Statistical Contraction Theory for Learning-based Control of Nonlinear Systems—Part 1: Contraction Theory

Hiroyasu Tsukamoto

Department of Aerospace Engineering
University of Illinois Urbana-Champaign

February 7, 2026

Table of contents

- In the following, $\|\cdot\|$ denotes the Euclidean norm for vectors and the induced 2-norm for matrices, i.e., $\|\cdot\| = \|\cdot\|_2$, unless otherwise noted.
- For a symmetric square matrix $S \in \mathbb{R}^{n\times n}$, we use the notation $S \succ 0$, $S \succeq 0$, $S \prec 0$, and $S \preceq 0$ for the positive definite, positive semi-definite, negative definite, and negative semi-definite matrices, respectively.
- This lecture is based on [1, 2].

Table of contents

A Linear System as an Analogy

Let's start with a very simple example. Consider a linear dynamical system

$$\dot{\mathbf{x}} = A\mathbf{x} \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^n$ is the system state and A is a constant matrix of appropriate dimension.

A Linear System as an Analogy

What is the condition for (1) to exponentially track an arbitrary desired trajectory $\mathbf{x}_d$ given by $\dot{\mathbf{x}}_d = A\mathbf{x}_d$?

A Linear System as an Analogy

We know that the following equivalence holds for linear systems:

(1) *exponentially tracks any* $\mathbf{x}_d \Leftrightarrow \exists P \succ 0$ *and* $\alpha > 0$ *s.t.* $PA + A^\top P \preceq -2\alpha P$.

We call $V(\mathbf{x}, \mathbf{x}_d) = (\mathbf{x} - \mathbf{x}_d)^\top P(\mathbf{x} - \mathbf{x}_d)$ a Lyapunov function.

ILLNOIS

A Linear System as an Analogy

Can we formulate an analogous result for nonlinear systems?

Introduction to Contraction Theory

Yes, to some extent.

The key observation is on how we constructed the Lyapunov function:

$$V(\mathbf{x}, \mathbf{x}_d) = \underbrace{(\mathbf{x} - \mathbf{x}_d)^\top}_{\text{difference of two trajectories}} \underbrace{P}_{\text{some positive definite matrix}} \underbrace{(\mathbf{x} - \mathbf{x}_d)}_{\text{difference of two trajectories}}$$

which is like a distance between the two system solution trajectories measured using a metric defined by P.

Based on this observation, we define what's called the incremental Lyapunov function defined as follows:

$$V(\mathbf{x}, \delta\mathbf{x}, t) = \delta\mathbf{x}^{\top} M(\mathbf{x}, t)\delta\mathbf{x}$$

where, as in the tracking example,

- $\delta\mathbf{x}$ is the differential state formalizing the "difference in the system solution trajectories when t is fixed,"
- $M(\mathbf{x}, t) \succ 0$ is a positive definite matrix that defines our metric.

Introduction to Contraction Theory

Contraction theory is about coming up with a metric and the matrix function M that satisfies the Lyapunov decreasing condition.

Such a metric is called the contraction metric.

Table of contents

Contraction Theory

Let's talk about the mathematical details.

Increment and Differential

Definition (Increment)

The increment $\Delta\phi(x,\delta x)$ of a function $\phi(x)$, where $x,\delta x \in \mathbb{R}^n$ (δx does not have to be infinitesimal) is defined as [3, p. 56]

$$\Delta\phi(x,\delta x) = \phi(x+\delta x) - \phi(x).$$

Definition (Differential)

The part of the increment linear in δx is called the differential of ϕ [3, p. 56], and it is often denoted by $\delta\phi(x,\delta x)$.

Note that for the function $\phi(x) = x$, the increment and the differential are just δx.

Increment and Differential

For example, if ϕ is differentiable, the increment $\Delta\phi$ and the differential $\delta\phi$ are

$$\Delta\phi(x,\delta x) = \frac{\partial \phi}{\partial x}(x)\delta x + \mathcal{O}(\|dx\|^2)$$
$$\delta\phi(x,\delta x) = \frac{\partial \phi}{\partial x}(x)\delta x.$$

And just so you know...

Definition (Variation)

If ϕ is a functional (a function of functions, roughly), the differential $\delta\phi$ is also called the variation, and δx is called the variation of the curve [4].

The variation is often used in optimal control.

Increment and Differential

Let's now consider the dynamical system

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

and a virtual displacement at a fixed time, denoted by $\delta\mathbf{x}(t)$.

Definition

Suppose that f is differentiable and the displacement is designed in a way that $\mathbf{x}(t) + \delta\mathbf{x}(t)$ also satisfies the dynamics. The part of the dynamics linear in $\delta\mathbf{x}(t)$ is called the differential dynamics (or variational dynamics) and is given by

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

How is this useful?

Virtual System

Let $\mathbf{x}_0$ and $\mathbf{x}_1$ be some particular solutions of the dynamics

$$\dot{\mathbf{x}} = f(\mathbf{x}(t), t). \tag{3}$$

and let $\mathbf{q}(\mu, t)$ be a smooth path connecting $\mathbf{x}_0$ and $\mathbf{x}_1$, parameterized by $\mu \in [0, 1]$, which also satisfies the dynamics

$$\dot{\mathbf{q}}(\mu, t) = f(\mathbf{q}(\mu, t), t). \tag{4}$$

with

- $\mathbf{q}(0, t) = \mathbf{x}_0$
- $\mathbf{q}(1, t) = \mathbf{x}_1$.

(4) is often called the virtual system.

Virtual System

Then, the dynamics of $\frac{\partial \mathbf{q}}{\partial \mu}$ is given as

$$\frac{d}{dt}\left(\frac{\partial \mathbf{q}}{\partial \mu}\right) = \frac{\partial f}{\partial \mathbf{q}}\frac{\partial \mathbf{q}}{\partial \mu} \tag{5}$$

and also, by construction,

$$\int_0^1 \frac{\partial \mathbf{q}}{\partial \mu} d\mu = \mathbf{q}(1,t) - \mathbf{q}(0,t) = \mathbf{x}_1(t) - \mathbf{x}_0(t). \tag{6}$$

Note that (5) has the same structure as the differential dynamics (2).

Virtual System

These relations imply that we can

(1) exploit the linear structure of (2) in analyzing the system performance as in Lyapunov theory for LTV systems

(2) and then utilize (6) to study the performance incrementally.

Contraction Theory

As implied in (6), the differential dynamics (2) intrinsically handles "incremental" behavior of the system.

Let's now generalize our ad-hoc idea of the incremental Lyapunov function

$$V(\mathbf{x},\mathbf{x}_d) = \underbrace{(\mathbf{x}-\mathbf{x}_d)^\top}_{\text{difference of two trajectories}} \underbrace{P}_{\text{some positive definite matrix}} \underbrace{(\mathbf{x}-\mathbf{x}_d)}_{\text{difference of two trajectories}}$$

and consider the following more intrinsic incremental, differential Lyapunov function:

$$V(\mathbf{x},\delta\mathbf{x},t) = \delta\mathbf{x}^\top M(\mathbf{x},t)\delta\mathbf{x}$$

What is the condition for the matrix weight $M(\mathbf{x},t)$ that achieves $\|\delta\mathbf{x}(t)\| \to 0$ as $t \to \infty$?

Contraction Theory

We have

$$\begin{aligned}\dot{V} &= \delta\mathbf{x}^\top \dot{M}(\mathbf{x},t)\delta\mathbf{x} + \delta\dot{\mathbf{x}}^\top M(\mathbf{x},t)\delta\mathbf{x} + \delta\mathbf{x}^\top M(\mathbf{x},t)\delta\dot{\mathbf{x}} \\ &= \delta\mathbf{x}^\top \left(\dot{M}(\mathbf{x},t) + M(\mathbf{x},t)\frac{\partial f}{\partial \mathbf{x}} + \frac{\partial f}{\partial \mathbf{x}}^\top M(\mathbf{x},t)\right)\delta\mathbf{x}.\end{aligned}$$

Inspired by the Lyapunov theory, we design M s.t. $\dot{V} \leq -2\alpha V$, which gives the contraction condition.

Contraction Theory

Theorem (Deterministic Contraction)

If there exists $M(\mathbf{x},t) = \Theta(\mathbf{x},t)^\top \Theta(\mathbf{x},t) \succ 0,\ \forall \mathbf{x}, t$, *where* $\Theta(\mathbf{x},t)$ *defines a smooth coordinate transformation of* $\delta\mathbf{x}$, *i.e.,* $\delta\mathbf{z} = \Theta(\mathbf{x},t)\delta\mathbf{x}$, *s.t. either of the following equivalent conditions holds for* $\exists \alpha > 0,\ \forall \mathbf{x}, t$*:*

$$\lambda_{\max}(\mathrm{sym}(F(\mathbf{x},t))) = \lambda_{\max}\left(\mathrm{sym}\left(\left(\dot{\Theta}(\mathbf{x},t) + \Theta(\mathbf{x},t)\frac{\partial f}{\partial \mathbf{x}}\right)\Theta(\mathbf{x},t)^{-1}\right)\right) \leq -\alpha \tag{7}$$

$$\dot{M}(\mathbf{x},t) + M(\mathbf{x},t)\frac{\partial f}{\partial \mathbf{x}} + \frac{\partial f}{\partial \mathbf{x}}^\top M(\mathbf{x},t) \preceq -2\alpha M(\mathbf{x},t) \tag{8}$$

then all the solution trajectories of (3) *converge to a single trajectory exponentially fast regardless of their initial conditions (i.e., contracting), with an exponential convergence rate* α. ***<u>The converse also holds</u>***.

Contraction Theory

Proof: The proof of this theorem can be found in [1, 2], but here we emphasize the use of the comparison lemma. Taking the time-derivative of $V = \delta\mathbf{z}^\top\delta\mathbf{z} = \delta\mathbf{x}^\top M(\mathbf{x},t)\delta\mathbf{x}$ along the differential dynamics (2), we have

$$\dot{V} = 2\delta\mathbf{z}^\top F\delta\mathbf{z} = \delta\mathbf{x}^\top\left(\dot{M} + \frac{\partial f}{\partial\mathbf{x}}^\top M + M\frac{\partial f}{\partial\mathbf{x}}\right)\delta\mathbf{x} \leq -2\alpha\delta\mathbf{z}^\top\delta\mathbf{z} = -2\alpha\delta\mathbf{x}^\top M(\mathbf{x},t)\delta\mathbf{x}$$

where the generalized Jacobian F in (7) obtained from $\delta\dot{\mathbf{z}} = \dot{\Theta}\delta\mathbf{x} + \Theta\delta\dot{\mathbf{x}} = F\delta\mathbf{z}$. Then we get

$$d\|\delta\mathbf{z}\|/dt \leq -\alpha\|\delta\mathbf{z}\|$$

which then yields

$$\|\delta\mathbf{z}(t)\| \leq \|\delta\mathbf{z}_0\|e^{-\alpha t}.$$

Contraction Theory

Hence, any infinitesimal length $\|\delta \mathbf{z}(t)\|$ and $\|\delta \mathbf{x}(t)\|$, as well as $\delta \mathbf{z}$ and $\delta \mathbf{x}$, tend to zero exponentially fast. By path integration (6), this implies

$$\|\mathbf{x}_1(t) - \mathbf{x}_0(t)\| = \|\mathbf{q}(1,t) - \mathbf{q}(0,t)\| = \left\| \int_0^1 \frac{\partial \mathbf{q}}{\partial \mu} d\mu \right\| = \left\| \int_{\mathbf{x}_0}^{\mathbf{x}_1} \delta \mathbf{x} \right\| \leq \int_{\mathbf{x}_0}^{\mathbf{x}_1} \|\delta \mathbf{x}\|$$

and

$$\|\mathbf{x}_1(t) - \mathbf{x}_0(t)\| \leq \int_{\mathbf{x}_0}^{\mathbf{x}_1} \|\delta \mathbf{x}\| \leq \int_{\mathbf{x}_0}^{\mathbf{x}_1} \frac{1}{\sqrt{\underline{m}}} \|\delta \mathbf{z}_0\| e^{-\alpha t}$$

for any two solution trajectories of $\mathbf{x}_0$ and $\mathbf{x}_1$ of (3), where $\underline{m}\mathbb{I} \preceq M$.

Contraction Theory

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

$$\|\delta \mathbf{x}(t)\|^2 \leq -k\|\delta \mathbf{x}(0)\|^2 e^{-2\beta t} \tag{9}$$

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

$$\dot{\Xi} = -2\beta\Xi - \Xi\frac{\partial f}{\partial \mathbf{x}} - \frac{\partial f}{\partial \mathbf{x}}^\top \Xi, \;\; \Xi(\mathbf{x}(0), 0) = k\mathbb{I}. \tag{10}$$

Note that, for $V = \delta \mathbf{x}^\top \Xi \delta \mathbf{x}$, (10) gives $\dot{V} = -2\beta V$, resulting in $V = -k\|\delta \mathbf{x}(0)\|^2 e^{-2\beta t}$. Substituting this into (9) yields

$$\|\delta \mathbf{x}(t)\|^2 = \delta \mathbf{x}(t)^\top \delta \mathbf{x}(t) \leq V = \delta \mathbf{x}(t)^\top \Xi(x(t), t)\delta \mathbf{x}(t) \tag{11}$$

which indeed implies that $\Xi \succeq \mathbb{I} \succ 0$ as (11) holds for any $\delta \mathbf{x}$. Thus, Ξ satisfies the contraction condition (8), i.e., Ξ defines a contraction metric [1]. $\square$

Contraction Theory

- The matrix function M defines how we measure the distance between two solution trajectories of (3), exponentially decreasing in time.
- This metric is called the contraction metric.
- If you see one of the solution trajectories being your "ideal" trajectory, the contraction condition provides a way to measure the distance of the system's current performance to the ideal in contracting manner.
- "ideal" here could mean various performance: stability, robustness, safety, optimality, etc.
- In our context, this is stability/consistency and robustness.

ILLINOIS

Table of contents

Example: Stable but not contracting

One of the distinct features of contraction theory in Theorem 1 is incremental stability with exponential convergence. Consider the example given in [5]:

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 & x_1 \\ -x_1 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}. \tag{12}$$

A Lyapunov function $V = \|\mathbf{x}\|^2/2$ for (12), where $\mathbf{x} = [x_1, x_2]^\top$, yields $\dot{V} \leq -2V$. Thus, (12) is exponentially stable with respect to $\mathbf{x} = 0$. The differential dynamics of (12) is given as

$$\frac{d}{dt}\begin{bmatrix} \delta x_1 \\ \delta x_2 \end{bmatrix} = \begin{bmatrix} -1 + x_2 & x_1 \\ -2x_1 & -1 \end{bmatrix}\begin{bmatrix} \delta x_1 \\ \delta x_2 \end{bmatrix}. \tag{13}$$

and the contraction condition (8) for (13) can no longer be proven by $V = \|\delta \mathbf{x}\|^2/2$, due to the lack of the skew-symmetric property of (12) in (13). More details can be found in [6].

Example: Contracting but not stable

Contraction defined by Theorem 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 $x : \mathbb{R}_{\geq 0} \mapsto \mathbb{R}$:

$$\dot{x} = -x + e^t. \tag{14}$$

Using the constraint (8) of Theorem 1, we can easily verify that (14) is contracting as $M = \mathbb{I}$ defines its contraction metric with the contraction rate $\alpha = 1$. However, since (14) 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: Connection to linear systems

Consider a Linear Time-Invariant (LTI) system

$$\dot{\mathbf{x}} = f(\mathbf{x}) = A\mathbf{x}.$$

As we saw earlier, the 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. [7, pp. 67-68]

$$\exists \epsilon > 0 \text{ s.t. } PA + A^\top P \preceq -\epsilon \mathbb{I}. \tag{15}$$

Now, let $\overline{p} = \|P\|$. Since $-\mathbb{I} \preceq -P/\overline{p}$, (15) implies that

$$PA + A^\top P \leq -\frac{\epsilon}{\overline{p}} P$$

which shows that $M = P$ with $\alpha = \epsilon/(2\overline{p})$ satisfies (8) due to the relation $\partial f/\partial \mathbf{x} = A$.

Example: Connection to linear systems

- The contraction condition (8) can thus be viewed also as a generalization of the Lyapunov stability condition (15) in a nonlinear non-autonomous system setting.

Furthermore, if $f(\mathbf{x}, t) = A(t)\mathbf{x}$ and $M = \mathbb{I}$, (8) results in

$$A(t) + A(t)^\top \preceq -2\alpha\mathbb{I}$$

(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 [8, pp. 114-115].

Table of contents

Example: Generalized gradient descent beyond convexity

Contraction theory provides a generalized view on the analysis of continuous-time gradient-based optimization algorithms (gradient flow) [9].

Example: Generalized gradient descent beyond convexity

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 $M(\mathbf{x}) \succ 0,\ \forall \mathbf{x} \in \mathbb{R}^n$, and the following natural gradient system [10]:

$$\dot{\mathbf{x}} = h(\mathbf{x}, t) = -M(\mathbf{x})^{-1} \nabla_{\mathbf{x}} f(\mathbf{x}, t). \tag{16}$$

Then, we have *(a)* $\Leftrightarrow$ *(b)*, where

(a) f *is geodesically* α*-strongly convex for each* t *in the metric defined by* $M(\mathbf{x})$

(i.e., $H(\mathbf{x}) \succeq \alpha M(\mathbf{x})$ with $H(\mathbf{x})$ being the Riemannian Hessian matrix of f with respect to M [11]), and

(b) (16) *is contracting with rate* α *in the metric defined by* M *as in* (8)

with $A = \partial h / \partial \mathbf{x}$. More specifically, the Riemannian Hessian verifies $H(\mathbf{x}) = -(\dot{M} + MA + A^\top M)/2$. See [9] for details.

Convexity and Optimization

Definition (Convex Sets)

A set $C \subseteq \mathbb{R}^n$ is convex if, for any $x, y \in C$ and $\theta \in [0, 1]$, we have

$$\theta x + (1 - \theta) y \in C$$

i.e., the line segment joining x, y lies entirely in C.

Definition (Convex Functions)

Let C be a convex set and A function $f : C \mapsto \mathbb{R}$ is convex if, $x, y \in C$ and $\theta \in [0, 1]$, we have

$$f(\theta x + (1 - \theta) y) \leq \theta f(x) + (1 - \theta) f(y).$$

Convexity and Optimization

Consider the following optimization problem:

$$\min_{x} f_0(x) \text{ s.t. } f_i(x) \leq 0,\ i = 1, \cdots, k \text{ and } g_j(x) = 0,\ j = 1, \cdots, m. \tag{17}$$

Definition

The problem (17) is convex if the function f_i are convex and h_j are affine transformations.

Convexity and Optimization

Theorem (KKT Conditions)

Suppose that the problem (17) is convex and the functions f_i and g_j are differentiable. A solution x^* of the problem (17) is optimal if and only if the Karush-Kuhn-Tucker (KKT) conditions (see [12]) hold.

Karush-Kuhn-Tucker (KKT) Conditions

Consider the nonlinear optimization problem:

$$\min_{\mathbf{x}} f(\mathbf{x}) \text{ s.t. } g_i(\mathbf{x}) \leq 0,\ i = 1, \cdots, m \ \&\ h_j(\mathbf{x}) = 0,\ j = 1, \cdots, p$$

KKT Conditions:

- Stationarity: $\nabla f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\mathbf{x}) + \sum_{j=1}^{p} \mu_j \nabla h_j(\mathbf{x}) = 0$
- Primal feasibility: $g_i(\mathbf{x}) \leq 0, \quad h_j(\mathbf{x}) = 0$
- Dual feasibility: $\lambda_i \geq 0, \quad i = 1, \ldots, m$
- Complementary slackness: $\lambda_i g_i(\mathbf{x}) = 0, \quad i = 1, \ldots, m.$

Reference

[1] W. Lohmiller and J.-J. E. Slotine. "On contraction analysis for nonlinear systems". In: *Automatica* 34.6 (1998), pp. 683–696. ISSN: 0005-1098. DOI: 10.1016/S0005-1098(98)00019-3.

[2] H. Tsukamoto, S.-J. Chung, and J.-J. E. Slotine. "Contraction theory for nonlinear stability analysis and learning-based control: A tutorial overview". In: *Annual Reviews in Control* 52 (2021), pp. 135–169. ISSN: 1367-5788. DOI: 10.1016/S0005-1098(98)00019-3.

[3] V. I. Arnold. *Mathematical Methods of Classical Mechanics*. 2nd. Vol. 60. Graduate Texts in Mathematics. New York: Springer-Verlag, 1989. ISBN: 978-0-387-96890-2. DOI: 10.1007/978-1-4757-2063-1.

[4] D. E. Kirk. *Optimal Control Theory: An Introduction*. Dover Publications, Apr. 2004. ISBN: 0486434842.

Reference

[5] J. Jouffroy and J.-J. E. Slotine. "Methodological remarks on contraction theory". In: *IEEE Conference on Decision and Control*. Vol. 3. Dec. 2004, pp. 2537–2543. DOI: 10.1109/CDC.2004.1428824.

[6] B. S. Rüffer, N. van de Wouw, and M. Mueller. "Convergent systems vs. incremental stability". In: *Systems & Control Letters* 62.3 (2013), pp. 277–285. ISSN: 0167-6911. DOI: 10.1016/j.sysconle.2012.11.015.

[7] W. J. Rugh. *Linear Systems Theory*. USA: Prentice-Hall, Inc., 1996. ISBN: 0134412052.

[8] J.-J. E. Slotine and W. Li. *Applied Nonlinear Control*. Upper Saddle River, NJ: Pearson, 1991.

[9] P. M. Wensing and J.-J. E. Slotine. "Beyond convexity – Contraction and global convergence of gradient descent". In: *PLOS ONE* 15 (Aug. 2020), pp. 1–29. DOI: 10.1371/journal.pone.0243330.

Reference

[10] S.-I. Amari. "Natural gradient works efficiently in learning". In: *Neural Computation* 10.2 (1998), pp. 251–276. DOI: 10.1162/089976698300017746.

[11] C. Udriste. *Convex functions and optimization methods on Riemannian manifolds*. Vol. 297. Springer Science & Business Media, 1994.

[12] S. P. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge, UK ; New York: Cambridge University Press, 2004. 716 pp. ISBN: 978-0-521-83378-3.