Deep-Neuro Control with Contraction Theory

Myeongseok Ryu *, Sesun You †, Kyunghwan Choi *

March 26, 2025
Version 0.0

Abstract

This project aims to develop control or estimator with deep neural network and contraction theory. Learning-based control methods have been widely studied as remarkable advancements in the deep learning field. However, naive application of deep learning to control problems often leads to poor performance and instability. Even though some literature has shown that the learning-based control methods have been successfully applied to various control problems, the theoretical analysis of the stability and performance in the control perspective is still lacking. Moreover, recent deep learning studies begins from the stochastic optimization algorithms, which make the analysis of the stability and performance more difficult. To overcome theses difficulties, the contraction theory is studied to analyze its capability to analysis the stability and performance of the learning-based control methods.

Contents

1 Introduction

1.1 Background

1.2 Research Objectives

The main objectives of this research are as follows:

- Mathematical stability analysis of the controller and estimator with deep neural networks using the contraction theory.

- Development of the controller and estimator with deep neural networks using the contraction theory.

---

*Myeongseok Ryu and Kyunghwan Choi are with the School of Mechanical and Robotics Engineering, Gwangju Institute of Science and Technology, 61005 Gwangju, Republic of Korea dding_98@gm.gist.ac.kr, khchoi@gist.ac.kr
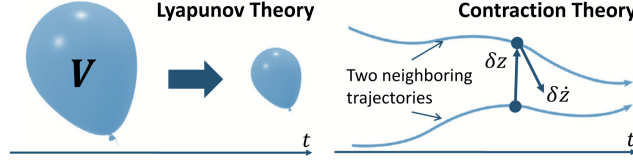†Sesun You is example@mail.com

Figure 1: Difference between Lyapunov and contraction theory [4, Fig. 1]. The Lyapunov theory investigates the convergence to a single point and the contraction theory does regarding a single trajectory.

2 Notations and Preliminaries

The following notations are used throughout this document:

- := denotes *defined as*.

- $(\cdot)^\top$ denotes the transpose of a matrix or a vector.

- $\boldsymbol{x} := [x_i]_{i \in \{1, \cdots, n\}} \in \mathbb{R}^n$ denotes the state vector.

- $\boldsymbol{A} := [a_{ij}]_{i,j \in \{1, \cdots, n\}} \in \mathbb{R}^{n \times n}$ denotes a matrix.

- $\lambda_i(\boldsymbol{A})$, $i \in \{\max, \min\}$ denotes the maximum and minimum singular value of $\boldsymbol{A}$, respectively.

- $\boldsymbol{I}_n$ denotes the identity matrix of size n and $\boldsymbol{0}_{n \times m}$ denotes the zero matrix of size $n \times m$.

- sym denotes the symmetric part of a matrix, i.e., $\mathrm{sym}(\boldsymbol{A}) := \boldsymbol{A} + \boldsymbol{A}^\top$ (see, [1]).

We introduce the following lemmas.

Lemma 2.1 (Comparison Lemma). *Suppose that a continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ satisfies the following inequality:*

$$\tfrac{\mathrm{d}}{\mathrm{d}t} f(t) \leq -a f(t) + b, \quad \forall t \in \mathbb{R}_{\geq 0},$$

where $a, b > 0$. Then, the following inequality holds:

$$f(t) \leq -a f(0) e^{-at} + \tfrac{b}{a}(1 - e^{-at}), \quad \forall t \in \mathbb{R}_{\geq 0}$$

and remains in a compact set $f(t) \in \{\|f(t)\| \mid \|f(0)\| \leq \tfrac{b}{a}\}$.

Proof. This is a simple special case of the comparison lemma [2, pp. 102-103]. See [2, pp. 659-660]. $\square$

3 Review of Contraction Theory

For your smooth start, we recommend you to begin with [3]. The overview of contraction theory is presented in a review paper [4].

3.1 Basic Results of Contraction Theory for Deterministic Systems

First, we start with the following deterministic systems:

$$\tfrac{\mathrm{d}}{\mathrm{d}t} \boldsymbol{x} = \boldsymbol{f}(\boldsymbol{x}, t), \tag{1}$$

where $\boldsymbol{f}(\boldsymbol{x}, t)$ is an $n \times 1$ sufficiently smooth non-linear vector function and $\boldsymbol{x} \in \mathbb{R}^n$ is the state vector. The smooth property of $\boldsymbol{f}(\boldsymbol{x}, t)$ is essential to ensure the existence and uniqueness of the solution to (1) [2, see, pp. 88-89].

The biggest difference between the traditional Lyapunov theory and the contraction theory is that the contraction theory investigates the convergence of the state trajectory to a single trajectory (contraction behavior), while the Lyapunov theory focuses on the convergence of the state trajectory to a single point i.e., see, Fig. 1. For this, motivated by the calculus of variations [5, Chap. 4], (1) can be rewritten as differential dynamics using *differential displacement* $\delta \boldsymbol{x}$ as follows:

$$\tfrac{\mathrm{d}}{\mathrm{d}t} \delta \boldsymbol{x} = \tfrac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}(\boldsymbol{x}, t) \delta \boldsymbol{x}. \tag{2}$$

For your information, $\delta \boldsymbol{x}$ is an infinitesimal displacement at *fixed time*.

3.1.1 Notable Definitions

Before we present the fundamental theorem of contraction theory, we introduce the following definitions. One can re-visit this section while reading further.

Definition 3.1 (see, Def. 2.2 [4]). *If any two trajectories $\boldsymbol{\xi}_1(t)$ and $\boldsymbol{\xi}_2(t)$ of (1) converge to a single trajectory, then the system (1) is said to be incrementally exponentially stable, if $\exists C, \alpha > 0$, subject to the following holds:*

$$\|\boldsymbol{\xi}_1(t) - \boldsymbol{\xi}_2(t)\| \leq C\|\boldsymbol{\xi}_1(0) - \boldsymbol{\xi}_2(0)\| \exp^{-\alpha t}, \ \forall t \in \mathbb{R}_{\geq 0}.$$

The result of Theorem 3.1 equivalently implies the incremental exponential stability, since we have $\|\boldsymbol{\xi}_1(t) - \boldsymbol{\xi}_2(t)\| = \|\int_{\boldsymbol{\xi}_1(t)}^{\boldsymbol{\xi}_2(t)} \delta\boldsymbol{x}(t)\|$.

Definition 3.2. *Let $\boldsymbol{\Theta}(\boldsymbol{x}, t)$ be a smooth coordinate transformation of $\delta\boldsymbol{x}$ to $\delta\boldsymbol{z}$, i.e., $\delta\boldsymbol{z} = \Theta(\boldsymbol{x}, t)\delta\boldsymbol{x}$. Then, a symmetric continuously differentiable matrix $\boldsymbol{M}(\boldsymbol{x}, t) := \boldsymbol{\Theta}(\boldsymbol{x}, t)^\top \boldsymbol{\Theta}(\boldsymbol{x}, t)$ is said to be a metric of the system (1).*

Definition 3.3. *The covariant derivative of $\boldsymbol{f}(\boldsymbol{x}, t)$ in $\delta\boldsymbol{x}$ coordinate is represented as*

$$\boldsymbol{F} := \left(\tfrac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\Theta} + \boldsymbol{\Theta}\tfrac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} \right) \boldsymbol{\Theta}^{-1},$$

and is called the generalized Jacobian. This can be easily derived by differentiating $\delta\boldsymbol{z} = \Theta(\boldsymbol{x}, t)\delta\boldsymbol{x}$ with respect to t, leading to $\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{z} = \boldsymbol{F}\boldsymbol{z}$.

3.1.2 Fundamental Theorem of Contraction Theory

The following theorem presents the fundamental theorem of contraction theory and corresponding necessary and sufficient condition for exponential convergence of the differential system (2).

Theorem 3.1 (see, T. 2.1 [4]). *If $\exists \boldsymbol{M}(\boldsymbol{x}, t) = \boldsymbol{\Theta}(\boldsymbol{x}, t)^\top \boldsymbol{\Theta}(\boldsymbol{x}, t) > 0, \forall\boldsymbol{x}, t$ where $\boldsymbol{\Theta}(\boldsymbol{x}, t)$ defines a smooth coordinate transformation of $\delta\boldsymbol{x}$ to $\delta\boldsymbol{z}$, i.e., $\delta\boldsymbol{z} = \Theta(\boldsymbol{x}, t)\delta\boldsymbol{x}$, subject to the following equivalent conditions holds for $\exists\alpha \in \mathbb{R}_{>0}, \ \forall\boldsymbol{x}, t$:*

$$\lambda_{\max}(\boldsymbol{F}(\boldsymbol{x}, t)) = \lambda_{\max}\left(\left(\tfrac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\Theta} + \boldsymbol{\Theta}\tfrac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} \right) \boldsymbol{\Theta}^{-1} \right) \leq -\alpha,$$

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{M} + \operatorname{sym}(\boldsymbol{M}\tfrac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}) \leq -2\alpha\boldsymbol{M},$$

where the arguments of $\boldsymbol{M}(\boldsymbol{x}, t)$ and $\boldsymbol{F}(\boldsymbol{x}, t)$ are omitted for simplicity, then, the system (1) is said to be contracting with an exponential rate α, i.e., all trajectories of (1) converge to a single trajectory. The converse is also true.

Proof. Taking time derivative of a differential Lyapunov function of $\delta\boldsymbol{x}$, $V = \delta\boldsymbol{z}^\top \boldsymbol{z} = \delta\boldsymbol{x}^\top \boldsymbol{M}\delta\boldsymbol{x}$, using the differential dynamics (2), we have

$$\tfrac{\mathrm{d}}{\mathrm{d}t}V(\boldsymbol{x}, \delta\boldsymbol{x}, t) = \delta\boldsymbol{x}^\top \left(\tfrac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{M} + \operatorname{sym}(\boldsymbol{M}\tfrac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}) \right) \delta\boldsymbol{x} = 2\delta\boldsymbol{z}^\top \boldsymbol{F}\delta\boldsymbol{z}$$

$$\leq -2\alpha\delta\boldsymbol{x}^\top \boldsymbol{M}\delta\boldsymbol{x} = -2\alpha\delta\boldsymbol{z}^\top\delta\boldsymbol{z} = -2\alpha V.$$

According to the comparison lemma (Lemma 2.1), we have $V(t) \leq V(0)e^{-2\alpha t}$, which then yields $\|\delta\boldsymbol{z}(t)\|^2 \leq \|\delta\boldsymbol{z}(0)\|^2 e^{-2\alpha t}$ and $\|\delta\boldsymbol{z}(t)\| \leq \|\delta\boldsymbol{z}(0)\|e^{-\alpha t}$. This implies that any infinitesimal displacement $\delta\boldsymbol{z}$ and $\delta\boldsymbol{x}$ converges to zero exponentially with rate α. Note that the initial conditions are exponentially "forgotten" as time goes on. The proof of the converse can be found in [3, Sec. 3.5]. □

It is notable that the unboundedness of the metric $\boldsymbol{M}(\boldsymbol{x}, t)$ does not create any problem in a technical sense. This is because, the dynamics of $\boldsymbol{M}(\boldsymbol{x}, t)$ is linear with infinite escape time. Therefore, it can be handled by renormalizing the metric $\boldsymbol{M}(\boldsymbol{x}, t)$.

Theorem 3.1 can also be proven by using the transformed squared length integrated over two arbitrary solutions of (1). The following theorem presents the alternative proof of Theorem 3.1.

Theorem 3.2 (see, p. 688 [3], T. 2.3 [4]). *Let $\boldsymbol{\xi}_1(t)$ and $\boldsymbol{\xi}_2(t)$ be two solutions of (1), and define the transformed squared length with $\boldsymbol{M}(\boldsymbol{x}, t)$ of Theorem 3.1 as follows:*

$$V_{sl}(\boldsymbol{x}, \delta\boldsymbol{x}, t) = \int_{\boldsymbol{\xi}_1(t)}^{\boldsymbol{\xi}_2(t)} \|\delta\boldsymbol{z}\|^2 = \int_0^1 \tfrac{\partial \boldsymbol{x}}{\partial \mu}^\top \boldsymbol{M}(\boldsymbol{x}, t)\tfrac{\partial \boldsymbol{x}}{\partial \mu}\mathrm{d}\mu, \tag{4}$$

where $\boldsymbol{x}$ is a smooth path parameterized as $\boldsymbol{x}(\mu = 0, t) = \boldsymbol{\xi}_1(t)$ and $\boldsymbol{x}(\mu = 1, t) = \boldsymbol{\xi}_2(t)$ by $\mu \in \{0, 1\}$. Also, define the path integral with the transformation $\boldsymbol{\Theta}(\boldsymbol{x}, t)$ for $\boldsymbol{M}(\boldsymbol{x}, t) = \boldsymbol{\Theta}(\boldsymbol{x}, t)^\top \boldsymbol{\Theta}(\boldsymbol{x}, t)$ as follows:

$$V_l(\boldsymbol{x}, \delta\boldsymbol{x}, t) = \int_{\boldsymbol{\xi}_1(t)}^{\boldsymbol{\xi}_2(t)} \|\delta\boldsymbol{z}\| = \int_{\boldsymbol{\xi}_1(t)}^{\boldsymbol{\xi}_2(t)} \|\boldsymbol{\Theta}(\boldsymbol{x}, t)\delta\boldsymbol{x}\|. \tag{5}$$

Then, (4) and (5) satisfy the following inequality:

$$\|\boldsymbol{\xi}_1(t) - \boldsymbol{\xi}_2(t)\| = \|\int_{\boldsymbol{\xi}_1(t)}^{\boldsymbol{\xi}_2(t)} \delta\boldsymbol{x}\| \leq \frac{V_l}{\sqrt{\underline{m}}} \leq \sqrt{\frac{V_{sl}}{\underline{m}}}, \tag{6}$$

where $\boldsymbol{M}(\boldsymbol{x}, t) \geq \underline{m}\boldsymbol{I}_n$, $\forall \boldsymbol{x}, t$ for $\exists \underline{m} \in \mathbb{R}_{>0}$ and Theorem 3.1 can also be proven by using (4) and (5) as a Lyapunov-like function, resulting in incremental exponential stability of the system (1) (see, Definition 3.1). Note that the shortest path integral V_l of (5) with a parameterized state $\boldsymbol{x}$ (i.e., $\inf V_l = \sqrt{\inf V_{sl}}$) defines the Riemannian distance and the path integral of a minimizing geodesic.

Proof. Note that $\boldsymbol{\xi}_i(t)$, $\forall i \in \{1, 2\}$ are parameterized $\boldsymbol{x}$. Therefore, left-side equailty of (6) holds. The right-side inequality of (6) can be proven by using the Cauchy-Schwarz inequality.

On the other hands, by taking time derivative of V_{sl} and V_l, we have $\frac{\mathrm{d}}{\mathrm{d}t}V_{sl} \leq -2\alpha V_{sl}$ and $\frac{\mathrm{d}}{\mathrm{d}t}V_l \leq -\alpha V_l$. As $\boldsymbol{M}(\boldsymbol{x}, t)$ is from Theorem 3.1, the incremental exponential stability of the system (1) is guaranteed using the comparison lemma of Lemma 2.1 as follows:

$$\|\boldsymbol{\xi}_1(t) - \boldsymbol{\xi}_2(t)\| \leq \frac{V_l(0)}{\sqrt{\underline{m}}} \exp(-\alpha t).$$

$\square$

Example 3.1. Consider the following system:

$$\frac{\mathrm{d}}{\mathrm{d}t} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 & x_1 \\ -x_1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

However, aforementioned theorems are limited to the convergence of the state trajectory to a single trajectory. In the next section, we introduce the partial contraction theory to investigate the convergence of the state trajectory to a desired trajectory.

3.1.3 Partial Contraction

3.2 Derterministic Perturbation

If there exists a bounded perturbation in (1) as follows:

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{x} = \boldsymbol{f}(\boldsymbol{x}, t) + \boldsymbol{q}(\boldsymbol{x}, t),$$

[6, 7]

3.3 Basic Results of Contraction Theory for Stochastic Systems

4 Conclusion

References

[1] H. Tsukamoto, S.-J. Chung, and J.-J. E. Slotine, "Neural stochastic contraction metrics for learning-based control and estimation," *IEEE Control Systems Letters*, vol. 5, no. 5, pp. 1825–1830, 2021.

[2] H. K. Khalil, *Nonlinear systems; 3rd ed.* Upper Saddle River, NJ: Prentice-Hall, 2002. The book can be consulted by contacting: PH-AID: Wallet, Lionel.

[3] W. LOHMILLER and J.-J. E. SLOTINE, "On contraction analysis for non-linear systems," *Automatica*, vol. 34, no. 6, pp. 683–696, 1998.

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

[5] D. Kirk, *Optimal Control Theory: An Introduction.* Dover Books on Electrical Engineering Series, Dover Publications, 2004.

[6] W. Wang and J.-J. E. Slotine, "On partial contraction analysis for coupled nonlinear oscillators," *Biological Cybernetics*, vol. 92, pp. 38–53, Dec. 2004.

[7] J. Jouffroy and J.-J. Slotine, "Methodological remarks on contraction theory," in *2004 43rd IEEE Conference on Decision and Control (CDC)*, vol. 3, pp. 2537–2543 Vol.3, 2004.