for $t > \tau$ and $(y(t))_\tau = y(t)$ for $0 \leq t \leq \tau$ with $\tau \in \mathbb{R}_{\geq 0}$. Furthermore, we use $f_x = \partial f/\partial x$, $M_{x_i} = \partial M/\partial x_i$, and $M_{x_i x_j} = \partial^2 M/(\partial x_i \partial x_j)$, where x_i and x_j ate the ith and jth elements of $x \in \mathbb{R}^n$, for describing partial derivatives in a limited space. The other notations are given in Table 2.

Part I: Nonlinear Stability Analysis

2. Contraction Theory

We present a brief review of the results from [7, 10, 11, 13, 16, 17, 24, 34]. They will be extensively used to provide formal robustness and stability guarantees for a variety of systems in the subsequent sections, simplifying and generalizing Lyapunov theory.

2.1. Fundamentals

Consider the following smooth non-autonomous (i.e., time-varying) nonlinear system:

$$\dot{x}(t) = f(x(t), t) \tag{1}$$

where $t \in \mathbb{R}_{\geq 0}$ is time, $x : \mathbb{R}_{\geq 0} \mapsto \mathbb{R}^n$ the system state, and $f : \mathbb{R}^n \times \mathbb{R}_{\geq 0} \mapsto \mathbb{R}^n$ is a smooth function. Note that the smoothness of $f(x, t)$ guarantees existence and uniqueness of the solution to (1) for a given $x(0) = x_0$ at least locally [3, pp. 88-95].

Definition 2.1. *A differential displacement, δx, is defined as an infinitesimal displacement at a fixed time as used in the calculus of variation [134, p. 107], and (1) yields the following differential dynamics:*

$$\delta\dot{x}(t) = \frac{\partial f}{\partial x}(x(t), t)\delta x(t) \tag{2}$$

where $f(x(t), t)$ is given in (1).

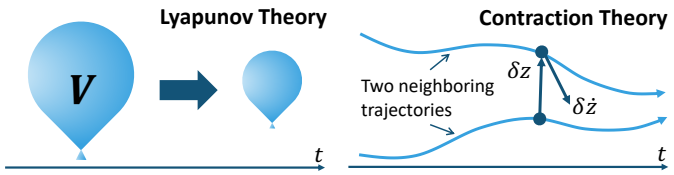

Figure 1: Lyapunov theory and contraction theory, where V is a Lyapunov function and $\delta z = \Theta(x, t)\delta x$ for $M(x, t) = \Theta(x, t)\Theta(x, t)^\top \succ 0$ that defines a contraction metric (see Theorem 2.1).

Let us first present a special case of the comparison lemma [3, pp. 102-103, pp. 350-353] to be used extensively throughout this paper.

Lemma 2.1. *Suppose that a continuously differentiable function $v \in \mathbb{R}_{\geq 0} \mapsto \mathbb{R}$ satisfies the following differential inequality:*

$$\dot{v}(t) \leq -\gamma v(t) + c, \ v(0) = v_0, \ \forall t \in \mathbb{R}_{\geq 0}$$

where $\gamma \in \mathbb{R}_{>0}$, $c \in \mathbb{R}$, and $v_0 \in \mathbb{R}$. Then we have

$$v(t) \leq v_0 e^{-\gamma t} + \frac{c}{\gamma}(1 - e^{-\gamma t}), \ \forall t \in \mathbb{R}_{\geq 0}.$$

Proof. See [3, pp. 659-660]. □

2.1.1. Contraction Theory and Contraction Metric

In Lyapunov theory, nonlinear stability of (1) is studied by constructing a Lyapunov function $V(x, t)$, one example of which is $V = x^\top P(x, t)x$. However, finding $V(x, t)$ for general nonlinear systems is challenging as $V(x, t)$ can be any scalar function of x (e.g., a candidate $V(x, t)$ can be obtained by solving a PDE [3, p. 211]). In contrast, as summarized in Table 1, contraction theory uses a differential Lyapunov function that is always a quadratic function of δx, i.e., $V(x, \delta x, t) = \delta x^\top M(x, t)\delta x$, thereby characterizing a necessary and sufficient condition for incremental exponential convergence of the multiple nonlinear system trajectories to one single trajectory. Thus, the problem of finding V for stability analysis boils down to finding a finite-dimensional positive-definite matrix M, as illustrated in Figure 1 [7]. These properties to be derived in Theorem 2.1, which hold both for autonomous (i.e. time-invariant) and non-autonomous systems, epitomize significant methodological simplifications of stability analysis in contraction theory.

Theorem 2.1. *If there exists a uniformly positive definite matrix $M(x, t) = \Theta(x, t)^\top \Theta(x, t) \succ 0$, $\forall x, t$, where $\Theta(x, t)$ defines a smooth coordinate transformation of δx,*

i.e., $\delta z = \Theta(x,t)\delta x$, s.t. either of the following equivalent conditions holds for $\exists \alpha \in \mathbb{R}_{>0}$, $\forall x, t$:

$$\lambda_{\max}(F(x,t)) = \lambda_{\max}\left(\left(\dot{\Theta} + \Theta\frac{\partial f}{\partial x}\right)\Theta^{-1}\right) \leq -\alpha \tag{3}$$

$$\dot{M} + M\frac{\partial f}{\partial x} + \frac{\partial f}{\partial x}^{\top} M \preceq -2\alpha M \tag{4}$$

where the arguments (x,t) of $M(x,t)$ and $\Theta(x,t)$ are omitted for notational simplicity, then all the solution trajectories of (1) *converge to a single trajectory exponentially fast regardless of their initial conditions (i.e., contracting, see Definition 2.3), with an exponential convergence rate α. The converse also holds.*

Proof. The proof of this theorem can be found in [7], but here we emphasize the use of the comparison lemma given in Lemma 2.1. Taking the time-derivative of a differential Lyapunov function of δx (or δz), $V = \delta z^{\top}\delta z = \delta x^{\top} M(x,t)\delta x$, using the differential dynamics (2), we have

$$\begin{aligned}\dot{V}(x,\delta x,t) &= 2\delta z^{\top} F \delta z = \delta x^{\top}\Big(\dot{M} + \frac{\partial f}{\partial x}^{\top} M + M\frac{\partial f}{\partial x}\Big)\delta x \\ &\leq -2\alpha\delta z^{\top}\delta z = -2\alpha\delta x^{\top} M(x,t)\delta x\end{aligned}$$

where the conditions (3) and (4) are used with the generalized Jacobian F in (3) obtained from $\delta\dot{z} = \dot{\Theta}\delta x + \Theta\delta\dot{x} = F\delta z$. We get $d\|\delta z\|/dt \leq -\alpha\|\delta z\|$ by $d\|\delta z\|^2/dt = 2\|\delta z\| d\|\delta z\|/dt$, which then yields $\|\delta z(t)\| \leq \|\delta z_0\| e^{-\alpha t}$ by the comparison lemma of Lemma 2.1. Hence, any infinitesimal length $\|\delta z(t)\|$ and $\|\delta x(t)\|$, as well as δz and δx, tend to zero exponentially fast. By path integration (see Definition 2.2 and Theorem 2.3), this immediately implies that the length of any finite path converges exponentially to zero from any initial conditions.

Conversely, consider an exponentially convergent system, which implies the following for $\exists \beta > 0$ and $\exists k \geq 1$:

$$\|\delta x(t)\|^2 \leq -k\|\delta x(0)\|^2 e^{-2\beta t} \tag{5}$$

and define a matrix-valued function $\Xi(x(t),t) \in \mathbb{R}^{n\times n}$ (not necessarily $\Xi \succ 0$) as

$$\dot{\Xi} = -2\beta\Xi - \Xi\frac{\partial f}{\partial x} - \frac{\partial f}{\partial x}^{\top}\Xi,\ \Xi(x(0),0) = k\mathrm{I}. \tag{6}$$

Note that, for $V = \delta x^{\top}\Xi\delta x$, (6) gives $\dot{V} = -2\beta V$, resulting in $V = -k\|\delta x(0)\|^2 e^{-2\beta t}$. Substituting this into (5) yields

$$\|\delta x(t)\|^2 = \delta x(t)^{\top}\delta x(t) \leq V = \delta x(t)^{\top}\Xi(x(t),t)\delta x(t) \tag{7}$$

which indeed implies that $\Xi \succeq \mathrm{I} \succ 0$ as (7) holds for any δx. Thus, Ξ satisfies the contraction condition (4), i.e., Ξ defines a contraction metric (see Definition 2.3) [7]. □

Remark 2.1. *Since M of Theorem 2.1 is positive definite, i.e., $v^{\top}Mv = \|\Theta v\|^2 \geq 0$, $\forall v \in \mathbb{R}^n$, and $\|\Theta v\|^2 = 0$ if and only if $v = 0$, the equation $\Theta v = 0$ only has a trivial solution $v = 0$. This implies that Θ is always non-singular (i.e., $\Theta(x,t)^{-1}$ always exists).*

Definition 2.2. *Let $\xi_0(t)$ and $\xi_1(t)$ denote some solution trajectories of* (1)*. We say that* (1) *is incrementally exponentially stable if $\exists C, \alpha > 0$ s.t. the following holds [12]:*

$$\|\xi_1(t) - \xi_0(t)\| \leq Ce^{-\alpha t}\|\xi_0(0) - \xi_1(0)\|$$

for any $\xi_0(t)$ and $\xi_1(t)$. Note that, since we have $\|\xi_1(t) - \xi_0(t)\| = \|\int_{\xi_0}^{\xi_1}\delta x\|$ (see Theorem 2.3), Theorem 2.1 implies incremental stability of the system (1)*.*

Definition 2.3. *The system* (1) *satisfying the conditions in Theorem 2.1 is said to be contracting, and a uniformly positive definite matrix M that satisfies* (4) *defines a contraction metric. As to be discussed in Theorem 2.3 of Sec. 2.2, a contracting system is incrementally exponentially stable in the sense of Definition 2.2.*

Example 2.1. *One of the distinct features of contraction theory in Theorem 2.1 is incremental stability with exponential convergence. Consider the example given in [135]:*

$$\frac{d}{dt}\begin{bmatrix}x_1\\x_2\end{bmatrix} = \begin{bmatrix}-1 & x_1\\ -x_1 & -1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}. \tag{8}$$

A Lyapunov function $V = \|x\|^2/2$ for (8)*, where $x = [x_1, x_2]^{\top}$, yields $\dot{V} \leq -2V$. Thus,* (8) *is exponentially stable with respect to $x = 0$. The differential dynamics of* (8) *is given as*

$$\frac{d}{dt}\begin{bmatrix}\delta x_1\\\delta x_2\end{bmatrix} = \begin{bmatrix}-1+x_2 & x_1\\ -2x_1 & -1\end{bmatrix}\begin{bmatrix}\delta x_1\\\delta x_2\end{bmatrix}. \tag{9}$$

and the contraction condition (4) *for* (9) *can no longer be proven by $V = \|\delta x\|^2/2$, due to the lack of the skew-symmetric property of* (8) *in* (9)*. This difficulty illustrates the difference between Lyapunov theory and contraction theory, where the former considers stability of* (8) *with respect to the equilibrium point, while the latter analyzes exponential convergence of any couple of trajectories in* (8) *with respect to each other (i.e., incremental stability in Definition 2.2) [7, 135].*

Example 2.2. *Contraction defined by Theorem 2.1 guarantees incremental stability of their solution trajectories but does not require the existence of stable fixed points. Let us consider the following system for $x : \mathbb{R}_{\geq 0} \mapsto \mathbb{R}$:*

$$\dot{x} = -x + e^t. \tag{10}$$

Using the constraint (4) *of Theorem 2.1, we can easily verify that* (10) *is contracting as $M = \mathrm{I}$ defines its contraction metric with the contraction rate $\alpha = 1$. However, since* (10) *has $x(t) = e^t/2 + (x(0) - 1/2)e^{-t}$ as its unique solution, it is not stable with respect to any fixed point.*

Example 2.3. *Consider a Linear Time-Invariant (LTI) system, $\dot{x} = f(x) = Ax$. Lyapunov theory states that the*

origin is globally exponentially stable if and only if there exists a constant positive-definite matrix $P \in \mathbb{R}^{n\times n}$ s.t. [136, pp. 67-68]

$$\exists \epsilon > 0 \text{ s.t. } PA + A^\top P \preceq -\epsilon \mathrm{I}. \tag{11}$$

Now, let $\overline{p} = \|P\|$. Since $-\mathrm{I} \preceq -P/\overline{p}$, (11) implies that $PA + A^\top P \leq -(\epsilon/\overline{p})P$, which shows that $M = P$ with $\alpha = \epsilon/(2\overline{p})$ satisfies (4) due to the relation $\partial f/\partial x = A$.

The contraction condition (4) can thus be viewed as a generalization of the Lyapunov stability condition (11) (see the generalized Krasovskii's theorem [5, pp. 83-86]) in a nonlinear non-autonomous system setting, expressed in the differential formulation that permits a non-constant metric and pure differential coordinate change [7]. Furthermore, if $f(x,t) = A(t)x$ and $M = \mathrm{I}$, (4) results in $A(t)+A(t)^\top \preceq -2\alpha\mathrm{I}$ (i.e., all the eigenvalues of the symmetric matrix $A(t)+A(t)^\top$ remain strictly in the left-half complex plane), which is a known sufficient condition for stability of Linear Time-Varying (LTV) systems [5, pp. 114-115].

Example 2.4. *Most of the learning-based techniques involving neural networks are based on optimizing their hyperparameters by gradient descent [109]. Contraction theory provides a generalized view on the analysis of such continuous-time gradient-based optimization algorithms [48].*

Let us consider a twice differentiable scalar output function $f : \mathbb{R}^n \times \mathbb{R} \mapsto \mathbb{R}$, a matrix-valued function $M : \mathbb{R}^n \mapsto \mathbb{R}^{n\times n}$ with $M(x) \succ 0$, $\forall x \in \mathbb{R}^n$, and the following natural gradient system [137]:

$$\dot{x} = h(x,t) = -M(x)^{-1}\nabla_x f(x,t). \tag{12}$$

Then, f is geodesically α-strongly convex for each t in the metric defined by $M(x)$ (i.e., $H(x) \succeq \alpha M(x)$ with $H(x)$ being the Riemannian Hessian matrix of f with respect to M [138]), if and only if (12) is contracting with rate α in the metric defined by M as in (4) of Theorem 2.1, where $A = \partial h/\partial x$. More specifically, the Riemannian Hessian verifies $H(x) = -(\dot{M} + MA + A^\top M)/2$. See [48] for details.

Remark 2.2. *Theorem 2.1 can be applied to other vector norms of $\|\delta z\|_p$ with, e.g., $p = 1$ or $p = \infty$ [7]. It can also be shown that for a contracting autonomous system of the form $\dot{x} = f(x)$, all trajectories converge to an equilibrium point exponentially fast.*

2.1.2. Partial Contraction

Although satisfying the condition (4) of Theorem 2.1 guarantees exponential convergence of any couple of trajectories in (1), proving their incremental stability with respect to a subset of these trajectories possessing a specific property could be sufficient for some cases [13–15, 17], leading to the concept of partial contraction [11].

Theorem 2.2. *Consider the following nonlinear system with the state $x \in \mathbb{R}_{\geq 0} \mapsto \mathbb{R}^n$ and the auxiliary or virtual system with the state $q \in \mathbb{R}_{\geq 0} \mapsto \mathbb{R}^n$:*

$$\dot{x}(t) = g(x(t), x(t), t) \tag{13}$$
$$\dot{q}(t) = g(q(t), x(t), t) \tag{14}$$

where $g : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}_{\geq 0} \to \mathbb{R}^n$ is a smooth function. Suppose that (14) is contracting with respect to q. If a particular solution of (14) verifies a smooth specific property, then all trajectories of (13) verify this property exponentially.

Proof. The theorem statement follows from Theorem 2.1 and the fact that $q = x$ and a trajectory with the specific property are particular solutions of (14) (see [11] for details). □

The importance of this theorem lies in the fact that we can analyze contraction of some specific parts of the system (13) while treating the rest as a function of the time-varying parameter $x(t)$. Strictly speaking, the system (13) is said to be partially contracting, but we will not distinguish partial contraction from contraction of Definition 2.3 in this paper for simplicity. Instead, we will use the variable q when referring to partial contraction of Theorem 2.2.

Examples of a trajectory with a specific property include the target trajectory x_d of feedback control in (57), or the trajectory of actual dynamics x for state estimation in (90) and (91) to be discussed in the subsequent sections. Note that contraction can be regarded as a particular case of partial contraction.

Example 2.5. *Let us illustrate the role of partial contraction using the following nonlinear system [135]:*

$$\dot{x} = -D(x)x + u,\ \dot{x}_d = -D(x_d)x_d$$

where x is the system state, x_d is the target state, u is the control input designed as $u = -K(x)(x - x_d) + (D(x) - D(x_d))x_d$, and $D(x)+K(x) \succ 0$. We could define a virtual system, which has $q = x$ and $q = x_d$ as its particular solutions, as follows:

$$\dot{q} = -(D(x) + K(x))(q - x_d) - D(x_d)x_d. \tag{15}$$

Since we have $\delta\dot{q} = -(D(x)+K(x))\delta q$ and $D(x)+K(x) \succ 0$, (15) is contracting with $M = \mathrm{I}$ in (4). However, if we consider the following virtual system:

$$\dot{q} = -(D(q) + K(q))(q - x_d) - D(x_d)x_d \tag{16}$$

which also has $q = x$ and $q = x_d$ as its particular solutions, proving contraction is no longer straightforward because of the terms $\partial D/\partial q_i$ and $\partial K/\partial q_i$ in the differential dynamics of (16). This is due to the fact that, in contrast to (15), (16) has particular solutions nonlinear in q in addition to $q = x$ and $q = x_d$, and the condition (4) becomes more involved for (16) as it is for guaranteeing exponential convergence of any couple of these particular solution trajectories [135].

Example 2.6. *As one of the key applications of partial contraction given in Theorem 2.2, let us consider the following closed-loop Lagrangian system [5, p. 392]:*

$$\mathcal{H}(\mathrm{q})\ddot{\mathrm{q}} + \mathcal{C}(\mathrm{q},\dot{\mathrm{q}})\dot{\mathrm{q}} + \mathcal{G}(\mathrm{q}) = u(\mathrm{q},\dot{\mathrm{q}},t) \quad (17)$$

$$u(\mathrm{q},\dot{\mathrm{q}},t) = \mathcal{H}(\mathrm{q})\ddot{\mathrm{q}}_r + \mathcal{C}(\mathrm{q},\dot{\mathrm{q}})\dot{\mathrm{q}}_r + \mathcal{G}(\mathrm{q}) - \mathcal{K}(t)(\dot{\mathrm{q}} - \dot{\mathrm{q}}_r) \quad (18)$$

where $\mathrm{q},\dot{\mathrm{q}} \in \mathbb{R}^n$, $\dot{\mathrm{q}}_r = \dot{\mathrm{q}}_d(t) - \Lambda(t)(\mathrm{q} - \mathrm{q}_d(t))$, $\mathcal{H} : \mathbb{R}^n \mapsto \mathbb{R}^{n\times n}$, $\mathcal{C} : \mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}^{n\times n}$, $\mathcal{G} : \mathbb{R}^n \mapsto \mathbb{R}^n$, $\mathcal{K} : \mathbb{R}_{\geq 0} \mapsto \mathbb{R}^{n\times n}$, $\Lambda : \mathbb{R}_{\geq 0} \mapsto \mathbb{R}^{n\times n}$, *and* $(\mathrm{q}_d, \dot{\mathrm{q}}_d)$ *is the target trajectory of the state* $(\mathrm{q},\dot{\mathrm{q}})$. *Note that* $\mathcal{K}, \Lambda \succ 0$ *are control gain matrices (design parameters), and* $\dot{\mathcal{H}} - 2\mathcal{C}$ *is skew-symmetric with* $\mathcal{H} \succ 0$ *by construction.*

By comparing with (17) *and* (18)*, we define the following virtual observer-like system of* q *(not* q*) that has* $q = \dot{\mathrm{q}}$ *and* $q = \dot{\mathrm{q}}_r$ *as its particular solutions:*

$$\mathcal{H}(\mathrm{q})\dot{q} + \mathcal{C}(\mathrm{q},\dot{\mathrm{q}})q + \mathcal{K}(t)(q - \dot{\mathrm{q}}) + \mathcal{G}(\mathrm{q}) = u(\mathrm{q},\dot{\mathrm{q}},t) \quad (19)$$

which gives $\mathcal{H}(\mathrm{q})\delta\dot{q} + (\mathcal{C}(\mathrm{q},\dot{\mathrm{q}}) + \mathcal{K}(t))\delta q = 0$. *We thus have that*

$$\frac{d}{dt}(\delta q^\top \mathcal{H}(\mathrm{q})\delta q) = \delta q^\top(\dot{\mathcal{H}} - 2\mathcal{C} - 2\mathcal{K})\delta q = -2\delta q^\top \mathcal{K}\delta q \quad (20)$$

where the skew-symmetry of $\dot{\mathcal{H}} - 2\mathcal{C}$ *is used to obtain the second equality. Since* $\mathcal{K} \succ 0$*,* (20) *indicates that the virtual system* (19) *is partially contracting in* q *with* $\mathcal{H}$ *defining its contraction metric. Contraction of the full state* $(\mathrm{q},\dot{\mathrm{q}})$ *will be discussed in Example 2.7.*

Note that if we treat the arguments $(\mathrm{q},\dot{\mathrm{q}})$ *of* $\mathcal{H}$ *and* $\mathcal{C}$ *also as the virtual state* q*, we end up having additional terms such as* $\partial\mathcal{H}/\partial q_i$*, which makes proving contraction analytically more involved as in Example 2.5.*

As can be seen from Examples 2.5 and 2.6, the role of partial contraction in theorem 2.2 is to provide some insight on stability even for cases where it is difficult to prove contraction for all solution trajectories as in Theorem 2.1. Although finding a contraction metric analytically for general nonlinear systems is challenging, we will see in Sec. 3 and Sec. 4 that the convex nature of the contraction condition (4) helps us find it numerically.

2.2. Path-Length Integral and Robust Incremental Stability Analysis

Theorem 2.1 can also be proven by using the transformed squared length integrated over two arbitrary solutions of (1) [7, 13, 16, 17, 24], which enables formalizing its connection to incremental stability discussed in Definition 2.3. Note that the integral forms (21) and (22) to be given in Theorem 2.3 are useful for handling perturbed systems with external disturbances as shall be seen in Theorems 2.4–2.8.

Theorem 2.3. *Let* ξ_0 *and* ξ_1 *be the two arbitrary solutions of* (1)*, and define the transformed squared length with* $M(x,t)$ *of Theorem 2.1 as follows:*

$$V_{s\ell}(x,\delta x,t) = \int_{\xi_0}^{\xi_1} \|\delta z\|^2 = \int_0^1 \frac{\partial x}{\partial \mu}^\top M(x,t)\frac{\partial x}{\partial \mu}d\mu \quad (21)$$

where x *is a smooth path parameterized as* $x(\mu = 0,t) = \xi_0(t)$ *and* $x(\mu = 1,t) = \xi_1(t)$ *by* $\mu \in [0,1]$*. Also, define the path integral with the transformation* $\Theta(x,t)$ *for* $M(x,t) = \Theta(x,t)^\top\Theta(x,t)$ *as follows:*

$$V_\ell(x,\delta x,t) = \int_{\xi_0}^{\xi_1} \|\delta z\| = \int_{\xi_0}^{\xi_1} \|\Theta(x,t)\delta x\|. \quad (22)$$

Then (21) *and* (22) *are related as*

$$\|\xi_1 - \xi_0\| = \left\|\int_{\xi_0}^{\xi_1} \delta x\right\| \leq \frac{V_\ell}{\sqrt{\underline{m}}} \leq \sqrt{\frac{V_{s\ell}}{\underline{m}}} \quad (23)$$

where $M(x,t) \succeq \underline{m}\mathrm{I}$, $\forall x,t$ *for* $\exists \underline{m} \in \mathbb{R}_{>0}$*, and Theorem 2.1 can also be proven by using* (21) *and* (22) *as a Lyapunov-like function, resulting in incremental exponential stability of the system* (1) *(see Definition 2.2). Note that the shortest path integral* V_ℓ *of* (22) *with a parameterized state* x *(i.e.,* $\inf V_\ell = \sqrt{\inf V_{s\ell}}$*) defines the Riemannian distance and the path integral of a minimizing minimizing geodesic [79].*

Proof. Using $M(x,t) \succeq \underline{m}\mathrm{I}$ which gives $\sqrt{\underline{m}}\|\xi_1 - \xi_0\| \leq V_\ell$, we have $\|\xi_1 - \xi_0\| = \|\int_{\xi_0}^{\xi_1} \delta x\| \leq V_\ell/\sqrt{\underline{m}}$ of (23). The inequality $V_\ell \leq \sqrt{V_{s\ell}}$ of (23) can be proven by applying the Cauchy–Schwarz inequality [139, p. 316] to the functions $\psi_1(\mu) = \|\Theta(x,t)(\partial x/\partial\mu)\|$ and $\psi_2(\mu) = 1$

We can also see that computing $\dot{V}_{s\ell}$ of (21) using the differential dynamics of (2) yields

$$\dot{V}_{s\ell} = \int_0^1 \frac{\partial x}{\partial \mu}^\top \left(\dot{M} + M\frac{\partial f}{\partial x} + \frac{\partial f}{\partial x}^\top M\right)\frac{\partial x}{\partial \mu}d\mu$$

to have $\dot{V}_{s\ell} \leq -2\alpha V_{s\ell}$ and $\dot{V}_\ell \leq -\alpha V_\ell$ by the contraction conditions (3) and (4). Since these hold for any ξ_0 and ξ_1, the incremental exponential stability results in Theorem 2.1 follow from the comparison lemma of Lemma 2.1 (see also [13, 16, 17], and [79, 80] for the discussion on the geodesic). □

2.2.1. Deterministic Perturbation

Let $\xi_0(t)$ be a solution of the system (1). It is now perturbed as

$$\dot{x} = f(x,t) + d(x,t) \quad (24)$$

and let $\xi_1(t)$ denote a trajectory of (24). Then a virtual system of a smooth path $q(\mu,t)$ parameterized by $\mu \in [0,1]$, which has $q(\mu = 0,t) = \xi_0$ and $q(\mu = 1,t) = \xi_1$ as its particular solutions is given as follows:

$$\dot{q}(\mu,t) = f(q(\mu,t),t) + d_\mu(\mu,\xi_1,t) \quad (25)$$

where $d_\mu(\mu,\xi_1,t) = \mu d(\xi_1,t)$. Since contraction means exponential convergence, a contracting system exhibits a superior property of robustness [7, 24].

Theorem 2.4. *If the system* (1) *satisfies* (3) *and* (4) *of Theorem 2.1 (i.e., the system* (1) *is contracting), then the path integral* $V_\ell(q,\delta q,t)=\int_{\xi_0}^{\xi_1}\|\Theta(q,t)\delta q\|$ *of* (22)*, where* ξ_0 *is a solution of the contracting system* (1)*,* ξ_1 *is a solution of the perturbed system* (24)*, and* q *is the virtual state of* (25)*, exponentially converges to a bounded error ball as long as* $\Theta d\in\mathcal{L}_\infty$ *(i.e.,* $\sup_{x,t}\|\Theta d\|<\infty$*). Specifically, if* $\exists\underline{m},\overline{m}\in\mathbb{R}_{>0}$ *and* $\exists\bar{d}\in\mathbb{R}_{\geq 0}$ *s.t.* $\bar{d}=\sup_{x,t}\|d(x,t)\|$ *and*

$$\underline{m}\mathrm{I}\preceq M(x,t)\preceq\overline{m}\mathrm{I},\ \forall x,t \tag{26}$$

then we have the following relation:

$$\|\xi_1(t)-\xi_0(t)\|\ \leq\frac{V_\ell(0)}{\sqrt{\underline{m}}}e^{-\alpha t}+\frac{\bar{d}}{\alpha}\sqrt{\frac{\overline{m}}{\underline{m}}}(1-e^{-\alpha t}) \tag{27}$$

where $V_\ell(t)=V_\ell(q(t),\delta q(t),t)$ *for notational simplicity.*

Proof. Using the contraction condition (4), we have for $M=\Theta^\top\Theta$ given in Theorem 2.1 that

$$\begin{aligned}\frac{d}{dt}\|\Theta(q,t)\partial_\mu q\|=&(2\|\Theta(q,t)\partial_\mu q\|)^{-1}\frac{d}{dt}\partial_\mu q^\top M(q,t)\partial_\mu q\\ \leq&-\alpha\|\Theta(q,t)\partial_\mu q\|+\|\Theta(q,t)\partial_\mu d_\mu\|\end{aligned}$$

where $\partial_\mu q=\partial q/\partial\mu$ and $\partial_\mu d_\mu=\partial d_\mu/\partial\mu=d(\xi_1,t)$. Taking the integral with respect to μ gives

$$\frac{d}{dt}\int_0^1\|\Theta\partial_\mu q\|d\mu\leq\int_0^1-\alpha\|\Theta\partial_\mu q\|+\|\Theta d(\xi_1,t)\|d\mu$$

which implies $\dot{V}_\ell(t)\leq-\alpha V_\ell(t)+\sup_{q,\xi_1,t}\|\Theta(q,t)d(\xi_1,t)\|$ for V_ℓ in (22) of Theorem 2.3. Thus, applying the comparison lemma (see Lemma 2.1) results in

$$V_\ell(t)\leq e^{-\alpha t}V_\ell(0)+\sup_{q,\xi_1,t}\|\Theta(q,t)d(\xi_1,t)\|\frac{1-e^{-\alpha t}}{\alpha}. \tag{28}$$

By using $\xi_1-\xi_0=\int_{\xi_0}^{\xi_1}\delta x$ and $\|\xi_1-\xi_0\|=\|\int_{\xi_0}^{\xi_1}\delta x\|\leq\int_{\xi_0}^{\xi_1}\|\delta x\|\leq\int_{\xi_0}^{\xi_1}\|\Theta^{-1}\|\|\delta z\|$, we obtain $\sqrt{\inf_t\lambda_{\min}(M)}\|\xi_1-\xi_0\|\leq V_\ell=\int_{\xi_0}^{\xi_1}\|\delta z\|$, and thus

$$\|\xi_1-\xi_0\|\ \leq\frac{e^{-\alpha t}V_\ell(0)}{\sqrt{\inf_t\lambda_{\min}(M)}}+\frac{\sup_{q,\xi_1,t}\|\Theta d\|}{\sqrt{\inf_t\lambda_{\min}(M)}}\frac{1-e^{-\alpha t}}{\alpha}.$$

This relation with the bounds on M and d gives (27). □

2.2.2. Stochastic Perturbation

Next, consider the following dynamical system modeled by the Itô stochastic differential equation:

$$dx=f(x,t)dt+G(x,t)d\mathscr{W}(t) \tag{29}$$

where $G:\mathbb{R}^n\times\mathbb{R}_{\geq 0}\to\mathbb{R}^{n\times w}$ is a matrix-valued function and $\mathscr{W}:\mathbb{R}_{\geq 0}\mapsto\mathbb{R}^w$ is a w-dimensional Wiener process [140, p. 100] (see also [140, p. xii] for the notations used). For the sake of existence and uniqueness of the solution, we assume in (29) that

$$\begin{aligned}\exists L\in\mathbb{R}_{\geq 0}\text{ s.t. }&\|f(x,t)-f(x',t)\|+\|G(x,t)-G(x',t)\|_F\\ &\leq L\|x-x'\|,\ \forall t\in\mathbb{R}_{\geq 0},\ \forall x,x'\in\mathbb{R}^n\end{aligned} \tag{30}$$

$$\begin{aligned}\exists\bar{L}\in\mathbb{R}_{\geq 0}\text{ s.t. }&\|f(x,t)\|^2+\|G(x,t)\|_F^2\\ &\leq\bar{L}(1+\|x\|^2),\ \forall t\in\mathbb{R}_{\geq 0},\ \forall x\in\mathbb{R}^n.\end{aligned} \tag{31}$$

In order to analyze the incremental stability property of (29) as in Theorem 2.3, we consider two trajectories $\xi_0(t)$ and $\xi_1(t)$ of stochastic nonlinear systems with Gaussian white noise, driven by two independent Wiener processes $\mathscr{W}_0(t)$ and $\mathscr{W}_1(t)$:

$$d\xi_i=f(\xi_i,t)dt+G_i(\xi_i,t)d\mathscr{W}_i(t),\ i=0,1. \tag{32}$$

One can show that (29) has a unique solution $x(t)$ which is continuous with probability one under the conditions (30) and (31) (see [140, p. 105] and [16, 34]), leading to the following lemma as in the comparison lemma of Lemma 2.1.

Lemma 2.2. *Suppose that* $V_{s\ell}$ *of* (21) *satisfies the following inequality:*

$$\mathscr{L}V_{s\ell}\leq-\gamma V_{s\ell}+c \tag{33}$$

where $\gamma\in\mathbb{R}_{>0}$*,* $c\in\mathbb{R}_{\geq 0}$*, and* $\mathscr{L}$ *denotes the infinitesimal differential generator of the Itô process given in [141, p. 15]. Then we have the following bound [34]:*

$$\mathbb{E}\left[\|\xi_1(t)-\xi_0(t)\|^2\right]\leq\frac{1}{\underline{m}}\left(\mathbb{E}[V_{s\ell}(0)]e^{-\gamma t}+\frac{c}{\gamma}\right) \tag{34}$$

where $V_{s\ell}(0)=V_{s\ell}(x(0),\delta x(0),0)$ *for* $V_{s\ell}$ *in* (21)*,* $\underline{m}$ *is given in* (26)*,* ξ_0 *and* ξ_1 *are given in* (32)*, and* $\mathbb{E}$ *denotes the expected value operator. Furthermore, the probability that* $\|\xi_1-\xi_0\|$ *is greater than or equal to* $\varepsilon\in\mathbb{R}_{>0}$ *is given as*

$$\mathbb{P}\left[\|\xi_1(t)-\xi_0(t)\|\geq\varepsilon\right]\leq\frac{1}{\varepsilon^2\underline{m}}\left(\mathbb{E}[V_{s\ell}(0)]e^{-\gamma t}+\frac{c}{\gamma}\right). \tag{35}$$

Proof. The bound (34) follows from Theorem 2 of [34] (see also [13, 16, 17, 142, 143] and [141, p. 10] (Dynkin's formula)). The probability tracking error bound (35) then follows from Markov's inequality [144, pp. 311-312]. □

Remark 2.3. *Although Lemma 2.2 considers the second moment of* $\|\xi_1(t)-\xi_0(t)\|$*, i.e.,* $\mathbb{E}[\|\xi_1(t)-\xi_0(t)\|^2]$*, it can be readily generalized to the p-th moment of* $\|\xi_1(t)-\xi_0(t)\|$*, i.e.,* $\mathbb{E}[\|\xi_1(t)-\xi_0(t)\|^p]$*, applying the Lyapunov-based technique proposed in [145].*

A virtual system of a smooth path $q(\mu,t)$ parameterized by $\mu\in[0,1]$, which has $q(\mu=0,t)=\xi_0$ and $q(\mu=1,t)=\xi_1$ of (32) as its particular solutions, is given as follows:

$$dq(\mu,t)=f(q(\mu,t),t)dt+G(\mu,\xi_0,\xi_1,t)d\mathscr{W}(t) \tag{36}$$

where $G(\mu,\xi_0,\xi_1,t) = [(1-\mu)G_0(\xi_0,t), \mu G_1(\xi_1,t)]$ and $\mathscr{W} = [\mathscr{W}_0^\top, \mathscr{W}_1^\top]^\top$. As a consequence of Lemma 2.2, showing stochastic incremental stability between ξ_0 and ξ_1 of (32) reduces to proving the relation (33), similar to the deterministic case in Theorems 2.1 and 2.4.

Theorem 2.5. *Suppose that $\exists \bar{g}_0 \in \mathbb{R}_{\geq 0}$ and $\exists \bar{g}_1 \in \mathbb{R}_{\geq 0}$ s.t. $\sup_{x,t} \|G_1(x,t)\|_F = \bar{g}_0$ and $\sup_{x,t} \|G_1(x,t)\|_F = \bar{g}_1$ in (32). Suppose also that there exists $M(x,t) \succ 0$, $\forall x,t$, s.t. $M_{x_i} = \partial M/\partial x_i$ is Lipschitz with respect to x for all $i = 1,\cdots,n$, i.e., $\exists L_m \in \mathbb{R}_{\geq 0}$ s.t.*

$$\|M_{x_i}(x,t) - M_{x_i}(x',t)\| \leq L_m \|x - x'\|, \ \forall x, x', t, i. \tag{37}$$

Also, suppose that M of (37) satisfies (26) and (4) with its right-hand side replaced by $-2\alpha M - \alpha_s \mathrm{I}$ for $\alpha_s = L_m(\bar{g}_0^2 + \bar{g}_1^2)(\alpha_G + 1/2)$, i.e.,

$$\dot{M} + M\frac{\partial f}{\partial x} + \frac{\partial f}{\partial x}^\top M \preceq -2\alpha M - \alpha_s \mathrm{I} \tag{38}$$

where $\alpha_G \in \mathbb{R}_{>0}$ is an arbitrary constant (see (42)). Then, the following error bound of incremental stability holds:

$$\mathbb{E}\left[\|\xi_1(t) - \xi_0(t)\|^2\right] \leq \frac{\mathbb{E}[V_{s\ell}(0)]}{\underline{m}} e^{-2\alpha t} + \frac{C}{2\alpha}\frac{\overline{m}}{\underline{m}} \tag{39}$$

where ξ_0 and ξ_1 are the trajectories given in (32), $V_{s\ell}(t) = V_{s\ell}(q(t),\delta q(t),t) = \int_{\xi_0}^{\xi_1} \delta q^\top M(q,t)\delta q$ is given in (21) with the virtual state q of (36), $\underline{m}$ and $\overline{m}$ are given in (26), $C = (\bar{g}_0^2 + \bar{g}_1^2)(2{\alpha_G}^{-1} + 1)$, and $\mathbb{E}$ denotes the expected value operator. Furthermore, the probability that $\|\xi_1 - \xi_0\|$ is greater than or equal to $\varepsilon \in \mathbb{R}_{>0}$ is given as

$$\mathbb{P}\left[\|\xi_1(t) - \xi_0(t)\| \geq \varepsilon\right] \leq \frac{1}{\varepsilon^2}\left(\frac{\mathbb{E}[V_{s\ell}(0)]}{\underline{m}} e^{-2\alpha t} + \frac{C}{2\alpha}\frac{\overline{m}}{\underline{m}}\right). \tag{40}$$

Proof. By definition of the infinitesimal differential generator given in Lemma 2.2 [141, p. 15], we have [13, 17]

$$\begin{aligned}\mathscr{L}V_{s\ell} = &\int_0^1 V_t + \sum_{i=1}^n \left(V_{q_i} f_i + V_{\partial_\mu q_i}\left(\frac{\partial f}{\partial q}\partial_\mu q\right)_i\right) \\ &+ \frac{1}{2}\sum_{i,j=1}^n \Big(V_{q_i q_j}(GG^\top)_{ij} + 2V_{q_i \partial_\mu q_j}(G\partial_\mu G^\top)_{ij} \\ &+ V_{\partial_\mu q_i \partial_\mu q_j}(\partial_\mu G \partial_\mu G^\top)_{ij}\Big) d\mu\end{aligned} \tag{41}$$

where $V = \partial_\mu q^\top M(q,t)\partial_\mu q$, $\partial_\mu q = \partial q/\partial\mu$, $\partial_\mu G = \partial G/\partial\mu$, $V_p = \partial V/\partial p$, and $V_{p_1 p_2} = \partial^2 V/(\partial p_1 \partial p_2)$.

Since M_{x_i} is Lipschitz as in (37), we have $\|M_{x_i x_j}\| \leq L_m$ and $\|M_{x_i}\| \leq \sqrt{2L_m\overline{m}}$ using (26) as derived [15]. Computing $\mathscr{L}V_{s\ell}$ of (41) using these bounds, the bounds of $\|G_0\|_F$ and $\|G_1\|_F$, and $\partial_\mu G = \partial G/\partial\mu = [-G_0, G_1]$ as in [13, 17] yields

$$\begin{aligned}\mathscr{L}V_{s\ell} &\leq \int_0^1 \partial_\mu q^\top(\dot{M} + 2\,\mathrm{sym}(Mf_x))\partial_\mu q d\mu \\ &\quad + (\bar{g}_0^2 + \bar{g}_1^2)(L_m\|\partial_\mu q\|^2/2 + 2\sqrt{2L_m\overline{m}}\|\partial_\mu q\| + \overline{m}) \\ &\leq \int_0^1 \partial_\mu q^\top(\dot{M} + 2\,\mathrm{sym}(Mf_x) + \alpha_s \mathrm{I})\partial_\mu q d\mu + C\overline{m}\end{aligned} \tag{42}$$

where $\alpha_s = L_m(\bar{g}_0^2 + \bar{g}_1^2)(\alpha_G + 1/2)$, $C = (\bar{g}_0^2 + \bar{g}_1^2)(2{\alpha_G}^{-1} + 1)$, and the relation $2ab \leq \alpha_G^{-1}a^2 + \alpha_G b^2$, which holds for any $a, b \in \mathbb{R}$ and $\alpha_G \in \mathbb{R}_{>0}$, is used with $a = \sqrt{2\overline{m}}$ and $b = \sqrt{L_m}\|\partial_\mu x\|$ to get the second inequality. This reduces to $\mathscr{L}V_{s\ell} \leq -2\alpha V_{s\ell} + \underline{m}C$ under the condition (38), resulting in (39) and (40) as a result of (34) and (35) in Lemma 2.2. □

Remark 2.4. *Although we consider the Gaussian white noise stochastic differential equation (32) when referring to stochastic systems in this paper, other types of stochastic noises, including compound Poisson shot noise and bounded-measure Lévy noise, could be considered as in Theorem 2.5 using contraction theory [35].*

2.3. Finite-Gain $\mathcal{L}_p$ Stability and Contraction of Hierarchically Combined Systems

Due to the simpler stability analysis of Theorem 2.1 when compared with Lyapunov theory, Input-to-State Stability (ISS) and input-output stability in the sense of finite-gain $\mathcal{L}_p$ stability can be easily studied using contraction theory [24].

Theorem 2.6. *If (24) is perturbed by $d(x,t) \in \mathcal{L}_{pe}$ (i.e., $\|(d)_\tau\|_{\mathcal{L}_p} < \infty$ for $\tau \in \mathbb{R}_{\geq 0}$ and $p \in [1,\infty]$, see Sec. 1.3) and Theorem 2.1 holds, then (24) is finite-gain $\mathcal{L}_p$ stable with $p \in [1,\infty]$ for an output $y = h(x,d,t)$ with $\int_{Y_0}^{Y_1}\|\delta y\| \leq \eta_0 \int_{\xi_0}^{\xi_1}\|\delta x\| + \eta_1\|d\|$, $\exists \eta_0, \eta_1 \geq 0$, i.e., $\forall \tau \in \mathbb{R}_{\geq 0}$ [24]*

$$\left\|\left(\int_{Y_0}^{Y_1}\|\delta y\|\right)_\tau\right\|_{\mathcal{L}_p} \leq \left(\frac{\eta_0}{\alpha} + \eta_1\right)\frac{\|(\Theta d)_\tau\|_{\mathcal{L}_p}}{\sqrt{\underline{m}}} + \frac{\eta_0\zeta V_\ell(0)}{\sqrt{\underline{m}}} \tag{43}$$

where Y_0 and Y_1 denote the output trajectories of the original contracting system (1) and its perturbed system (24), respectively, $\underline{m}$ is defined as $M(x,t) \succeq \underline{m}\mathrm{I}$, $\forall x,t$, as in (26), and $V_\ell(0) = V_\ell(x(0),\delta x(0),0)$ for V_ℓ in (22). Also, $\zeta = 1$ if $p = \infty$ and $\zeta = 1/(\alpha p)^{1/p}$ if $p \in [1,\infty)$. The perturbed system (24) also exhibits ISS.

Proof. In keeping with Theorem 5.1 of [3], (28) also implies the following relation for $M = \Theta^\top\Theta$:

$$\begin{aligned}\|(V_\ell)_\tau\|_{\mathcal{L}_p} &\leq V_\ell(0)\|(e^{-\alpha t})_\tau\|_{\mathcal{L}_p} + \|e^{-\alpha t}\|_{\mathcal{L}_1}\|(\Theta d)_\tau\|_{\mathcal{L}_p} \\ &\leq V_\ell(0)\zeta + \|(\Theta d)_\tau\|_{\mathcal{L}_p}/\alpha.\end{aligned} \tag{44}$$

Since we have $\|\Theta^{-1}\| \leq 1/\sqrt{\underline{m}}$ for $M = \Theta^\top\Theta$, (43) can be obtained by using both (44) and the known bound of $\|\delta y\|$, thereby yielding a finite $\mathcal{L}_p$ gain independently of τ. ISS can be guaranteed by Lemma 4.6 of [3, p. 176], which states that exponential stability of an unperturbed system results in ISS. □

Theorems 2.1–2.6 are also applicable to the hierarchically combined system of two contracting dynamics due to Theorem 2.7 [9, 10].

Theorem 2.7. *Consider the following hierarchically combined system of two contracting dynamics:*

$$\frac{d}{dt}\begin{bmatrix}\delta z_0\\ \delta z_1\end{bmatrix} = \begin{bmatrix}F_{00} & 0\\ F_{10} & F_{11}\end{bmatrix}\begin{bmatrix}\delta z_0\\ \delta z_1\end{bmatrix} \tag{45}$$

where $F = F_{00}$ and $F = F_{11}$ both satisfy the contraction condition (3)*, and suppose that it is subject to perturbation $[d_0, d_1]^\top$. Then the path length integral $V_{\ell,i}(t) = \int_{\xi_0}^{\xi_1} \|\delta z_i\|$ with $i = 0, 1$ between the original and perturbed dynamics trajectories, ξ_0 and ξ_1, respectively, verifies [7]*

$$\dot{V}_{\ell,0} + \alpha_0 V_{\ell,0} \leq \|\Theta_0 d_0\| \tag{46}$$

$$\dot{V}_{\ell,1} + \alpha_1 V_{\ell,1} \leq \|\Theta_1 d_1\| + \int_{\xi_0}^{\xi_1} \|F_{10}\|\|\delta z_1\| \tag{47}$$

where $\alpha_i = \sup_{x,t} |\lambda_{\max}(F_{ii})|$, $i = 0, 1$. Hence, the error bounds of $V_{\ell,0}(t)$ and $V_{\ell,1}(t)$ can be obtained using Theorem 2.4 if $\|F_{10}\|$ is bounded. In particular, if $\|\Theta_i d_i\| \leq \sqrt{\overline{m}_i}\bar{d}_i$, $i = 0, 1$, the relations (46) *and* (47) *yield $V_{\ell,0}(t) \leq V_{\ell,0}(0)e^{-\alpha_0 t} + (\sqrt{\overline{m}_0}\bar{d}_0/\alpha_0)(1 - e^{-\alpha_0 t})$ as in* (27)*, and*

$$V_{\ell,1}(t) \leq V_{\ell,1}(0)e^{-\alpha_1 t} + \frac{\sqrt{\overline{m}_1}\bar{d}_1 + \bar{f}_{10}\bar{V}_0}{\alpha_1}(1 - e^{-\alpha_1 t}) \tag{48}$$

where $\bar{f}_{10} = \sup_{t\in\mathbb{R}_{\geq 0}} \|F_{10}\|$ and $\bar{V}_0 = V_{\ell,0}(0) + \sqrt{\overline{m}_0}\bar{d}_0/\alpha_0$.

Also, similar to Theorem 2.6, we have $\|(V_{\ell,0})_\tau\|_{\mathcal{L}_p} \leq V_{\ell,0}(0)\zeta_0 + \|(\Theta_0 d_0)_\tau\|_{\mathcal{L}_p}/\alpha_0$, and thus a hierarchical connection for finite-gain $\mathcal{L}_p$ stability can be established as follows [24]:

$$\|(V_{\ell,1})_\tau\|_{\mathcal{L}_p} \leq V_{\ell,1}(0)\zeta_1 + \frac{\|(\Theta_1 d_1)_\tau\|_{\mathcal{L}_p} + \bar{f}_{10}\|(V_{\ell,0})_\tau\|_{\mathcal{L}_p}}{\alpha_1} \tag{49}$$

where $\zeta_i = 1$ if $p = \infty$ and $\zeta_i = 1/(\alpha_i p)^{1/p}$ if $p \in [1, \infty)$ for $i = 0, 1$. By recursion, this result can be extended to an arbitrary number of hierarchically combined groups.

Proof. Applying the comparison lemma of Lemma 2.1 to (46) and (47), which follow from the differenital dynamics (45) with the condition (3), we get (48) as in Theorem 2.4. Similarly, (49) follows from obtaining $\|(V_{\ell,0})_\tau\|_{\mathcal{L}_p}$ by (44) using (46), and then recursively obtaining $\|(V_{\ell,1})_\tau\|_{\mathcal{L}_p}$ using (47) as in Theorem 2.6 [24]. □

Example 2.7. *As demonstrated in Example 2.6, the Lagrangian virtual system* (19) *is contracting with respect to q, having $q = \dot{\mathtt{q}}$ and $q = \dot{\mathtt{q}}_r = \dot{\mathtt{q}}_d - \Lambda(\mathtt{q} - \mathtt{q}_d)$ as its particular solutions. Let $q_0 = q$ for such q. The virtual system of q_1 which has $[q_0^\top, q_1^\top]^\top = [\dot{\mathtt{q}}^\top, \mathtt{q}^\top]^\top$ and $[\dot{\mathtt{q}}_r^\top, \mathtt{q}_d^\top]^\top$ as its particular solutions is given as*

$$\dot{q}_1 = q_0 - \Lambda(q_1 - \mathtt{q})$$

resulting in $\delta\dot{q}_1 = \mathtt{I}\delta q_0 - \Lambda\delta q_1$. Since the virtual system of q_0 is contracting as in Example 2.6 and the virtual system $\delta\dot{q}_1 = -\Lambda\delta q_1$ is contracting in q_1 due to $\Lambda \succ 0$, (48) *of Theorem 2.7 implies that the whole system* (17) *for $[q_0^\top, q_1^\top]^\top$ is hierarchically contracting and robust against perturbation in the sense of Theorem 2.7 ($F_{10} = \mathtt{I}$ in this case). Also, see [24, 40] for the hierarchical multi-timescale separation of tracking and synchronization control for multiple Lagrangian systems.*

2.4. Contraction Theory for Discrete-time Systems

Let us consider the following nonlinear system with bounded deterministic perturbation $d_k : \mathbb{R}^n \times \mathbb{N} \mapsto \mathbb{R}^n$ with $\bar{d} \in \mathbb{R}_{\geq 0}$ s.t. $\bar{d} = \sup_{x,k} \|d_k(x, k)\|$:

$$x(k+1) = f_k(x(k), k) + d_k(x(k), k) \tag{50}$$

where $k \in \mathbb{N}$, $x : \mathbb{N} \mapsto \mathbb{R}^n$ is the discrete system state, and $f_k : \mathbb{R}^n \times \mathbb{N} \mapsto \mathbb{R}^n$ is a smooth function. Although this tutorial paper focuses mainly on continuous-time nonlinear systems, let us briefly discuss contraction theory for (50) to imply that the techniques in the subsequent sections are applicable also to discrete-time nonlinear systems.

Let $\xi_0(k)$ and $\xi_1(k)$ be solution trajectories of (50) with $d_k = 0$ and $d_k \neq 0$, respectively. Then a virtual system of $q(\mu, k)$ parameterized by $\mu \in [0, 1]$, which has $q(\mu = 0, k) = \xi_0(k)$ and $q_k(\mu = 1, k) = \xi_1(k)$ as its particular solutions, can be expressed as follows:

$$q(\mu, k+1) = f_k(q(\mu, k), k) + \mu d_k(\xi_1(k), k). \tag{51}$$

The discrete version of robust contraction in Theorem 2.4 is given in the following theorem.

Theorem 2.8. *Let $x(k) = x_k$ and $q(\mu, k) = q_k$ for any $k \in \mathbb{N}$ for simplicity. If there exists a uniformly positive definite matrix $M_k(x_k, k) = \Theta_k(x_k, k)^\top\Theta_k(x_k, k) \succ 0$, $\forall x_k, k$, where Θ_k defines a smooth coordinate transformation of δx_k, i.e., $\delta z_k = \Theta_k(x_k, k)\delta x_k$, s.t. either of the following equivalent conditions holds for $\exists\alpha \in (0, 1)$, $\forall x_k, k$:*

$$\left\|\Theta_{k+1}(x_{k+1}, k+1)\frac{\partial f_k}{\partial x_k}\Theta_k(x_k, k)^{-1}\right\| \preceq \alpha \tag{52}$$

$$\frac{\partial f_k}{\partial x_k}^\top M_{k+1}(x_{k+1}, k+1)\frac{\partial f_k}{\partial x_k} \preceq \alpha^2 M_k(x_k, k) \tag{53}$$

then we have the following bound as long as we have $\underline{m}\mathrm{I} \preceq M_x(x_k, k) \preceq \overline{m}\mathrm{I}$, $\forall x_k, k$, as in (26)*:*

$$\|\xi_1(k) - \xi_0(k)\| \leq \frac{V_\ell(0)}{\sqrt{\underline{m}}}\alpha^k + \frac{\bar{d}(1 - \alpha^k)}{1 - \alpha}\sqrt{\frac{\overline{m}}{\underline{m}}} \tag{54}$$

where $V_\ell(k) = \int_{\xi_0}^{\xi_1} \|\Theta_k(q_k,k)\delta q_k\|$ *as in* (22) *for the unperturbed trajectory* ξ_0*, perturbed trajectory* ξ_1*, and virtual state* $q_k = q(k)$ *given in* (51).

PROOF. If (52) or (53) holds, we have that

$$V_\ell(k+1) \le \int_0^1 \|\Theta_{k+1}(\partial_{q_k} f_k(q_k,k)\partial_\mu q_k + d_k(x_k,k))\|d\mu$$
$$\le \alpha \int_0^1 \|\Theta_k(q_k,k)\partial_\mu q_k\|d\mu + \bar{d}\sqrt{m} = \alpha V_\ell(k) + \bar{d}\sqrt{m}$$

where $\Theta_{k+1} = \Theta_{k+1}(q_{k+1}, k+1)$, $\partial_{q_k} f_k(q_k,k) = \partial f_k/\partial q_k$, and $\partial_\mu q_k = \partial q_k/\partial\mu$. Applying this inequality iteratively results in (54).

Theorem 2.8 can be used with Theorem 2.4 for stability analysis of hybrid nonlinear systems [8, 37, 42], or with Theorem 2.5 for stability analysis of discrete-time stochastic nonlinear systems [13, 17, 42]. For example, it is shown in [13] that if the time interval in discretizing (1) as (50) is sufficiently small, contraction of discrete-time systems with stochastic perturbation reduces to that of continuous-time systems.

We finally remark that the steady-state upper bounds of (27) in Theorem 2.4, (39) in Theorem 2.5, and (54) in Theorem 2.8 are all functions of $\overline{m}/\underline{m}$. This property is to be used extensively in Sec. 4 for designing a convex optimization-based control and estimation synthesis algorithm via contraction theory.

3. Robust Nonlinear Control and Estimation via Contraction Theory

As shown in Theorem 2.4 for deterministic disturbance and in Theorem 2.5 for stochastic disturbance, contraction theory provides explicit bounds on the distance of any couple of perturbed system trajectories. This property is useful in designing robust and optimal feedback controllers for a nonlinear system such as $\mathcal{H}_\infty$ control [25–33, 146, 147], which attempts to minimize the system $\mathcal{L}_2$ gain for optimal disturbance attenuation.

Since most of such feedback control and estimation schemes are based on the assumption that we know a Lyapunov function candidate, this section delineates one approach to solve a nonlinear optimal feedback control problem via contraction theory [13, 17], thereby proposing one explicit way to construct a Lyapunov function and contraction metric for general nonlinear systems for the sake of robustness. This approach is also utilizable for optimal state estimation problems as shall be seen in Sec. 4.

We consider the following smooth nonlinear system, perturbed by bounded deterministic disturbances $d_c(x,t)$ with $\sup_{x,t}\|d_c(x,t)\| = \bar{d}_c \in \mathbb{R}_{\ge 0}$ or by Gaussian white noise $\mathscr{W}(t)$ with $\sup_{x,t}\|G_c(x,t)\|_F = \bar{g}_c \in \mathbb{R}_{\ge 0}$:

$$\dot{x} = f(x,t) + B(x,t)u + d_c(x,t) \tag{55}$$
$$dx = (f(x,t) + B(x,t)u)dt + G_c(x,t)d\mathscr{W}(t) \tag{56}$$
$$\dot{x}_d = f(x_d,t) + B(x_d,t)u_d \tag{57}$$

where $x : \mathbb{R}_{\ge 0} \mapsto \mathbb{R}^n$ is the system state, $u \in \mathbb{R}^m$ is the system control input, $f : \mathbb{R}^n \times \mathbb{R}_{\ge 0} \mapsto \mathbb{R}^n$ and $B : \mathbb{R}^n \times \mathbb{R}_{\ge 0} \mapsto \mathbb{R}^{n\times m}$ are known smooth functions, $d_c : \mathbb{R}^n \times \mathbb{R}_{\ge 0} \mapsto \mathbb{R}^n$ and $G_c : \mathbb{R}^n \times \mathbb{R}_{\ge 0} \mapsto \mathbb{R}^{n\times w}$ are unknown bounded functions for external disturbances, and $\mathscr{W} : \mathbb{R}_{\ge 0} \mapsto \mathbb{R}^w$ is a w-dimensional Wiener process. Also, for (57), $x_d : \mathbb{R}_{\ge 0} \mapsto \mathbb{R}^n$ and $u_d : \mathbb{R}_{\ge 0} \mapsto \mathbb{R}^m$ denote the desired target state and control input trajectories, respectively.

Remark 3.1. *We consider control-affine nonlinear systems* (55)–(57) *in Sec. 3, 4, and 6–8. This is primarily because the controller design techniques for control-affine nonlinear systems are less complicated than those for control non-affine systems (which often result in* u *given implicitly by* $u = k(x,u,t)$ *[82, 83]), but still utilizable even for the latter, e.g., by treating* $\dot{u}$ *as another control input (see Example 3.1), or by solving the implicit equation* $u = k(x,u,t)$ *iteratively with a discrete-time controller (see Example 3.2 and Remark 3.3).*

Example 3.1. *By using* $\dot{u}$ *instead of* u *in* (55) *and* (56)*, a control non-affine system,* $\dot{x} = f(x,u,t)$*, can be rewritten as*

$$\frac{d}{dt}\begin{bmatrix} x \\ u \end{bmatrix} = \begin{bmatrix} f(x,u,t) \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \mathrm{I} \end{bmatrix} \dot{u}$$

which can be viewed as a control-affine nonlinear system with the state $[x^\top, u^\top]^\top$ *and control* $\dot{u}$*.*

Example 3.2. *One drawback of the technique in Example 3.1 is that we have to control* $\dot{u}$ *instead of* u*, which could be difficult in practice. In this case, we can utilize the following control non-affine nonlinear system decomposed into control-affine and non-affine parts:*

$$\dot{x} = f(x,u,t) = f_a(x,t) + B_a(x,t)u + r(x,u,t)$$

where $r(x,u,t) = f(x,u,t) - f_a(x,t) - B_a(x,t)u$*. The controller* u *can now be designed implicitly as*

$$B_a(x,t)u = B_a(x,t)u^* - r(x,u,t) \tag{58}$$

where u^* *is a stabilizing controller for the control-affine system* $\dot{x} = f_a(x,t) + B_a(x,t)u^*$*. Since solving such an implicit equation in* (58) *in real-time could be unrealistic in practice, we will derive a learning-based approach to solve it iteratively for unknown* $r(x,u,t)$*, without deteriorating its stability performance (see Lemma 9.1 and Theorem 9.2 of Sec. 9).*

3.1. Overview of Nonlinear Control and Estimation

We briefly summarize the advantages and disadvantages of existing nonlinear feedback control and state estimation schemes, so that one can identify which strategy is appropriate for their study and refer to the relevant parts of this tutorial paper.

Table 3: Comparison between SDC and CCM Formulation (note that $\gamma(\mu=0,t)=x_d$ and $\gamma(\mu=1,t)=x$).

	SDC formulation (Theorem 4.2) [13–17]	CCM formulation (Theorem 4.6) [79, 80]
Control law	$u=u_d-K(x,x_d,u_d,t)(x-x_d)$ or $u_d-K(x,t)(x-x_d)$	$u=u_d+\int_0^1 k(\gamma(\mu,t),\partial_\mu\gamma(\mu,t),u,t)d\mu$
Computation	Evaluates K for given (x,x_d,u_d,t) as in LTV systems	Computes geodesics γ for given (x,x_d,t) and integrates k
Generality	Captures nonlinearity by (multiple) SDC matrices	Handles general differential dynamics
Contraction	Depends on (x,x_d,u_d,t) or (x,t) (partial contraction)	Depends on (x,t) (contraction)

3.1.1. Systems with Known Lyapunov Functions

As discussed in Sec. 1.2, there are several nonlinear systems equipped with a known contraction metric/Lyapunov function, such as Lagrangian systems [5, p. 392], whose inertia matrix $\mathcal{H}(\mathtt{q})$ defines its contraction metric (see Example 2.6), or the nonlinear SLAM problem [16, 47] with virtual synthetic measurements, which can be reduced to an LTV estimation problem [47]. Once we have a contraction metric/Lyapunov function, stabilizing control and estimation laws can be easily derived by using, e.g., [3, 70, 71]. Thus, those dealing primarily with such nonlinear systems should skip this section and proceed to Part II of this paper (Sec. 5–9) on learning-based and data-driven control using contraction theory. Note that these known contraction metrics are not necessarily optimal, and the techniques to be derived in Sec. 3 and Sec. 4 are for obtaining contraction metrics with an optimal disturbance attenuation property [13, 17].

3.1.2. Linearization of Nonlinear Systems

If a contraction metric of a given nonlinear system is unknown, we could linearize it to apply methodologies inspired by LTV systems theory such as $\mathcal{H}_\infty$ control [27–32], iterative Linear Quadratic Regulator (iLQR) [148, 149], or Extended Kalman Filter (EKF). Their stability is typically analyzed by decomposing $f(x,t)$ as $f(x,t)=Ax+(f(x,t)-Ax)$ assuming that the nonlinear part $f(x,t)-Ax$ is bounded, or by finding a local contraction region for the sake of local exponential stability as in [14, 45]. Since the decomposition $f(x,t)=Ax+(f(x,t)-Ax)$ allows applying the result of Theorem 2.4, we could exploit the techniques in Sec. 3 and Sec. 4 for providing formal robustness and optimality guarantees for the LTV systems-type approaches. For systems whose nonlinear part $f(x,t)-Ax$ is not necessarily bounded, Sec. 9.1 elucidates how contraction theory can be used to stabilize them with the learned residual dynamics for control synthesis.

3.1.3. State-Dependent Coefficient (SDC) Formulation

It is shown in [13–17] that the SDC-based control and estimation [76–78, 150], which capture nonlinearity using a state-dependent matrix $A(x,t)$ s.t. $f(x,t)=A(x,t)x$ (e.g., we have $A(x,t)=\cos x$ for $f(x,t)=x\cos x$), result in exponential boundedness of system trajectories both for deterministic and stochastic systems due to Theorems 2.4 and 2.5 [14]. Because of the extended linear form of SDC (see Table 3), the results to be presented in Sec. 3–4 based on the SDC formulation are applicable to linearized dynamics that can be viewed as an LTV system with some modifications (see Remark 3.2).

This idea is slightly generalized in [15] to explicitly consider incremental stability with respect to a target trajectory (e.g., x_d for control and x for estimation) instead of using $A(x,t)x=f(x,t)$. Let us derive the following lemma for this purpose [13, 15–17, 78].

Lemma 3.1. *Let $f:\mathbb{R}^n\times\mathbb{R}_{\geq 0}\mapsto\mathbb{R}^n$ and $B:\mathbb{R}^n\times\mathbb{R}_{\geq 0}\mapsto\mathbb{R}^{n\times m}$ be piecewise continuously differentiable functions. Then there exists a matrix-valued function $A:\mathbb{R}^n\times\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}_{\geq 0}\mapsto\mathbb{R}^{n\times n}$ s.t., $\forall s\in\mathbb{R}^n$, $\bar{s}\in\mathbb{R}^n$, $\bar{u}\in\mathbb{R}^m$, and $t\in\mathbb{R}_{\geq 0}$,*

$$A(s,\bar{s},\bar{u},t)\mathtt{e}=f(s,t)+B(s,t)\bar{u}-f(\bar{s},t)-B(\bar{s},t)\bar{u}$$

where $\mathtt{e}=s-\bar{s}$, and one such A is given as follows:

$$A(s,\bar{s},\bar{u},t)=\int_0^1\frac{\partial\bar{f}}{\partial s}(cs+(1-c)\bar{s},\bar{u},t)dc \tag{59}$$

where $\bar{f}(s,\bar{u},t)=f(s,t)+B(s,t)\bar{u}$. We call A an SDC matrix if it is constructed to satisfy the controllability (or observability for estimation) condition. Also, the choice of A is not unique for $n\geq 2$, where n is the number of states, and the convex combination of such non-unique SDC matrices also verifies extended linerization as follows:

$$\begin{aligned}&f(s,t)+B(s,t)\bar{u}-f(\bar{s},t)-B(\bar{s},t)\bar{u}\\&=A(\varrho,s,\bar{s},\bar{u},t)(s-\bar{s})=\sum_{i=1}^{s_A}\varrho_i A_i(s,\bar{s},\bar{u},t)(s-\bar{s})\end{aligned} \tag{60}$$

where $\varrho=(\varrho_1,\cdots,\varrho_{s_A})$, $\sum_{i=1}^{s_A}\varrho_i=1$, $\varrho_i\geq 0$, and each A_i satisfies the relation $\bar{f}(s,\bar{u},t)-\bar{f}(\bar{s},\bar{u},t)=A_i(s,\bar{s},\bar{u},t)(s-\bar{s})$.

Proof. The first statement on (59) follows from the integral relation given as

$$\int_0^1\frac{d\bar{f}}{dc}(cs+(1-c)\bar{s},\bar{u},t)dc=\bar{f}(s,\bar{u},t)-\bar{f}(\bar{s},\bar{u},t).$$

If there are multiple SDC matrices A_i, we clearly have $\varrho_i A_i(s,\bar{s},\bar{u},t)(s-\bar{s})=\varrho_i(\bar{f}(s,\bar{u},t)-\bar{f}(\bar{s},\bar{u},t))$, $\forall i$, and therefore, the relation $\sum_{i=1}^{s_A}\varrho_i=1$, $\varrho_i\geq 0$ gives (60). □

Example 3.3. *Let us illustrate how Lemma 3.1 can be used in practice taking the following nonlinear system as an example:*

$$\dot{x} = [x_2, -x_1 x_2]^\top + [0, \cos x_1]^\top u \tag{61}$$

where $x = [x_1, x_2]^\top$. *If we use* $(s, \bar{s}, \bar{u}) = (x, x_d, u_d)$ *in Lemma 3.1 for a given target trajectory* (x_d, u_d) *that satisfies* (61)*, evaluating the integral of* (59) *gives*

$$A_1(x, x_d, u_d, t) = -\begin{bmatrix} 0 & 1 \\ \frac{x_2 + x_{2d}}{2} - \frac{u_d(\cos x_1 - \cos x_{d1})}{x_1 - x_{d1}} & \frac{x_1 + x_{1d}}{2} \end{bmatrix} \tag{62}$$

due to the relation $\partial \bar{f}/\partial s = \left[\begin{smallmatrix} 0 & 1 \\ -s_2 & -s_1 \end{smallmatrix}\right] + \left[\begin{smallmatrix} 0 & 0 \\ -u_d \sin s_1 & 0 \end{smallmatrix}\right]$ *for* $\bar{f}(s, u_d, t) = f(s, t) + B(s, t)u_d$, *where* $x_d = [x_{1d}, x_{2d}]^\top$. *Note that we have*

$$\frac{(\cos x_1 - \cos x_{d1})}{x_1 - x_{d1}} = -\sin\left(\frac{x_1 + x_{1d}}{2}\right) \operatorname{sinc}\left(\frac{x_1 - x_{1d}}{2}\right)$$

and thus $A(x, x_d, u_d, t)$ *is defined for all* x, x_d, u_d, *and* t. *The SDC matrix* (62) *indeed verifies* $A_1(x, x_d, u_d, t)(x - x_d) = \bar{f}(x, t) - \bar{f}(x_d, t)$.

We can see that the following is also an SDC matrix of the nonlinear system (61)*:*

$$A_2(x, x_d, u_d, t) = -\begin{bmatrix} 0 & 1 \\ x_2 - \frac{u_d(\cos x_1 - \cos x_{d1})}{x_1 - x_{d1}} & x_{1d} \end{bmatrix}. \tag{63}$$

Therefore, the convex combination of A_1 *in* (62) *and* A_2 *in* (63)*,* $A = \varrho_1 A_1 + \varrho_2 A_2$ *with* $\varrho_1 + \varrho_2 = 1$, $\varrho_1, \varrho_2 \geq 0$, *is also an SDC matrix due to Lemma 3.1.*

The major advantage of the formalism in Lemma 3.1 lies in its systematic connection to LTV systems based on uniform controllability and observability, adequately accounting for the nonlinear nature of underlying dynamics through $A(\varrho, x, x_d, u_d, t)$ for global stability, as shall be seen in Sec. 3 and Sec. 4. Since A depends also on (x_d, u_d) in this case unlike the original SDC matrix, we could consider contraction metrics using a positive definite matrix $M(x, x_d, u_d, t)$ instead of $M(x, t)$ in Definition 2.3, to improve the representation power of M at the expense of computational efficiency. Another interesting point is that the non-uniqueness of A in Lemma 3.1 for $n \geq 2$ creates additional degrees of freedom for selecting the coefficients ϱ, which can also be treated as decision variables in constructing optimal contraction metrics as proposed in [13, 16, 17].

We focus mostly on the generalized SDC formulation in Sec. 3 and Sec. 4, as it yields optimal control and estimation laws with global stability [15] while keeping the analysis simple enough to be understood as in LTV systems theory.

Remark 3.2. *This does not mean that contraction theory works only for the SDC parameterized nonlinear systems but implies that it can be used with the other techniques discussed in Sec. 3.1. For example, due to the extended linear form given in Table 3, the results to be presented in Sec. 3 and in Sec. 4 based on the SDC formulation are applicable to linearized dynamics that can be viewed as an LTV system with some modifications, regarding the dynamics modeling error term as an external disturbance as in Sec. 3.1.2. Also, the original SDC formulation with respect to a fixed point (e.g.,* $(s, \bar{s}, \bar{u}) = (x, 0, 0)$ *in Lemma 3.1) can still be used to obtain contraction conditions independent of a target trajectory* (x_d, u_d) *(see Theorem 3.2 for details).*

3.1.4. Control Contraction Metric (CCM) Formulation

We could also consider using the partial derivative of f of the dynamical system directly for control synthesis through differential state feedback $\delta u = k(x, \delta x, u, t)$. This idea, formulated as the concept of a CCM [33, 79–83], constructs contraction metrics with global stability guarantees independently of target trajectories, achieving greater generality while requiring added computation in evaluating integrals involving minimizing geodesics. Similar to the CCM, we could design a state estimator using a general formulation based on geodesics distances between trajectories [50, 84]. These approaches are well compatible with the convex optimization-based schemes in Sec. 4, and hence will be discussed in Sec. 4.3.

The differences between the SDC and CCM formulation are summarized in Table 3. Considering such trade-offs would help determine which form of the control law is the best fit when using contraction theory for nonlinear stabilization.

Remark 3.3. *For control non-affine nonlinear systems, we could find* $f(x, u, t) - f(x_d, u_d, t) = A(x, x_d, u, u_d, t)(x - x_d) + B(x, x_d, u, u_d, t)(u - u_d)$ *by Lemma 3.1 on the SDC formulation and use it in Theorem 4.2, although* (64) *has to be solved implicitly as* B *depends on* u *in this case. A similar approach for the CCM formulation can be found in [82, 83]. As discussed in Example 3.2, designing such implicit control laws will be discussed in Lemma 9.1 and Theorem 9.2 of Sec. 9.1.*

3.2. LMI Conditions for Contraction Metrics

We design a nonlinear feedback tracking control law parameterized by a matrix-valued function $M(x, x_d, u_d, t)$ (or $M(x, t)$, see Theorem 3.2) as follows:

$$\begin{aligned} u &= u_d - K(x, x_d, u_d, t)(x - x_d) \\ &= u_d - R(x, x_d, u_d, t)^{-1} B(x, t)^\top M(x, x_d, u_d, t)(x - x_d) \end{aligned} \tag{64}$$

where $R(x, x_d, u_d, t) \succ 0$ is a weight matrix on the input u and $M(x, x_d, u_d, t) \succ 0$ is a positive definite matrix (which satisfies the matrix inequality constraints for a contraction metric, to be given in Theorem 3.1). As discussed in Sec. 3.1.3, the extended linear form of the tracking control (64) enables LTV systems-type approaches to Lyapunov

function construction, while being general enough to capture the nonlinearity of the underlying dynamics due to Lemma 3.2 [105].

Lemma 3.2. *Consider a general feedback controller u defined as $u = k(x, x_d, u_d, t)$ with $k(x_d, x_d, u_d, t) = u_d$, where $k : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_{\geq 0} \mapsto \mathbb{R}^m$. If k is piecewise continuously differentiable, then $\exists K : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_{\geq 0} \mapsto \mathbb{R}^{m\times n}$ s.t. $u = k(x, x_d, u_d, t) = u_d - K(x, x_d, u_d, t)(x - x_d)$.*

Proof. Using $k(x_d, x_d, u_d, t) = u_d$, u can be decomposed as $u = u_d + (k(x, x_d, u_d, t) - k(x_d, x_d, u_d, t))$. Since we have $k(x, x_d, u_d, t) - k(x_d, x_d, u_d, t) = \int_0^1 (dk(cx + (1 - c)x_d, x_d, u_d, t)/dc)dc$, selecting K as

$$K = -\int_0^1 \frac{\partial k}{\partial x}(cx + (1 - c)x_d, x_d, u_d, t)dc$$

gives the desired relation [105]. □

Remark 3.4. *Lemma 3.2 implies that designing optimal k of $u = k(x, x_d, u_d, t)$ reduces to designing the optimal gain $K(x, x_d, u_d, t)$ of $u = u_d - K(x, x_d, u_d, t)(x - x_d)$. We could also generalize this idea further using the CCM-based differential feedback controller $\delta u = k(x, \delta x, u, t)$ [33, 79–83] (see Theorem 4.6).*

Substituting (64) into (55) and (56) yields the following virtual system of a smooth path $q(\mu, t)$, parameterized by $\mu \in [0, 1]$ to have $q(0, \mu) = x_d$ and $q(1, t) = x$, for partial contraction in Theorem 2.2:

$$\dot{q}(\mu, t) = \zeta(q(\mu, t), x, x_d, u_d, t) + d(\mu, x, t) \tag{65}$$
$$dq(\mu, t) = \zeta(q(\mu, t), x, x_d, u_d, t)dt + G(\mu, x, t)d\mathscr{W}(t) \tag{66}$$

where $d(\mu, x, t) = \mu d_c(x, t)$, $G(\mu, x, t) = \mu G_c(x, t)$, and $\zeta(q, x, x_d, u_d, t)$ is defined as

$$\zeta = (A(\varrho, x, x_d, u_d, t) - B(x, t)K(x, x_d, u_d, t))(q - x_d) + f(x_d, t) + B(x_d, t)u_d \tag{67}$$

where A is the SDC matrix of Lemma 3.1 with $(s, \bar{s}, \bar{u}) = (x, x_d, u_d)$. Setting $\mu = 1$ in (65) and (66) results in (55) and (56), respectively, and setting $\mu = 0$ simply results in (57). Consequently, both $q = x$ and $q = x_d$ are particular solutions of (65) and (66). If there is no disturbance acting on the dynamics (55) and (56), the differential dynamics of (65) and (66) for $\partial_\mu q = \partial q/\partial \mu$ is given as

$$\partial_\mu \dot{q} = (A(\varrho, x, x_d, u_d, t) - B(x, t)K(x, x_d, u_d, t))\,\partial_\mu q. \tag{68}$$

In [13–15, 17], it is proposed that the contraction conditions of Theorems 2.1 and 2.5 for the closed-loop dynamics (65) and (66) can be expressed as convex constraints as summarized in Theorem 3.1.

Theorem 3.1. *Let β be defined as $\beta = 0$ for deterministic systems* (55) *and*

$$\beta = \alpha_s = L_m \bar{g}_c^2(\alpha_G + 1/2)$$

for stochastic systems (56)*, respectively, where $\bar{g}_c$ is given in* (56)*, L_m is the Lipschitz constant of $\partial M/\partial x_i$ for M of* (64)*, and $\alpha_G \in \mathbb{R}_{>0}$ is an arbitrary constant as in Theorem 2.5. Also, let $W = M(x, x_d, u_d, t)^{-1}$ (or $W = M(x, t)^{-1}$, see Theorem 3.2), $\bar{W} = \nu W$, and $\nu = \overline{m}$. Then the following three matrix inequalities are equivalent:*

$$\dot{M} + M\frac{\partial \zeta}{\partial q} + \frac{\partial \zeta}{\partial q}^\top M \preceq -2\alpha M - \beta \mathrm{I},\ \forall \mu \in [0, 1] \tag{69}$$
$$\dot{M} + 2\,\mathrm{sym}(MA) - 2MBR^{-1}B^\top M \preceq -2\alpha M - \beta \mathrm{I} \tag{70}$$
$$-\dot{\bar{W}} + 2\,\mathrm{sym}(A\bar{W}) - 2\nu BR^{-1}B^\top \preceq -2\alpha \bar{W} - \frac{\beta}{\nu}\bar{W}^2 \tag{71}$$

where ζ is as defined in (67)*. For stochastic systems with $\beta = \alpha_s > 0$, these inequalities are also equivalent to*

$$\begin{bmatrix} -\dot{\bar{W}} + 2\,\mathrm{sym}\,(A\bar{W}) - 2\nu BR^{-1}B^\top + 2\alpha \bar{W} & \bar{W} \\ \bar{W} & -\frac{\nu}{\beta}\mathrm{I} \end{bmatrix} \preceq 0. \tag{72}$$

Note that ν and $\bar{W}$ are required for (71) *and* (72) *and the arguments (x, x_d, u_d, t) for each matrix are suppressed for notational simplicity.*

Furthermore, under these equivalent contraction conditions, Theorems 2.4 and 2.5 hold for the virtual systems (65) *and* (66)*, respectively. In particular, if $\underline{m}\mathrm{I} \preceq M \preceq \overline{m}\mathrm{I}$ of* (26) *holds, or equivalently*

$$\mathrm{I} \preceq \bar{W}(x, x_d, u_d, t) \preceq \chi \mathrm{I} \tag{73}$$

holds for $\chi = \overline{m}/\underline{m}$, then we have the following bounds:

$$\|x(t) - x_d(t)\| \leq \frac{V_\ell(0)}{\sqrt{\underline{m}}}e^{-\alpha t} + \frac{\bar{d}_c}{\alpha}\sqrt{\chi}(1 - e^{-\alpha t}) \tag{74}$$
$$\mathbb{E}\left[\|x(t) - x_d(t)\|^2\right] \leq \frac{\mathbb{E}[V_{s\ell}(0)]}{\underline{m}}e^{-2\alpha t} + \frac{C_C}{2\alpha}\chi \tag{75}$$

where $V_{s\ell} = \int_{x_d}^{x} \delta q^\top M \delta q$ and $V_\ell = \int_{x_d}^{x} \|\Theta \delta q\|$ are as given in Theorem 2.3 with $M = \Theta^\top \Theta$, the disturbance bounds $\bar{d}_c$ and $\bar{g}_c$ are given in (55) *and* (56)*, respectively, and $C_C = \bar{g}_c^2(2{\alpha_G}^{-1} + 1)$. Note that for stochastic systems, the probability that $\|x - x_d\|$ is greater than or equal to $\varepsilon \in \mathbb{R}_{>0}$ is given as*

$$\mathbb{P}\left[\|x(t) - x_d(t)\| \geq \varepsilon\right] \leq \frac{1}{\varepsilon^2}\left(\frac{\mathbb{E}[V_{s\ell}(0)]}{\underline{m}}e^{-2\alpha t} + \frac{C_C}{2\alpha}\chi\right). \tag{76}$$

Proof. Substituting (67) into (69) gives (70). Since $\nu > 0$ and $W \succ 0$, multiplying (70) by ν and then by W from both sides preserves matrix definiteness. Also, the resultant inequalities are equivalent to the original ones [21, p. 114]. These operations performed on (70) yield (71). If $\beta = \alpha_s > 0$ for stochastic systems, applying Schur's complement lemma [21, p. 7] to (71) results in the Linear Matrix Inequality (LMI) constraint (72) in terms of $\bar{W}$ and ν. Therefore, (69)–(72) are indeed equivalent.

Also, since we have $\|\partial_\mu d(\mu,x,t)\| \le \bar{d}_c$ for d in (65) and $\|\partial_\mu G(\mu,x,t)\|_F^2 \le \bar{g}_c^2$ for G in (66), the virtual systems in (65) and (66) clearly satisfy the conditions of Theorems 2.4 and 2.5 if it is equipped with (69), which is equivalent to (70)–(72). This implies the exponential bounds (74)–(76) rewritten using $\chi = \overline{m}/\underline{m}$, following the proofs of Theorems 2.4 and 2.5. □

Because of the control and estimation duality in differential dynamics similar to that of the Kalman filter and Linear Quadratic Regulator (LQR) in LTV systems, we have an analogous robustness result for the contraction theory-based state estimator as to be derived in Sec. 4.2.

Although the conditions (69)–(72) depend on (x_d, u_d), we could also use the SDC formulation with respect to a fixed point [13, 17] in Lemma 3.1 to make them independent of the target trajectory as in the following theorem.

Theorem 3.2. *Let $(\bar{x},\bar{u})$ be a fixed point selected arbitrarily in $\mathbb{R}^n \times \mathbb{R}^m$, e.g., $(\bar{x},\bar{u}) = (0,0)$, and let $A(x,t)$ be an SDC matrix constructed with $(s,\bar{s},\bar{u}) = (x,\bar{x},\bar{u})$ in Lemma 3.1, i.e.,*

$$A(\varrho,x,t)(x-\bar{x}) = f(x,t) + B(x,t)\bar{u} - f(\bar{x},t) - B(\bar{x},t)\bar{u}. \tag{77}$$

Suppose that the contraction metric of Theorem 3.1 is designed by $M(x,t)$ with A of (77), independently of the target trajectory (x_d,u_d), and that the systems (55) and (56) are controlled by

$$u = u_d - R(x,t)^{-1}B(x,t)^\top M(x,t)(x-x_d) \tag{78}$$

with such $M(x,t)$, where $R(x,t) \succ 0$ is a weight matrix on u. If the function $\phi(x,x_d,u_d,t) = A(\varrho,x,t)(x_d-\bar{x}) + B(x,t)(u_d-\bar{u})$ is Lipschitz in x with its Lipschitz constant $\bar{L}$, then Theorem 3.1 still holds with α of the conditions (69)–(72) replaced by $\alpha + \bar{L}\sqrt{\overline{m}/\underline{m}}$. The same argument holds for state estimation of Theorem 4.3 to be discussed in Sec. 4.2.

Proof. The unperturbed virtual system of (55), (56), and (57) with A of (77) and u of (78) is given as follows:

$$\begin{aligned}\dot{q} &= (A(\varrho,x,t) - B(x,t)K(x,t))(q-x_d) \\ &+ A(\varrho,q,t)(x_d-\bar{x}) + B(q,t)(u_d-\bar{u}) + f(\bar{x},t) + B(\bar{x},t)\bar{u}\end{aligned} \tag{79}$$

where $K(x,t) = R(x,t)^{-1}B(x,t)^\top M(x,t)$. Following the proof of Theorem 3.1, the computation of $\dot{V}$, where $V = \delta q^\top M(x,t)\delta q$, yields an extra term

$$2\delta q^\top M \frac{\partial \phi}{\partial q}(q,x_d,u_d,t)\delta q \le 2\bar{L}\sqrt{\frac{\overline{m}}{\underline{m}}}\delta q^\top M \delta q \tag{80}$$

due to the Lipschitz condition on ϕ, where $\phi(q,x_d,u_d,t) = A(q,t)(x_d-\bar{x}) + B(q,t)(u_d-\bar{u})$. This indeed implies that the system (79) is contracting as long as the conditions (69)–(72) hold with α replaced by $\alpha + \bar{L}\sqrt{\overline{m}/\underline{m}}$. The last statement on state estimation follows from the nonlinear control and estimation duality to be discussed in Sec. 4.2.

Remark 3.5. *As demonstrated in [13], we could directly use the extra term $2\delta q^\top M(\partial\phi/\partial q)\delta q$ of (80) in (69)–(72) without upper-bounding it, although now the constraints of Theorem 3.2 depend on (x,q,t) instead of (x,t). Also, the following two inequalities given in [13] with $\bar{\gamma} = \nu\gamma$, $\gamma \in \mathbb{R}_{\ge 0}$:*

$$\begin{aligned}&-\dot{\bar{W}} + A\bar{W} + \bar{W}A^\top + \bar{\gamma}\mathrm{I} - \nu B R^{-1}B^\top \preceq 0 \\ &\begin{bmatrix} \bar{\gamma}\mathrm{I} + \nu B R^{-1}B^\top - \bar{W}\phi^\top - \phi\bar{W} - 2\alpha\bar{W} & \bar{W} \\ \bar{W} & \frac{\nu}{2\alpha_s}\mathrm{I}\end{bmatrix} \succeq 0.\end{aligned}$$

are combined as one LMI (72) in Theorems 3.1 and 3.2.

Example 3.4. *The inequalities in Theorem 3.1 can be interpreted as in the Riccati inequality in $\mathcal{H}_\infty$ control. Consider the following system:*

$$\dot{x} = Ax + B_u u + B_w w,\ z = C_z x \tag{81}$$

where $A \in \mathbb{R}^{n\times n}$, $B_u \in \mathbb{R}^{n\times m}$, $B_w \in \mathbb{R}^{n\times w}$, and $C_z \in \mathbb{R}^{o\times n}$ are constant matrices, $w \in \mathbb{R}^w$ is an exogenous input, and $z \in \mathbb{R}^o$ is a system output. As shown in [25] and [21, p. 109], there exists a state feedback gain $K = R^{-1}B_u^\top P$ such that the $\mathcal{L}_2$ gain of the closed-loop system (81), $\sup_{\|w\|\neq 0}\|z\|/\|w\|$, is less than or equal to γ if

$$2\,\mathrm{sym}\,(PA) - 2PB_uR^{-1}B_u^\top P + \frac{PB_wB_w^\top P}{\gamma^2} + C_z^\top C_z \preceq 0 \tag{82}$$

has a solution $P \succ 0$, where $R \succ 0$ is a constant weight matrix on the input u. If we select B_w and C_z to have $B_wB_w^\top \succeq (P^{-1})^2$ and $C_z^\top C_z \succeq 2\alpha P$ for some $\alpha > 0$, the contraction condition (70) in Theorem 3.1 can be satisfied with $M = P$, $B = B_u$, and $\beta = 1/\gamma^2$ due to (82).

In Sections 3.3 and 3.4, we will discuss the relationship to input-output stability theory as in Example 3.4, using the results of Theorem 3.1.

3.3. Bounded Real Lemma in Contraction Theory

The LMI (72) of Theorem 3.1 can be interpreted as the LMI condition for the bounded real lemma [151, p. 369]. Let us consider the following Hurwitz linear system with $\varsigma(t) \in \mathcal{L}_{2e}$, (i.e., $\|(\varsigma)_\tau\|_{\mathcal{L}_2} < \infty$ for $\tau \in \mathbb{R}_{\ge 0}$, see Sec. 1.3):

$$\dot{\kappa} = \mathcal{A}\kappa + \mathcal{B}\varsigma(t),\quad \upsilon = \mathcal{C}\kappa + \mathcal{D}\varsigma(t). \tag{83}$$

Setting $\kappa = \delta q$ and viewing ς as external disturbance, this system can be interpreted as the differential closed-loop dynamics defined earlier in (68). The bounded real lemma states that this system is $\mathcal{L}_2$ gain stable with its $\mathcal{L}_2$ gain less than or equal to γ, i.e., $\|(\upsilon)_\tau\|_{\mathcal{L}_2} \le \gamma\|(\varsigma)_\tau\|_{\mathcal{L}_2} + const.$, or equivalently, the $\mathcal{H}_\infty$ norm of the transfer function of (83) is less than or equal to γ, if the following LMI for $\mathcal{P} \succ 0$ holds (see [151, p. 369]):

$$\begin{bmatrix} \dot{\mathcal{P}} + 2\,\mathrm{sym}\,(\mathcal{P}\mathcal{A}) + \mathcal{C}^\top\mathcal{C} & \mathcal{P}\mathcal{B} + \mathcal{C}^\top\mathcal{D} \\ \mathcal{B}^\top\mathcal{P} + \mathcal{D}^\top\mathcal{C} & \mathcal{D}^\top\mathcal{D} - \gamma^2\mathrm{I}\end{bmatrix} \preceq 0. \tag{84}$$

Theorem 3.3 introduces the bounded real lemma in the context of contraction theory.

Theorem 3.3. *Let $\mathcal{A} = A - BR^{-1}B^\top M$, $\mathcal{B} = M^{-1}$, $C = \sqrt{2\alpha}\Theta$, and $\mathcal{D} = 0$ in* (83)*, where $M = \Theta^\top\Theta$, and the other variables are as defined in Theorem 3.1. Then* (84) *with $\mathcal{P} = M$ and $\gamma = 1/\sqrt{\beta}$ is equivalent to* (72)*, and thus* (72) *implies $\mathcal{L}_2$ gain stability of* (83) *with its $\mathcal{L}_2$ gain less than or equal to γ.*

PROOF. Multiplying (72) by ν^{-1} and then by $\left[\begin{smallmatrix} M & 0 \\ 0 & \mathrm{I} \end{smallmatrix}\right]$ from both sides gives the following matrix inequality:

$$\nu \begin{bmatrix} \dot{M} + 2\,\mathrm{sym}\left(M(A - BR^{-1}B^\top M)\right) + 2\alpha M & \mathrm{I} \\ \mathrm{I} & -\frac{1}{\beta}\mathrm{I} \end{bmatrix} \preceq 0.$$

This is indeed equivalent to (84) if $\mathcal{A} = A - BR^{-1}B^\top M$, $\mathcal{B} = M^{-1}$, $C = \sqrt{2\alpha}\Theta$, $\mathcal{D} = 0$, $\mathcal{P} = M$, and $\gamma = 1/\sqrt{\beta}$. Now, multiplying (84) by $[\delta q^\top, \varsigma^\top]^\top = [\kappa^\top, \varsigma^\top]^\top$ for such $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$ gives

$$\kappa^\top(\dot{\mathcal{P}} + 2\mathcal{P}\mathcal{A} + \mathcal{C}^\top\mathcal{C})\kappa + 2\kappa^\top\mathcal{P}\mathcal{B}\varsigma - \gamma^2\|\varsigma\|^2 \preceq 0$$

resulting in $\dot{V} + \|\upsilon\|^2 - \gamma^2\|\varsigma\|^2 \preceq 0$ for $V = \kappa^\top\mathcal{P}\kappa = \delta q^\top M \delta q$. This implies $\mathcal{L}_2$ gain stability with its $\mathcal{L}_2$ gain less than or equal to $\gamma = 1/\sqrt{\beta}$ [3, p. 209]. □

3.4. KYP Lemma in Contraction Theory

Analogously to Theorem 3.3, the LMI (72) of Theorem 3.1 can be understood using the Kalman-Yakubovich-Popov (KYP) lemma [22, p. 218]. Consider the quadratic Lyapunov function $V = \kappa^\top\mathcal{Q}\kappa$ with $\mathcal{Q} \succ 0$, satisfying the following output strict passivity (dissipativity) condition:

$$\dot{V} - 2\varsigma^\top\upsilon + (2/\gamma)\upsilon^\top\upsilon \leq 0 \tag{85}$$

which can be expanded by completing the square to have

$$\begin{aligned}\dot{V} \leq& -\|\gamma(\varsigma - \upsilon/\gamma)\|^2 + \gamma\|\varsigma\|^2 - (1/\gamma)\|\upsilon\|^2 \\ \leq& \gamma\|\varsigma\|^2 - (1/\gamma)\|\upsilon\|^2.\end{aligned} \tag{86}$$

This implies that we have $\|(\upsilon)_\tau\|_{\mathcal{L}_2} \leq \gamma\|(\varsigma)_\tau\|_{\mathcal{L}_2} + \sqrt{\gamma V(0)}$ by the comparison lemma [3, pp. 102-103, pp. 350-353], leading to $\mathcal{L}_2$ gain stability with its $\mathcal{L}_2$ gain less than or equal to γ [22, p. 218]. The condition (85) can be expressed equivalently as an LMI form as follows:

$$\begin{bmatrix} \dot{\mathcal{Q}} + 2\mathrm{sym}(\mathcal{Q}\mathcal{A}) + \frac{2\mathcal{C}^\top\mathcal{C}}{\gamma} & \mathcal{Q}\mathcal{B} + \mathcal{C}^\top\left(\frac{2\mathcal{D}}{\gamma} - \mathrm{I}\right) \\ \mathcal{B}^\top\mathcal{Q} + \left(\frac{2\mathcal{D}^\top}{\gamma} - \mathrm{I}\right)\mathcal{C} & -(\mathcal{D}^\top + D) + \frac{2\mathcal{D}^\top\mathcal{D}}{\gamma} \end{bmatrix} \preceq 0 \tag{87}$$

where $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{D}$ are as defined in (83).

Theorem 3.4. *If* (87) *holds, the LMI* (84) *for the bounded real lemma holds with $\mathcal{P} = \gamma\mathcal{Q}$, i.e., the system* (83) *is $\mathcal{L}_2$ gain stable with its $\mathcal{L}_2$ gain less than or equal to γ. Thus, for systems with $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{D}$ defined in Theorem 3.3, the condition* (87) *guarantees the contraction condition* (72) *of Theorem 3.1.*

PROOF. Writing the inequality in (86) in a matrix form, we have that

$$\begin{aligned}&\begin{bmatrix} \dot{\mathcal{Q}} + 2\mathrm{sym}(\mathcal{Q}\mathcal{A}) + \frac{2\mathcal{C}^\top\mathcal{C}}{\gamma} & \mathcal{Q}\mathcal{B} + \mathcal{C}^\top\left(\frac{2\mathcal{D}}{\gamma} - \mathrm{I}\right) \\ \mathcal{B}^\top\mathcal{Q} + \left(\frac{2\mathcal{D}^\top}{\gamma} - \mathrm{I}\right)\mathcal{C} & -(\mathcal{D}^\top + D) + \frac{2\mathcal{D}^\top\mathcal{D}}{\gamma} \end{bmatrix} \\ &\succeq \begin{bmatrix} \dot{\mathcal{Q}} + 2\mathrm{sym}(\mathcal{Q}\mathcal{A}) & \mathcal{Q}\mathcal{B} \\ \mathcal{B}^\top\mathcal{Q} & 0 \end{bmatrix} + \begin{bmatrix} \frac{\mathcal{C}^\top\mathcal{C}}{\gamma} & \frac{\mathcal{C}^\top\mathcal{D}}{\gamma} \\ \frac{\mathcal{D}^\top\mathcal{C}}{\gamma} & \frac{\mathcal{D}^\top\mathcal{D}}{\gamma} \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0 & \gamma\mathrm{I} \end{bmatrix} \\ &= \begin{bmatrix} \dot{\mathcal{Q}} + 2\mathrm{sym}(\mathcal{Q}\mathcal{A}) + \frac{\mathcal{C}^\top\mathcal{C}}{\gamma} & \mathcal{Q}\mathcal{B} + \frac{\mathcal{C}^\top\mathcal{D}}{\gamma} \\ \mathcal{B}^\top\mathcal{Q} + \frac{\mathcal{D}^\top\mathcal{C}}{\gamma} & \frac{\mathcal{D}^\top\mathcal{D}}{\gamma} - \gamma\mathrm{I} \end{bmatrix}.\end{aligned}$$

Therefore, a necessary condition of (87) reduces to (84) if $\mathcal{Q} = \mathcal{P}/\gamma$. The rest follows from Theorem 3.3. □

Example 3.5. *Theorems 3.3 and 3.4 imply the following statements.*

- *If* (84) *holds, the system is finite-gain $\mathcal{L}_2$ stable with γ as its $\mathcal{L}_2$ gain.*
- *If* (87) *holds, the system is finite-gain $\mathcal{L}_2$ stable with γ as its $\mathcal{L}_2$ gain.*
- (85) *is equivalent to* (87)*, and to the output strict passivity condition with dissipation $1/\gamma$ [3, p. 231].*
- *If* (87) *holds (KYP lemma),* (84) *holds (bounded real lemma). The converse is not necessarily true.*

4. Convex Optimality in Robust Nonlinear Control and Estimation via Contraction Theory

Theorem 3.1 indicates that the problem of finding contraction metrics for general nonlinear systems could be formulated as a convex feasibility problem. This section thus delineates one approach, called the method of Con-Vex optimization-based Steady-state Tracking Error Minimization (CV-STEM) [13–15, 17], to optimally design M of Theorem 3.1 that defines a contraction metric and minimizes an upper bound of the steady-state tracking error in Theorem 2.4 or in Theorem 2.5 via convex optimization. In particular, we present an overview of the SDC- and CCM-based CV-STEM frameworks for provably stable and optimal feedback control and state estimation of nonlinear systems, perturbed by deterministic and stochastic disturbances. It is worth noting that the steady-state bound is expressed as a function of the condition number $\chi = \overline{m}/\underline{m}$ as to be seen in (88) and (104), which renders the CV-STEM applicable and effective even to learning-based and data-driven control frameworks as shall be seen in Sec. 5.

4.1. CV-STEM Control

As a result of Theorems 2.4, 2.5, and 3.1, the control law $u = u_d - K(x - x_d)$ of (64) gives a convex steady-state upper bound of the Euclidean distance between the system trajectories, which can be used as an objective function for the CV-STEM control framework in Theorem 4.2 [13–15, 17].

Theorem 4.1. *If one of the matrix inequalities of Theorem 3.1 holds, we have the following bound for* $\chi = \overline{m}/\underline{m}$*:*

$$\lim_{t\to\infty} \sqrt{\mathbb{E}\left[\|x - x_d\|^2\right]} \leq c_0(\alpha, \alpha_G)\sqrt{\chi} \leq c_0(\alpha, \alpha_G)\chi \qquad (88)$$

where $c_0(\alpha, \alpha_G) = \bar{d}_c/\alpha$ *for deterministic systems and* $c_0(\alpha, \alpha_G) = \bar{g}_c\sqrt{(2\alpha_G^{-1}+1)/(2\alpha)}$ *for stochastic systems, with the variables given in Theorem 3.1.*

PROOF. Taking $\lim_{t\to\infty}$ in the exponential tracking error bounds (74) and (75) of Theorem 3.1 gives the first inequality of (88). The second inequality follows from the relation $1 \leq \sqrt{\overline{m}/\underline{m}} \leq \overline{m}/\underline{m} = \chi$ due to $\underline{m} \leq \overline{m}$. □

The convex optimization problem of minimizing (88), subject to the contraction constraint of Theorem 3.1, is given in Theorem 4.2, thereby introducing the CV-STEM control [13–15, 17] for designing an optimal contraction metric and contracting control policy as depicted in Fig. 2.

Theorem 4.2. *Suppose that* α, α_G, $\bar{d}_c$, *and* $\bar{g}_c$ *in* (88) *and the Lipschitz constant* L_m *of* $\partial M/\partial x_i$ *in Theorem 3.1 are given. If the pair* (A, B) *is uniformly controllable, the non-convex optimization problem of minimizing the upper bound* (88) *is equivalent to the following convex optimization problem, with the convex contraction constraint* (71) *or* (72) *of Theorem 3.1 and* $\mathrm{I} \preceq \bar{W} \preceq \chi\mathrm{I}$ *of* (73)*:*

$$J^*_{CV} = \min_{\nu\in\mathbb{R}_{>0}, \chi\in\mathbb{R}, \bar{W}\succ 0} c_0\chi + c_1\nu + c_2 P(\chi, \nu, \bar{W}) \qquad (89)$$

s.t. (71) *and* (73) *for deterministic systems*

(72) *and* (73) *for stochastic systems*

where $W = M(x, x_d, u_d, t)^{-1}$ *for Theorem 3.1 or* $W = M(x, t)^{-1}$ *for Theorem 3.2,* $\bar{W} = \nu W$, $\nu = \overline{m}$, $\chi = \overline{m}/\underline{m}$, c_0 *is as defined in* (88) *of Theorem 4.1,* $c_1, c_2 \in \mathbb{R}_{\geq 0}$, *and* P *is a performance-based convex cost function.*

The weight c_1 *for* ν *can be viewed as a penalty on the* 2*-norm of the feedback control gain* K *of* $u = u_d - K(x - x_d)$ *in* (64)*. Using non-zero* c_1 *and* c_2 *thus enables finding contraction metrics optimal in a different sense. Furthermore, the coefficients of the SDC parameterizations* ϱ *in Lemma 3.1 (i.e.,* $A = \sum \varrho_i A_i$ *in* (60)*) can also be treated as decision variables by convex relaxation, thereby adding a design flexibility to mitigate the effects of external disturbances while verifying the system controllability.*

PROOF. As proven in Theorems 3.1 and 4.1, $\underline{m}\mathrm{I} \preceq M \preceq \overline{m}\mathrm{I}$ of (26), the contraction constraint (69), and the objective (88) reduce to (89) with $c_1 = 0$ and $c_2 = 0$. Since the resultant constraints are convex and the objective is affine in terms of the decision variables ν, χ, and $\bar{W}$, the problem (89) is indeed convex. Since we have $\sup_{x,t}\|M\| \leq \overline{m} = \nu$, c_1 can be used as a penalty to optimally adjust the induced 2-norm of the control gain (see Example 4.1). The SDC coefficients ϱ can also be utilized as decision variables for controllability due to Proposition 1 of [13]. □

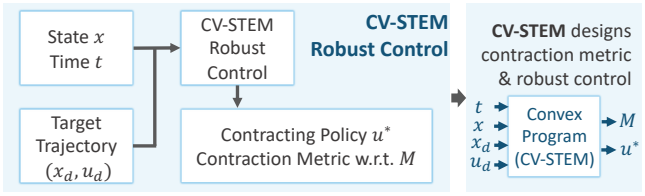

Figure 2: Block diagram of CV-STEM.

Example 4.1. *The weights* c_0 *and* c_1 *of the CV-STEM control of Theorem 4.2 establish an analogous trade-off to the case of the LQR with the cost weight matrices of* Q *for state and* R *for control, since the upper bound of the steady-state tracking error* (88) *is proportional to* χ*, and an upper bound of* $\|K\|$ *for the control gain* K *in* (64)*,* $u = u_d - K(x - x_d)$*, is proportional to* ν*. In particular [14, 15],*

- *if* c_1 *is much greater than* c_0 *in* (89) *of Theorem 4.2, we get smaller control effort but with a large steady-state tracking error, and*
- *if* c_1 *is much smaller than* c_0 *in* (89) *of Theorem 4.2, we get a smaller steady-state state tracking error but with larger control effort.*

This is also because the solution $P \succ 0$ of the LQR Riccati equation [136, p. 89], $-\dot{P} = PA + A^\top P - PBR^{-1}B^\top P + Q$, can be viewed as a positive definite matrix M that defines a contraction metric as discussed in Example 2.3.

4.2. CV-STEM Estimation

We could also design an optimal state estimator analogously to the CV-STEM control of Theorem 4.2, due to the differential nature of contraction theory that enables LTV systems-type approaches to stability analysis. In particular, we exploit the estimation and control duality in differential dynamics similar to that of the Kalman filter and LQR in LTV systems.

Let us consider the following smooth nonlinear systems with a measurement $y(t)$, perturbed by deterministic disturbances $d_{e0}(x,t)$ and $d_{e1}(x,t)$ with $\sup_{x,t}\|d_{e0}(x,t)\| = \bar{d}_{e0} \in \mathbb{R}_{\geq 0}$ and $\sup_{x,t}\|d_{e1}(x,t)\| = \bar{d}_{e1} \in \mathbb{R}_{\geq 0}$, or by Gaussian white noise $\mathscr{W}_0(t)$ and $\mathscr{W}_1(t)$ with $\sup_{x,t}\|G_{e0}(x,t)\|_F = \bar{g}_{e0} \in \mathbb{R}_{\geq 0}$ and $\sup_{x,t}\|G_{e1}(x,t)\|_F = \bar{g}_{e1} \in \mathbb{R}_{\geq 0}$:

$$\dot{x} = f(x,t) + d_{e0}(x,t),\ y = h(x,t) + d_{e1}(x,t) \qquad (90)$$

$$dx = f(x,t)dt + G_{e0}d\mathscr{W}_0,\ ydt = h(x,t)dt + G_{e1}d\mathscr{W}_1 \qquad (91)$$

where $t \in \mathbb{R}_{\geq 0}$ is time, $x : \mathbb{R}_{\geq 0} \mapsto \mathbb{R}^n$ is the system state, $y : \mathbb{R}_{\geq 0} \mapsto \mathbb{R}^m$ is the system measurement, $f : \mathbb{R}^n \times \mathbb{R}_{\geq 0} \mapsto \mathbb{R}^n$ and $h : \mathbb{R}^n \times \mathbb{R}_{\geq 0} \mapsto \mathbb{R}^m$ are known smooth functions, $d_{e0} : \mathbb{R}^n \times \mathbb{R}_{\geq 0} \mapsto \mathbb{R}^n$, $d_{e1} : \mathbb{R}^n \times \mathbb{R}_{\geq 0} \mapsto \mathbb{R}^0$, $G_{e0} : \mathbb{R}^n \times \mathbb{R}_{\geq 0} \mapsto \mathbb{R}^{n\times w_0}$, and $G_{e1} : \mathbb{R}^n \times \mathbb{R}_{\geq 0} \mapsto \mathbb{R}^{n\times w_1}$ are unknown bounded functions for external disturbances, $\mathscr{W}_0 : \mathbb{R}_{\geq 0} \mapsto \mathbb{R}^{w_0}$ and $\mathscr{W}_1 : \mathbb{R}_{\geq 0} \mapsto \mathbb{R}^{w_1}$ are two independent Wiener processes, and the arguments of $G_{e0}(x,t)$ and $G_{e1}(x,t)$ are suppressed for notational convenience. Let $A(\varrho_a, x, \hat{x}, t)$ and $C(\varrho_c, x, \hat{x}, t)$ be the SDC matrices

given by Lemma 3.1 with $(f, s, \bar{s}, \bar{u})$ replaced by $(f, \hat{x}, x, 0)$ and $(h, \hat{x}, x, 0)$, respectively, i.e.

$$A(\varrho_a, x, \hat{x}, t)(\hat{x} - x) = f(\hat{x}, t) - f(x, t) \tag{92}$$
$$C(\varrho_c, x, \hat{x}, t)(\hat{x} - x) = h(\hat{x}, t) - h(x, t). \tag{93}$$

We design a nonlinear state estimation law parameterized by a matrix-valued function $M(\hat{x}, t)$ as follows:

$$\begin{aligned}\dot{\hat{x}} =& f(\hat{x}, t) + L(\hat{x}, t)(y - h(\hat{x}, t)) \\ =& f(\hat{x}, t) + M(\hat{x}, t)\bar{C}(\varrho_c, \hat{x}, t)^\top R(\hat{x}, t)^{-1}(y - h(\hat{x}, t))\end{aligned} \tag{94}$$

where $\bar{C}(\varrho_c, \hat{x}, t) = C(\varrho_c, \hat{x}, \bar{x}, t)$ for a fixed trajectory $\bar{x}$ (e.g., $\bar{x} = 0$, see Theorem 3.2), $R(\hat{x}, t) \succ 0$ is a weight matrix on the measurement y, and $M(\hat{x}, t) \succ 0$ is a positive definite matrix (which satisfies the matrix inequality constraint for a contraction metric, to be given in (99) of Theorem 4.5). Note that we could use other forms of estimation laws such as the EKF [14, 45, 152], analytical SLAM [47], or SDC with respect to a fixed point [13, 16, 17], depending on the application of interest, which result in a similar stability analysis as in Theorem 3.2.

4.2.1. Nonlinear Stability Analysis of SDC-based State Estimation using Contraction Theory

Substituting (94) into (90) and (91) yields the following virtual system of a smooth path $q(\mu, t)$, parameterized by $\mu \in [0, 1]$ to have $q(\mu = 0, t) = x$ and $q(\mu = 1, t) = \hat{x}$:

$$\dot{q}(\mu, t) = \zeta(q(\mu, t), x, \hat{x}, t) + d(\mu, x, \hat{x}, t) \tag{95}$$
$$dq(\mu, t) = \zeta(q(\mu, t), x, \hat{x}, t)dt + G(\mu, x, \hat{x}, t)d\mathscr{W}(t) \tag{96}$$

where $d(\mu, x, \hat{x}, t) = (1 - \mu)d_{e0}(x, t) + \mu L(\hat{x}, t)d_{e1}(x, t)$, $G(\mu, x, \hat{x}, t) = [(1 - \mu)G_{e0}(x, t), \mu L(\hat{x}, t)G_{e1}(x, t)]$, $\mathscr{W} = [\mathscr{W}_0^\top, \mathscr{W}_1^\top]^\top$, and $\zeta(q, x, x_d, u_d, t)$ is defined as

$$\begin{aligned}\zeta(q, x, \hat{x}, t) =& (A(\varrho_a, x, \hat{x}, t) - L(\hat{x}, t)C(\varrho_c, x, \hat{x}, t))(q - x) \\ &+ f(x, t).\end{aligned} \tag{97}$$

Note that (97) is constructed to contain $q = \hat{x}$ and $q = x$ as its particular solutions of (95) and (96). If $d = 0$ and $\mathscr{W} = 0$, the differential dynamics of (95) and (96) for $\partial_\mu q = \partial q/\partial\mu$ is given as

$$\partial_\mu \dot{q} = \left(A(\varrho_a, x, \hat{x}, t) - L(\hat{x}, t)C(\varrho_c, x, \hat{x}, t)\right)\partial_\mu q. \tag{98}$$

The similarity between (68) ($\partial_\mu \dot{q} = (A - BK)\partial_\mu q$) and (98) leads to the following theorem [13–15, 17]. Again, note that we could also use the SDC formulation with respect to a fixed point as delineated in Theorem 3.2 and as demonstrated in [13, 16, 17].

Theorem 4.3. *Suppose $\exists \bar{\rho}, \bar{c} \in \mathbb{R}_{\geq 0}$ s.t. $\|R^{-1}(\hat{x}, t)\| \leq \bar{\rho}$, $\|C(\varrho_c, x, \hat{x}, t)\| \leq \bar{c}$, $\forall x, \hat{x}, t$. Suppose also that $\underline{m}\mathrm{I} \preceq M \preceq \overline{m}I$ of (26) holds, or equivalently, $\mathrm{I} \preceq \bar{W} \preceq \chi I$ of (73) holds with $W = M(\hat{x}, t)^{-1}$, $\bar{W} = \nu W$, $\nu = \overline{m}$, and $\chi = \overline{m}/\underline{m}$. As in Theorem 3.1, let β be defined as $\beta = 0$ for deterministic systems (90) and*

$$\begin{aligned}\beta =& \alpha_s = \alpha_{e0} + \nu^2\alpha_{e1} \\ =& L_m\bar{g}_{e0}^2(\alpha_G + 1/2)/2 + \nu^2 L_m\bar{\rho}^2\bar{c}^2\bar{g}_{e1}^2(\alpha_G + 1/2)/2\end{aligned}$$

for stochastic systems (91), where $2\alpha_{e0} = L_m\bar{g}_{e0}^2(\alpha_G + 1/2)$, $2\alpha_{e1} = L_m\bar{\rho}^2\bar{c}^2\bar{g}_{e1}^2(\alpha_G + 1/2)$, L_m is the Lipschitz constant of $\partial W/\partial x_i$, $\bar{g}_{e0}$ and $\bar{g}_{e1}$ are given in (91), and $\exists\alpha_G \in \mathbb{R}_{>0}$ is an arbitrary constant as in Theorem 2.5.

If $M(\hat{x}, t)$ in (94) is constructed to satisfy the following convex constraint for $\exists\alpha \in \mathbb{R}_{>0}$:

$$\dot{\bar{W}} + 2\,\mathrm{sym}\left(\bar{W}A - \nu\bar{C}^\top R^{-1}C\right) \preceq -2\alpha\bar{W} - \nu\beta\mathrm{I} \tag{99}$$

then Theorems 2.4 and 2.5 hold for the virtual systems (95) and (96), respectively, i.e., we have the following bounds for $\mathsf{e} = \hat{x} - x$ with $\nu = \overline{m}$ and $\chi = \overline{m}/\underline{m}$:

$$\|\mathsf{e}(t)\| \leq \sqrt{\overline{m}}V_\ell(0)e^{-\alpha t} + \frac{\bar{d}_{e0}\sqrt{\chi} + \bar{\rho}\bar{c}\bar{d}_{e1}\nu}{\alpha}(1 - e^{-\alpha t}) \tag{100}$$

$$\mathbb{E}\left[\|\mathsf{e}(t)\|^2\right] \leq \overline{m}\mathbb{E}[V_{s\ell}(0)]e^{-2\alpha t} + \frac{C_{e0}\chi + C_{e1}\chi\nu^2}{2\alpha} \tag{101}$$

where $V_{s\ell} = \int_x^{\hat{x}} \delta q^\top W \delta q$ and $V_\ell = \int_x^{\hat{x}} \|\Theta\delta q\|$ are given in Theorem 2.3 with $W = M^{-1} = \Theta^\top\Theta$ defining a contraction metric, the disturbance bounds $\bar{d}_{e0}$, $\bar{d}_{e1}$, $\bar{g}_{e0}$, and $\bar{g}_{e1}$ are given in (90) and (91), respectively, $C_{e0} = \bar{g}_{e0}^2(2{\alpha_G}^{-1} + 1)$, and $C_{e1} = \bar{\rho}^2\bar{c}^2\bar{g}_{e1}^2(2{\alpha_G}^{-1} + 1)$. Note that for stochastic systems, the probability that $\|\mathsf{e}\|$ is greater than or equal to $\varepsilon \in \mathbb{R}_{>0}$ is given as

$$\mathbb{P}\left[\|\mathsf{e}(t)\| \geq \varepsilon\right] \leq \frac{1}{\varepsilon^2}\left(\overline{m}\mathbb{E}[V_{s\ell}(0)]e^{-2\alpha t} + \frac{C_E}{2\alpha}\right) \tag{102}$$

where $C_E = C_{e0}\chi + C_{e1}\chi\nu^2$.

Proof. Theorem 3.1 indicates that (99) is equivalent to

$$\dot{W} + 2\,\mathrm{sym}\left(WA - \bar{C}^\top R^{-1}C\right) \preceq -2\alpha W - \beta\mathrm{I}. \tag{103}$$

Computing the time derivative of a Lyapunov function $V = \partial_\mu q^\top W \partial_\mu q$ with $\partial_\mu q = \partial q/\partial\mu$ for the unperturbed virtual dynamics (98), we have using (103) that

$$\begin{aligned}\dot{V} =& \partial_\mu q^\top W\partial_\mu q = \partial_\mu q^\top(\dot{W} + 2WA - 2\bar{C}^\top R^{-1}C)\partial_\mu q \\ \leq& -2\alpha V - \beta\|\partial_\mu q\|^2\end{aligned}$$

which implies that $W = M^{-1}$ defines a contraction metric. Since we have $\overline{m}^{-1}\mathrm{I} \preceq W \preceq \underline{m}^{-1}\mathrm{I}$, $V \geq \overline{m}^{-1}\|\partial_\mu q\|^2$, and

$$\begin{aligned}\|\Theta(\hat{x}, t)\partial_\mu d\| \leq& \bar{d}_{e0}/\sqrt{\underline{m}} + \bar{d}_{e1}\bar{\rho}\bar{c}\sqrt{\overline{m}} \\ \|\partial_\mu G\|_F^2 \leq& \bar{g}_{e0}^2 + \bar{\rho}^2\bar{c}^2\bar{g}_{e1}^2\overline{m}^2\end{aligned}$$

for d in (95) and G in (96), the bounds (100)–(102) follow from the proofs of Theorems 2.4 and 2.5 [14, 15]. □

Remark 4.1. *Although (99) is not an LMI due to the nonlinear term $-\nu\beta\mathrm{I}$ on its right-hand side for stochastic systems (91), it is a convex constraint as $-\nu\beta = -\nu\alpha_s = -\nu\alpha_{e0} - \nu^3\alpha_{e1}$ is a concave function for $\nu \in \mathbb{R}_{>0}$ [15, 153].*

4.2.2. CV-STEM Formulation for State Estimation

The estimator (94) gives a convex steady-state upper bound of the Euclidean distance between x and $\hat{x}$ as in Theorem 4.1 [13–15, 17].

Theorem 4.4. *If* (99) *of Theorem 4.3 holds, then we have the following bound:*

$$\lim_{t\to\infty} \sqrt{\mathbb{E}\left[\|\hat{x}-x\|^2\right]} \le c_0(\alpha,\alpha_G)\chi + c_1(\alpha,\alpha_G)\nu^s \tag{104}$$

where $c_0 = \bar{d}_{e0}/\alpha$, $c_1 = \bar{\rho}\bar{c}\bar{d}_{e1}/\alpha$, $s = 1$ *for deterministic systems* (95), *and* $c_0 = \sqrt{C_{e0}/(2\alpha)}$, $c_1 = C_{e1}/(2\sqrt{2\alpha C_{e0}})$, *and* $s = 2$ *for stochastic systems* (96), *with* C_{e0} *and* C_{e0} *given as* $C_{e0} = \bar{g}_{e0}^2(2\alpha_G^{-1}+1)$ *and* $C_{e1} = \bar{\rho}^2\bar{c}^2\bar{g}_{e1}^2(2\alpha_G^{-1}+1)$.

Proof. The upper bound (104) for deterministic systems (95) follows from (100) with the relation $1 \le \sqrt{\chi} \le \chi$ due to $\underline{m} \le \overline{m}$. For stochastic systems, we have using (101) that

$$C_{e0}\chi + C_{e1}\nu^2\chi \le C_{e0}(\chi + (C_{e1}/(2C_{e0}))\nu^2)^2$$

due to $1 \le \chi \le \chi^2$ and $\nu \in \mathbb{R}_{>0}$. This gives (104) for stochastic systems (96). □

Finally, the CV-STEM estimation framework is summarized in Theorem 4.5 [13–15, 17].

Theorem 4.5. *Suppose that* α, α_G, $\bar{d}_{e0}$, $\bar{d}_{e1}$, $\bar{g}_{e0}$, $\bar{g}_{e1}$, *and* L_m *in* (99) *and* (104) *are given. If the pair* (A, C) *is uniformly observable, the non-convex optimization problem of minimizing the upper bound* (104) *is equivalent to the following convex optimization problem with the contraction constraint* (99) *and* $\mathrm{I} \preceq \bar{W} \preceq \chi\mathrm{I}$ *of* (73)*:*

$$J_{CV}^* = \min_{\nu\in\mathbb{R}_{>0},\chi\in\mathbb{R},\bar{W}\succ 0} c_0\chi + c_1\nu^s + c_2 P(\chi,\nu,\bar{W}) \tag{105}$$

s.t. (99) *and* (73)

where c_0, c_1, *and* s *are as defined in* (104) *of Theorem 4.4,* $c_2 \in \mathbb{R}_{\ge 0}$, *and* P *is some performance-based cost function as in Theorem 4.2.*

The weight c_1 *for* ν^s *indicates how much we trust the measurement* $y(t)$*. Using non-zero* c_2 *enables finding contraction metrics optimal in a different sense in terms of* P*. Furthermore, the coefficients of the SDC parameterizations* ϱ_a *and* ϱ_c *in Lemma 3.1 (i.e.,* $A = \sum \varrho_{a,i}A_i$ *and* $C = \sum \varrho_{c,i}C_i$ *in* (92) *and* (93)*) can also be treated as decision variables by convex relaxation [16], thereby adding a design flexibility to mitigate the effects of external disturbances while verifying the system observability.*

Proof. The proposed optimization problem is convex as its objective and constraints are convex in terms of decision variables χ, ν, and $\bar{W}$ (see Remark 4.1). Also, larger $\bar{d}_{e1}$ and $\bar{g}_{e1}$ in (90) and (91) imply larger measurement uncertainty. Thus by definition of c_1 in Theorem 4.4, the larger the weight of ν, the less confident we are in $y(t)$ (see Example 4.2). The last statement on the SDC coefficients for guaranteeing observability follows from Proposition 1 of [16] and Proposition 1 of [13]. □

Example 4.2. *The weights* c_0 *and* c_1 *of the CV-STEM estimation of Theorem 4.5 has an analogous trade-off to the case of the Kalman filter with the process and sensor noise covariance matrices,* Q *and* P*, respectively, since the term* $c_0\chi$ *in upper bound of the steady-state tracking error in* (104) *becomes dominant if measurement noise is much smaller than process noise (*$\bar{d}_{e0} \gg \bar{d}_{e1}$ *or* $\bar{g}_{e0} \gg \bar{g}_{e1}$*), and the term* $c_1\nu^s$ *becomes dominant if measurement noise is much greater than process noise (*$\bar{d}_{e0} \ll \bar{d}_{e1}$ *or* $\bar{g}_{e0} \ll \bar{g}_{e1}$*). In particular [14, 15],*

- *if* c_1 *is much greater than* c_0*, large measurement noise leads to state estimation that responds slowly to unexpected changes in the measurement* y *(i.e. small estimation gain due to* $\nu = \overline{m} \ge \|M\|$*), and*
- *if* c_1 *is much smaller than* c_0*, large process noise leads to state estimation that responds fast to changes in the measurement (i.e. large* $\nu = \overline{m} \ge \|M\|$*).*

This is also because the solution $Q = P^{-1} \succ 0$ *of the Kalman filter Riccati equation [152, p. 375],* $\dot{P} = AP + PA^\top - PC^\top R^{-1}CP + Q$*, can be viewed as a positive definite matrix that defines a contraction metric as discussed in Example 2.3.*

4.3. Control Contraction Metrics (CCMs)

As briefly discussed in Remark 3.4, the concept of a CCM [33, 79–83] is introduced to extend contraction theory to design differential feedback control laws for control-affine deterministic nonlinear systems (55), $\dot{x} = f(x,t) + B(x,t)u$. Let us show that the CCM formulation could also be considered in the CV-STEM control of Theorem 4.2, where similar ideas have been investigated in [80, 119, 121].

Theorem 4.6. *Suppose the CV-STEM control of Theorem 4.2 is designed with its contraction condition replaced by the following set of convex constraints along with* $\mathrm{I} \preceq \bar{W} \preceq \chi\mathrm{I}$ *of* (73)*:*

$$B_\perp^\top\left(-\frac{\partial \bar{W}}{\partial t} - \partial_f \bar{W} + 2\,\mathrm{sym}\left(\frac{\partial f}{\partial x}\bar{W}\right) + 2\alpha\bar{W}\right)B_\perp \prec 0 \tag{106}$$

$$B_\perp^\top\left(\partial_{b_i}\bar{W} - 2\,\mathrm{sym}\left(\frac{\partial b_i}{\partial x}\bar{W}\right)\right)B_\perp = 0, \forall x,t,i \tag{107}$$

where $B_\perp(x,t)$ *is a matrix whose columns span the cokernel of* $B(x,t)$ *defined as* $\mathrm{coker}(B) = \{a \in \mathbb{R}^n | B^\top a = 0\}$ *satisfying* $B^\top B_\perp = 0$, $b_i(x,t)$ *is the* i*th column of* $B(x,t)$, $\bar{W}(x,t) = \nu W(x,t)$, $W(x,t) = M(x,t)^{-1}$ *for* $M(x,t)$ *that defines a contraction metric,* $\nu = \overline{m}$, $\chi = \overline{m}/\underline{m}$, *and* $\partial_p F = \sum_{k=1}^n (\partial F/\partial x_k)p_k$ *for* $p(x,t) \in \mathbb{R}^n$ *and* $F(x,t) \in \mathbb{R}^{n\times n}$*. Then the controlled system* (55) *with* $d_c = 0$*, i.e.,* $\dot{x} = f(x,t) + B(x,t)u$*, is*

1) *universally exponentially open-loop controllable*
2) *universally exponentially stabilizable via sampled-data feedback with arbitrary sample times*

3) universally exponentially stabilizable using continuous feedback defined almost everywhere, and everywhere in a neighborhood of the target trajectory

all with rate α *and overshoot* $R = \sqrt{\overline{m}/\underline{m}}$ *for* $\underline{m}$ *and* $\overline{m}$ *given in* $\underline{m}\mathrm{I} \preceq M \preceq \overline{m}\mathrm{I}$ *of* (26)*. Such a positive definite matrix* $M(x,t)$ *defines a CCM.*

Given a CCM, there exists a differential feedback controller $\delta u = k(x, \delta x, u, t)$ *that stabilizes the following differential dynamics of* (55) *with* $d_c = 0$ *along all solutions (i.e., the closed-loop dynamics is contracting as in Definition 2.3):*

$$\delta\dot{x} = A(x,u,t)\delta x + B(x,t)\delta u \tag{108}$$

where $A = \partial f/\partial x + \sum_{i=1}^{m}(\partial b_i/\partial x)u_i$ *and* δu *is a tangent vector to a smooth path of controls at* u*. Furthermore, it can be computed as follows [33, 79, 80]:*

$$u(x(t),t) = u_d + \int_0^1 k(\gamma(\mu,t), \partial_\mu\gamma(\mu,t), u(\gamma(\mu,t),t), t)d\mu \tag{109}$$

where $\partial_\mu\gamma = \partial\gamma/\partial\mu$*,* γ *is the minimizing geodesic with* $\gamma(0,t) = x_d$ *and* $\gamma(1,t) = x$ *for* $\mu \in [0,1]$*,* (x_d, u_d) *is a given target trajectory in* $\dot{x}_d = f(x_d,t) + B(x_d,t)u_d$ *of* (57)*, and the computation of* k *is given in [79, 80] (see Remark 4.2). Furthermore, Theorem 2.4 for deterministic disturbance rejection still holds and the CCM controller* (109) *thus possesses the same sense of oplimality as for the CV-STEM control in Theorem 4.2.*

Proof. Since the columns of $B_\perp(x,t)$ span the cokernel of $B(x,t)$, the constraints (106) and (107) can be equivalently written as $B_\perp(x,t)$ replaced by a in $\mathrm{coker}(B) = \{a \in \mathbb{R}^n | B^\top a = 0\}$. Let $a = M\delta x = W^{-1}\delta x$. Then multiplying (106) and (107) by $\nu^{-1} > 0$ and rewriting them using a yields

$$\delta x^\top MB = 0 \;\Rightarrow\; \delta x^\top(\dot{M} + 2\,\mathrm{sym}(MA) + 2\alpha M)\delta x < 0 \tag{110}$$

where $A = \partial f/\partial x + \sum_{i=1}^{m}(\partial b_i/\partial x)u_i$ for the differential dynamics (108). The relation (110) states that δx orthogonal to the span of actuated directions b_i is naturally contracting, thereby implying the stabilizability of the system (55) with $d_c = 0$, i.e., $\dot{x} = f(x,t) + B(x,t)u$. See [33, 79, 80] (and [119] on the CV-STEM formulation) for the rest of the proof. □

Example 4.3. *Let us consider a case where* k *of the differential feedback controller does not depend on* u*, and is defined explicitly by* $M(x,t) \succ 0$ *as follows [119]:*

$$u = u_d - \int_{x_d}^{x} R(q,t)^{-1}B(q,t)^\top M(q,t)\delta q \tag{111}$$

where $R(x,t) \succ 0$ *is a given weight matrix and* q *is a smooth path that connects* x *to* x_d*. Since* (111) *yields* $\delta u = -R(x,t)^{-1}B(x,t)^\top M(x,t)\delta x$*, the contraction conditions* (106) *and* (107) *could be simplified as*

$$-\frac{\partial\bar{W}}{\partial t} - \partial_f\bar{W} + 2\,\mathrm{sym}\left(\frac{\partial f}{\partial x}\bar{W}\right) - \nu BR^{-1}B^\top \preceq -2\alpha\bar{W}$$
$$-\partial_{b_i}\bar{W} + 2\,\mathrm{sym}\left(\frac{\partial b_i}{\partial x}\bar{W}\right) = 0 \tag{112}$$

yielding the convex optimization-based control synthesis algorithm of Theorem 4.6 independently of (x_d, u_d) *similar to that of Theorem 4.2 (see [119] for details).*

Example 4.4. *If* B *is of the form* $[0, \mathrm{I}]^\top$ *for the zero matrix* $0 \in \mathbb{R}^{n_1\times m}$ *and identity matrix* $\mathrm{I} \in \mathbb{R}^{n_2\times m}$ *with* $n = n_1 + n_2$*, the condition* (112) *says that* M *should not depend on the last* n_2 *state variables [79].*

Remark 4.2. *We could consider stochastic perturbation in Theorem 4.6 using Theorem 2.5, even with the differential control law of the form* (109) *or* (111) *as demonstrated in [119]. Also, although the relation* (107) *or* (112) *is not included as a constraint in Theorem 4.2 for simplicity of discussion, the dependence of* $\dot{\bar{W}}$ *on* u *in Theorem 4.2 can be removed by using it in a similar way to [119].*

As stated in (109)*, the computation of the differential feedback gain* $k(x, \delta x, u, t)$ *and minimizing geodesics* γ *is elaborated in [79, 80]. For example, if* M *is state-independent, then geodesics are just straight lines.*

Let us again emphasize that, as delineated in Sec. 3.1, the differences between the SDC- and CCM-based CV-STEM frameworks in Theorems 4.2 and 4.6 arise only from their different form of controllers in $u = u_d - K(x - x_d)$ of (64) and $u = u_d + \int_0^1 k d\mu$ of (109), leading to the trade-offs outlined in Table 3.

4.4. Remarks in CV-STEM Implementation

We propose some useful techniques for the practical application of the CV-STEM in Theorems 4.2, 4.5, and 4.6. Note that the CV-STEM requires solving a convex optimization problem at each time instant, but its solution can be approximated with formal stability guarantees to enable faster computation using machine learning techniques as shall be seen in Sec. 6.

4.4.1. Performance-based Cost Function

Selecting $c_2 = 0$ in Theorems 4.2, 4.5, and 4.6 yields a convex objective function that allows for a systematic interpretation of its weights as seen earlier in Examples 4.1 and 4.2. One could also select $c_2 > 0$ to augment the CV-STEM with other control and estimation performances of interest, as long as $P(\nu, \chi, \bar{W})$ in (89) and (105) is convex. For example, we could consider a steady-state tracking error bound of parametric uncertain systems as P, using the adaptive control technique to be discussed in Sec. 8 [104]. Such a modification results in a contraction metric optimal in a different sense.

4.4.2. Selecting and Computing CV-STEM Parameters

The CV-STEM optimization problems derived in Theorems 4.2, 4.5, and 4.6 are convex if we assume that α, α_G, and L_m are given. However, these parameters would also affect the optimality of resultant contraction metrics. In [14, 15, 80], a line search algorithm is performed to find optimal α and α_G, while the Lipschitz constraint given with L_m is guaranteed by spectrally-normalization [15, 132] as shall be seen in detail in Sec. 6.3 (see [62–64] for contraction theory-based techniques for obtaining Lipschitz bounds). Also, the CV-STEM can be formulated as a finite-dimensional problem by using backward difference approximation on $\dot{\bar{W}}$, where we can then use $-\bar{W} \preceq -\mathrm{I}$ to get a sufficient condition of its constraints, or we could alternatively solve it along pre-computed trajectories $\{x(t_i)\}_{i=0}^{M}$ as in [14]. In Sec. 6, we use a parameterized function such as neural networks [106–108] for approximating M to explicitly compute $\dot{M}$.

Part II: Learning-based Control

5. Contraction Theory for Learning-based Control

Machine learning techniques, e.g., reinforcement learning [110–113], imitation learning [114–118], and neural networks [106–108], have gained popularity due to their ability to achieve a large variety of innovative engineering and scientific tasks which have been impossible heretofore. Starting from this section, we will see how contraction theory enhances learning-based and data-driven automatic control frameworks providing them with formal optimality, stability, and robustness guarantees. In particular, we present Theorems 5.2 and 5.3 for obtaining robust exponential bounds on trajectory tracking errors of nonlinear systems in the presence of learning errors, whose steady-state terms are again written as a function of the condition number $\chi = \overline{m}/\underline{m}$ of a positive definite matrix M that defines a contraction metric, consistently with the CV-STEM frameworks of Sec. 4.

5.1. Problem Formulation

Let us consider the following virtual nonlinear system as in Theorem 2.2:

$$\dot{q}(\mu,t) = g(q(\mu,t),\varpi,t) + \mu\Delta_L(\varpi,t) + d(\mu,\varpi,t) \tag{113}$$

where $\mu \in [0,1]$, $q : [0,1]\times\mathbb{R}_{\geq 0} \mapsto \mathbb{R}^n$ is a smooth path of the system states, $\varpi : \mathbb{R}_{\geq 0} \mapsto \mathbb{R}^p$ is a time-varying parameter, $g : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}_{\geq 0} \mapsto \mathbb{R}^n$ is a known smooth function which renders $\dot{q} = g(q,\varpi,t)$ contracting with respect to q, $d : [0,1]\times\mathbb{R}^p\times\mathbb{R}_{\geq 0} \mapsto \mathbb{R}^n$ with $\sup_{\mu,\varpi,t}\|\partial d/\partial\mu\| = \bar{d} \in \mathbb{R}_{\geq 0}$ is unknown external disturbance parameterized by μ as in Theorem 2.4, and $\Delta_L : \mathbb{R}^p\times\mathbb{R}_{\geq 0} \mapsto \mathbb{R}^n$ is the part to be learned with machine learning-based methodologies such as neural networks [106–108]. We can also formulate this in a stochastic setting as follows:

$$dq(\mu,t) = (g(q(\mu,t),\varpi,t) + \mu\Delta_L(\varpi,t))dt + G(\mu,\varpi,t)d\mathscr{W} \tag{114}$$

where $\mathscr{W} : \mathbb{R}_{\geq 0} \mapsto \mathbb{R}^w$ is a Wiener process, and $G : [0,1]\times\mathbb{R}^p\times\mathbb{R}_{\geq 0} \mapsto \mathbb{R}^{n\times w}$ is a matrix-valued function parameterized by μ with $\sup_{\mu,\varpi,t}\|\partial G/\partial\mu\|_F = \bar{g} \in \mathbb{R}_{\geq 0}$ as in Theorem 2.5. The objective of learning is to find Δ_L which satisfies the following, assuming that $\varpi \in \mathcal{S}_\varpi \subseteq \mathbb{R}^p$ and $t \in \mathcal{S}_t \subseteq \mathbb{R}_{\geq 0}$ for some compact sets $\mathcal{S}_\varpi$ and $\mathcal{S}_t$:

$$\|\Delta_L(\varpi,t)\| \leq \epsilon_{\ell 0} + \epsilon_{\ell 1}\|\xi_1 - \xi_0\|,\ \forall(\varpi,t) \in \mathcal{S} \tag{115}$$

where $\mathcal{S} = \mathcal{S}_\varpi \times \mathcal{S}_t$, $\xi_0 = q(0,t)$ and $\xi_1 = q(1,t)$ are particular solutions of (113) and (114), and $\epsilon_{\ell 0}, \epsilon_{\ell 1} \in \mathbb{R}_{\geq 0}$ are given learning errors.

Examples 5.1–5.3 illustrate how the problem (115) with the nonlinear systems (113) and (114) can be used to describe several learning-based and data-driven control problems to be discussed in the subsequent sections, regarding ϖ as x, x_d, $\hat{x}$, u_d, etc.

Example 5.1 (Tracking control). *Let us consider the problem of learning a computationally-expensive (or unknown) feedback control law $u^*(x,x_d,u_d,t)$ that tracks a target trajectory (x_d,u_d) given by $\dot{x}_d = f(x_d,t)+B(x_d,t)u_d$ as in* (57)*, assuming that f and B are known. If the dynamics is perturbed by $d_c(x,t)$ as in* (55)*, we have that*

$$\begin{aligned}\dot{x} =& f(x,t) + B(x,t)u_L + d_c(x,t)\\ =& f(x,t) + B(x,t)u^* + B(x,t)(u_L - u^*) + d_c(x,t)\end{aligned}$$

where $u_L(x,x_d,u_d,t)$ denotes a learning-based control law that models $u^*(x,x_d,u_d,t)$. As long as u^* renders the closed-loop dynamics $\dot{x} = f(x,t) + B(x,t)u^*(x,x_d,u_d,t)$ contracting with respect to x [13, 14, 79, 80], we can define the functions of (113) *as follows:*

$$\begin{aligned}g(q,\varpi,t) =& f(q,t) + B(q,t)u^*(q,x_d,u_d,t)\\ \Delta_L(\varpi,t) =& B(x,t)(u_L(x,x_d,u_d,t) - u^*(x,x_d,u_d,t))\end{aligned} \tag{116}$$

with $d(\mu,\varpi,t) = \mu d_c(x,t)$, where $\varpi = [x^\top, x_d^\top, u_d^\top]^\top$ in this case. It can be easily verified that (113) *indeed has $q(\mu=0,t) = x_d$ and $q(\mu=1,t) = x$ as particular solutions if $u^*(x_d,x_d,u_d,t) = u_d$. Similarly, we can use* (114) *with $G = \mu G_c(x,t)$ if the dynamics is stochastically perturbed by $G_c(x,t)d\mathscr{W}$ as in* (56)*.*

The learning objective here is to make $\|u_L(x,x_d,u_d,t) - u^*(x,x_d,u_d,t)\|$ as small as possible, which aligns with the aforementioned objective in (115)*. Note that if $\|B\|$ is bounded in the compact set $\mathcal{S}$ of* (115)*, we can bound $\|\Delta_L\|$ of* (116) *with $\epsilon_{\ell 0} \neq 0$ and $\epsilon_{\ell 1} = 0$ as to be explained in Remark 5.1. See Theorems 6.1, 6.3, and 7.1 for details.*

Example 5.2 (State estimation). *Next, let us consider the problem of learning a computationally-expensive (or unknown) state estimator, approximating its estimation*

gain $L(\hat{x},t)$ by $L_L(\hat{x},t)$. If there is no disturbance in (90) *and* (91)*, we have that*

$$\begin{aligned}\dot{\hat{x}} =& f(\hat{x},t) + L_L(h(x,t) - h(\hat{x},t)) = f(\hat{x},t) \\ &+ L(h(x,t) - h(\hat{x},t)) + (L_L - L)(h(x,t) - h(\hat{x},t))\end{aligned}$$

where $\dot{x} = f(x,t)$ is the true system, $y = h(x,t)$ is the measurement, and we assume that f and h are known. If L is designed to render $\dot{q} = f(q,t)+L(\hat{x},t)(h(x,t)-h(q,t))$ contracting with respect to q [13, 14, 16, 45], we could define the functions of (113) *and* (114) *as follows:*

$$g(q,\varpi,t) = f(q,t) + L(\hat{x},t)(h(x,t) - h(q,t)) \quad (117)$$

$$\Delta_L(q,\varpi,t) = (L_L(\hat{x},t) - L(\hat{x},t))(h(x,t) - h(\hat{x},t)) \quad (118)$$

where $\varpi = [x^\top, \hat{x}^\top]^\top$. It can be seen that (113) *and* (114) *with the relations* (117) *and* (118) *indeed has $q(\mu = 0, t) = x$ and $q(\mu = 1, t) = \hat{x}$ as its particular solutions when perturbed by deterministic and stochastic disturbances as in* (90) *and* (91)*, respectively. We can bound $\|\Delta_L\|$ of* (118) *in the compact set of* (115) *with $\epsilon_{\ell 0} = 0$ and $\epsilon_{\ell 1} \neq 0$ if h is Lipschitz, and with $\epsilon_{\ell 0} \neq 0$ and $\epsilon_{\ell 1} = 0$ if h is bounded in $\mathcal{S}$, using the techniques to be discussed in Remark 5.1. See Theorems 6.2 and 6.3 for details.*

Example 5.3 (System identification). *We can use the problem* (115) *with the systems* (113) *and* (114) *also if f_{true} of the underlying dynamics $\dot{x}^* = f_{\text{true}}(x^*,t)$ is unknown and learned by $\dot{x} = f_L(x,t)$. Since we have*

$$\dot{x}^* = f_{\text{true}}(x^*,t) = f_L(x^*,t) + (f_{\text{true}}(x^*,t) - f_L(x,t))$$

we could define g and Δ_L of (113) *and* (114) *as follows, to have $q(\mu = 0, t) = x$ and $q(\mu = 1, t) = x^*$ as its particular solutions:*

$$g(q,\varpi,t) = f_L(q,t),\ \Delta_L(\varpi,t) = f_{\text{true}}(x^*,t) - f_L(x^*,t) \quad (119)$$

where $\varpi = x^*$, as long as $\dot{x} = f_L(x,t)$ is contracting with respect to x [81, 133]. Since Δ_L of (119) *is the learning error itself, $\|\Delta_L\|$ can be bounded in $\mathcal{S}$ using the techniques of Remark 5.1. See Theorems 9.1 and 9.2 for details.*

Remark 5.1. *As seen in Examples 5.1–5.2, Δ_L of* (115) *is typically given by a learning error, $\phi(z) - \varphi(z)$, multiplied by a bounded or Lipschitz continuous function, where $z = (\varpi, t) \in \mathcal{S}$ for a compact set $\mathcal{S}$, and ϕ is the true computationally-expensive/unknown function to be learned by φ as in* (116)*,* (118)*,* (119)*, etc.*

Let $\mathcal{T}$ be a set of training data $\{(z_i, \phi(z_i))\}_{i=1}^N$ sampled in $\mathcal{S}$. For systems with the true function ϕ and its approximation φ being Lipschitz (e.g., by spectral normalization [132] to be discussed in Definition 6.3 or by using contraction theory as in [62–64]), we can analytically find a bound for $\|\phi(z) - \varphi(z)\|$ as follows if target data samples z are in $B(r) = \{z \in \mathcal{S} | \sup_{z' \in \mathcal{T}} \|z - z'\| \leq r\}$ for some $r \in \mathbb{R}_{\geq 0}$:

$$\sup_{z \in B(r)} \|\phi(z) - \varphi(z)\| \leq \sup_{z' \in \mathcal{T}} \|\phi(z') - \varphi(z')\| + (L_\phi + L_\varphi) r \quad (120)$$

where L_ϕ and L_φ are the Lipschitz constants of ϕ and φ, respectively. The term $\sup_{z' \in \mathcal{T}} \|\phi(z') - \varphi(z')\|$ can then be bounded by a constant, e.g., by using a deep robust regression model as proven in [154, 155] with spectral normalization, under standard training data distribution assumptions.

Deep Neural Networks (DNNs) have been shown to generalize well to the set of unseen events that are from almost the same distribution as their training set [61, 156–159], and consequently, obtaining a tighter and more general upper bound for the learning error as in (120) *has been an active field of research [154, 155, 160, 161]. Thus, the condition* (115) *has become a common assumption in analyzing the performance of learning-based and data-driven control techniques [61, 104, 105, 119, 162–166].*

5.2. Formal Stability Guarantees via Contraction Theory

One drawback of naively using existing learning-based and data-driven control approaches for the perturbed nonlinear systems (113) and (114) without analyzing contraction is that, as shall be seen in the following theorem, we can only guarantee the trajectory error to be bounded by a function that increases exponentially with time.

Theorem 5.1. *Suppose that Δ_L of* (113) *is learned to satisfy* (115)*, and that $g(q,\varpi,t)$ of* (113) *is Lipschitz with respect to q with its 2-norm Lipschitz constant $L_g \in \mathbb{R}_{\geq 0}$, i.e.,*

$$\|g(q,\varpi,t) - g(q',\varpi,t)\| \leq L_g \|q - q'\|,\ \forall (\varpi,t) \in \mathcal{S}$$

where $q, q' \in \mathbb{R}^n$ and $\mathcal{S}$ is the compact set of (115)*. Then we have the following upper bound for all $(\varpi,t) \in \mathcal{S}$:*

$$\|\mathsf{e}(t)\| \leq \|\mathsf{e}(0)\| e^{(L_g + \epsilon_{\ell 1})t} + \frac{\epsilon_{\ell 0} + \bar{d}}{L_g + \epsilon_{\ell 1}} (e^{(L_g + \epsilon_{\ell 1})t} - 1) \quad (121)$$

where $\mathsf{e}(t) = \xi_1(t) - \xi_0(t)$, $\xi_0(t) = q(0,t)$ and $\xi_1(t) = q(1,t)$ for q of (113)*, and $\bar{d} = \sup_{\mu,\varpi,t} \|\partial d / \partial \mu\|$ for d of* (115)*.*

Proof. See the Gronwall-Bellman inequality [3, p. 651] and Theorem 3.4 of [3, pp. 96-97]. □

The bound obtained in Theorem 5.1 is useful in that it gives mathematical guarantees even for naive learning-based frameworks without a contracting property (e.g., it can be used to prove safety in the learning-based Model Predictive Control (MPC) framework [125]). However, the exponential term $e^{(L_g + \epsilon_{\ell 1})t}$ in (121) causes the upper bound to diverge, which could result in more conservative automatic control designs than necessary.

In contrast, contraction theory gives an upper bound on the trajectory tracking error $\|\mathsf{e}(t)\|$ which is exponentially bounded linearly in the learning error, even under the presence of external disturbances [14, 104, 105, 119].

Theorem 5.2. *Let us consider the virtual system of a smooth path $q(\mu,t)$ in* (113) *and suppose that Δ_L of* (113)

is learned to satisfy (115). *If the condition* $\underline{m}\mathrm{I} \preceq M \preceq \overline{m}\mathrm{I}$ *of* (26) *holds and the system* (113) *with* $\Delta_L = 0$ *and* $d = 0$ *is contracting, i.e.,*

$$\dot{M} + M(\partial g/\partial q) + (\partial g/\partial q)^\top M \preceq -2\alpha M$$

of Theorem 2.1 or 2.2 holds for $M \succ 0$ *that defines a contraction metric with the contraction rate* α*, and if the learning error* $\epsilon_{\ell 1}$ *of* (115) *is sufficiently small to satisfy*

$$\exists \alpha_\ell \in \mathbb{R}_{>0} \text{ s.t. } \alpha_\ell = \alpha - \epsilon_{\ell 1}\sqrt{\overline{m}/\underline{m}} > 0 \tag{122}$$

then we have the following bound for all $(\varpi, t) \in \mathcal{S}$*:*

$$\|\mathsf{e}(t)\| \leq \frac{V_\ell(0)}{\sqrt{\underline{m}}} e^{-\alpha_\ell t} + \frac{\epsilon_{\ell 0} + \bar{d}}{\alpha_\ell}\sqrt{\frac{\overline{m}}{\underline{m}}}(1 - e^{-\alpha_\ell t}) \tag{123}$$

where $\mathsf{e}(t) = \xi_1(t) - \xi_0(t)$ *with* $\xi_0(t) = q(0,t)$ *and* $\xi_1(t) = q(1,t)$ *for* q *of* (113), $\bar{d} = \sup_{\mu,\varpi,t} \|\partial d/\partial \mu\|$ *as in* (113), $\mathcal{S}$ *is the compact set of Theorem 5.1,* $\epsilon_{\ell 0}$ *and* $\epsilon_{\ell 1}$ *are the learning errors of* (115)*, and* $V_\ell(t) = \int_{\xi_0}^{\xi_1} \|\Theta(q(t),t)\delta q(t)\|$ *for* $M = \Theta^\top\Theta$ *as in Theorem 2.3.*

Proof. Since (113) with $\Delta_L = 0$ and $d = 0$ is contracting and we have that

$$\begin{aligned}\|\partial(\Delta_L + d)/\partial\mu\| &\leq \epsilon_{\ell 0} + \epsilon_{\ell 1}\|\xi_1 - \xi_0\| + \bar{d}\\ &\leq \epsilon_{\ell 0} + \frac{\epsilon_{\ell 1}}{\sqrt{\underline{m}}}\int_{\xi_0}^{\xi_1} \|\Theta\delta q\| + \bar{d} = \epsilon_{\ell 0} + \frac{\epsilon_{\ell 1}}{\sqrt{\underline{m}}} V_\ell + \bar{d}\end{aligned} \tag{124}$$

for all $\mu \in [0,1]$ and $(\varpi, t) \in \mathcal{S}$ due to $\sup_{\mu,\varpi,t} \|\partial d/\partial\mu\| = \bar{d}$, the direct application of Theorem 2.4 to the system (113), along with the condition (122), yields (123). □

Using Theorem 2.5, we can easily derive a stochastic counterpart of Theorem 5.2 for the system (114) [15, 119].

Theorem 5.3. *Consider the virtual system of a smooth path* $q(\mu, t)$ *in* (114) *and suppose* Δ_L *of* (114) *is learned to satisfy* (115). *If* $\underline{m}\mathrm{I} \preceq M \preceq \overline{m}\mathrm{I}$ *of* (26) *holds and the system* (114) *is contracting, i.e.,*

$$\dot{M} + M(\partial g/\partial q) + (\partial g/\partial q)^\top M \preceq -2\alpha M - \alpha_s \mathrm{I}$$

of Theorem 2.5 holds, and if the learning error $\epsilon_{\ell 1}$ *of* (115) *and an arbitrary constant* $\alpha_d \in \mathbb{R}_{>0}$ *(see* (128)*) are selected to satisfy*

$$\exists \alpha_\ell \in \mathbb{R}_{>0} \text{ s.t. } \alpha_\ell = \alpha - (\alpha_d/2 + \epsilon_{\ell 1}\sqrt{\overline{m}/\underline{m}}) > 0 \tag{125}$$

then we have the following bound for all $(\varpi, t) \in \mathcal{S}$*:*

$$\mathbb{E}\left[\|\mathsf{e}(t)\|^2\right] \leq \frac{\mathbb{E}[V_{s\ell}(0)]}{\underline{m}} e^{-2\alpha_\ell t} + \frac{C}{2\alpha_\ell}\frac{\overline{m}}{\underline{m}} \tag{126}$$

where $\mathsf{e}(t) = \xi_1(t) - \xi_0(t)$ *with* $\xi_0(t) = q(0,t)$ *and* $\xi_1(t) = q(1,t)$ *for* q *of* (114), $\mathcal{S}$ *is the compact set given in Theorem 5.1,* $\epsilon_{\ell 0}$ *and* $\epsilon_{\ell 1}$ *are the learning errors of* (115), $V_{s\ell}(t) = \int_{\xi_0}^{\xi_1} \|\Theta(q(t),t)\delta q(t)\|^2$ *for* $M = \Theta^\top\Theta$ *as given in Theorem 2.3, and* $C = \bar{g}^2(2{\alpha_G}^{-1} + 1) + \epsilon_{\ell 0}^2\alpha_d^{-1}$ *for* $\bar{g} = \sup_{\mu,\varpi,t} \|\partial G/\partial\mu\|_F$ *of* (114) *with an arbitrary constant* $\alpha_G \in \mathbb{R}_{>0}$ *as in Theorem 2.5. Furthermore, the probability that* $\|\mathsf{e}\|$ *is greater than or equal to* $\varepsilon \in \mathbb{R}_{>0}$ *is given as*

$$\mathbb{P}\left[\|\mathsf{e}(t)\| \geq \varepsilon\right] \leq \frac{1}{\varepsilon^2}\left(\frac{\mathbb{E}[V_{s\ell}(0)]}{\underline{m}} e^{-2\alpha_\ell t} + \frac{C}{2\alpha_\ell}\frac{\overline{m}}{\underline{m}}\right). \tag{127}$$

Proof. Computing $\mathscr{L}V_{s\ell}$ with the virtual system (114) as in the proof of Theorem 2.5, we get an additional term $2\int_0^1 \partial_\mu q^\top M\Delta_L$. Using the learning error assumption (115) to have $\|\Delta_L\| \leq \epsilon_{\ell 0} + (\epsilon_{\ell 1}/\sqrt{\underline{m}})\int_0^1 \|\Theta\partial_\mu q\| d\mu$ as in (124), we have that

$$\begin{aligned}2\int_0^1 \partial_\mu q^\top M\Delta_L \leq& 2\sqrt{\overline{m}}V_\ell(\epsilon_{\ell 0} + (\epsilon_{\ell 1}/\sqrt{\underline{m}})V_\ell)\\ \leq& \alpha_d^{-1}\epsilon_{\ell 0}^2\overline{m} + (\alpha_d + 2\epsilon_{\ell 1}\sqrt{\overline{m}/\underline{m}})V_\ell^2\end{aligned} \tag{128}$$

where $V_\ell = \int_0^1 \|\Theta\partial_\mu q\| d\mu$ as in Theorem 2.3, and the relation $2ab \leq \alpha_d^{-1}a^2 + \alpha_d b^2$, which holds for any $a, b \in \mathbb{R}$ and $\alpha_d \in \mathbb{R}_{>0}$ ($a = \epsilon_{\ell 0}\sqrt{\overline{m}}$ and $b = V_\ell$ in this case), is used to obtain the second inequality. Since we have $V_\ell^2 \leq V_{s\ell}$ as proven in Theorem 2.3, selecting α_d and $\epsilon_{\ell 1}$ sufficiently small to satisfy (125) gives the desired relations (126) and (127) due to Theorem 2.5. □

As discussed in Theorems 5.2 and 5.3, using contraction theory for learning-based and data-driven control, we can formally guarantee the system trajectories to stay in a tube with an exponentially convergent bounded radius centered around the target trajectory, even with the external disturbances and learning errors $\epsilon_{\ell 0}$ and $\epsilon_{\ell 1}$. The exponential bounds (123), (126), and (127) become tighter as we achieve smaller $\epsilon_{\ell 0}$ and $\epsilon_{\ell 1}$ using more training data for verifying (115) (see Remark 5.1). It is also worth noting that the steady-state bounds of (123) and (126) are again some functions of the condition number $\chi = \overline{m}/\underline{m}$ as in Theorems 4.1 and 4.4, which renders the aforementioned CV-STEM approach valid and effective also in the learning-based frameworks. Theorems 5.2 and 5.3 play a central role in providing incremental exponential stability guarantees for learning-based and data-driven automatic control to be introduced in Sec. 6–8.

5.3. Motivation for Neural Contraction Metrics

The CV-STEM schemes in Theorems 4.2, 4.5, and 4.6 permit the construction of optimal contraction metrics via convex optimization for synthesizing optimal, provably stable, and robust feedback control and state estimation. It is also shown in Theorems 5.2 and 5.3 that these contraction metrics are useful for obtaining stability guarantees for machine learning-based and data-driven control techniques while retaining the CV-STEM-type optimality due to the tracking error bounds (123) and (126), which are comparable to those in Theorems 4.1 and 4.4. However,

the CV-STEM requires that a nonlinear system of equations or an optimization problem be solved at each time instant, which is not suitable for systems with limited online computational capacity.

The Neural Contraction Metric (NCM) and its extensions [14, 15, 104, 105] have been developed to address such a computational issue by modeling the CV-STEM optimization scheme using a DNN [106–108], making it implementable in real-time. We will see in the subsequent sections that Theorems 5.2 and 5.3 provide one formal approach to prove robustness and stability properties of the NCM methodologies [119].

6. Learning-based Robust Control and Estimation

Let us again consider the systems (55)–(57) for control, i.e.,

$$\dot{x} = f(x,t) + B(x,t)u + d_c(x,t) \tag{129}$$
$$dx = (f(x,t) + B(x,t)u)dt + G_c(x,t)d\mathscr{W}(t) \tag{130}$$
$$\dot{x}_d = f(x_d,t) + B(x_d,t)u_d \tag{131}$$

with $\sup_{x,t}\|d_c\| = \bar{d}_c \in \mathbb{R}_{\geq 0}$ and $\sup_{x,t}\|G_c\|_F = \bar{g}_c \in \mathbb{R}_{\geq 0}$, and (90)–(91) for estimation, i.e.,

$$\dot{x} = f(x,t) + d_{e0}(x,t),\ y = h(x,t) + d_{e1}(x,t) \tag{132}$$
$$dx = f(x,t)dt + G_{e0}d\mathscr{W}_0,\ ydt = h(x,t)dt + G_{e1}d\mathscr{W}_1 \tag{133}$$

with $\sup_{x,t}\|d_{e0}\| = \bar{d}_{e0} \in \mathbb{R}_{\geq 0}$, $\sup_{x,t}\|d_{e1}\| = \bar{d}_{e1} \in \mathbb{R}_{\geq 0}$, $\sup_{x,t}\|G_{e0}\|_F = \bar{g}_{e0} \in \mathbb{R}_{\geq 0}$, and $\sup_{x,t}\|G_{e1}\|_F = \bar{g}_{e1} \in \mathbb{R}_{\geq 0}$. The following definitions of a DNN [106–108] and NCM [14, 15, 119] will be utilized extensively hereinafter.

Definition 6.1. *A neural network is a nonlinear mathematical model to approximately represent training samples $\{(x_i, y_i)\}_{i=1}^N$ of $y = \phi(x)$ by optimally tuning its hyperparameters W_ℓ, and is given as follows:*

$$y_i = \varphi(x_i; W_\ell) = T_{L+1} \circ \sigma \circ T_L \circ \cdots \circ \sigma \circ T_1(x_i) \tag{134}$$

where $T_\ell(x) = W_\ell x$, $\circ$ denotes composition of functions, and σ is an activation function (e.g. $\sigma(x) = \tanh(x)$, note that $\varphi(x)$ is smooth in this case). We call a neural network that has more than two layers (i.e., $L \geq 2$) a Deep Neural Network (DNN).

Definition 6.2. *A Neural Contraction Metric (NCM) is a DNN model of a contraction metric given in Theorem 2.1. Its training data could be sampled by solving the CV-STEM presented in Theorem 4.2 for control (or Theorem 4.6 for differential feedback control) and Theorem 4.5 for estimation. The NCM framework is summarized in Fig. 3.*

Remark 6.1. *Although a stochastic version of the NCM is called a Neural Stochastic Contraction Metric (NSCM) in [15], we also denote it as an NCM in this paper for simplicity of presentation.*

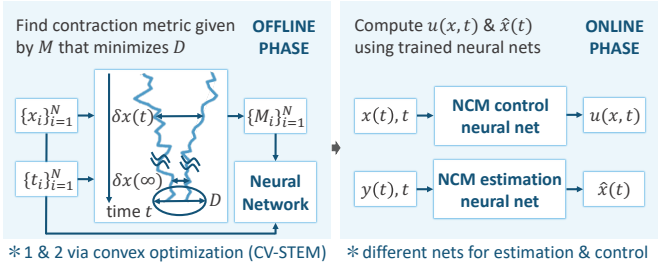

Figure 3: Illustration of NCM (x: system state; M: positive definite matrix that defines optimal contraction metric; x_i and M_i: sampled x and M; $\hat{x}$: estimated system state; y: measurement; and u: system control input). Note that the target trajectory (x_d, u_d) is omitted in the figure fore simplicity.

Since the NCM only requires one function evaluation at each time instant to get M that defines a contraction metric, without solving any optimization problems unlike the CV-STEM approaches, it enables real-time optimal feedback control and state estimation in most engineering and scientific applications.

6.1. Stability of NCM-based Control and Estimation

The NCM exhibits superior incremental robustness and stability due to its internal contracting property.

Theorem 6.1. *Let $\mathcal{M}$ define the NCM in Definition 6.2, and let $u^* = u_d - R^{-1}B^\top M(x - x_d)$ for M and R given in Theorem 4.2. Suppose that the systems (129) and (130) are controlled by u_L, which computes the CV-STEM controller u^* of Theorem 4.2 replacing M by $\mathcal{M}$, i.e.,*

$$u_L = u_d - R(x, x_d, u_d, t)^{-1}B(x,t)^\top \mathcal{M}(x, x_d, u_d, t)\mathtt{e} \tag{135}$$

for $\mathtt{e} = x - x_d$. Note that we could use the differential feedback control $u = u_d + \int_0^1 k d\mu$ of Theorem 4.6 for u^* and u_L. Define Δ_L of Theorems 5.2 and 5.3 as follows:

$$\Delta_L(\varpi, t) = B(x,t)(u_L(x, x_d, u_d, t) - u^*(x, x_d, u_d, t)) \tag{136}$$

where we use $q(0,t) = \xi_0(t) = x_d(t)$, $q(1,t) = \xi_1(t) = x(t)$, and $\varpi = [x^\top, x_d^\top, u_d^\top]^\top$ in the learning-based control formulation (113) and (114). If the NCM is learned to satisfy the learning error assumptions of Theorems 5.2 and 5.3 for Δ_L given by (136), then (123), (126), and (127) of Theorems 5.2 and 5.3 hold as follows:

$$\|\mathtt{e}(t)\| \leq \frac{V_\ell(0)}{\sqrt{\underline{m}}}e^{-\alpha_\ell t} + \frac{\epsilon_{\ell 0} + \bar{d}_c}{\alpha_\ell}\sqrt{\frac{\overline{m}}{\underline{m}}}(1 - e^{-\alpha_\ell t}) \tag{137}$$

$$\mathbb{E}\left[\|\mathtt{e}(t)\|^2\right] \leq \frac{\mathbb{E}[V_{s\ell}(0)]}{\underline{m}}e^{-2\alpha_\ell t} + \frac{C_C}{2\alpha_\ell}\frac{\overline{m}}{\underline{m}} \tag{138}$$

$$\mathbb{P}\left[\|\mathtt{e}(t)\| \geq \varepsilon\right] \leq \frac{1}{\varepsilon^2}\left(\frac{\mathbb{E}[V_{s\ell}(0)]}{\underline{m}}e^{-2\alpha_\ell t} + \frac{C_C}{2\alpha_\ell}\frac{\overline{m}}{\underline{m}}\right) \tag{139}$$

where $\mathtt{e} = x - x_d$, $C_C = \bar{g}_c^2(2\alpha_G{}^{-1} + 1) + \epsilon_{\ell 0}^2\alpha_d^{-1}$, the disturbance upper bounds $\bar{d}_c$ and $\bar{g}_c$ are given in (129) and (130), respectively, and the other variables are defined in Theorems 4.2, 5.2 and 5.3.

PROOF. Let g of the virtual systems given in Theorems 5.2 and 5.3 be $g = \zeta$, where ζ is given by (67) (i.e., $\zeta = (A - BR^{-1}B^\top M)(q - x_d) + \dot{x}_d$). By definition of Δ_L, this has $q = x$ and $q = x_d$ as its particular solutions (see Example 5.1). As defined in the proof of Theorems 3.1, we have $d = \mu d_c$ and $G = \mu G_c$ for these virtual systems, resulting in the upper bounds of disturbances in the learning-based control formulation (123) and (126) given as

$$\bar{d} = \bar{d}_c \text{ and } \bar{g} = \bar{g}_c. \tag{140}$$

Since M and u^* are constructed to render $\dot{q} = \zeta(q, \varpi, t)$ contracting as presented in Theorem 4.2, and we have $\|\Delta_L\| \leq \epsilon_{\ell 0} + \epsilon_{\ell 1}\|\xi_1 - \xi_0\|$ due to the learning assumption (115), Theorem 5.2 implies the deterministic bound (137) and Theorem 5.3 implies the stochastic bounds (138) and (139). Note that using the differential feedback control of Theorem 4.6 as u^* also gives the bounds (137)–(139) following the same argument [119]. □

Due to the estimation and control duality in differential dynamics observed in Sec. 4, we have a similar result to Theorem 6.1 for the NCM-based state estimator.

Theorem 6.2. *Let $\mathcal{M}$ define the NCM in Definition 6.2, and let $L(\hat{x}, t) = M\bar{C}^\top R^{-1}$ for M, $\bar{C}$ and R given in Theorem 4.5. Suppose that the systems* (132) *and* (133) *are estimated with an estimator gain L_L, which computes L of the CV-STEM estimator $\dot{\hat{x}} = f(\hat{x}, t) + L(\hat{x}, t)(y - h(\hat{x}, t))$ in Theorem 4.5 replacing M by $\mathcal{M}$, i.e.,*

$$\begin{aligned}\dot{\hat{x}} =& f(\hat{x}, t) + L_L(\hat{x}, t)(y - h(\hat{x}, t)) \\ =& f(\hat{x}, t) + \mathcal{M}(\hat{x}, t)\bar{C}(\varrho_c, \hat{x}, t)^\top R(\hat{x}, t)^{-1}(y - h(\hat{x}, t))\end{aligned} \tag{141}$$

Define Δ_L of Theorems 5.2 and 5.3 as follows:

$$\Delta_L(\varpi, t) = (L_L(\hat{x}, t) - L(\hat{x}, t))(h(x, t) - h(\hat{x}, t)) \tag{142}$$

where we use $q(0, t) = \xi_0(t) = x(t)$, $q(1, t) = \xi_1(t) = \hat{x}(t)$, and $\varpi = [x^\top, \hat{x}^\top]^\top$ in the learning-based control formulation (113) *and* (114)*. If the NCM is learned to satisfy the learning error assumptions of Theorems 5.2 and 5.3 for Δ_L given by* (142)*, then* (123)*,* (126)*, and* (127) *of Theorems 5.2 and 5.3 hold as follows:*

$$\|\mathsf{e}(t)\| \leq \sqrt{\overline{m}} V_\ell(0) e^{-\alpha_\ell t} + \frac{\bar{d}_a \sqrt{\frac{\overline{m}}{\underline{m}}} + \bar{d}_b \overline{m}}{\alpha_\ell}(1 - e^{-\alpha_\ell t}) \tag{143}$$

$$\mathbb{E}\left[\|\mathsf{e}(t)\|^2\right] \leq \overline{m}\mathbb{E}[V_{s\ell}(0)] e^{-2\alpha_\ell t} + \frac{C_E}{2\alpha_\ell}\frac{\overline{m}}{\underline{m}} \tag{144}$$

$$\mathbb{P}\left[\|\mathsf{e}(t)\| \geq \varepsilon\right] \leq \frac{1}{\varepsilon^2}\left(\overline{m}\mathbb{E}[V_{s\ell}(0)] e^{-2\alpha_\ell t} + \frac{C_E}{2\alpha_\ell}\frac{\overline{m}}{\underline{m}}\right) \tag{145}$$

where $\mathsf{e} = \hat{x} - x$, $\bar{d}_a = \epsilon_{\ell 0} + \bar{d}_{e0}$, $\bar{d}_b = \bar{\rho}\bar{c}\bar{d}_{e1}$, $C_E = (\bar{g}_{e0}^2 + \bar{\rho}^2\bar{c}^2\bar{g}_{e1}^2\overline{m}^2)(2{\alpha_G}^{-1} + 1) + \epsilon_{\ell 0}^2\alpha_d^{-1}$, the disturbance upper bounds $\bar{d}_{e0}$ $\bar{d}_{e1}$, $\bar{g}_{e0}$, and $\bar{g}_{e1}$ are given in (132) *and* (133)*, and the other variables are defined in Theorems 4.5, 5.2 and 5.3.*

PROOF. Let g of the virtual systems given in Theorems 5.2 and 5.3 be $g = \zeta$, where ζ is given by (97) (i.e., $\zeta = (A - LC)(q - x) + f(x, t)$). By definition of Δ_L, this has $q = \hat{x}$ and $q = x$ as its particular solutions (see Example 5.2). As defined in the proof Theorem 4.3, we have

$$d = (1 - \mu)d_{e0} + \mu L_L d_{e1} \text{ and } G = [(1 - \mu)G_{e0}, \mu L_L G_{e1}]$$

for these virtual systems, resulting in the upper bounds of external disturbances in the learning-based control formulation (123) and (126) given as

$$\bar{d} = \bar{d}_{e0} + \bar{\rho}\bar{c}\bar{d}_{e1}\sqrt{\overline{m}\underline{m}} \text{ and } \bar{g} = \sqrt{\bar{g}_{e0}^2 + \bar{\rho}^2\bar{c}^2\bar{g}_{e1}^2\overline{m}^2}.$$

Since $W = M^{-1}$ is constructed to render $\dot{q} = \zeta(q, \varpi, t)$ contracting as presented in Theorem 4.5, and we have $\|\Delta_L\| \leq \epsilon_{\ell 0} + \epsilon_{\ell 1}\|\xi_1 - \xi_0\|$ due to the learning assumption (115), Theorem 5.2 implies the bound (143), and Theorem 5.3 implies the bounds (144) and (145). □

Remark 6.2. *As discussed in Examples 5.1 and 5.2 and in Remark 5.1, we have $\epsilon_{\ell 1} = 0$ for the learning error $\epsilon_{\ell 1}$ of* (115) *as long as B and h are bounded a compact set by the definition of Δ_L in* (136) *and* (142)*. If h and u^* are Lipschitz with respect to x in the compact set, we have non-zero $\epsilon_{\ell 1}$ with $\epsilon_{\ell 0} = 0$ in* (115) *(see Theorem 6.3).*

Using Theorems 6.1 and 6.2, we can relate the learning error of the matrix that defines the NCM, $\|\mathcal{M} - M\|$, to the error bounds (137)–(139) and (143)–(145), if M is given by the CV-STEM of Theorems 4.2 and 4.5.

Theorem 6.3. *Let $\mathcal{M}$ define the NCM in Definition 6.2, and let $\mathcal{S}_s \subseteq \mathbb{R}^n$, $\mathcal{S}_u \subseteq \mathbb{R}^m$, and $t \in \mathcal{S}_t \subseteq \mathbb{R}_{\geq 0}$ be some compact sets. Suppose that the systems* (129)–(133) *are controlled by* (135) *and estimated by* (141)*, respectively, as in Theorems 6.1 and 6.2. Suppose also that $\exists \bar{b}, \bar{c}, \bar{\rho} \in \mathbb{R}_{\geq 0}$ s.t. $\|B(x, t)\| \leq \bar{b}$, $\|C(\varrho_c, x, \hat{x}, t)\| \leq \bar{c}$, and $\|R^{-1}\| \leq \bar{\rho}$, $\forall x, \hat{x} \in \mathcal{S}_s$ and $t \in \mathcal{S}_t$, for B in* (129)*, C in Theorem 4.5, and R in* (135) *and* (141)*. If we have for all $x, x_d, \hat{x} \in \mathcal{S}_s$, $u_d \in \mathcal{S}_u$, and $t \in \mathcal{S}_t$ that*

$$\begin{aligned}\|\mathcal{M}(x, x_d, u_d, t) - M(x, x_d, u_d, t)\| \leq& \epsilon_\ell \text{ for control} \\ \|\mathcal{M}(\hat{x}, t) - M(\hat{x}, t)\| \leq& \epsilon_\ell \text{ for estimation}\end{aligned} \tag{146}$$

where $M(x, x_d, u_d, t)$ and $M(\hat{x}, t)$ (or $M(x, t)$ and $M(\hat{x}, t)$, see Theorem 3.2) are the CV-STEM solutions of Theorems 4.2 and 4.5 with $\exists \epsilon_\ell \in \mathbb{R}_{\geq 0}$ being the learning error, then the bounds (137)–(139) *of Theorem 6.1 and* (143)–(145) *of Theorem 6.2 hold with $\epsilon_{\ell 0} = 0$ and $\epsilon_{\ell 1} = \bar{\rho}\bar{b}^2\epsilon_\ell$ for control, and $\epsilon_{\ell 0} = 0$ and $\epsilon_{\ell 1} = \bar{\rho}\bar{c}^2\epsilon_\ell$ for estimation, as long as ϵ_ℓ is sufficiently small to satisfy the conditions* (122) *and* (125) *of Theorems 5.2 and 5.3 for deterministic and stochastic systems, respectively.*

PROOF. For Δ_L defined in (136) and (142) with the controllers and estimators given as $u^* = u_d - R^{-1}B^\top M(x -$

$x_d)$, $u_L = u_d - R^{-1}B^\top \mathcal{M}(x - x_d)$, $L = MC^\top R^{-1}$, and $L_L = \mathcal{M}C^\top R^{-1}$, we have the following upper bounds:

$$\|\Delta_L\| \leq \begin{cases} \bar{\rho}\bar{b}^2\epsilon_\ell \|x - x_d\| & \text{(controller)} \\ \bar{\rho}\bar{c}^2\epsilon_\ell \|\hat{x} - x\| & \text{(estimator)} \end{cases}$$

where the relation $\|h(x,t) - h(\hat{x},t)\| \leq \bar{c}\|\hat{x} - x\|$, which follows from the equality $h(x,t) - h(\hat{x},t) = C(\varrho_c, x, \hat{x}, t)(x - \hat{x})$ (see Lemma 3.1), is used to obtain the second inequality. This implies that we have $\epsilon_{\ell 0} = 0$ and $\epsilon_{\ell 1} = \bar{\rho}\bar{b}^2\epsilon_\ell$ for control, and $\epsilon_{\ell 0} = 0$ and $\epsilon_{\ell 1} = \bar{\rho}\bar{c}^2\epsilon_\ell$ for estimation in the learning error of (115). The rest follows from Theorems 6.1 and 6.2. □

Example 6.1. *The left-hand side of Fig. 4 shows the state estimation errors of the following Lorenz oscillator perturbed by process noise d_0 and measurement noise d_1 [14]:*

$$\begin{aligned} \dot{x} =& [\sigma(x_2 - x_1), x_1(\rho - x_3) - x_2, x_1x_2 - \beta x_3]^\top + d_0(x,t) \\ y =& [1,0,0]x + d_2(x,t) \end{aligned}$$

where $x = [x_1, x_2, x_3]^\top$, $\sigma = 10$, $\beta = 8/3$, and $\rho = 28$. It can be seen that the steady-state estimation errors of the NCM and CV-STEM are below the optimal steady-state upper bound of Theorem 4.5 (dotted line in the left-hand side of Fig. 4) while the EKF has a larger error compared to the other two. In this example, training data is sampled along 100 trajectories with uniformly distributed initial conditions ($-10 \leq x_i \leq 10$, $i = 1,2,3$) using Theorem 4.5, and then modeled by a DNN as in Definition 6.2. Additional implementation and network training details can be found in [14].

Example 6.2. *Consider a robust feedback control problem of the planar spacecraft perturbed by deterministic disturbances, the unperturbed dynamics of which is given as follows [167]:*

$$\begin{bmatrix} m & 0 & 0 \\ 0 & m & 0 \\ 0 & 0 & I \end{bmatrix} \ddot{x} = \begin{bmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & -1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & -1 & -1 & 0 & 0 & 1 & 1 \\ -\ell & \ell & -b & b & -\ell & \ell & -b & b \end{bmatrix} u$$

where $x = [p_x, p_y, \phi, \dot{p}_x, \dot{p}_y, \dot{\phi}]^T$ with p_x, p_y, and ϕ being the horizontal coordinate, vertical coordinate, and yaw angle of the spacecraft, respectively, with the other variables defined as m = mass of spacecraft, I = yaw moment of inertia, u = thruster force vector, ℓ = half-depth of spacecraft, and b = half-width of spacecraft (see Fig. 8 of [167] for details).

As shown in the right-hand side of Fig. 4, the NCM keeps their trajectories within the bounded error tube of Theorem 4.2 (shaded region) around the target trajectory (dotted line) avoiding collision with the circular obstacles, even under the presence of disturbances, whilst requiring much smaller computational cost than the CV-STEM. Data sampling and the NCM training are performed as in Example 6.1 [14].

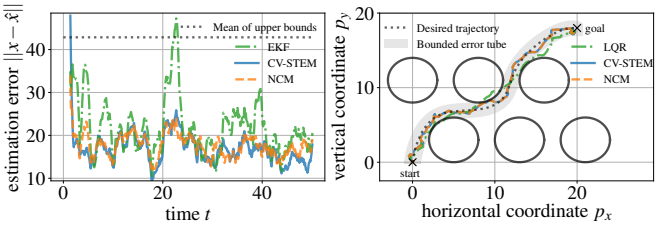

Figure 4: Lorenz oscillator state estimation error smoothed using a 15-point moving average filter in Example 6.1 (left), and spacecraft motion (p_x, p_y) on a planar field in Example 6.2 (right).

Remark 6.3. *We could also simultaneously synthesize a controller u and the NCM, and Theorem 6.1 can provide formal robustness and stability guarantees even for such cases [119, 120]. In Sec. 7, we generalize the NCM to learn contraction theory-based robust control using a DNN only taking x, t, and a vector containing local environment information as its inputs, to avoid the online computation of x_d for the sake of automatic guidance and control implementable in real-time.*

6.2. NCMs as Lyapunov Functions

We could also utilize the NCM constructed by the CV-STEM contraction metric for designing a Control Lyapunov function (CLF) [3, 70, 71], resulting in a stability result different from Theorems 6.1 and 6.3. To this end, let us recall that designing optimal k of $u = k(x, x_d, u_d, t)$ reduces to designing optimal $K(x, x_d, u_d, t)$ of $u = u_d - K(x, x_d, u_d, t)(x - x_d)$ due to Lemma 3.2 of Sec. 3. For such u, the virtual system of (129)/(130) and (131) without any perturbation, which has $q = x$ and $q = x_d$ as its particular solutions, is given as follows:

$$\begin{aligned} \dot{q} =& (A(\varrho, x, x_d, u_d, t) - B(x,t)K(x, x_d, u_d, t))(q - x_d) \\ &+ f(x_d, t) + B(x_d, t)u_d \end{aligned} \tag{147}$$

where A is the SDC matrix of the system (129) given by Lemma 3.1, i.e., $A(\varrho, x, x_d, u_d, t)(x - x_d) = f(x,t) + B(x,t)u_d - f(x_d,t) - B(x_d,t)u_d$. Using the result of Theorem 2.3, one of the Lyapunov functions for the virtual system (147) with unknown K can be given as the following transformed squared length integrated over x of (129)/(130) and x_d of (131):

$$V_{\text{NCM}} = \int_{x_d}^{x} \delta q^\top \mathcal{M}(x, x_d, u_d, t)\delta q \tag{148}$$

where $\mathcal{M}$ defines the NCM of Definition 6.2 modeling M of the CV-STEM contraction metric in Theorem 4.2.

Theorem 6.4. *Suppose* (129) *and* (130) *are controlled by $u = u_d - K^*(x, x_d, u_d, t)\mathtt{e}$, where $\mathtt{e} = x - x_d$, and that K^* is designed by the following convex optimization for given*

(x, x_d, u_d, t):

$$(K^*, p^*) = \arg \min_{K \in \mathcal{C}_K, p \in \mathbb{R}} \|K\|_F^2 + p^2 \tag{149}$$

$$\begin{aligned} s.t.\ & \int_{x_d}^{x} \delta q^\top (\dot{\mathcal{M}} + 2\,\mathrm{sym}\,(\mathcal{M}A - \mathcal{M}BK))\delta q \\ & \leq -2\alpha \int_{x_d}^{x} \delta q^\top (\mathcal{M} + \beta \mathrm{I})\delta q + p \end{aligned} \tag{150}$$

where $\mathcal{M}$ defines the NCM of Definition 6.2 that models M of the CV-STEM contraction metric in Theorem 4.2 as in (146)*, $\mathcal{C}_K$ is a convex set containing admissible K, α is the contraction rate, and β is as defined in theorem 3.1 (i.e., $\beta = 0$ for deterministic systems and $\beta = \alpha_s$ for stochastic systems). Then Theorems 2.4 and 2.5 still hold for the Lyapunov function defined in* (148)*, yielding the following bounds:*

$$\|\mathsf{e}(t)\|^2 \leq \frac{V_{\mathrm{NCM}}(0)}{\underline{m}} e^{-2\alpha_p t} + \frac{\frac{\bar{p}}{\underline{m}} + \frac{\bar{d}_c^2 \overline{m}}{\alpha_d \underline{m}}}{2\alpha_p}(1 - e^{-2\alpha_p t}) \tag{151}$$

$$\mathbb{E}\left[\|\mathsf{e}(t)\|^2\right] \leq \frac{\mathbb{E}[V_{\mathrm{NCM}}(0)]}{\underline{m}} e^{-2\alpha t} + \frac{C_C \overline{m} + \bar{p}}{2\alpha \underline{m}} \tag{152}$$

$$\mathbb{P}\left[\|\mathsf{e}(t)\| \geq \varepsilon\right] \leq \frac{1}{\varepsilon^2}\left(\frac{\mathbb{E}[V_{\mathrm{NCM}}(0)]}{\underline{m}} e^{-2\alpha t} + \frac{C_C \overline{m} + \bar{p}}{2\alpha \underline{m}}\right) \tag{153}$$

where $\bar{p} = \sup_{x, x_d, u_d, t} p$, $\underline{m}\mathrm{I} \preceq \mathcal{M} \preceq \overline{m}\mathrm{I}$ as in (26)*, $\alpha_d \in \mathbb{R}_{>0}$ is an arbitrary positive constant (see* (154)*) selected to satisfy $\alpha_p = \alpha - \alpha_d/2 > 0$, $C_C = \bar{g}_c^2(2\alpha_G{}^{-1} + 1)$, $\alpha_G \in \mathbb{R}_{>0}$ is an arbitrary constant as in Theorem 2.5, $\varepsilon \in \mathbb{R}_{\geq 0}$, and the disturbance terms $\bar{d}_c$ and $\bar{g}_c$ are given in* (129) *and* (130)*, respectively. Note that the problem* (149) *is feasible due to the constraint relaxation variable p.*

Furthermore, if $\epsilon_\ell = 0$ in (146)*,* (149) *with $p = 0$ is always feasible, and the optimal feedback gain K^* minimizes its Frobenius norm under the contraction constraint* (150)*.*

Proof. Computing $\dot{V}_{\mathrm{NCM}}$ for the deterministic system (129) along with the virtual dynamics (147) and the stability condition (150) yields

$$\begin{aligned} \dot{V}_{\mathrm{NCM}} &\leq -2\alpha V_{\mathrm{NCM}} + \bar{d}_c \sqrt{\overline{m}} \sqrt{V_{\mathrm{NCM}}} + \bar{p} \\ &\leq -2(\alpha - \alpha_d/2) V_{\mathrm{NCM}} + \bar{d}_c^2 \overline{m}/\alpha_d + \bar{p} \end{aligned} \tag{154}$$

where the relation $2ab \leq \alpha_d^{-1} a^2 + \alpha_d b^2$, which holds for any $a, b \in \mathbb{R}$ and $\alpha_d \in \mathbb{R}_{>0}$ ($a = \bar{d}_c \sqrt{\overline{m}}$ and $b = \sqrt{V_{\mathrm{NCM}}}$ in this case), is used to obtain the second inequality. Since we can arbitrarily select α_d to have $\alpha_p = \alpha - \alpha_d/2 > 0$, (154) gives the exponential bound (151) due to the comparison lemma of Lemma 2.1. Similarly, computing $\mathscr{L} V_{\mathrm{NCM}}$ for the stochastic system (130) yields

$$\mathscr{L} V_{\mathrm{NCM}} \leq -2\alpha V_{\mathrm{NCM}} + C_c \overline{m} + \bar{p}$$

which gives (152) and (153) due to Lemma 2.2. If $\epsilon_\ell = 0$ in (146), $K = R^{-1} B^\top M$ satisfies (150) with $p = 0$ as we have $\mathcal{M} = M$ and M defines the contraction metric of Theorem 4.2, which implies that (149) with $p = 0$ is always feasible. □

We could also use the CCM-based differential feedback formulation of Theorem 4.6 in Theorem 6.4 (see [79, 80]). To this end, define E_{NCM} (Riemannian energy) as follows:

$$E = \int_0^1 \gamma_\mu(\mu, t)^\top \mathcal{M}(\gamma(\mu, t), t) \gamma_\mu(\mu, t) d\mu \tag{155}$$

where $\mathcal{M}$ defines the NCM of Definition 6.2 that models M of the CCM in Theorem 4.6, γ is the minimizing geodesic connecting $x(t) = \gamma(1, t)$ of (129) and $x_d(t) = \gamma(0, t)$ of (131) (see Theorem 2.3), and $\gamma_\mu = \partial \gamma / \partial \mu$.

Theorem 6.5. *Suppose $u = u^*$ of* (129) *is designed by the following convex optimization for given (x, x_d, u_d, t):*

$$(u^*, p^*) = \arg \min_{u \in \mathcal{C}_u, p \in \mathbb{R}} \|u - u_d\|^2 + p^2 \tag{156}$$

$$\begin{aligned} s.t.\ & \frac{\partial E}{\partial t} + 2\gamma_\mu(1, t)^\top \mathcal{M}(x, t)(f(x, t) + B(x, t)u) \\ & - 2\gamma_\mu(0, t)^\top \mathcal{M}(x_d, t)(f(x_d, t) + B(x_d, t)u_d) \leq -2\alpha E + p \end{aligned} \tag{157}$$

where $\mathcal{M}$ defines the NCM of Definition 6.2 that models the CCM of Theorem 4.6, and $\mathcal{C}_u$ is a convex set containing admissible u. Then the bound (151) *of Theorem 6.4 holds.*

Furthermore, if $\epsilon_\ell = 0$ for the NCM learning error ϵ_ℓ in (146)*, then the problem* (156) *with $p = 0$ is always feasible, and u^* minimizes the deviation of u from u_d under the contraction constraint* (157)*.*

Proof. Since the right-hand side of (157) represents $\dot{E}$ of (155) if $d_c = 0$ in (129), the comparison lemma of Lemma 2.1 gives (151) as in Theorem 6.4. The rest follows from the fact that u of Theorem 4.6 is a feasible solution of (156) if $p = 0$ [79, 80].

Remark 6.4. *As discussed in Remark 4.2, Theorem 6.5 can also be formulated in a stochastic setting as in Theorem 6.4 using the technique discussed in [119].*

Theorems 6.4 and 6.5 provide a perspective on the NCM stability different from Theorems 6.1 and 6.3 written in terms of the constraint relaxation variable p, by directly using the NCM as a contraction metric instead of the CV-STEM contraction metric or the CCM. Since the CLF problems are feasible with $p = 0$ when we have zero NCM modeling error, it is implied that the upper bound of p (i.e., $\bar{p}$ in (151) and (152)) decreases as we achieve a smaller learning error ϵ_ℓ using more training data for verifying (146) (see Remark 5.1). This intuition will be formalized later in Theorem 9.3 of Sec. 9 [133]. See [79, 80] for how we compute the minimizing geodesic γ and its associated Riemannian energy E in Theorem 6.5.

Example 6.3. *If $\mathcal{C}_K = \mathbb{R}^{m \times n}$ and $\mathcal{C}_u = \mathbb{R}^m$ in* (149) *and* (156)*, respectively, they can both be expressed as the following quadratic optimization problem:*

$$\begin{aligned} & v^* = \arg \min_{v \in \mathbb{R}^v} \frac{1}{2} v^\top v + c(x, x_d, u_d, t)^\top v \\ & s.t.\ \varphi_0(x, x_d, u_d, t) + \varphi_1(x, x_d, u_d, t)^\top v \leq 0 \end{aligned}$$

by defining $c \in \mathbb{R}^{\mathrm{v}}$, $\varphi_0 \in \mathbb{R}$, $\varphi_1 \in \mathbb{R}^{\mathrm{v}}$, and $v \in \mathbb{R}^{\mathrm{v}}$ appropriately. Applying the KKT condition [153, pp. 243-244] to this problem, we can show that

$$v^* = \begin{cases} -c - \frac{\varphi_0 - \varphi_1^\top c}{\varphi_1^\top \varphi_1}\varphi_1 & \text{if } \varphi_0 - \varphi_1^\top c > 0 \\ -c & \text{otherwise.} \end{cases}$$

This implies that the controller utilizes u_d with the feedback term $((\varphi_0 - \varphi_1^\top c)/\varphi_1^\top \varphi_1)\varphi_1$ only if $\dot{x} = f(x,t) + B(x,t)u$ with $u = u_d$ is not contracting, i.e., $\varphi_0 - \varphi_1^\top c > 0$.

Example 6.4. *When q is parameterized linearly by μ in Theorem 6.4, i.e., $q = \mu x + (1-\mu)x_d$, we have $V_{\mathrm{NCM}} = (x - x_d)^\top \mathcal{M}(x - x_d)$ as in the standard Lyapunov function formulation [3, 70, 71]. Also, a linear input constraint, $u_{\min} \le u \le u_{\max}$ can be implemented by using*

$$\mathcal{C}_K = \{K \in \mathbb{R}^{m \times n} | u_d - u_{\max} \le K\mathsf{e} \le u_d - u_{\min}\}$$
$$\mathcal{C}_u = \{u \in \mathbb{R}^m | u_{\min} \le u \le u_{\max}\}$$

in (149) *and* (156)*, respectively.*

Remark 6.5. *For the CV-STEM and NCM state estimation in Theorems 4.5, 6.2, and 6.3, we could design a similar Lyapunov function-based estimation scheme if its contraction constraints of Theorem 4.5 do not explicitly depend on the actual state x. This can be achieved by using $A(\varrho_A, \hat{x}, t)$ and $C(\varrho_C, \hat{x}, t)$ instead of $A(\varrho_A, x, \hat{x}, t)$ and $C(\varrho_C, x, \hat{x}, t)$ in Theorem 4.5 as in Theorem 3.2 [16], leading to exponential boundedness of system trajectories as derived in [13, 16, 17].*

6.3. Remarks in NCM Implementation

This section briefly summarizes several practical techniques in constructing NCMs using the CV-STEM of Theorems 4.2, 4.5, and 4.6.

6.3.1. Neural Network Training

We model M of the CV-STEM contraction metric sampled offline by the DNN of Definition 6.1, optimizing its hyperparameters by stochastic gradient descent [48, 109] to satisfy the learning error condition (115) (see Remark 5.1). The positive definiteness of M could be exploited to reduce the dimension of the DNN target output due to the following lemma [14].

Lemma 6.1. *A matrix $A \succ 0$ has a unique Cholesky decomposition, i.e., there exists a unique upper triangular matrix $U \in \mathbb{R}^{n \times n}$ with strictly positive diagonal entries s.t. $A = U^T U$.*

Proof. See [139, pp. 441].

Selecting Θ of $M = \Theta^\top \Theta$ as the unique Cholesky decomposition of M and training the DNN using only the nonzero entries of Θ, we can reduce the dimension of the target output by $n(n-1)/2$ without losing any information on M. The pseudocode to obtain NCMs, using this approach depicted in Fig. 3, can be found in [14].

Instead of solving the convex optimization in Theorems 4.2, 4.5, and 4.6 to sample training data, we could also use them directly for training and optimizing the DNN hyperparameters, treating the constraints and the objective functions as the DNN loss functions as demonstrated in [120]. Although this approach no longer preserves the convexity and can lead only to a sub-optimal solution, this still gives the exponential tracking error bounds as long as we can get sufficiently small $\epsilon_{\ell 0}$ and $\epsilon_{\ell 1}$ in the learning error assumption (115), as discussed in Theorems 5.2, 5.3, and 6.1–6.3. See [119] for more details on the robustness and stability properties of this approach.

6.3.2. Lipschitz Condition and Spectral Normalization

Let us first define a spectrally-normalized DNN, a useful mathematical tool designed to overcome the instability of network training by constraining (134) of Definition 6.1 to be Lipschitz, i.e., $\exists\ L_{nn} \in \mathbb{R}_{\ge 0}$ s.t. $\|\varphi(x) - \varphi(x')\| \le L_{nn}\|x - x'\|,\ \forall x, x'$ [61, 132].

Definition 6.3. *A spectrally-normalized DNN is a DNN of Definition 6.1 with its weights W_ℓ normalized as $W_\ell = (C_{nn}\Omega_\ell)/\|\Omega_\ell\|$, where $C_{nn} \in \mathbb{R}_{\ge 0}$ is a given constant.*

Lemma 6.2. *A spectrally-normalized DNN given in Definition 6.3 is Lipschitz continuous with its 2-norm Lipschitz constant $C_{nn}^{L+1} L_\sigma^L$, where L_σ is the Lipschitz constant of the activation function σ in* (134)*. Also, it is robust to perturbation in its input.*

Proof. Let $\|f\|_{\mathrm{Lip}}$ represent the Lipschitz constant of a function f and let $W_\ell = (C_{nn}\Omega_\ell)/\|\Omega_\ell\|$. Using the property $\|f_1 \circ f_2\|_{\mathrm{Lip}} \le \|f_1\|_{\mathrm{Lip}}\|f_2\|_{\mathrm{Lip}}$, we have for a spectrally-normalized DNN in Definition 6.3 that

$$\|\varphi(x; W_\ell)\| \le \|\sigma\|_{\mathrm{Lip}}^L \prod_{\ell=1}^{L+1} \|T_\ell\|_{\mathrm{Lip}} = L_\sigma^L \prod_{\ell=1}^{L+1} \|W_\ell\|$$

where we used the fact that $\|f\|_{\mathrm{Lip}}$ of a differentiable function f is equal to the maximum spectral norm (induced 2-norm) of its gradient over its domain to obtain the last equality. Since $\|W_\ell\| = C_{nn}$ holds, the Lipschitz constant of $\varphi(x_i; W_\ell)$ is $C_{nn}^{L+1} L_\sigma^L$ as desired. Since φ is Lipschitz, we have for small perturbation x_ϵ that

$$\begin{aligned} \|\varphi(x + x_\epsilon; W_\ell) - \varphi(x; W_\ell)\| &\le C_{nn}^{L+1} L_\sigma^L \|(x + x_\epsilon) - x\| \\ &= C_{nn}^{L+1} L_\sigma^L \|x_\epsilon\| \end{aligned}$$

implying robustness to input perturbation. □

In general, a spectrally-normalized DNN is useful for improving the robustness and generalization properties of DNNs, and for obtaining the learning error bound as in (115), as delineated in Remark 5.1 [132, 160, 161]. Using it for system identification yields a DNN-based nonlinear

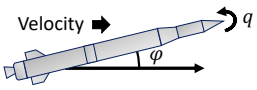

Figure 5: Rocket model (angle of attack φ, pitch rate q).

Table 4: Notations in Example 6.5.

φ	angle of attack (state variable)
q	pitch rate (state variable)
u	tail fin deflection (control input)
y	measurement
M	Mach number
p_0	static pressure at 20,000 ft
S	reference area
d	reference diameter
I_y	pitch moment of inertia
g	acceleration of gravity
m	rocket mass
V	rocket speed

feedback controller with a formal stability guarantee as shall be seen in Sec. 9.1.

For the NCM framework with stochastic perturbation, we can utilize the spectrally-normalized DNN of Definition 6.3 to guarantee the Lipschitz condition on $\partial M/\partial x_i$, appeared in the stochastic contraction condition of Theorem 2.5 (see Proposition 1 of [15]). We could also utilize the technique proposed in [62–64], which designs Lipschitz-bounded equilibrium neural networks using contraction theory, for obtaining a result analogous to Lemma 6.2. The pseudocode for the NCM construction in a stochastic setting can be found in [15].

Example 6.5. *Let us demonstrate how having the Lipschitz condition on $\partial M/\partial x_i$ in Theorem 2.5 would affect the NCM-based control and estimation performance. We consider a rocket autopilot problem perturbed by stochastic disturbances as in [15], the unperturbed dynamics of which is depicted in Fig. 5 and given as follows, assuming that the pitch rate q and specific normal force are available as a measurement via rate gyros and accelerometers [168, 169]:*

$$\frac{d}{dt}\begin{bmatrix}\varphi\\ q\end{bmatrix} = \begin{bmatrix}\frac{gC(M)\cos(\varphi)\phi_z(\varphi,M)}{mV} + q\\ \frac{C(M)d\phi_m(\varphi)}{I_y}\end{bmatrix} + \begin{bmatrix}\frac{gC(M)\bar{d}_n\cos(\varphi)}{mV}\\ \frac{C(M)\bar{d}_m}{I_y}\end{bmatrix} u$$

$$y = \begin{bmatrix} q \\ \frac{C(M)\phi_z(\varphi,M)}{m}\end{bmatrix} + \begin{bmatrix} 0 \\ \frac{C(M)\bar{d}_n}{m}\end{bmatrix} u$$

where $C(M) = p_0M^2S$, $\phi_z(\varphi,M) = 0.7(a_n\varphi^3 + b_n\varphi|\varphi| + c_n(2+M/3)\varphi)$, $\phi_m(\varphi,M) = 0.7(a_m\varphi^3 + b_m\varphi|\varphi| - c_m(7-8M/3)\varphi)$, $\bar{d}_n = 0.7d_n$, $\bar{d}_m = 0.7d_m$, (a_n,b_n,c_n,d_n) and (a_m,b_m,c_m,d_m) are given in [169], and the notations are defined in Table 4. Since this example explicitly takes into account stochastic perturbation, the spectrally-normalized DNN of Definition 6.3 is used to guarantee the Lipschitz condition of Theorem 2.5 by Lemma 6.2.

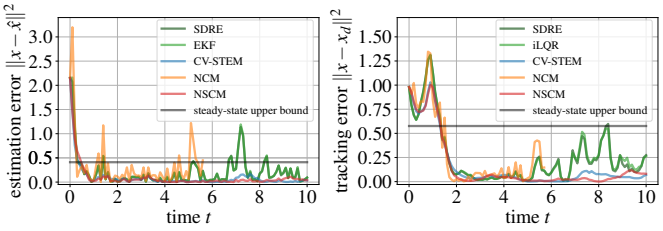

Figure 6: Rocket state estimation and tracking errors in Example 6.5 ($x = [\varphi, q]^T$).

Figure 6 shows the state estimation and tracking error performance of each estimator and controller, where the NSCM is the stochastic counterpart of the NCM aforementioned in Remark 6.1. It is demonstrated that the steady-state errors of the NSCM and CV-STEM are indeed smaller than its steady-state upper bounds of Theorems 4.2 and 4.5 (solid black line), while the other controllers violate this condition. In particular, the optimal contraction rate of the deterministic NCM for state estimation is turned out to be much larger than the NSCM as it does not account for stochastic perturbation, which makes the NCM trajectory diverge around $t = 5.8$ in Fig. 6. The NSCM circumvents this difficulty by imposing the Lipschitz condition on $\partial M/\partial x_i$ as in Theorem 2.5 using spectral normalization of Lemma 6.2. See [15] for additional details.

7. Learning-based Robust Motion Planning

The CV-STEM and NCM are also useful for designing a real-time robust motion planning algorithm for systems perturbed by deterministic and stochastic disturbances as in (129) and (130). The purpose of this section is not for proposing a new motion planner to compute x_d and u_d of (131), but for augmenting existing motion planners with real-time learning-based robustness and stability guarantees via contraction theory.

7.1. Overview of Existing Motion Planners

Let us briefly review the following standard motion planning techniques and their limitations:

(a) Learning-based motion planner (see, e.g., [112–118]): $(o_\ell(x,o_g),t) \mapsto u_d(o_g,t)$, modeled by a DNN, where $o_g \in \mathbb{R}^g$ is a vector containing global environment information, and $o_\ell : \mathbb{R}^n \times \mathbb{R}^g \mapsto \mathbb{R}^\ell$ with $\ell \leq g$ is local environment information extracted from $o_g \in \mathbb{R}^g$ [115].

(b) Robust tube-based motion planner (see, e.g., [13–15, 17, 79, 80, 88–91, 120–124, 170, 171]): $(x, x_d, u_d, t) \mapsto u^*$, where u^* is a robust feedback tracking controller given by Theorems 4.2, 4.6, 6.4, or 6.5.

The robust tube-based motion planner (b) ensures that the perturbed trajectories x of (129) and (130) stay in an

Table 5: Comparison of LAG-ROS [105] with the Learning-based and Robust Tube-based Motion Planners.

Motion planning scheme	Control policy	State tracking error $\|x - x_d\|$	Computational load
(a) Learning-based	$(o_\ell, t) \mapsto u_d$	Increases exponentially (Theorem 5.1)	One DNN evaluation
(b) Robust tube-based	$(x, x_d, u_d, t) \mapsto u^*$	Exponentially bounded (Theorems 4.2, 4.6, 6.4, 6.5)	Computation of (x_d, u_d)
(c) LAG-ROS	$(x, o_\ell, t) \mapsto u^*$	Exponentially bounded (Theorem 7.1)	One DNN evaluation

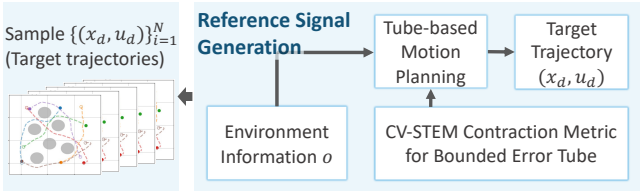

Figure 7: Block diagram of tube-based motion planning using contraction theory.

exponentially bounded error tube around the target trajectory x_d of (131) [13–15, 17, 79, 80, 88–91, 120, 121] given as in Fig. 7, using Theorems 2.4 and 2.5. However, it requires the online computation of (x_d, u_d) as an input to their control policy given in Theorems 4.2, 4.6, 6.4, and 6.5, which is not realistic for systems with limited computational resources. The learning-based motion planner (a) circumvents this issue by modeling the target policy $(o_\ell, t) \mapsto u_d$ by a DNN. In essence, our approach, to be proposed in Theorem 7.1, is for providing (a) with the contraction theory-based stability guarantees of (b), thereby significantly enhancing the performance of (a) that only assures the tracking error $\|x - x_d\|$ to be bounded by a function which increases exponentially with time, as derived previously in Theorem 5.1 of Sec. 5.

7.2. Stability Guarantees of LAG-ROS

The method of LAG-ROS – for Learning-based Autonomous Guidance with RObustness and Stability guarantees [105], bridges the gap between (a) and (b) by ensuring that the distance between the target and controlled trajectories to be exponentially bounded.

(c) LAG-ROS (see Fig. 12 and [105]): $(x, o_\ell(x, o_g), t) \mapsto u^*(x, x_d(o_g, t), u_d(o_g, t), t)$, modeled by u_L to be given in Theorem 7.1, where $u^*(x, x_d, u_d, t)$ of (b) is viewed as a function of (x, o_ℓ, t).

Table 5 summarizes the differences of the existing motion planners (a) and (b) from (c), which is defined as follows.

Definition 7.1. *Learning-based Autonomous Guidance with RObustness and Stability guarantees (LAG-ROS) is a DNN model for approximating a function which maps state x, local environment information o_ℓ, and time t to an optimal robust feedback control input u^* given by* (64) *of Theorem 4.2, i.e., $u^* = u_d - K(x - x_d)$, or* (109) *of Theorem 4.6, i.e., $u = u_d + \int_0^1 k d\mu$, where its training data is sampled as explained in Sec. 7.3 (see Figures 9, 10, and 12).*

The LAG-ROS in Definition 7.1 achieves online computation of u without solving a motion planning problem, and it still possesses superior robustness and stability properties due to its internal contracting architecture [105].

Theorem 7.1. *Let $u_L = u_L(x, o_\ell(x, o_g), t)$ be the LAG-ROS in Definition 7.1, and let u^* be given by* (64) *of Theorem 4.2, i.e., $u^* = u_d - K(x - x_d)$, or* (109) *of Theorem 4.6, i.e., $u^* = u_d + \int_0^1 k d\mu$. Define Δ_L of Theorems 5.2 and 5.3 as follows:*

$$\Delta_L = B(u_L(x, o_\ell(x, o_g), t) - u^*(x, x_d(o_g, t), u_d(o_g, t), t))$$

where $B = B(x, t)$ is the actuation matrix given in (129) *and* (130). *Note that we use $q(0, t) = \xi_0(t) = x_d(o_g, t)$ and $q(1, t) = \xi_1(t) = x(t)$ in the learning-based control formulation of Theorems 5.2 and 5.3.*

If the LAG-ROS u_L is learned to satisfy (115) *with $\epsilon_{\ell 1} = 0$, i.e., $\|\Delta_L\| \le \epsilon_{\ell 0}$ for all $x \in \mathcal{S}_s$, $o_g \in \mathcal{S}_o$, and $t \in \mathcal{S}_t$, where $\mathcal{S}_s \subseteq \mathbb{R}^n$, $\mathcal{S}_o \subseteq \mathbb{R}^g$, and $\mathcal{S}_t \subseteq \mathbb{R}_{\ge 0}$ are some compact sets, and if the control-affine nonlinear systems* (129) *and* (130) *are controlled by $u = u_L$, then* (123) *of Theorem 5.2 and* (126) *of Theorem 5.3 hold with $\bar{d}$ and $\bar{g}$ replaced as in* (140), *i.e., we have the following for $\mathsf{e}(t) = x(t) - x_d(o_g, t)$:*

$$\|\mathsf{e}(t)\| \le \frac{V_\ell(0)}{\sqrt{\underline{m}}} e^{-\alpha t} + \frac{\bar{d}_{\epsilon_\ell}}{\alpha} \sqrt{\frac{\overline{m}}{\underline{m}}} (1 - e^{-\alpha t}) = r_\ell(t) \tag{158}$$

$$\mathbb{E}\left[\|\mathsf{e}(t)\|^2\right] \le \frac{\mathbb{E}[V_{s\ell}(0)]}{\underline{m}} e^{-2\alpha t} + \frac{C_{\epsilon_\ell}}{2\alpha} \frac{\overline{m}}{\underline{m}} = r_{s\ell}(t) \tag{159}$$

$$\mathbb{P}\left[\|\mathsf{e}(t)\| \ge \varepsilon\right] \le \frac{1}{\varepsilon^2} \left(\frac{\mathbb{E}[V_{s\ell}(0)]}{\underline{m}} e^{-2\alpha t} + \frac{C_{\epsilon_\ell}}{2\alpha} \frac{\overline{m}}{\underline{m}} \right) \tag{160}$$

where $x_d = x_d(o_g, t)$ and $u_d = u_d(o_g, t)$ are as given in (131), *$\bar{d}_{\epsilon_\ell} = \epsilon_{\ell 0} + \bar{d}_c$, $C_{\epsilon_\ell} = \bar{g}_c^2 (2\alpha_G{}^{-1} + 1) + \epsilon_{\ell 0}^2 \alpha_d^{-1}$, and the other variables are as defined in Theorems 5.2 and 5.3 with the disturbance terms in* (140) *of Theorem 6.1.*

Proof. We have $f(x, t) + B(x, t)u_L = f(x, t) + B(x, t)u^* + B(x, t)(u_L - u^*)$ and $\|\Delta_L\| = \|B(x, t)(u_L - u^*)\| \le \epsilon_{\ell 0}$ by (115) with $\epsilon_{\ell 1} = 0$ for $x_d = x_d(o_g, t)$ and $u_d = u_d(o_g, t)$. Since the virtual system which has x of $\dot{x} = f(x, t) + B(x, t)u^*$ and x_d of (131) as its particular solutions is contracting due to Theorems 4.2 or 4.6, the application of Theorems 5.2 and 5.3 yields the desired relations (158)–(160). □

Due to its internal feedback structure, the LAG-ROS achieves exponential boundedness of the trajectory track-

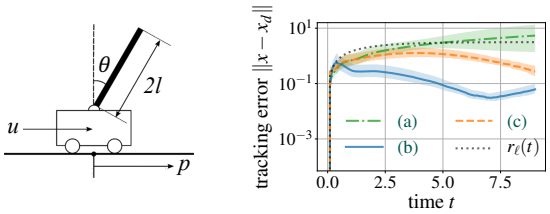

Figure 8: Cart-pole balancing task: $x = [p, \theta, \dot{p}, \dot{\theta}]^\top$, x and x_d are given in (129) and (131), (a)–(c) are given in Table 5, and r_ℓ is given in (161). The shaded area denotes the standard deviation ($+1\sigma$ and -0.5σ).

ing error as in Theorem 7.1, thereby improving the performance of existing learning-based feedforward control laws to model $(o_\ell, t) \mapsto u_d$ as in (a), whose tracking error bound increases exponentially with time as in Theorem 5.1 (see [105]). This property enables its use in safety-critical guidance and control tasks which we often encounter in modern machine learning applications.

Example 7.1. *Let us consider the cart-pole balancing task in Fig. 8, perturbed externally by deterministic perturbation, to demonstrate the differences of (a)–(c) summarized in Table 5. Its dynamical system is given in [104, 172] with its values defined in [172].*

The target trajectories (x_d, u_d) are sampled with the objective function $\int_0^T \|u\|^2 dt$, and the LAG-ROS of Theorem 7.1 is modeled by a DNN independently of the target trajectory as detailed in [105]. The environment information o_g is selected as random initial and target terminal states, and we let $V_\ell(0) = 0$ and $\epsilon_{\ell 0} + \bar{d}_c = 0.75$ in (158) of Theorem 7.1, yielding the following exponential bound:

$$r_\ell(t) = \frac{V_\ell(0)}{\sqrt{\underline{m}}} + \frac{\bar{d}_{\epsilon_\ell}}{\alpha}\sqrt{\frac{\overline{m}}{\underline{m}}}(1 - e^{-\alpha t}) = 3.15(1 - e^{-0.60t}) \tag{161}$$

where $\bar{d}_{\epsilon_\ell} = \epsilon_{\ell 0} + \bar{d}_c$.

Figure 8 shows the tracking errors of each motion planner averaged over 50 simulations at each time instant t. The LAG-ROS (c) and robust tube-based planner (b) indeed satisfy the bound (161) (dotted black curve) for all t with a small standard deviation σ, unlike learning-based motion planner (a) with a diverging bound in Theorem 5.1 and increasing deviation σ. This example demonstrates one of the major advantages of contraction theory, which enables such quantitative analysis on robustness and stability of learning-based planners.

As implied in Example 7.1, the LAG-ROS indeed possesses the robustness and stability guarantees of u^* as proven in Theorem 7.1, unlike (a), while retaining significantly lower computational cost than that of (b). See [105] for a more detailed discussion of this simulation result.

7.3. Tube-based State Constraint Satisfaction

We exploit the result of Theorem 7.1 in the tube-based motion planning [80] for generating LAG-ROS training data, which satisfies given state constraints even under the existence of the learning error and disturbance of Theorem 5.2. In particular, we sample target trajectories (x_d, u_d) of (131) in $\mathcal{S}_s \times \mathcal{S}_u$ of Theorem 7.1 by solving the following, assuming the learning error $\epsilon_{\ell 0}$ and disturbance upper bound $\bar{d}_c$ of Theorem 7.1 are selected *a priori* [61]:

$$\min_{\substack{\bar{x}=\bar{x}(o_g,t)\\ \bar{u}=\bar{u}(o_g,t)}} \int_0^T c_0\|\bar{u}\|^2 + c_1 P(\bar{x}, \bar{u}, t)dt \tag{162}$$

$$\text{s.t. } \dot{\bar{x}} = f(\bar{x}, t) + B(\bar{x}, t)\bar{u},\ \bar{x} \in \bar{\mathcal{X}}(o_g, t), \text{ and } \bar{u} \in \bar{\mathcal{U}}(o_g, t)$$

where $c_0 > 0$, $c_1 \geq 0$, $P(\bar{x}, \bar{u}, t)$ is some performance-based cost function (e.g., information-based cost [173]), $\bar{\mathcal{X}}$ is robust admissible state space defined as $\bar{\mathcal{X}}(o_g, t) = \{v(t) \in \mathbb{R}^n | \forall \xi(t) \in \{\xi(t) | \|v(t) - \xi(t)\| \leq r_\ell(t)\},\ \xi(t) \in \mathcal{X}(o_g, t)\}$, $\mathcal{X}(o_g, t)$ is given admissible state space, $r_\ell(t)$ is given by (158) of Theorem 7.1, and $\bar{u} \in \bar{\mathcal{U}}(o_g, t)$ is an input constraint. The following theorem shows that the LAG-ROS ensures the perturbed state x of (129) to satisfy $x \in \mathcal{X}$, due to the contracting property of u^*.

Lemma 7.1. *If the solution of (162) gives (x_d, u_d), the LAG-ROS u_L of Theorem 7.1 ensures the perturbed solution $x(t)$ of $\dot{x} = f(x,t) + B(x,t)u_L + d_c(x,t)$ in (129) to stay in the admissible state space $\mathcal{X}(o_g, t)$, i.e., $x(t) \in \mathcal{X}(o_g, t)$, even with the learning error $\epsilon_{\ell 0}$ of u_L in Theorem 7.1.*

Proof. See [105]. □

Lemma 7.1 implies that the perturbed trajectory (129) controlled by LAG-ROS will not violate the given state constraints as long as x_d is sampled by (162), which helps greatly reduce the need for safety control schemes such as [174]. The localization method in [115] allows extracting o_ℓ of (a) by o_g of (162), to render LAG-ROS applicable to different environments in a distributed way with a single policy.

We therefore sample artificially perturbed states x in the tube $S(x_d(o_g, t)) = \{\xi(t) \in \mathbb{R}^n | \|\xi(t) - x_d(o_g, t)\| \leq r_\ell(t)\}$ as depicted in Figures 9 and 10 [105]. The robust control inputs u^* for training LAG-ROS of Theorem 7.1 is then sampled by computing the CV-STEM of Theorem 4.2 or Theorem 4.6. The LAG-ROS framework is summarized in Fig. 12, and a pseudocode for its offline construction can be found in [105].

Remark 7.1. *For stochastic systems, we could sample x around x_d using a given probability distribution of the disturbance using (159) and (160) of Theorem 7.1.*

Example 7.2. *Let us again consider the spacecraft dynamics of Example 6.2, but now for designing a real-time robust motion planner using the LAG-ROS in a multi-agent setting. In this example, each agent is assumed to*

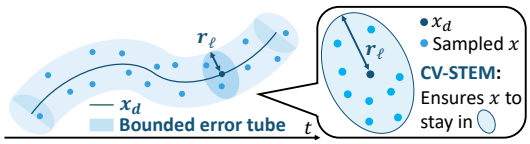

Figure 9: Illustration of state sampling in robust bounded error tube.

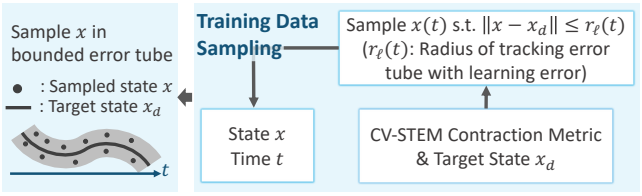

Figure 10: Block diagram of state sampling in robust bounded error tube.

have access only to the states of the other agents/obstacles located in a circle with a limited radius, 2.0m, centered around it. The performance of LAG-ROS is compared with (a), (b), and the following centralized motion planner (d) which is not computable in real-time:

(d) Centralized robust motion planner:
$(x, x_d, u_d, t) \mapsto u^*$, *offline centralized solution of (b).*

The objective function for sampling target trajectories of (131) *is selected as* $\int_0^T \|u\|^2 dt$ *with their target terminal positions* $x_f = [p_{xf}, p_{yf}, 0, 0]^\top$, *where* (p_{xf}, p_{yf}) *is a random position in* $(0,0) \leq (p_{xf}, p_{yf}) \leq (5,5)$. *For the state constraint satisfaction in Lemma 7.1, the following error tube* (158)*:*

$$r_\ell(t) = (\bar{d}_{\epsilon_\ell}/\alpha)\sqrt{\chi}(1 - e^{-\alpha t}) = 0.125(1 - e^{-0.30t})$$

is used with an input constraint $u_i \geq 0$, $\forall i$, *to avoid collisions with a random number of multiple circular obstacles and of other agents, even under the learning error and disturbances. See [105] for the LAG-ROS training details [105].*

Figure 11 shows one example of the trajectories of the motion planners (a)–(d) under deterministic external disturbances. It implies the following:

- *For (a), the tracking error accumulates exponentially with time due to the lack of robustness as proven in Theorem 5.1.*
- *(b) only yields locally optimal* (x_d, u_d) *of* (131)*, as its time horizon has to be small enough to make the problem solvable online with limited computational capacity, only with local environment information. This renders some agents stuck in local minima as depicted in Fig. 11.*
- *LAG-ROS (c) tackles these two problems by providing formal robustness and stability guarantees of Theorem 7.1, whilst implicitly knowing the global solution (only from the local information* o_ℓ *as in [115]) without computing it online. It indeed satisfies the state constraints due to Lemma 7.1 as can be seen from Fig. 11. See [105] for the in-depth discussion on this simulation result.*

7.4. Safe Exploration and Learning of Disturbances

All the theorems presented so far assume that the sizes of unknown external disturbances are fixed (i.e., $\bar{d}$ in Theorem 2.4 and $(\bar{g}_0, \bar{g}_1)$ in Theorem 2.5 are known and fixed). Since such an assumption could yield conservative state constraints, e.g., in utilizing motion planning of Lemma 7.1 with the tube of Theorem 7.1, we could consider better estimating the unknown parts, $d(x,t)$ in Theorem 2.4 or $G_0(x,t)$ and $G_1(x,t)$ in Theorem 2.5, also using contraction theory and machine learning. For example, it is demonstrated in [171] that the stochastic bounds of Theorem 2.5 can be used to ensure safe exploration and reduction of uncertainty over epochs in learning unknown, control non-affine residual dynamics. Here, its optimal motion plans are designed by solving chance-constrained trajectory optimization and executed using a feedback controller such as Theorem 4.2 with a control barrier function-based safety filter [174, 175]. It is also shown in [89–91] that the state-dependent disturbance $d(x)$ in Theorem 2.5 can be learned adaptively with the Gaussian process while ensuring safety.

These techniques are based on the robustness and stability properties of contraction theory-based control and estimation schemes, which guarantee their state trajectories to stay in a tube around the target trajectory as in Theorems 2.4, 2.5, 5.2, and 5.3, and at the same time, enable safely sampling training data for learning the unknown part of the dynamics. Sections 8 and 9 are for giving an overview of these concepts in the context of contraction theory-based adaptive control [85, 86, 104] and robust control synthesis for learned models [61, 64, 81, 131, 133].

8. Learning-based Adaptive Control

In this section, we consider the following smooth nonlinear system with an uncertain parameter $\theta \in \mathbb{R}^c$:

$$\dot{x} = f(x,\theta) + B(x,\theta)u,\ \dot{x}_d = f(x_d,\theta) + B(x_d,\theta)u_d \quad (163)$$

where x, u, x_d, and u_d are as defined in (129) and (131), and $f : \mathbb{R}^n \times \mathbb{R}^c \mapsto \mathbb{R}^n$ and $B : \mathbb{R}^n \times \mathbb{R}^c \mapsto \mathbb{R}^{n\times m}$ are known smooth functions with the uncertain parameter $\theta \in \mathbb{R}^c$. Due to Theorems 2.4 and 5.2, we can see that the robust control techniques presented earlier in Theorems 4.2, 4.6, 6.4, and 6.5 are still useful in this case if the modeling errors $\|f(x,\theta) - f(x,\theta_n)\|$ and $\|B(x,\theta) - B(x,\theta_n)\|$ are bounded, where θ_n is a nominal guess of the true parameter θ. However, there are situations where such assumptions are not necessarily true.

We present a method of deep learning-based adaptive control for nonlinear systems with parametric uncertainty, thereby further improving the real-time performance of robust control in Theorems 5.2 and 5.3 for model-based systems, and Theorems 9.1 and 9.2 for model-free systems.

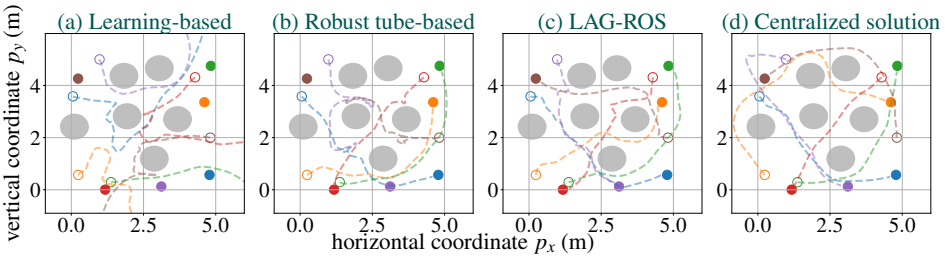

Figure 11: Trajectories for the learning-based planner (a), robust tube-based planner (b), LAG-ROS (c), and offline centralized solution (d) (○: start, ●: goal).

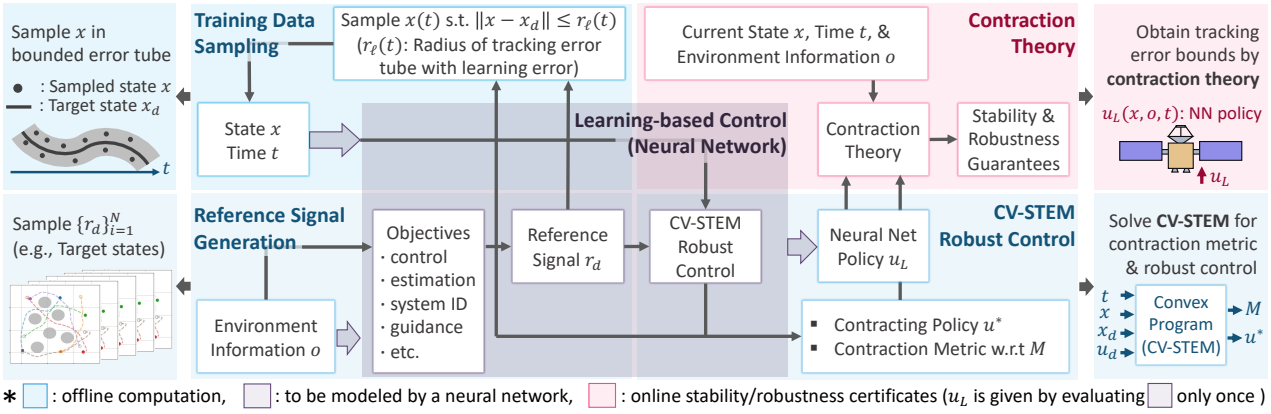

Figure 12: Detailed block diagram of machine learning-based control using contraction theory, including LAG-ROS, where Fig. 2, Fig. 7, and Fig. 10 are utilized as building blocks.

Although we consider continuous-time dynamical systems in this section, discrete-time changes could be incorporated in this framework using [87, 130]. Also, note that the techniques in this section [104] can be used with differential state feedback frameworks [33, 79–83] as described in Theorem 4.6, trading off added computational cost for generality (see Table 3 and [85, 86]).

8.1. Adaptive Control with CV-STEM and NCM

Let us start with the following simple case.

Assumption 8.1. *The matrix B in (163) does not depend on θ, and $\exists \Pi(x) \in \mathbb{R}^{c\times n}$ s.t. $\Pi(x)^\top \vartheta = f(x,\theta_n) - f(x,\theta)$, where $\vartheta = \theta_n - \theta$ and θ_n is a nominal guess of the uncertain parameter θ.*

Under Assumption 8.1, we can write (163) as

$$\dot{x} = f(x,\theta_n) + B(x)u - \Pi(x)^\top \vartheta$$

leading to the following theorem [104] for the NCM-based adaptive control. Note that we could also use the SDC formulation with respect to a fixed point as delineated in Theorem 3.2 [13, 16, 17].

Theorem 8.1. *Suppose that Assumption 8.1 holds and let $\mathcal{M}$ defines the NCM of Definition 6.2, which models M of the CV-STEM contraction metric in Theorem 4.2 for the nominal system (163) with $\theta = \theta_n$, constructed with an additional convex constraint given as $\partial_{b_i(x)}\bar{W} + \partial_{b_i(x_d)}\bar{W} = 0$, where $\partial_{b_i(q)}\bar{W} = \sum_i (\partial \bar{W}/\partial q_i) b_i(q)$ for $B = [b_1, \cdots, b_m]$ (see [79, 85]). Suppose also that the matched uncertainty condition [85] holds, i.e., $(\Pi(x) - \Pi(x_d))^\top \vartheta \in span(B(x))$ for $\Pi(x)$, and that (163) is controlled by the following adaptive control law:*

$$u = u_L + \varphi(x, x_d)^\top \hat{\vartheta} \tag{164}$$

$$\dot{\hat{\vartheta}} = -\Gamma(\varphi(x,x_d)B(x)^\top \mathcal{M}(x,x_d,u_d)(x - x_d) + \sigma\hat{\vartheta}) \tag{165}$$

where u_L is given by (135) of Theorem 6.3, i.e., $u_L = u_d - R^{-1}B^\top \mathcal{M}(x - x_d)$, $\Gamma \in \mathbb{R}^{c\times c}$ is a diagonal matrix with positive elements that governs the rate of adaptation, $\sigma \in \mathbb{R}_{>0}$, and $(\Pi(x) - \Pi(x_d))^\top \vartheta = B(x)\varphi(x,x_d)^\top \vartheta$.

If $\exists \underline{\gamma}, \overline{\gamma}, \bar{b}, \bar{\rho}, \bar{\phi}, \bar{\vartheta} \in \mathbb{R}_{>0}$ s.t. $\underline{\gamma}\mathrm{I} \preceq \Gamma \preceq \overline{\gamma}\mathrm{I}$, $\|B(x)\| \le \bar{b}$, $\|R^{-1}(x,x_d,u_d)\| \le \bar{\rho}$, $\|\varphi(x,x_d)\| \le \bar{\phi}$, and $\|\vartheta\| \le \bar{\vartheta}$, and if Γ and σ of (165) are selected to satisfy the following relation for the learning error $\|\mathcal{M} - M\| \le \epsilon_\ell$ in some compact set $\mathcal{S}$ as in (146) of Theorem 6.3:

$$\begin{bmatrix} -2\alpha_\ell \underline{m} & \bar{\phi}\bar{b}\epsilon_\ell \\ \bar{\phi}\bar{b}\epsilon_\ell & -2\sigma \end{bmatrix} \preceq -2\alpha_a \begin{bmatrix} \overline{m} & 0 \\ 0 & 1/\underline{\gamma} \end{bmatrix} \tag{166}$$

for $\exists \alpha_a \in \mathbb{R}_{>0}$, α_ℓ given in Theorem 6.3, and $\underline{m}$ and $\overline{m}$ given in $\underline{m}\mathrm{I} \preceq M \preceq \overline{m}\mathrm{I}$ of (26), then the system (163) is robust against bounded deterministic and stochastic disturbances with $\sigma \neq 0$, and we have the following bound in the compact set $\mathcal{S}$:

$$\|\mathsf{e}(t)\| \le (V_\ell(0)e^{-\alpha_a t} + \alpha_a^{-1}\sigma\sqrt{\overline{\gamma}}\bar{\vartheta}(1 - e^{-\alpha_a t}))/\sqrt{\underline{m}} \tag{167}$$

where $\mathtt{e} = x - x_d$, *and* $V_\ell = \int_{\xi_0}^{\xi_1} \|\Theta \delta q\|$ *is defined in Theorem 2.3 with* $M = \Theta^\top \Theta$ *replaced by* $\mathrm{diag}(M, \Gamma^{-1})$ *for* $\xi_0 = [x_d^\top, \vartheta^\top]^\top$ *and* $\xi_1 = [x^\top, \hat{\vartheta}^\top]^\top$. *Furthermore, if the learning error* $\epsilon_\ell = 0$ *(CV-STEM control),* (165) *with* $\sigma = 0$ *guarantees asymptotic stability of* x *to* x_d *in* (163).

PROOF. The proof can be found in [104], but here we emphasize the use of contraction theory. For u given by (164), the virtual system of a smooth path $q(\mu, t) = [q_x^\top, q_\vartheta^\top]^\top$ parameterized by $\mu \in [0, 1]$, which has $q(\mu = 0, t) = [x_d^\top, \vartheta^\top]^\top$ and $q(\mu = 1, t) = [x^\top, \hat{\vartheta}^\top]^\top$ as its particular solutions, is given as follows:

$$\dot{q} = \begin{bmatrix} \zeta(q_x, x, x_d, u_d) + B\varphi^\top q_\vartheta - \Pi(x)^\top \vartheta \\ -\Gamma(\varphi B^\top \mathcal{M}(q_x - x_d) + \sigma(q_\vartheta - \vartheta)) + d_{q\vartheta}(\mu, \vartheta) \end{bmatrix}$$

where $d_{q\vartheta}(\mu, \vartheta) = -\mu\sigma\vartheta$. Note that ζ is as given in (67), i.e., $\zeta = (A - BR^{-1}B^\top \mathcal{M})(q_x - x_d) + \dot{x}_d$, where the SDC matrix A is defined as (see Lemma 3.1)

$$A(x - x_d) = f(x, \theta_n) + B(x)u_d - f(x_d, \theta_n) - B(x_d)u_d.$$

The arguments of $A(\varrho, x, x_d, u_d)$, $B(x)$, $\mathcal{M}(x, x_d, u_d)$, and $\varphi(x, x_d)$ are omitted for notational simplicity. Since $\dot{q}_x = \zeta(q_x, x, x_d, u_d, t)$ is contracting due to Theorems 4.2 and 6.3 with a contraction rate given by α_ℓ in Theorem 6.3, we have for a Lyapunov function $V = \delta q_x^\top M \delta q_x + \delta q_\vartheta^\top \Gamma^{-1} \delta q_\vartheta$ that

$$\begin{aligned} \dot{V}/2 \leq & -\alpha_\ell \delta q_x^\top M \delta q_x + \delta q_x^\top (M - \mathcal{M}) B \varphi^\top \delta q_\vartheta \\ & - \sigma \|\delta q_\vartheta\|^2 + \delta q_\vartheta^\top \delta d_{q\vartheta}. \end{aligned} \tag{168}$$

Applying (166) with $\|\mathcal{M} - M\| \leq \epsilon_\ell$ of (146), we get

$$\begin{aligned} &\dot{V}/2 - \delta q_\vartheta^\top \delta d_{q\vartheta} \leq -(\alpha_\ell \underline{m}) \|\delta q_x\|^2 + \bar{\phi}\bar{b}\epsilon_\ell \|\delta q_x\| \|\delta q_\vartheta\| \\ &- \sigma \|\delta q_\vartheta\|^2 \leq -\alpha_a (\overline{m} \|\delta q_x\|^2 + \|\delta q_\vartheta\|^2 / \underline{\gamma}) \leq -\alpha_a V. \end{aligned}$$

Since we have $\|\partial d_{q\vartheta}/\partial \mu\| = \sigma\bar{\vartheta}$, this implies $\dot{V}_\ell \leq -\alpha_a V_\ell + \sigma\sqrt{\bar{\gamma}}\bar{\vartheta}$, yielding the bound (167) due to Lemma 2.1 [104]. Robustness against deterministic and stochastic disturbances follows from Theorem 3.1 if $\sigma \neq 0$. If $\epsilon_\ell = 0$ and $\sigma = 0$, the relation (168) reduces to $\dot{V}/2 \leq -\alpha \delta q_x^\top M \delta q_x$, which results in asymptotic stability of x to x_d in (163) due to Barbalat's lemma [3, p. 323] as in the proof of Theorem 2 in [85]. □

Remark 8.1. *Although Theorem 8.1 is for the case where* $f(x)$ *is affine in its parameter, it is also useful for the following types of systems with an uncertain parameter* $\theta \in \mathbb{R}^c$ *and a control input* u *(see [104]):*

$$H(x)p^{(n)} + h(x) + \Pi(x)\theta = u$$

where $p \in \mathbb{R}^n$, $u \in \mathbb{R}^n$, $h : \mathbb{R}^n \mapsto \mathbb{R}^n$, $H : \mathbb{R}^n \mapsto \mathbb{R}^{n \times n}$, $\Pi : \mathbb{R}^n \mapsto \mathbb{R}^{n \times c}$, $x = [(p^{(n-2)})^\top, \cdots, (p)^\top]^\top$, *and* $p^{(k)}$ *denotes the* k*th time derivative of* p. *In particular, adaptive sliding control [176] designs* u *to render the system of the composite variable* s *given as* $s = p^{(n-1)} - p_r^{(n-1)}$ *to be contracting, where* $p_r^{(n-1)} = p_d^{(n-1)} - \sum_{i=0}^{n-2} \lambda_i \mathtt{e}^{(i)}$, $\mathtt{e} = p - p_d$, p_d *is a target trajectory, and* $\kappa^{n-1} + \lambda_{n-2}\kappa^{n-2} + \cdots + \lambda_0$ *is a stable (Hurwitz) polynomial in the Laplace variable* κ *(see Example 2.6). Since we have* $\mathtt{e}^{(n-1)} = s - \sum_{i=0}^{n-2} \lambda_i \mathtt{e}^{(i)}$ *and the system for* $[\mathtt{e}^{(0)}, \cdots, \mathtt{e}^{(n-2)}]$ *is also contracting if* $s = 0$ *due to the Hurwitz property, the hierarchical combination property [9, 10] of contraction in Theorem 2.7 guarantees* $\lim_{t \to \infty} \|p - p_d\| = 0$ *[5, p. 352] (see Example 2.7).*

Example 8.1. *Using the technique of Remark 8.1, we can construct adaptive control for the Lagrangian system in Example 2.6 as follows:*

$$\mathcal{H}(\mathtt{q})\ddot{\mathtt{q}} + \mathcal{C}(\mathtt{q}, \dot{\mathtt{q}})\dot{\mathtt{q}} + \mathcal{G}(\mathtt{q}) = u(\mathtt{q}, \dot{\mathtt{q}}, t) \tag{169}$$

$$u(\mathtt{q}, \dot{\mathtt{q}}, t) = -\mathcal{K}(t)(\dot{\mathtt{q}} - \dot{\mathtt{q}}_r) + \hat{\mathcal{H}}(\mathtt{q})\ddot{\mathtt{q}}_r + \hat{\mathcal{C}}(\mathtt{q}, \dot{\mathtt{q}})\dot{\mathtt{q}}_r + \hat{\mathcal{G}}(\mathtt{q}) \tag{170}$$

where $\hat{\mathcal{H}}$, $\hat{\mathcal{C}}$, *and* $\hat{\mathcal{G}}$ *are the estimates of* $\mathcal{H}$, $\mathcal{C}$, *and* $\mathcal{G}$, *respectively, and the other variables are as defined in Example 2.6. Suppose that the terms* $\mathcal{H}$, $\mathcal{C}$, *and* $\mathcal{G}$ *depend linearly on the unknown parameter vector* θ *as follows [5, p. 405]:*

$$\mathcal{H}(\mathtt{q})\ddot{\mathtt{q}}_r + \mathcal{C}(\mathtt{q}, \dot{\mathtt{q}})\dot{\mathtt{q}}_r + \mathcal{G}(\mathtt{q}) = Y(\mathtt{q}, \dot{\mathtt{q}}, \dot{\mathtt{q}}_r, \ddot{\mathtt{q}}_r)\theta.$$

Updating the parameter estimate $\hat{\theta}$ *using the adaptation law,* $\dot{\hat{\theta}} = -\Gamma Y(\mathtt{q}, \dot{\mathtt{q}}, \dot{\mathtt{q}}_r, \ddot{\mathtt{q}}_r)^\top (\dot{\mathtt{q}} - \dot{\mathtt{q}}_r)$, *as in* (165) *where* $\Gamma \succ 0$, *we can define the following virtual system which has* $q = \xi_0 = [\dot{\mathtt{q}}_r^\top, \theta^\top]^\top$ *and* $q = \xi_1 = [\dot{\mathtt{q}}^\top, \hat{\theta}^\top]^\top$ *as its particular solutions:*

$$\begin{bmatrix} \mathcal{H} & 0 \\ 0 & \Gamma^{-1} \end{bmatrix} (\dot{q} - \dot{\xi}_0) + \begin{bmatrix} \mathcal{C} + \mathcal{K} & -Y \\ Y^\top & 0 \end{bmatrix} (q - \xi_0) = 0$$

where the relation $u = -\mathcal{K}(t)(\dot{\mathtt{q}} - \dot{\mathtt{q}}_r) + Y(\mathtt{q}, \dot{\mathtt{q}}, \dot{\mathtt{q}}_r, \ddot{\mathtt{q}}_r)\hat{\theta}$ *is used. Thus, for a Lyapunov function* $V = \delta q^\top \left[\begin{smallmatrix} \mathcal{H} & 0 \\ 0 & \Gamma^{-1} \end{smallmatrix}\right] \delta q$, *we have that*

$$\dot{V} = \delta q^\top \begin{bmatrix} \mathcal{K} & Y \\ -Y^\top & 0 \end{bmatrix} \delta q = \delta q^\top \begin{bmatrix} \mathcal{K} & 0 \\ 0 & 0 \end{bmatrix} \delta q$$

which results in asymptotic stability of the differential state δq *(i.e., semi-contraction [11], see also Barbalat's lemma [5, p. 405-406]).*

8.2. Parameter-Dependent Contraction Metric (aNCM)

Although Theorem 8.1 utilizes the NCM designed for the nominal system (163) with $\theta = \theta_n$, we could improve its representational power by explicitly taking the parameter estimate $\hat{\theta}$ as one of the NCM arguments [104], leading to the concept of an adaptive NCM (aNCM).

In this section, we consider multiplicatively-separable nonlinear systems given in Assumption 8.2, which holds for many types of systems including robotics systems [5], spacecraft high-fidelity dynamics [126, 127], and systems modeled by basis function approximation and DNNs [128, 129].

Assumption 8.2. *The dynamical system* (163) *is multiplicatively separable in terms of x and θ, i.e., $\exists\, Y_f : \mathbb{R}^n \mapsto \mathbb{R}^{n\times c_z}$, $Y_{b_i} : \mathbb{R}^n \mapsto \mathbb{R}^{n\times c_z}$, and $Z : \mathbb{R}^c \mapsto \mathbb{R}^{c_z}$ s.t.*

$$Y_f(x)Z(\theta) = f(x,\theta),\ Y_{b_i}(x)Z(\theta) = b_i(x,\theta)$$

where $B(x,\theta) = [b_1(x,\theta),\cdots,b_m(x,\theta)]$. We could define θ as $[\theta^\top, Z(\theta)^\top]^\top$ so we have $Y_f(q)\theta = f(q,\theta)$ and $Y_{b_i}(q)\theta = b_i(q;\theta)$. Such an over-parameterized system could be regularized using the Bregman divergence as in [177] (see Example 8.4), and thus we denote $[\theta^\top, Z(\theta)^\top]^\top$ as θ in the subsequent discussion.

Under Assumption 8.2, the dynamics for $\mathsf{e} = x - x_d$ of (163) can be expressed as follows:

$$\dot{\mathsf{e}} = A(\varrho, x, x_d, u_d, \hat{\theta})\mathsf{e} + B(x,\hat{\theta})(u - u_d) - \tilde{Y}(\hat{\theta} - \theta) \quad (171)$$

where A is the SDC matrix of Lemma 3.1, $\hat{\theta}$ is the current estimate of θ, and $\tilde{Y}$ is defined as

$$\tilde{Y} = Y - Y_d = (Y_f(x) + Y_b(x,u)) - (Y_f(x_d) + Y_b(x_d,u_d)) \quad (172)$$

where $Y_b(x,u) = \sum_{i=1}^m Y_{b_i}(q)u_i$.

Definition 8.1. *The adaptive Neural Contraction Metric (aNCM) in Fig. 13 is a DNN model for the optimal parameter-dependent contraction metric, given by solving the adaptive CV-STEM, i.e.,* (89) *of Theorem 4.2 (or Theorem 4.6 for differential feedback) with its contraction constraint replaced by the following convex constraint:*

$$-\Xi + 2\,\mathrm{sym}\left(A\bar{W}\right) - 2\nu BR^{-1}B^\top \preceq -2\alpha\bar{W} \quad (173)$$

for deterministic systems, and

$$\begin{bmatrix} -\Xi + 2\,\mathrm{sym}\left(A\bar{W}\right) - 2\nu BR^{-1}B^\top + 2\alpha\bar{W} & \bar{W} \\ \bar{W} & -\frac{\nu}{\alpha_s}\mathrm{I} \end{bmatrix} \preceq 0 \quad (174)$$

for stochastic systems, where $W = M(x, x_d, u_d, \hat{\theta})^{-1} \succ 0$ (or $W = M(x,\hat{\theta})^{-1} \succ 0$, see Theorem 3.2), $\bar{W} = \nu W$, $\nu = \overline{m}$, $R = R(x, x_d, u_d) \succ 0$ is a weight matrix on u, $\Xi = (d/dt)|_{\hat{\theta}}\bar{W}$ is the time derivative of $\bar{W}$ computed along (163) *with $\theta = \hat{\theta}$, A and B are given in* (171)*, α, $\underline{m}$, $\overline{m}$, and α_s are as given in Theorem 4.2, and the arguments $(x, x_d, u_d, \hat{\theta})$ are omitted for notational simplicity.*

The aNCM given in Definition 8.1 has the following stability property along with its optimality due to the CV-STEM of Theorem 4.2 [104].

Theorem 8.2. *Suppose that Assumption 8.2 holds and let $\mathcal{M}$ define the aNCM of Definition 8.1. Suppose also that the true dynamics* (163) *is controlled by the following adaptive control law:*

$$u = u_d - R(x, x_d, u_d)^{-1}B(x,\hat{\theta})^\top \mathcal{M}(x, x_d, u_d, \hat{\theta})\mathsf{e} \quad (175)$$

$$\dot{\hat{\theta}} = \Gamma((Y^\top d\mathcal{M}_x + Y_d^\top d\mathcal{M}_{x_d} + \tilde{Y}^\top\mathcal{M})\mathsf{e} - \sigma\hat{\theta}) \quad (176)$$

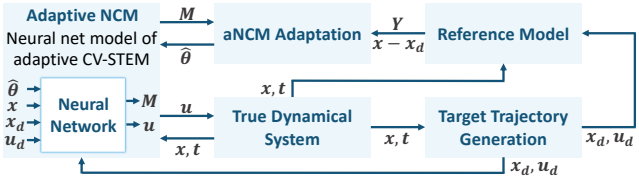

Figure 13: Illustration of aNCM (M: positive definite matrix that defines aNCM; $\hat{\theta}$: estimated parameter; Y: error signal, see (172); x: system state; u: system control input; and (x_d, u_d): target state and control input trajectory).

where $d\mathcal{M}_q = [(\partial\mathcal{M}/\partial q_1)\mathsf{e},\cdots,(\partial\mathcal{M}/\partial q_n)\mathsf{e}]^\top/2$, $\mathsf{e} = x - x_d$, $\Gamma \succ 0$, $\sigma \in \mathbb{R}_{\geq 0}$, and Y, Y_d, $\tilde{Y}$ are given in (172)*. Suppose further that the learning error in $\|\mathcal{M} - M\| \leq \epsilon_\ell$ of Theorem 6.3 additionally satisfies $\|d\mathcal{M}_{x_d} - dM_{x_d}\| \leq \epsilon_\ell$ and $\|d\mathcal{M}_x - dM_x\| \leq \epsilon_\ell$ in some compact set $\mathcal{S}$.*

If $\exists \bar{b}, \bar{\rho}, \bar{y} \in \mathbb{R}_{>0}$ s.t. $\|B(x,\hat{\theta})\| \leq \bar{b}$, $\|R^{-1}(x, x_d, u_d)\| \leq \bar{\rho}$, $\|Y\| \leq \bar{y}$, $\|Y_d\| \leq \bar{y}$, and $\|\tilde{Y}\| \leq \bar{y}$ in (175) *and* (176)*, and if Γ and σ of* (176) *are selected to satisfy the following as in Theorem 8.1:*

$$\begin{bmatrix} -2\alpha_\ell\underline{m} & \bar{y}\epsilon_\ell \\ \bar{y}\epsilon_\ell & -2\sigma \end{bmatrix} \preceq -2\alpha_a \begin{bmatrix} \overline{m} & 0 \\ 0 & 1/\underline{\gamma} \end{bmatrix} \quad (177)$$

for $\exists \alpha_a \in \mathbb{R}_{>0}$, α_ℓ given in Theorem 6.3, and $\underline{m}$ and $\overline{m}$ given in $\underline{m}\mathrm{I} \preceq M \preceq \overline{m}\mathrm{I}$ of (26)*, then the system* (163) *is robust against bounded deterministic and stochastic disturbances, and we have the exponential bound* (167) *in the compact set $\mathcal{S}$. Furthermore, if $\epsilon_\ell = 0$ (adaptive CV-STEM control),* (176) *with $\sigma = 0$ guarantees asymptotic stability of x to x_d in* (163)*.*

Proof. Replacing the contraction constraints of the CV-STEM in Theorem 4.2 by (173) and (174), the bound (167) and the asymptotic stability result can be derived as in the proof of Theorem 8.1 (see Theorem 4 and Corollary 2 of [104] for details). Theorems 2.4 and 2.5 guarantee robustness of (163) against bounded deterministic and stochastic disturbances for $\sigma \neq 0$. □

Let us again emphasize that, by using Theorem 4.6, the results of Theorems 8.1 and 8.2 can be extended to adaptive control with CCM-based differential feedback [85, 86] (see Table 3 for the trade-offs).

Since the adaptation laws (165) and (176) in Theorems 8.1 and 8.2 yield an explicit bound on the steady-state error as in (167), it could be used as the objective function of the CV-STEM in Theorem 4.2, regarding Γ and σ as extra decision variables to get M optimal in a sense different from Theorem 4.2. Smaller ϵ_ℓ would lead to a weaker condition on them in (166) and (177). Also, the size of $\|\vartheta\| \leq \bar{\vartheta}$ in (167) can be adjusted simply by rescaling it (e.g., $\vartheta \to \theta/\bar{\vartheta}$). However, such a robustness guarantee comes with a drawback of having $\lim_{t\to\infty}\|\hat{\theta}(t)\| = 0$ for $\sigma \neq 0$ in (165), leading to the trade-offs in different types of adaptation laws, some of which are given in Examples 8.2–8.4 (see also Remark 8.2).

Example 8.2. *Let us briefly describe the following projection operator-based adaptation law, again for the Lagrangian system* (169) *of Example 8.1 with the unknown parameter vector* θ*:*

$$\dot{\hat{\theta}} = \mathrm{Proj}\,(\hat{\theta}, -\Gamma Y(\mathsf{q}, \dot{\mathsf{q}}, \dot{\mathsf{q}}_r, \ddot{\mathsf{q}}_r)^\top(\dot{\mathsf{q}} - \dot{\mathsf{q}}_r), p) \tag{178}$$

where Proj *is the projection operator and* p *is a convex boundary function (e.g.,* $p(\hat{\theta}) = (\hat{\theta}^\top\hat{\theta} - \theta_{\max}^2)/(\epsilon_\theta\theta_{\max}^2)$ *for given positive constants* $\theta_{\max}$ *and* ϵ_θ*). If* $p(\hat{\theta}) > 0$ *and* $\nabla p(\hat{\theta})^\top\xi > 0$*,*

$$\mathrm{Proj}\,(\hat{\theta}, \xi, p) = \xi - (\nabla p(\hat{\theta})\nabla p(\hat{\theta})^\top/\|\nabla p(\hat{\theta})\|^2)\xi p(\hat{\theta}),$$

otherwise, $\mathrm{Proj}\,(\hat{\theta}, \xi, p) = \xi$*. The projection operator has the following useful property which allows bounding the parameter estimate* $\hat{\theta}$*.*

Lemma 8.1. *If* $\hat{\theta}$ *is given by* (178) *with* $\hat{\theta}(0) \in \Omega_\theta = \{\theta \in \mathbb{R}^c | p(\theta) \leq 1\}$ *for a convex function* $p(\theta)$*, then* $\hat{\theta}(t) \in \Omega_\theta$*,* $\forall t \geq 0$ *(e.g.,* $\|\hat{\theta}\| \leq \theta_{\max}\sqrt{1+\epsilon_\theta}$ *if* $p(\hat{\theta}) = (\hat{\theta}^\top\hat{\theta} - \theta_{\max}^2)/(\epsilon_\theta\theta_{\max}^2)$*).*

Proof. *See [178].* □

Since Lemma 8.1 guarantees the boundedness of $\|\hat{\theta}\|$*, the adaptive controller* (170)*,* $u = -\mathcal{K}(\dot{\mathsf{q}} - \dot{\mathsf{q}}_r) + Y\theta + Y\tilde{\theta}$*, can be viewed as the exponentially stabilizing controller* $-\mathcal{K}(\dot{\mathsf{q}} - \dot{\mathsf{q}}_r) + Y\theta$ *(see Example 2.6) plus a bounded external disturbance* $Y\tilde{\theta}$*, implying robustness due to Theorem 2.4 [24, 86, 178, 179].*

Let us also remark that, as in Example 8.1, the projection operator-based adaptation law (178) *still achieves asymptotic stabilization. Applying the control law* (170) *with the adaptation* (178) *yields the following virtual system which has* $q = \xi_0 = [\dot{\mathsf{q}}_r^\top, \theta^\top]^\top$ *and* $q = \xi_1 = [\dot{\mathsf{q}}^\top, \hat{\theta}^\top]^\top$ *as its particular solutions:*

$$\begin{bmatrix} \mathcal{H} & 0 \\ 0 & \mathrm{I} \end{bmatrix}(\dot{q} - \dot{\xi}_0) + \begin{bmatrix} (\mathcal{C}+\mathcal{K})(q_s - \dot{\mathsf{q}}_r) - Y(q_\theta - \theta) \\ \mathrm{Proj}\,(\hat{\theta}, \Gamma Y^\top(q_s - \dot{\mathsf{q}}_r), p) \end{bmatrix} = 0$$

where $q = [q_s^\top, q_\theta^\top]^\top$*. Thus, for a Lyapunov function* $V_{s\ell} = \frac{1}{2}\int_{\xi_0}^{\xi_1}\delta q^\top \begin{bmatrix} \mathcal{H}(\mathsf{q}) & 0 \\ 0 & \Gamma^{-1} \end{bmatrix}\delta q$*, we have that*

$$\dot{V}_{s\ell} = \int_{\xi_0}^{\xi_1} -\delta q_s^\top\mathcal{K}\delta q_s + \delta q_\theta^\top(\mathrm{Proj}\,(\hat{\theta}, Y^\top\delta q_s, p) - Y^\top\delta q_s).$$

Using the convex property of the projection operator, i.e., $\tilde{\theta}^\top(\mathrm{Proj}\,(\hat{\theta}, \xi, p) - \xi) \leq 0$ *[178, 179], this gives* $\dot{V}_{s\ell} \leq -\int_{\xi_0}^{\xi_1}\delta q_s^\top\mathcal{K}\delta q_s$*, which results in asymptotic stability of* δq *due to Barbalat's lemma [5, p. 405-406] as in Example 8.1.*

Example 8.3. *The dependence on* u *and* $\dot{\hat{\theta}}$ *in* $(d/dt)|_{\hat{\theta}}M$ *can be removed by using* $\partial_{b_i(x)}M + \partial_{b_i(x_d)}M = 0$ *as in Theorem 8.1, and by using adaptation rate scaling introduced in [86]. In essence, the latter multiplies the adaptation law* (176) *by any strictly-increasing and strictly-positive scalar function* $v(2\rho)$*, and update* ρ *as*

$$\dot{\rho} = \frac{1}{2}\frac{v(2\rho)}{v_\rho(2\rho)}\sum_{i=1}^{c}\frac{1}{V_{\mathsf{e}} + \eta}\frac{\partial V_{\mathsf{e}}}{\partial\hat{\theta}_i}\dot{\theta}_i \tag{179}$$

where $v_\rho = \partial v/\partial\rho$*,* $\eta \in \mathbb{R}_{>0}$*, and* $V_{\mathsf{e}} = \mathsf{e}^\top M(x, x_d, u_d, \hat{\theta})\mathsf{e}$ *for* $\mathsf{e} = x - x_d$ *and* M *given in Definition 8.1, so the additional term due to* (179) *cancels out the term involving* $\dot{\hat{\theta}}$ *in* $(d/dt)|_{\hat{\theta}}\bar{W}$ *of* (173) *(see [86] for details). Its robustness property follows from Theorem 8.2 also in this case.*

Example 8.4. *Using the Bregman divergence-based adaptation in [177], we could implicitly regularize the parameter estimate* $\hat{\theta}$ *as follows:*

$$\lim_{t\to\infty}\hat{\theta} = \arg\min_{\vartheta\in A} d_\psi(\vartheta\|\theta^*) = \arg\min_{\vartheta\in A} d_\psi(\vartheta\|\hat{\theta}(0))$$

where d_ψ *is the Bregman divergence defined as* $d_\psi(x\|y) = \psi(x) - \psi(y) - (x-y)^\top\nabla\psi(y)$ *for a convex function* ψ*, and* $\mathcal{A}$ *is a set containing only parameters that interpolate the dynamics along the entire trajectory. If* $\hat{\theta}(0) = \arg\min_{b\in\mathbb{R}^p}\psi(b)$*, we have* $\lim_{t\to\infty}\hat{\theta} = \arg\min_{\vartheta\in A}\psi(\vartheta)$*, which regularizes* $\hat{\theta}$ *to converge to a parameter that minimizes* ψ*. For example, using* 1*-norm for* ψ *would impose sparsity on the steady-state parameter estimate* $\hat{\theta}$ *[177].*

These extensions of adaptive control techniques described in Examples 8.2–8.4 could be utilized with contraction theory and learning-based control as in Theorems 8.1 and 8.2.

Remark 8.2. *Note that the results presented earlier in this section do not necessarily mean parameter convergence,* $\lim_{t\to\infty}\|\tilde{\theta}\| = 0$*, as the adaptation objective is to drive the tracking error* $\|x - x_d\|$ *to zero [5, p. 331], not to find out the true parameter* θ *out of the many that achieve perfect tracking.*

Asymptotic parameter convergence, $\lim_{t\to\infty}\|\tilde{\theta}\| = 0$ *for* θ *of* (163) *under Assumption 8.2, could be guaranteed if there is no disturbance and learning error with* $\sigma = 0$*, and we have* $\exists T, \alpha_{PE} \in \mathbb{R}_{>0}$ *s.t.* $\int_t^{t+T}\tilde{Y}^\top\tilde{Y}d\tau \succeq \alpha_{PE}\mathrm{I}$*,* $\forall t$ *for* $\tilde{Y}$ *given in* (172) *(the persistent excitation condition [5, p. 366]). We could also utilize the Bregman divergence-based adaptation to regularize the behavior of* $\lim_{t\to\infty}\|\hat{\theta}\|$ *as in Example 8.4.*

9. Contraction Theory for Learned Models

In recent applications of learning-based and data-driven automatic control frameworks, we often encounter situations where we only have access to a large amount of system trajectory data. This section, therefore, considers the cases where the true underlying dynamical system is poorly modeled or completely unknown, and assumptions in Sec. 8 for using learning-based adaptive control techniques are no longer valid.

A typical approach is to perform system identification [61, 64, 81, 131] using trajectory data generated by (181):

$$\dot{x} = f_L(x, u(x,t), t) \tag{180}$$
$$\dot{x}^* = f_{\text{true}}(x^*, u(x^*,t), t) \tag{181}$$

where $x : \mathbb{R}_{\geq 0} \mapsto \mathbb{R}^n$ is the system state, $u : \mathbb{R}^n \times \mathbb{R}_{\geq 0} \mapsto \mathbb{R}^m$ is the system control input, $f_{\text{true}} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_{\geq 0} \mapsto \mathbb{R}^n$ is a smooth function of the true dynamical system (181), which is unknown and thus modeled by a learned smooth function $f_L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_{\geq 0} \mapsto \mathbb{R}^n$ of (180). If we can learn f_L to render the system (180) contracting, contraction theory still allows us to ensure robustness and stability of these systems.

Theorem 9.1. *Let $q(0,t) = \xi_0(t) = x(t)$, $q(1,t) = \xi_1(t) = x^*(t)$, and $g = f_L$ in Theorems 5.2 and 5.3 as in Example 5.3, and define Δ_L as*

$$\Delta_L = f_{\text{true}}(x^*, u(x^*,t), t) - f_L(x^*, u(x^*,t), t) \tag{182}$$

for the learning error condition $\|\Delta_L\| \leq \epsilon_{\ell 0} + \epsilon_{\ell 1}\|\xi_1 - \xi_0\|$ in (115)*. If the function f_L is learned to satisfy* (115) *with $\epsilon_{\ell 1} = 0$, i.e., $\|\Delta_L\| \leq \epsilon_{\ell 0}$ in some compact set $\mathcal{S}$, and if there exists a contraction metric defined by M bounded as $\underline{m}\mathrm{I} \preceq M \preceq \overline{m}\mathrm{I}$ as in* (26)*, which renders* (180) *contracting as in* (4) *of Theorem 2.1 for deterministic systems, i.e.,*

$$\dot{M} + M(\partial f_L/\partial x) + (\partial f_L/\partial x)^\top M \preceq -2\alpha M, \tag{183}$$

and (38) *of Theorem 2.5 for stochastic systems, i.e.,*

$$\dot{M} + M(\partial f_L/\partial x) + (\partial f_L/\partial x)^\top M \preceq -2\alpha M - \alpha_s \mathrm{I}, \tag{184}$$

then we have the following in the compact set $\mathcal{S}$:

$$\|x(t) - x^*(t)\| \leq \frac{V_\ell(0)}{\sqrt{\underline{m}}} e^{-\alpha t} + \frac{\epsilon_{\ell 0}}{\alpha}\sqrt{\frac{\overline{m}}{\underline{m}}}(1 - e^{-\alpha t}) \tag{185}$$

for x and x^* in (180) *and* (181)*, where $V_\ell = \int_{x^*}^{x} \|\Theta \delta x\|$ as in Theorem 2.3 with $M = \Theta^\top \Theta$. Furthermore, the systems* (180) *and* (181) *are robust against bounded deterministic and stochastic disturbances.*

Proof. Let $p_t^* = (x^*, u(x^*(t),t), t)$ for notational simplicity. Since (181) can be written as $\dot{x}^* = f_{\text{true}}(p_t^*) = f_L(p_t^*) + (f_{\text{true}}(p_t^*) - f_L(p_t^*))$ (see Example 5.3) and $\|\Delta_L\| = \|f_{\text{true}}(p_t^*) - f_L(p_t^*)\| \leq \epsilon_{\ell 0}$, Theorem 2.4 holds with $\bar{d} = \epsilon_{\ell 0}$ as (180) is contracting, resulting in (185). Also, defining Δ_L as (182) in Theorems 5.2 and 5.3 results in robustness of (180) and (181) against bounded deterministic and stochastic disturbances due to (183) and (184), respectively. □

Theorem 9.1 is the theoretical foundation for stability analysis of model-free nonlinear dynamical systems. The bound (185) becomes tighter as we achieve smaller $\epsilon_{\ell 0}$ using more training data for verifying $\|\Delta_L\| \leq \epsilon_{\ell 0}$ (see Remark 5.1). From here onwards, we utilize contraction theory to provide stability guarantees to such model-free systems, partially enabling the use of the aforementioned model-based techniques.

9.1. Robust Control of Systems Modeled by DNNs

One challenge in applying Theorem 9.1 in practice is to find contraction metrics for the control non-affine nonlinear systems (180). This section delineates one way to construct a contraction metric in Theorem 9.1 for provably-stable feedback control, using the CV-STEM and NCM of Theorems 4.2, 4.6, 6.1, 6.3–6.5, and 7.1, along with the spectrally-normalized DNN of Definition 6.3.

To this end, let us assume that f_{true} of the dynamics (181) can be decomposed into a known control-affine part $f(x^*,t) + B(x^*,t)u$ and an unknown control non-affine residual part $r(x^*,u,t)$ as follows:

$$\dot{x}^* = f_{\text{true}} = f(x^*,t) + B(x^*,t)(u + r(x^*,u,t)). \tag{186}$$

Ideally, we would like to design u as

$$u = u^*(x,t) - r_L(x,u,t) \tag{187}$$

to cancel out the unknown term $r(x,u,t)$ of the dynamical system (186) by the model $r_L(x,u,t)$ learned using trajectory data, where u^* is a nominal stabilizing control input for $\dot{x} = f(x,t) + B(x,t)u$ given by, e.g., Theorems 4.2 and 4.6. However, the equation (187) depends implicitly on u, which brings extra computational burden especially if the learned model $r_L(x,u,t)$ is highly nonlinear as in DNNs. In [61], a discrete-time nonlinear controller is proposed to iteratively solve (187).

Lemma 9.1. *Define a mapping $\mathcal{F}$ as $\mathcal{F}(u) = u^*(x,t) - r_L(x,u,t)$, where u^* and r_L are given in* (187)*. If r_L is Lipschitz in u with a 2-norm Lipschitz constant $L_u < 1$, i.e., $\|r_L(x,u,t) - r_L(x,u',t)\| \leq L_u\|u - u'\|$, $\forall u, u'$, then $\mathcal{F}$ is a contraction mapping for fixed x,t. Therefore, if x,t are fixed, discrete-time nonlinear control u_k defined as*

$$u_k = \mathcal{F}(u_{k-1}) = u^*(x,t) - r_L(x,u_{k-1},t) \tag{188}$$

converges to a unique solution u given by $u = \mathcal{F}(u)$.

Proof. Since r_L is Lipschitz, we have that

$$\|\mathcal{F}(u) - \mathcal{F}(u')\| \leq \|r_L(x,u,t) - r_L(x,u',t)\| \leq L_u\|\Delta u\|$$

where $\Delta u = u - u'$. Thus, the assumption $L_u < 1$ ensures that $\mathcal{F}$ is a contraction mapping for fixed x,t [61]. □

By applying contraction theory to the discrete-time controller (188) of Lemma 9.1, we can guarantee the stability of (186) if r_L is modeled by a spectrally-normalized DNN of Definition 6.3.

Theorem 9.2. *Let x be the trajectory of the following ideal system without the unknown part r of the dynamics* (186)*:*

$$\dot{x} = f(x,t) + B(x,t)u^*(x,t) \tag{189}$$

where u^* is a stabilizing controller that renders (189) *contracting as in Theorem 9.1 for M which satisfies $\underline{m}\mathrm{I} \preceq$*

$M \preceq \overline{m}\mathrm{I}$ of (26)*. Note that such u^* can be designed by using, e.g., Theorems 4.2 and 4.6. Suppose that the true dynamics* (186) *is controlled by* (188) *and*

$$\exists \rho \in \mathbb{R}_{\geq 0} \ s.t. \ \|u_k - u_{k-1}\| \leq \rho \|x - x^*\| \tag{190}$$

for x^* in (186) *[61]. If $\exists \bar{b} \in \mathbb{R}_{\geq 0}$ s.t. $\|B(x,t)\| \leq \bar{b}$, and if r_L is modeled by a spectrally-normalized DNN of Definition 6.3 to have*

$$\|r_L(x,u,t) - r(x,u,t)\| \leq \epsilon_\ell \tag{191}$$

for all $x \in \mathcal{S}_s$, $u \in \mathcal{S}_u$, and $t \in \mathcal{S}_t$, where $\mathcal{S}_s \subseteq \mathbb{R}^n$, $\mathcal{S}_u \subseteq \mathbb{R}^m$, and $\mathcal{S}_t \subseteq \mathbb{R}_{\geq 0}$ are some compact sets, then r_L is Lipschitz continuous, and the controller (188) *applied to* (186) *gives the following bound in the compact set:*

$$\|x(t) - x^*(t)\| \leq \frac{V_\ell(0)}{\sqrt{\underline{m}}} e^{-\alpha_\ell t} + \frac{\bar{b}\epsilon_\ell}{\alpha_\ell} \sqrt{\frac{\overline{m}}{\underline{m}}} (1 - e^{-\alpha_\ell t}) \tag{192}$$

as long as the Lipschitz constant of r_L is selected to have

$$\exists \alpha_\ell \in \mathbb{R}_{>0} \ s.t. \ \alpha_\ell = \alpha - \bar{b} L_u \rho \sqrt{\overline{m}/\underline{m}} > 0 \tag{193}$$

where $V_\ell = \int_{x^*}^{x} \|\Theta \delta x\|$ as in Theorem 2.3 with $M = \Theta^\top \Theta$. Furthermore, the system (186) *with the controller* (188) *is robust against deterministic and stochastic disturbances.*

Proof. If r_L is modeled by a spectrally-normalized DNN, we can arbitrarily choose its Lipschitz constant L_u by Lemma 6.2. Applying (188) to (186) yields

$$\dot{x}^* = f_{cl}(x^*,t) + B(x^*,t)(r(x^*,u_k,t) - r_L(x^*,u_{k-1},t))$$

where $f_{cl}(x^*,t) = f(x^*,t) + B(x^*,t)u^*$. Using the Lipschitz condition on r_L and the learning error assumption (191), we have that

$$\begin{aligned}\|r(x^*,u_k,t) - r_L(x^*,u_{k-1},t)\| \leq& \epsilon_\ell + L_u \|u_k - u_{k-1}\| \\ \leq& \epsilon_\ell + L_u \rho \|x - x^*\|\end{aligned}$$

where (190) is used to obtain the second inequality. Since we have $\|B(x,t)\| \leq \bar{b}$ and the closed-loop system $\dot{x} = f_{cl}(x,t)$ is contracting, Theorem 5.2 holds with $\epsilon_{\ell 0} = \bar{b}\epsilon_\ell$ and $\epsilon_{\ell 1} = \bar{b} L_u \rho$ as long as we select L_u to satisfy (193), resulting in the bound (192). The final robustness statement follows from Theorems 5.2 and 5.3. □

Theorem 9.2 implies that the control synthesis algorithms via contraction theory, including robust control of Theorems 4.2 and 4.6 (CV-STEM), learning-based robust control of Theorems 6.1, 6.3–6.5, and 7.1 (NCM, LAG-ROS), can be enhanced to provide explicit robustness and stability guarantees even for systems partially modeled by DNNs that depend nonlinearly on u.

Example 9.1. *Let us consider the following Lagrangian system of Example 2.6 perturbed externally by unknown control non-affine residual disturbance:*

$$\mathcal{H}(\mathtt{q})\ddot{\mathtt{q}} + \mathcal{C}(\mathtt{q},\dot{\mathtt{q}})\dot{\mathtt{q}} + \mathcal{G}(\mathtt{q}) = u + r(x,u) \tag{194}$$

where $x = [\mathtt{q}^\top, \dot{\mathtt{q}}^\top]^\top$ and the other variables are as given in Example 2.6. Using the result of Theorem 9.2, we can design a discrete-time nonlinear controller by augmenting the exponentially stabilizing controller of Example 2.6 with a learned residual part $r_L(x,u)$ as follows:

$$\begin{aligned}u_k =& -\mathcal{K}(t)(\dot{\mathtt{q}} - \dot{\mathtt{q}}_r) + \mathcal{H}(\mathtt{q})\ddot{\mathtt{q}}_r + \mathcal{C}(\mathtt{q},\dot{\mathtt{q}})\dot{\mathtt{q}}_r + \mathcal{G}(\mathtt{q}) \\ & - r_L(x,u_{k-1})\end{aligned}$$

where $\dot{\mathtt{q}}_r = \dot{\mathtt{q}}_d(t) - \Lambda(t)(\mathtt{q} - \mathtt{q}_d(t))$, $\mathcal{K} : \mathbb{R}_{\geq 0} \mapsto \mathbb{R}^{n\times n}$, $\Lambda : \mathbb{R}_{\geq 0} \mapsto \mathbb{R}^{n\times n}$, and $(\mathtt{q}_d, \dot{\mathtt{q}}_d)$ is the target trajectory of the state $(\mathtt{q}, \dot{\mathtt{q}})$, and $\mathcal{K}, \Lambda \succ 0$ are control gain matrices (design parameters). Again, note that $\dot{\mathcal{H}} - 2\mathcal{C}$ is skew-symmetric with $\mathcal{H} \succ 0$ by construction. This gives us the following closed-loop virtual system of a smooth path $q(\mu,t)$ parameterized by $\mu \in [0,1]$, which has $q(\mu = 0,t) = \dot{\mathtt{q}}_r$ and $q(\mu = 1,t) = \dot{\mathtt{q}}$ as its particular solutions as in Example 2.6, but now with non-zero perturbation due to $r(x,u)$:

$$\mathcal{H}(\dot{q} - \ddot{\mathtt{q}}_r) + (\mathcal{C} + \mathcal{K})(q - \dot{\mathtt{q}}_r) = \mu(r(x,u_k) - r_L(x,u_{k-1})).$$

After some algebra as in the proof of Theorem 9.2, we can show that

$$\frac{d}{dt}\int_0^1 \|\Theta \partial_\mu q\| \leq -\left(\frac{k_\ell}{h_u} - \frac{L_u \rho}{h_\ell}\right)\int_0^1 \|\Theta \partial_\mu q\| + \frac{\epsilon_\ell}{\sqrt{h_\ell}}$$

where $\mathcal{H} = \Theta^\top \Theta$, $h_\ell \mathrm{I} \preceq \mathcal{H} \preceq h_u \mathrm{I}$, $k_\ell \mathrm{I} \preceq \mathcal{K}$, $\|u_k - u_{k-1}\| \leq \rho\|\dot{\mathtt{q}} - \dot{\mathtt{q}}_r\|$, ϵ_ℓ is the learning error of r_L, and L_u is the Lipschitz constant of r_L (assuming r_L is modeled by a spectrally-normalized DNN of Definition 6.3). This indeed indicates that the tracking error of the Lagrangian system (194) *is exponentially bounded as proven in Theorem 9.2.*

In [61], the technique in Theorem 9.2 and in Example 9.1 is used to perform precise near-ground trajectory control of multi-rotor drones, by learning complex aerodynamic effects caused by high-order interactions between multi-rotor airflow and the environment. It is demonstrated that it significantly outperforms a baseline nonlinear tracking controller in both landing and cross-table trajectory tracking tasks. Theorem 9.2 enables applying it to general nonlinear systems with state and control dependent uncertainty, as long as we have a nominal exponentially stabilizing controller (which can be designed by Theorems 4.2 and 4.6, or approximately by Theorems 6.1, 6.3–6.5, and 7.1).

9.2. Learning Contraction Metrics from Trajectory Data

This section presents a data-driven method to design contraction metrics for stability guarantees directly from trajectory data, assuming that its underlying dynamical system is completely unknown unlike (186). We specifically consider the situation where the underlying system is given by a continuous-time autonomous system, $\dot{x} = f(x)$ with f being unknown, and the state $x \in \mathbb{R}^n$ being fully observed.

Table 6: Notations in Sec. 9.2 [133].

$\mathbb{B}_2^n(r)$	Closed ℓ_2-ball in $\mathbb{R}^n$ of radius r centered at 0
$\mathbb{B}_2^n(\xi, r)$	Closed ℓ_2-ball in $\mathbb{R}^n$ of radius r centered at ξ
$\mathbb{S}^{n-1}$	Sphere in $\mathbb{R}^n$
$\mu_{\text{Leb}}(\cdot)$	Lebesgue measure on $\mathbb{R}^n$
$\psi_t(\cdot)$	Induced flow on prolongated system given as $p(x, \delta x) = (f(x), (\partial f/\partial x)\delta x)^\top$
$\theta_t(\delta\xi, \xi)$	Second element of $\psi_t(\xi, \delta\xi)$
$\zeta_n(r)$	Haar measure of a spherical cap in $\mathbb{S}^{n-1}$ with arc length r
ν	Uniform probability measure on $\mathcal{S}_s \times \mathbb{S}^{n-1}$
$\mathcal{T}S$	Tangent bundle of $S = \cup_{t\in\mathcal{S}_t} \varphi_t(\mathcal{S}_s)$

Let $\mathcal{S}_s \subseteq \mathbb{R}^n$ be a compact set and $\mathcal{S}_t \subseteq \mathbb{R}_{\geq 0}$ be the maximal interval starting at zero for which a unique solution $\varphi_t(\xi)$ exists for all initial conditions $\xi \in \mathcal{S}_s$ and $t \in \mathcal{S}_t$. Let us also assume access to sampled trajectories generated from random initial conditions. Notations to be used in this section are given in Table 6.

The following theorem states that a contraction metric defined by $\mathcal{M}$, learned from trajectory data, indeed guarantees contraction of the unknown underlying dynamical system [133].

Theorem 9.3. *Suppose that $\mathcal{S}_s \subseteq \mathbb{R}^n$ is full-dimensional, $n \geq 2$, and $\dot{x} = f(x)$ is contracting in a metric defined by a uniformly bounded $M(x) \succ 0$, i.e.,*

$$\dot{M} + M(\partial f/\partial x) + (\partial f/\partial x)^\top M \preceq -2\alpha M$$

of (4) and $\underline{m}\mathrm{I} \preceq M \preceq \overline{m}\mathrm{I}$ of (26) hold for M. Suppose also that a matrix-valued function $\mathcal{M}(x)$ with $\mathcal{M}(x) \succeq \underline{m}_L\mathrm{I}$ is learned to satisfy $\nu(Z_b) \leq \epsilon_\ell$, where the contraction violation set Z_b for $p(x, \delta x) = (f(x), (\partial f/\partial x)\delta x)^\top$ is defined as

$$Z_b = \{(\xi, \delta\xi) \in \mathcal{S}_s \times \mathbb{S}^{n-1} : \max_{t\in\mathcal{S}_t} (\nabla\mathcal{V}(\psi_t(\xi, \delta\xi))^\top p(\psi_t(\xi, \delta\xi)) - \alpha\mathcal{V}(\psi_t(\xi, \delta\xi)) > 0)\}$$

with $\mathcal{V} : \mathcal{T}S \mapsto \mathbb{R}$ being a learned differential Lyapunov function $\mathcal{V}(x, \delta x) = \delta x^\top \mathcal{M}(x)\delta x$ for $S = \cup_{t\in\mathcal{S}_t} \varphi_t(\mathcal{S}_s)$. Define $\bar{B}$, B_H, $B_{\nabla q}$, and $B_{\nabla\mathcal{V}}$ as

$$\bar{B} = B_H(B_{\nabla q} + \alpha B_{\nabla\mathcal{V}})(\overline{m}/\underline{m})^{3/2},\ B_H = \sup_{x\in S} \left\|\partial^2 f/\partial x^2\right\|$$

$$B_{\nabla q} = \sup_{x\in S} \|\nabla q(x)\|,\ B_{\nabla\mathcal{V}} = \sup_{x\in S} \|\nabla\mathcal{V}(x)\|$$

where $q = \nabla\mathcal{V}(x)^\top f(x)$. Also, we define r_{ϵ_ℓ} and r_b as

$$r_{\epsilon_\ell} = \sup\{r \in \mathbb{R}_{>0} : r^n \zeta_n(r) \leq \epsilon_\ell \mu_{\text{Leb}}(\mathcal{S}_s)/\mu_{\text{Leb}}(\mathbb{B}_2^n(1))\}$$

$$r_b = \sqrt{r_{\epsilon_\ell}\bar{B}/(\eta\alpha\underline{m}_L)}$$

where $\eta \in (0, 1)$. Finally, define $\tilde{X}_t(r_b)$ as

$$\tilde{X}_t(r_b) = \{\xi \in \tilde{X} : \inf_{\delta\xi\in\mathbb{S}^{n-1}} \|\theta_t(\delta\xi, \xi)\| \geq r_b\}$$

for $t \in \mathcal{S}_t$, where $\tilde{X} = \{\xi \in \mathcal{S}_s : \mathbb{B}_2^n(\xi, r_{\epsilon_\ell}) \subset \mathcal{S}_s\}$. Then the system is contracting in the learned metric defined by $\mathcal{M}(x)$ at the rate $(1-\eta)\alpha$ with $\eta \in (0, 1)$, for every $x \in \tilde{S}(r_b) = \cup_{t\in\mathcal{S}_t} \varphi_t(\tilde{X}_t(r_b))$.

Proof. See Theorem 5.2 of [133]. □

Agreeing with our intuition, Theorem 9.3 ensures that the region $\tilde{S}(r_b)$ where the contraction condition holds becomes larger, as the upper bound ϵ_ℓ on $\nu(Z_b)$ for the contraction violation set Z_b becomes smaller, yielding the learned contraction metric of higher quality. In [133], it is proven that if the samples are drawn uniformly from $\mathcal{S}_s \times \mathbb{S}^{n-1}$, we have $\epsilon_\ell \leq O(k \cdot \text{polylog}(N)/N)$ decay rates for various parametric and nonparametric function classes, where N is the number of sampled trajectories and k is the effective number of parameters of the function class of interest. Such a learned contraction metric can be used in Theorems 9.1 and 9.2 for certifying robustness and stability of model-free systems.

10. Concluding Remarks

The main contribution of this paper is twofold. First, a tutorial overview of contraction theory is presented to generalize and simplify Lyapunov-based stability methods for incremental exponential stability analysis of nonlinear non-autonomous systems. The use of differential dynamics and its similarity to an LTV system allow for LMI and convex optimization formulations that are useful for systematic nonlinear control and estimation synthesis. Second, various methods of machine learning-based control using contraction theory are presented to augment the existing learning frameworks with formal robustness and stability guarantees, extensively using the results of the first part of the paper. Such formal guarantees are essential for their real-world applications but could be difficult to obtain without accounting for a contracting property. It is also emphasized that, especially in situations where ISS and uniform asymptotic stability-based arguments render nonlinear stability analysis unnecessarily complicated, the use of exponential stability and the comparison lemma in contraction theory helps to achieve significant conceptual and methodological simplifications. A connection to the KYP and bounded real lemmas is also shown in the context of contraction-based incremental stability analysis.

Considering the promising outcomes on its utilization for model-based learning in Sec. 5–8 and for model-free data-driven learning in Sec. 9, the methods of contraction theory that are surveyed in this paper provide important mathematical tools for formally providing safety and stability guarantees of learning-based and data-driven control, estimation, and motion planning techniques for high-performance robotic and autonomous systems. Examples are elucidated to provide clear guidelines for its use in deep learning-based stability analysis and its associated control design for various nonlinear systems.

Acknowledgments

This work was in part funded by the Jet Propulsion Laboratory, California Institute of Technology, and benefited from discussions with Nick Boffi, Winfried Lohmiller, Brett Lopez, Ian Manchester, Quang Cuong Pham, Sumeet Singh, Stephen Tu, and Patrick Wensing. We also thank Chuchu Fan and Guanya Shi.

References

[1] A. Isidori, Nonlinear Control Systems, 3rd ed., Springer-Verlag, Berlin, Heidelberg, 1995.

[2] R. Marino, P. Tomei, Nonlinear Control Design: Geometric, Adaptive, and Robust, Prentice-Hall information and system sciences series, Prentice Hall, 1995.

[3] H. K. Khalil, Nonlinear Systems, 3rd ed., Prentice-Hall, Upper Saddle River, NJ, 2002.

[4] M. Vidyasagar, Nonlinear Systems Analysis, second ed., Society for Industrial and Applied Mathematics, 2002.

[5] J.-J. E. Slotine, W. Li, Applied Nonlinear Control, Pearson, Upper Saddle River, NJ, 1991.

[6] H. Nijmeijer, A. Van der Schaft, Nonlinear Dynamical Control Systems, Springer-Verlag, Berlin, Heidelberg, 1990.

[7] W. Lohmiller, J.-J. E. Slotine, On contraction analysis for nonlinear systems, Automatica 34 (1998) 683 – 696.

[8] W. Lohmiller, J.-J. E. Slotine, Nonlinear process control using contraction theory, AIChE Journal 46 (2000) 588 – 596.

[9] J.-J. E. Slotine, W. Lohmiller, Modularity, evolution, and the binding problem: a view from stability theory, Neural Networks 14 (2001) 137 – 145.

[10] J.-J. E. Slotine, Modular stability tools for distributed computation and control, Int. J. Adaptive Control Signal Process. 17 (2003) 397–416.

[11] W. Wang, J.-J. E. Slotine, On partial contraction analysis for coupled nonlinear oscillators, Biol. Cybern. 92 (2005) 38–53.

[12] D. Angeli, A Lyapunov approach to incremental stability properties, IEEE Trans. Autom. Control 47 (2002) 410–421.

[13] H. Tsukamoto, S.-J. Chung, Robust controller design for stochastic nonlinear systems via convex optimization, IEEE Trans. Autom. Control 66 (2021) 4731–4746.

[14] H. Tsukamoto, S.-J. Chung, Neural contraction metrics for robust estimation and control: A convex optimization approach, IEEE Control Syst. Lett. 5 (2021) 211–216.

[15] H. Tsukamoto, S.-J. Chung, J.-J. E. Slotine, Neural stochastic contraction metrics for learning-based control and estimation, IEEE Control Syst. Lett. 5 (2021) 1825–1830.

[16] A. P. Dani, S.-J. Chung, S. Hutchinson, Observer design for stochastic nonlinear systems via contraction-based incremental stability, IEEE Trans. Autom. Control 60 (2015) 700–714.

[17] H. Tsukamoto, S.-J. Chung, Convex optimization-based controller design for stochastic nonlinear systems using contraction analysis, in: IEEE CDC, 2019, pp. 8196–8203.

[18] P. Hartman, Ordinary Differential Equations, second ed., Soc. Ind. Appl. Math., 2002.

[19] B. P. Demidovich, Dissipativity of a system of nonlinear differential equations in the large, in: Uspekhi Mat. Nauk, volume 16, 1961, p. 216.

[20] D. C. Lewis, Metric properties of differential equations, Amer. J. Math. 71 (1949) 294–312.

[21] S. Boyd, L. El Ghaoui, E. Feron, V. Balakrishnan, Linear Matrix Inequalities in System and Control Theory, volume 15 of *Studies in Applied Mathematics*, SIAM, Philadelphia, PA, 1994.

[22] G. E. Dullerud, F. Paganini, A Course in Robust Control Theory, Texts in Applied Mathematics, 1st ed., Springer, 2000.

[23] A. Rantzer, A dual to Lyapunov's stability theorem, Syst. Control Lett. 42 (2001) 161 – 168.

[24] S.-J. Chung, S. Bandyopadhyay, I. Chang, F. Y. Hadaegh, Phase synchronization control of complex networks of Lagrangian systems on adaptive digraphs, Automatica 49 (2013) 1148–1161.

[25] A. J. Van der Schaft, $\mathcal{L}_2$-gain analysis of nonlinear systems and nonlinear state-feedback $\mathcal{H}_\infty$ control, IEEE Trans. Autom. Control 37 (1992) 770–784.

[26] A. Isidori, A. Astolfi, Disturbance attenuation and $\mathcal{H}_\infty$-control via measurement feedback in nonlinear systems, IEEE Trans. Autom. Control 37 (1992) 1283–1293.

[27] J. C. Doyle, K. Glover, P. P. Khargonekar, B. A. Francis, State-space solutions to standard $\mathcal{H}_2$ and $\mathcal{H}_\infty$ control problems, IEEE Trans. Autom. Control 34 (1989) 831–847.

[28] P. P. Khargonekar, M. A. Rotea, Mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control: A convex optimization approach, IEEE Trans. Autom. Control 36 (1991) 824–837.

[29] L. Xie, E. de Souza Carlos, Robust $\mathcal{H}_\infty$ control for linear systems with norm-bounded time-varying uncertainty, IEEE Trans. Autom. Control 37 (1992) 1188–1191.

[30] C. Scherer, P. Gahinet, M. Chilali, Multiobjective output-feedback control via LMI optimization, IEEE Trans. Autom. Control 42 (1997) 896–911.

[31] D. Hinrichsen, A. Pritchard, Stochastic $\mathcal{H}_\infty$, SIAM J. Control Optim. 36 (1998) 1504–1538.

[32] B.-S. Chen, W. Zhang, Stochastic $\mathcal{H}_2/\mathcal{H}_\infty$ control with state-dependent noise, IEEE Trans. Autom. Control 49 (2004) 45–57.

[33] I. R. Manchester, J.-J. E. Slotine, Robust control contraction metrics: A convex approach to nonlinear state-feedback $\mathcal{H}^\infty$ control, IEEE Control Systems Letters 2 (2018) 333–338.

[34] Q. Pham, N. Tabareau, J.-J. E. Slotine, A contraction theory approach to stochastic incremental stability, IEEE Trans. Autom. Control 54 (2009) 816–820.

[35] S. Han, S.-J. Chung, Incremental nonlinear stability analysis for stochastic systems perturbed by Lévy noise, arXiv:2103.13338, 2021.

[36] W. Wang, J.-J. E. Slotine, Contraction analysis of time-delayed communications and group cooperation, IEEE Trans. Autom. Control 51 (2006) 712–717.

[37] J.-J. E. Slotine, W. Wang, K. El Rifai, Contraction analysis of synchronization in networks of nonlinearly coupled oscillators, in: Int. Symp. Math. Theory Netw. Syst., 2004.

[38] Q.-C. Pham, J.-J. E. Slotine, Stable concurrent synchronization in dynamic system networks, Neural Networks 20 (2007) 62 – 77.

[39] G. Russo, M. di Bernardo, E. D. Sontag, A contraction approach to the hierarchical analysis and design of networked systems, IEEE Trans. Autom. Control 58 (2013) 1328–1331.

[40] S.-J. Chung, J.-J. E. Slotine, Cooperative robot control and concurrent synchronization of Lagrangian systems, IEEE Trans. Robot. 25 (2009) 686–700.

[41] V. M. Popov, R. Georgescu, Hyperstability of Control Systems, Springer-Verlag, Berlin, Heidelberg, 1973.

[42] Q. Pham, Analysis of discrete and hybrid stochastic systems by nonlinear contraction theory, in: Int. Conf. Control Automat. Robot. Vision, 2008, pp. 1054–1059.

[43] R. Wang, R. Tóth, I. R. Manchester, Discrete-time contraction-based control of nonlinear systems with parametric uncertainties using neural networks, arXiv:2105.05432, 2021.

[44] I. R. Manchester, J.-J. E. Slotine, Transverse contraction criteria for existence, stability, and robustness of a limit cycle, Syst. Control Lett. 63 (2014) 32–38.

[45] S. Bonnabel, J.-J. E. Slotine, A contraction theory-based analysis of the stability of the deterministic extended Kalman filter, IEEE Trans. Autom. Control 60 (2015) 565–569.

[46] Y. Zhao, J.-J. E. Slotine, Discrete nonlinear observers for inertial navigation, Syst. Control Lett. 54 (2005) 887–898.

[47] F. Tan, W. Lohmiller, J.-J. E. Slotine, Analytical SLAM without linearization, Int. J. Robot. Res. 36 (2017) 1554–1578.

[48] P. M. Wensing, J.-J. E. Slotine, Beyond convexity – Contraction and global convergence of gradient descent, PLOS ONE

15 (2020) 1–29.
[49] F. Forni, R. Sepulchre, A differential Lyapunov framework for contraction analysis, IEEE Trans. Autom. Control 59 (2014) 614–628.
[50] Q. C. Pham, J.-J. E. Slotine, Stochastic contraction in Riemannian metrics, arXiv:1304.0340, 2013.
[51] J. W. Simpson-Porco, F. Bullo, Contraction theory on Riemannian manifolds, Syst. Control Lett. 65 (2014) 74–80.
[52] S. Y. Shafi, Z. Aminzare, M. Arcak, E. D. Sontag, Spatial uniformity in diffusively-coupled systems using weighted L^2 norm contractions, in: Proc. Amer. Control Conf., 2013, pp. 5619–5624.
[53] S. Bonnabel, A. Astolfi, R. Sepulchre, Contraction and observer design on cones, in: IEEE CDC, 2011, pp. 7147–7151.
[54] P. Cisneros-Velarde, S. Jafarpour, F. Bullo, Contraction theory for dynamical systems on Hilbert spaces, arXiv:2010.01219, 2020.
[55] A. Davydov, S. Jafarpour, F. Bullo, Non-Euclidean contraction theory via semi-inner products, arXiv:2103.12263, 2021.
[56] W. Lohmiller, P. Gassert, J.-J. E. Slotine, Notes on stable learning with piecewise-linear basis functions, arXiv:1804.10085, 2018.
[57] M. C. Luis D'Alto, Incremental quadratic stability, Numer. Algebr. Control Optim. 3 (2013) 175–201.
[58] M. Margaliot, E. D. Sontag, T. Tuller, Contraction after small transients, Automatica 67 (2016) 178–184.
[59] L. Wang, F. Forni, R. Ortega, Z. Liu, H. Su, Immersion and invariance stabilization of nonlinear systems via virtual and horizontal contraction, IEEE Trans. Autom. Control 62 (2017) 4017–4022. doi:10.1109/TAC.2016.2614888.
[60] A. Astolfi, R. Ortega, Immersion and invariance: a new tool for stabilization and adaptive control of nonlinear systems, IEEE Trans. Autom. Control 48 (2003) 590–606. doi:10.1109/TAC.2003.809820.
[61] G. Shi, X. Shi, M. O'Connell, R. Yu, K. Azizzadenesheli, A. Anandkumar, Y. Yue, S.-J. Chung, Neural lander: Stable drone landing control using learned dynamics, in: IEEE ICRA, 2019, pp. 9784–9790.
[62] M. Revay, R. Wang, I. R. Manchester, Lipschitz bounded equilibrium networks, arXiv:2010.01732, 2020.
[63] M. Revay, R. Wang, I. R. Manchester, Recurrent equilibrium networks: Flexible dynamic models with guaranteed stability and robustness, arXiv:2104.05942, 2021.
[64] I. R. Manchester, M. Revay, R. Wang, Contraction-based methods for stable identification and robust machine learning: A tutorial, in: IEEE CDC, 2021.
[65] R. A. Freeman, P. V. Kokotovic, Optimal nonlinear controllers for feedback linearizable systems, in: Proc. Amer. Control Conf., volume 4, 1995, pp. 2722–2726.
[66] J. A. Primbs, V. Nevistic, J. C. Doyle, Nonlinear optimal control: A control Lyapunov function and receding horizon perspective, Asian J. Control 1 (1999) 14–24.
[67] S. Prajna, A. Papachristodoulou, F. Wu, Nonlinear control synthesis by sum of squares optimization: A Lyapunov-based approach, in: Proc. Asian Control Conf., volume 1, 2004, pp. 157–165.
[68] E. M. Aylward, P. A. Parrilo, J.-J. E. Slotine, Stability and robustness analysis of nonlinear systems via contraction metrics and SOS programming, Automatica 44 (2008) 2163 – 2170.
[69] G. Chesi, LMI techniques for optimization over polynomials in control: A Survey, IEEE Trans. Autom. Control 55 (2010) 2500–2510.
[70] E. Sontag, A Lyapunov-like characterization of asymptotic controllability, SIAM J. Control Optim. 21 (1983) 462–471.
[71] E. D. Sontag, A 'universal' construction of Artstein's theorem on nonlinear stabilization, Syst. Control Lett. 13 (1989) 117 – 123.
[72] M. Krstic, P. V. Kokotovic, I. Kanellakopoulos, Nonlinear and Adaptive Control Design, 1st ed., John Wiley & Sons, Inc., USA, 1995.
[73] H. Deng, M. Krstic, Stochastic nonlinear stabilization – I: A backstepping design, Syst. Control Lett. 32 (1997) 143 – 150.
[74] H. Deng, M. Krstic, Output-feedback stochastic nonlinear stabilization, IEEE Trans. Autom. Control 44 (1999) 328–333.
[75] B. B. Sharma, I. N. Kar, Contraction based adaptive control of a class of nonlinear systems, in: Proc. Amer. Control Conf., 2009, pp. 808–813.
[76] J. R. Cloutier, State-dependent Riccati equation techniques: An overview, in: Proc. Amer. Control Conf., volume 2, 1997, pp. 932–936.
[77] H. T. Banks, B. M. Lewis, H. T. Tran, Nonlinear feedback controllers and compensators: A state-dependent Riccati equation approach, Comput. Optim. Appl. 37 (2007) 177–218.
[78] T. Çimen, Survey of state-dependent Riccati equation in nonlinear optimal feedback control synthesis, J. Guid. Control Dyn. 35 (2012) 1025–1047.
[79] I. R. Manchester, J.-J. E. Slotine, Control contraction metrics: Convex and intrinsic criteria for nonlinear feedback design, IEEE Trans. Autom. Control 62 (2017) 3046–3053.
[80] S. Singh, A. Majumdar, J.-J. E. Slotine, M. Pavone, Robust online motion planning via contraction theory and convex optimization, in: IEEE ICRA, 2017, pp. 5883–5890.
[81] S. Singh, S. M. Richards, V. Sindhwani, J.-J. E. Slotine, M. Pavone, Learning stabilizable nonlinear dynamics with contraction-based regularization, Int. J. Robot. Res. (2020).
[82] R. Wang, R. Tóth, I. R. Manchester, A comparison of LPV gain scheduling and control contraction metrics for nonlinear control, in: 3rd IFAC Workshop on LPVS, volume 52, 2019, pp. 44–49.
[83] R. Wang, R. Tóth, I. R. Manchester, Virtual control contraction metrics: Convex nonlinear feedback design via behavioral embedding, arXiv:2003.08513, 2020.
[84] I. R. Manchester, Contracting nonlinear observers: Convex optimization and learning from data, in: Proc. Amer. Control Conf., 2018, pp. 1873–1880.
[85] B. T. Lopez, J.-J. E. Slotine, Adaptive nonlinear control with contraction metrics, IEEE Control Syst. Lett. 5 (2021) 205–210.
[86] B. T. Lopez, J.-J. E. Slotine, Universal adaptive control of nonlinear systems, arXiv:2012.15815, 2021.
[87] N. M. Boffi, S. Tu, J.-J. E. Slotine, Regret bounds for adaptive nonlinear control, arXiv:2011.13101, 2020.
[88] B. T. Lopez, J.-J. E. Slotine, J. P. How, Dynamic tube MPC for nonlinear systems, in: Proc. Amer. Control Conf., 2019, pp. 1655–1662.
[89] A. Lakshmanan, A. Gahlawat, N. Hovakimyan, Safe feedback motion planning: A contraction theory and $\mathcal{L}_1$-adaptive control based approach, in: IEEE CDC, 2020, pp. 1578–1583.
[90] A. Gahlawat, A. Lakshmanan, L. Song, A. Patterson, Z. Wu, N. Hovakimyan, E. A. Theodorou, Contraction $\mathcal{L}_1$-adaptive control using Gaussian processes, in: Proc. Conf. Learn. Dyn. Control, volume 144 of *Proc. Mach. Learn. Res.*, 2021, pp. 1027–1040.
[91] D. Sun, M. J. Khojasteh, S. Shekhar, C. Fan, Uncertainty-aware safe exploratory planning using Gaussian process and neural control contraction metric, arXiv:2105.06567, 2021.
[92] M. Johansson, A. Rantzer, Computation of piecewise quadratic Lyapunov functions for hybrid systems, IEEE Trans. Autom. Control 43 (1998) 555–559.
[93] T. A. Johansen, Computation of Lyapunov functions for smooth nonlinear systems using convex optimization, Automatica 36 (2000) 1617 – 1626.
[94] P. Giesl, Construction of Global Lyapunov Functions Using Radial Basis Functions, volume 1904 of *Lecture Notes in Mathematics*, Springer-Verlag Berlin Heidelberg, 2007.
[95] A. Papachristodoulou, S. Prajna, On the construction of Lyapunov functions using the sum of squares decomposition, in: IEEE CDC, volume 3, 2002, pp. 3482–3487.
[96] S. M. Richards, F. Berkenkamp, A. Krause, The Lyapunov neural network: Adaptive stability certification for safe learning of dynamical systems, in: CoRL, volume 87, 2018, pp. 466–476.

[97] Y.-C. Chang, N. Roohi, S. Gao, Neural Lyapunov control, in: Adv. Neural Inf. Process. Syst., 2019, pp. 3245–3254.

[98] D. Han, D. Panagou, Chebyshev approximation and higher order derivatives of Lyapunov functions for estimating the domain of attraction, in: IEEE Conf. Decis. Control, 2017, pp. 1181–1186. doi:10.1109/CDC.2017.8263816.

[99] Z. Wang, R. M. Jungers, Scenario-based set invariance verification for black-box nonlinear systems, IEEE Control Syst. Lett. 5 (2021) 193–198. doi:10.1109/LCSYS.2020.3001882.

[100] A. Chakrabarty, C. Danielson, S. Di Cairano, A. Raghunathan, Active learning for estimating reachable sets for systems with unknown dynamics, IEEE Trans. Cybern. (2020) 1–12. doi:10.1109/TCYB.2020.3000966.

[101] A. Chakrabarty, K. Berntorp, S. Di Cairano, Learning-based parameter-adaptive reference governors, in: Amer. Control Conf., 2020, pp. 956–961. doi:10.23919/ACC45564.2020.9147615.

[102] N. Gaby, F. Zhang, X. Ye, Lyapunov-net: A deep neural network architecture for Lyapunov function approximation, arXiv:2109.13359, 2021.

[103] W. Tan, Nonlinear control analysis and synthesis using sum-of-squares programming, Ph.D. thesis, Univ. California, Berkeley, 2006.

[104] H. Tsukamoto, S.-J. Chung, J.-J. E. Slotine, Learning-based adaptive control via contraction theory, in: IEEE CDC, 2021.

[105] H. Tsukamoto, S.-J. Chung, Learning-based robust motion planning with guaranteed stability: A contraction theory approach, IEEE Robot. Automat. Lett. 6 (2021) 6164–6171. doi:10.1109/LRA.2021.3091019.

[106] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, Neural Networks 2 (1989) 359–366.

[107] G. Cybenko, Approximation by superpositions of a sigmoidal function, Math. Control Signals Syst 2 (1989) 303–314.

[108] K.-I. Funahashi, On the approximate realization of continuous mappings by neural networks, Neural Networks 2 (1989) 183–192.

[109] H. Robbins, S. Monro, A stochastic approximation method, Ann. Math. Statist. 22 (1951) 400 – 407.

[110] R. S. Sutton, A. G. Barto, Reinforcement Learning: An Introduction, MIT Press, 1998.

[111] D. P. Bertsekas, J. N. Tsitsiklis, Neuro-Dynamic Programming, Athena Scientific, 1996.

[112] M. Everett, Y. F. Chen, J. P. How, Motion planning among dynamic, decision-making agents with deep reinforcement learning, in: IEEE/RSJ IROS, 2018, pp. 3052–3059.

[113] F. Berkenkamp, M. Turchetta, A. Schoellig, A. Krause, Safe model-based reinforcement learning with stability guarantees, in: Adv. Neural Inf. Process. Syst., volume 30, Curran Associates, Inc., 2017, pp. 908–918.

[114] J. Carius, F. Farshidian, M. Hutter, Mpc-net: A first principles guided policy search, IEEE Robot. Automat. Lett. 5 (2020) 2897–2904.

[115] B. Rivière, W. Hönig, Y. Yue, S. Chung, GLAS: Global-to-local safe autonomy synthesis for multi-robot motion planning with end-to-end learning, IEEE Robot. Automat. Lett. 5 (2020) 4249–4256.

[116] J. Ho, S. Ermon, Generative Adversarial Imitation Learning, in: D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, R. Garnett (Eds.), Adv. Neural Inf. Process. Syst., volume 29, Curran Associates, Inc., 2016, pp. 4565–4573.

[117] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, A. Alahi, Social GAN: Socially acceptable trajectories with generative adversarial networks, in: IEEE/CVF Conf. Comput. Vision Pattern Recognit., 2018, pp. 2255–2264.

[118] A. Kuefler, J. Morton, T. Wheeler, M. Kochenderfer, Imitating driver behavior with generative adversarial networks, in: IEEE Intell. Vehicles Symp., 2017, pp. 204–211.

[119] H. Tsukamoto, S.-J. Chung, J.-J. E. Slotine, C. Fan, A theoretical overview of neural contraction metrics for learning-based control with guaranteed stability, in: IEEE CDC, 2021.

[120] D. Sun, S. Jha, C. Fan, Learning certified control using contraction metric, arXiv:2011.12569, 2020.

[121] P. Zhao, A. Lakshmanan, K. Ackerman, A. Gahlawat, M. Pavone, N. Hovakimyan, Tube-certified trajectory tracking for nonlinear systems with robust control contraction metrics, arXiv:2109.04453, 2021.

[122] D. Q. Mayne, E. C. Kerrigan, E. J. van Wyk, P. Falugi, Tube-based robust nonlinear model predictive control, Int. J. Robust Nonlinear Control 21 (2011) 1341–1353.

[123] D. Mayne, M. Seron, S. Raković, Robust model predictive control of constrained linear systems with bounded disturbances, Automatica 41 (2005) 219 – 224.

[124] A. Bemporad, M. Morari, Robust model predictive control: A survey, in: Robustness in identification and control, Springer London, London, 1999, pp. 207–226.

[125] L. Hewing, K. P. Wabersich, M. Menner, M. N. Zeilinger, Learning-based model predictive control: Toward safe learning in control, Annu. Rev. Control Robot. Auton. Syst. 3 (2020) 269–296.

[126] R. H. Battin, An Introduction to the Mathematics and Methods of Astrodynamics, AIAA Education Series, 1999.

[127] D. Morgan, S.-J. Chung, L. Blackmore, B. Acikmese, D. Bayard, F. Y. Hadaegh, Swarm-keeping strategies for spacecraft under J_2 and atmospheric drag perturbations, Journal of Guidance, Control, and Dynamics 35 (2012) 1492–1506.

[128] O. Nelles, Nonlinear Dynamic System Identification, Springer Berlin Heidelberg, 2001.

[129] R. M. Sanner, J.-J. E. Slotine, Gaussian networks for direct adaptive control, IEEE Trans. Neur. Netw. 3 (1992) 837–863.

[130] K. El Rifai, O. El Rifai, K. Youcef-Toumi, On robust adaptive switched control, in: Proc. Amer. Control Conf., volume 1, 2005, pp. 18–23.

[131] C. Blocher, M. Saveriano, D. Lee, Learning stable dynamical systems using contraction theory, in: IEEE URAI, 2017, pp. 124–129.

[132] T. Miyato, T. Kataoka, M. Koyama, Y. Yoshida, Spectral normalization for generative adversarial networks, in: Int. Conf. Learn. Representations, 2018.

[133] N. M. Boffi, S. Tu, N. Matni, J.-J. E. Slotine, V. Sindhwani, Learning stability certificates from data, arXiv:2008.05952, 2020.

[134] D. E. Kirk, Optimal Control Theory: An Introduction, Dover Publications, 2004.

[135] J. Jouffroy, J.-J. E. Slotine, Methodological remarks on contraction theory, in: IEEE CDC, volume 3, 2004, pp. 2537–2543.

[136] W. J. Rugh, Linear Systems Theory, Prentice-Hall, Inc., USA, 1996.

[137] S.-I. Amari, Natural gradient works efficiently in learning, Neural Computation 10 (1998) 251–276.

[138] C. Udriste, Convex functions and optimization methods on Riemannian manifolds, volume 297, Springer Science & Business Media, 1994.

[139] R. A. Horn, C. R. Johnson, Matrix Analysis, 2nd ed., Cambridge University Press, 2012.

[140] L. Arnold, Stochastic Differential Equations: Theory and Applications, Wiley, 1974.

[141] H. J. Kushner, Stochastic Stability and Control, Academic Press New York, 1967.

[142] T.-J. Tarn, Y. Rasis, Observers for nonlinear stochastic systems, IEEE Trans. Autom. Control 21 (1976) 441–448.

[143] M. Zakai, On the ultimate boundedness of moments associated with solutions of stochastic differential equations, SIAM J. Control 5 (1967) 588–593.

[144] D. S. Geoffrey R. Grimmett, Probability and Random Processes, 3rd ed., Oxford University Press, United Kingdom, 2001.

[145] E. Mazumdar, T. Westenbroek, M. I. Jordan, S. Shankar Sastry, High confidence sets for trajectories of stochastic time-varying nonlinear systems, in: IEEE CDC, 2020, pp. 4275–4280.

[146] T. Basar, P. Bernhard, $\mathcal{H}_\infty$ Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach, Birkhauser, 1995.
[147] W. Zhang, B. Chen, State feedback $\mathcal{H}_\infty$ control for a class of nonlinear stochastic systems, SIAM J. Control Optim. 44 (2006) 1973–1991.
[148] W. Li, E. Todorov, Iterative linear quadratic regulator design for nonlinear biological movement systems, in: Int. Conf. Inform. Control Automat. Robot., 2004, pp. 222–229.
[149] W. Li, E. Todorov, An iterative optimal control and estimation design for nonlinear stochastic system, in: IEEE CDC, 2006, pp. 3242–3247.
[150] I. Chang, S.-J. Chung, Exponential stability region estimates for the state-dependent Riccati Equation controllers, in: IEEE CDC, 2009, pp. 1974–1979.
[151] T. Başar, P. Bernhard, $\mathcal{H}_\infty$ Optimal Control and Related Minimax Design Problems, Modern Birkhäuser Classics, 2nd ed., Birkhäuser Basel, 2008.
[152] R. G. Brown, P. Y. C. Hwang, Introduction to Random Signals and Applied Kalman Filtering: with MATLAB Exercises and Solutions, 4th ed., Wiley, New York, NY, 1997.
[153] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.
[154] X. Chen, M. Monfort, A. Liu, B. D. Ziebart, Robust Covariate Shift Regression, in: A. Gretton, C. C. Robert (Eds.), 19th Int. Conf. Artif. Intell. Statist., volume 51 of *Proc. Mach. Learn. Res.*, Cadiz, Spain, 2016, pp. 1270–1279.
[155] A. Liu, G. Shi, S.-J. Chung, A. Anandkumar, Y. Yue, Robust regression for safe exploration in control, in: CoRL, 2020.
[156] C. Zhang, S. Bengio, M. Hardt, B. Recht, O. Vinyals, Understanding deep learning requires rethinking generalization, in: 5th Int. Conf. Learn. Representations, OpenReview.net, Toulon, France, 2017.
[157] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: IEEE Conf. Comput. Vision Pattern Recognit., 2016, pp. 770–778.
[158] G. K. Dziugaite, D. M. Roy, Computing nonvacuous generalization bounds for deep (stochastic) neural networks with many more parameters than training data, in: 33rd Conf. Uncertainty Artif. Intell., AUAI Press, Sydney, Australia, 2017.
[159] B. Neyshabur, S. Bhojanapalli, D. Mcallester, N. Srebro, Exploring generalization in deep learning, in: Adv. Neural Inf. Process. Syst., volume 30, Curran Associates, Inc., 2017.
[160] P. L. Bartlett, D. J. Foster, M. J. Telgarsky, Spectrally-normalized margin bounds for neural networks, in: Adv. Neural Inf. Process. Syst., volume 30, 2017.
[161] B. Neyshabur, S. Bhojanapalli, N. Srebro, A PAC-Bayesian approach to spectrally-normalized margin bounds for neural networks, in: Int. Conf. Learn. Representations, 2018.
[162] A. J. Taylor, V. D. Dorobantu, H. M. Le, Y. Yue, A. D. Ames, Episodic learning with control Lyapunov functions for uncertain robotic systems, in: IEEE/RSJ IROS, 2019, pp. 6878–6884.
[163] R. Cheng, A. Verma, G. Orosz, S. Chaudhuri, Y. Yue, J. Burdick, Control regularization for reduced variance reinforcement learning, in: K. Chaudhuri, R. Salakhutdinov (Eds.), 36th Int. Conf. Mach. Learn., volume 97 of *Proc. Mach. Learn. Res.*, PMLR, 2019, pp. 1141–1150.
[164] C. D. McKinnon, A. P. Schoellig, Learn fast, forget slow: Safe predictive learning control for systems With unknown and changing dynamics performing repetitive tasks, IEEE Robot. Automat. Lett. 4 (2019) 2180–2187.
[165] G. Shi, W. Hönig, Y. Yue, S.-J. Chung, Neural-swarm: Decentralized close-proximity multirotor control using learned interactions, in: IEEE ICRA, 2020, pp. 3241–3247.
[166] G. Shi, W. Hönig, X. Shi, Y. Yue, S.-J. Chung, Neural-swarm2: Planning and control of heterogeneous multirotor swarms using learned interactions, IEEE Trans. Robot. (2021) 1–17.
[167] Y. Nakka, R. Foust, E. Lupu, D. Elliott, I. Crowell, S.-J. Chung, F. Hadaegh, A six degree-of-freedom spacecraft dynamics simulator for formation control research, in: Astrodyn. Special. Conf., 2018.
[168] J. S. Shamma, J. R. Cloutier, Gain-scheduled missile autopilot design using linear parameter varying transformations, J. Guid. Control Dyn. 16 (1993) 256–263. doi:10.2514/3.20997.
[169] S. Bennani, D. Willemsen, C. Scherer, C. Scherer, S. Bennani, D. Willemsen, Robust LPV control with bounded parameter rates, in: Guid. Navigation Control Conf., 1997.
[170] J. Schulman, Y. Duan, J. Ho, A. Lee, I. Awwal, H. Bradlow, J. Pan, S. Patil, K. Goldberg, P. Abbeel, Motion planning with sequential convex optimization and convex collision checking, Int. J. Robot. Res. 33 (2014) 1251–1270.
[171] Y. K. Nakka, A. Liu, G. Shi, A. Anandkumar, Y. Yue, S. J. Chung, Chance-constrained trajectory optimization for safe exploration and learning of nonlinear systems, IEEE Robot. Automat. Lett. 6 (2021) 389–396.
[172] A. G. Barto, R. S. Sutton, C. W. Anderson, Neuronlike adaptive elements that can solve difficult learning control problems, IEEE Trans. Syst. Man Cybern. SMC-13 (1983) 834–846.
[173] Y. K. K. Nakka, et al., Information-based guidance and control architecture for multi-spacecraft on-orbit inspection, in: AIAA Scitech, 2021.
[174] S. Kolathaya, A. D. Ames, Input-to-state safety with control barrier functions, IEEE Control Syst. Lett. 3 (2019) 108–113.
[175] A. D. Ames, J. W. Grizzle, P. Tabuada, Control barrier function based quadratic programs with application to adaptive cruise control, in: IEEE CDC, 2014, pp. 6271–6278.
[176] J.-J. E. Slotine, J. A. Coetsee, Adaptive sliding controller synthesis for non-linear systems, Int. J. Control 43 (1986) 1631–1651.
[177] N. M. Boffi, J.-J. E. Slotine, Implicit regularization and momentum algorithms in nonlinear adaptive control and prediction, arXiv:1912.13154, 2020.
[178] E. Lavretsky, T. E. Gibson, Projection operator in adaptive systems, arXiv:1112.4232, 2012.
[179] Z. Cai, M. de Queiroz, D. Dawson, A sufficiently smooth projection operator, IEEE Trans. Autom. Control 51 (2006) 135–139.