Learning Stable Dynamical Systems using Contraction Theory

eingereichte
MASTERARBEIT
von

cand. ing. Caroline Blocher

geb. am 19.08.1990
wohnhaft in:
Leonrodstrasse 72
80636 München
Tel.: 0176 27250499

Lehrstuhl für
STEUERUNGS- und REGELUNGSTECHNIK
Technische Universität München

Univ.-Prof. Dr.-Ing./Univ. Tokio Martin Buss

Fachgebiet für
DYNAMISCHE MENSCH-ROBOTER-INTERAKTION
für AUTOMATISIERUNGSTECHNIK
Technische Universität München

Prof. Dongheui Lee, Ph.D.

Betreuer: M.Sc. Matteo Saveriano
Beginn: 03.08.2015
Zwischenbericht: 25.09.2015
Abgabe: 19.01.2016

In your final hardback copy, replace this page with the signed exercise sheet.

Abstract

This report discusses the learning of robot motion via non-linear dynamical systems and Gaussian Mixture Models while optimizing the trade-off between global stability and accurate reproduction. Contrary to related work, the approach used in this thesis seeks to guarantee the stability via Contraction Theory. This point of view allows the use of results in robust control theory and switched linear systems for the analysis of the global stability of the dynamical system. Furthermore, a modification of existing approaches to learn a globally stable system and an approach to locally stabilize an already learned system are proposed. Both approaches are based on Contraction Theory and are compared to existing methods.

Zusammenfassung

Diese Arbeit behandelt das Lernen von stabilen dynamischen Systemen über eine Gauss'sche Mischverteilung. Im Gegensatz zu bisherigen Arbeiten wird die Stabilität des Systems mit Hilfe der Contraction Theory untersucht. Ergebnisse aus der robusten Regulung und der Stabilität von schaltenden Systemen können so übernommen werden. Um die Stabilität des dynamischen Systems zu garantieren und gleichzeitig die Bewegung des gelernten Systems möglichst wenig zu beeinflussen, wird eine Anpassung der bereits bestehenden Methode an die Bewegung vorgeschlagen. Darüber hinaus wird ein Regler, der lokale Stabilität in der Nähe des Ruhepunkts garantiert, entworfen. Beide Ansätze basieren auf Contraction Theory und werden mit bereits bestehenden Methoden verglichen.

Contents

Chapter 1

Introduction

Programming by Demonstration provides a useful tool to teach a robot i.e. a point-to-point motion. Instead of having to explicitly program the robot, the user has to provide several demonstrations of the task. To be of use in a dynamic environment, the learned trajectory should then reproduce accurately the motion while being robust to perturbations and adaptable to changes in the environment, i.e. a moved target or obstacles. Representing a motion via a dynamical system $\dot{x} = f(x)$ yields an interesting approach to face these challenges [JIS02, SKS00, SL13b, PHAS09].
To encode the main features of the demonstrations, different probabilistic approaches such as Gaussian Mixture Models, Hidden Markov Models, Gaussian Processes and Support-Vector-Machines may be used to estimate a dynamical system of the form $\dot{x} = \hat{f}(x)$ [CDS+10, KKB15, RW06, SS04].
At the latest, if the dynamics of the system are non-linear, stability of the system becomes an important issue to guarantee the correct convergence of the motion, as spurious attractors might cause problems. To guarantee that i.e. a robot arm guided by the estimated dynamical system actually reaches the target of the motion, it is therefore necessary to consider the stability properties of $\hat{f}$ (see section 1.1).

A good reproduction of the motion via a dynamical system requires hence that the estimated system $\hat{f}$ is stable, while being as close to the demonstrations as possible. This Master's thesis addresses the problem of finding a good trade-off between accuracy and stability when using a Gaussian Mixture Model to encode the motion.

Previous approaches considered the stability problem with respect to Lyapunov Theory. Throughout this report, we investigate on the stability when considering Contraction Theory instead. In chapter 5, we show the equivalence of the sufficient condition for global asymptotic stability in [KZB11a] with respect to Lyapunov and Contraction Theory. Results from robust control and switched linear systems are used to modify the approach in [KZB11a]. Additionally, we develop a simple control law that guarantees local stability of the target without influencing the overall motion. Simulation results of the proposed approaches for a set of handwriting motions are given and compared to existing methods.

1.1 Related Work

Learning stable systems

In [KZB10], a method is proposed that guarantees asymptotic convergence of trajectories that remain within the demonstration area to the target. Gaussian Mixture Regression is used. In [KZB11a], Khansari-Zadeh and Billard propose the Stable Estimator of Dynamical Systems (SEDS) that guarantees global asymptotic stability of an arbitray dynamical system of the form

$$\dot{x} = \sum_{k=1}^{K} h^k(x)(A^k x + b^k)$$

where $x \in \mathbb{R}^n$, $A^k \in \mathbb{R}^{n\times n}$, $b^k \in \mathbb{R}^n$ and $h^k \in [0,1]$ such that $\sum_{k=1}^{K} h^k = 1$. The differential equation represents i.e. a robot motion. The parameters of the system result from a Gaussian Mixture Model learned via a set of demonstrations of the motion. To ensure the stability of the system, the authors formulate an optimization problem that on the one hand maximizes the likelihood of the demonstrations given the parameters of the learned model and that on the other hand searches to satisfy the following stability constraints:

- Ensuring $b^k = -A^k x^*$ guarantees that there exists an equilibrium point x^*.
- Ensuring that all A^k are negative definite corresponds to guaranteeing the stability in the sense of Lyapunov.

The approach is based on the use of a quadratic Lyapunov-function of the form $V(x) = \frac{1}{2}(x - x^*)^T (x - x^*)$ and provides inherent global stability of the motion.
In some cases, the form of the Lyapunov-function can limit the accuracy of the reproduction of the demonstrated data. Neumann et al. address this problem in [NLS13]. They propose a neural learning approach to (i) learn a Lyapunov Candidate that is well suited for the demonstrated motion and (ii) use this candidate to learn the stable motion.
In [KZB14], Khansari-Zadeh and Billard propose to learn a suitable Lyapunov-Candidate via a parametrization as weighted sum of asymmetric quadratic functions. The approach is called SEDS II. A stabilizing command is defined by the means of this function to guarantee global convergence to the equilibrium. This command forces the system to follow the negative gradient of the Lyapunov-function.

Finding a suitable contraction metric

To prove that a dynamical system is stable in the sense of contraction it is necessary to prove the existence of a contraction metric. The theory is presented in detail in chapter 2. Finding such a metric is a difficult problem that requires experience. However, in [RdBS11], an algorithm is proposed that can provide sufficient conditions for the existence of a such a metric. In [APS08], a numerical approach originally used to find a Lyapunov-function is proposed in the context of finding a contraction metric.

As results from robust control and switched linear systems may be applied to the stability problem addressed in this report, the remainder of this section presents related work in these fields.

Consider systems of the form:

$$\dot{x} = Ax, \ \ A \in \mathbf{A}, \ \ \mathbf{A} \subset \mathbb{R}^{n \times n}$$

The set $\mathbf{A}$ is the convex hull of matrices

$$\mathbf{A} = \sum_{k=1}^{K} \gamma^k A^k$$

with $\sum_{k=1}^{K} \gamma^k = 1$ and $\gamma^k > 0$ and the vertex matrics $\{A^1 ... A^K\}$.

In [Zah03], Zahreddine analyses the stability of a convex hull of matrices in the context of interval dynamical systems, hence linear systems where the uncertainties in the state matrix can be modeled by intervals. The convex hull $\mathbf{A}$ is stable if each matrix that is an element of $\mathbf{A}$ has only eigenvalues with negative real parts. Zahreddine shows that a sufficient condition for the stability of the convex hull is the existence of a matrix measure μ, such that $\mu(A^k) < 0$ for any $k = 1...K$.

Pastravanu and Matcovschi consider linear systems with parametric uncertainty [PM09]. The systems are of the form $\dot{x} = Ax$ where A is fixed but taken from the convex hull of matrices $\mathbf{A}$.
Given a vector norm $||...||$ and a corresponding matrix measure $\mu_{||...||}$, one of their main results is the equivalence of the two following statements:

- The vertices of $\mathbf{A}$ satisfy: $\mu_{||...||}(A^k) < 0 \ \ \forall k = 1...K$
- The function $V(x) = ||x||$ is a common Lyapunov-function for the uncertain system.

This result provides the direct link from the stability of a convex hull of matrices to common Lyapunov-functions.

Note that the authors of both articles make the assumption that the uncertainties of the linear dynamical system are time-invariant. Once A is chosen from $\mathbf{A}$, it remains constant.

Switched linear systems are systems of the form

$$\dot{x}(t) = A_i x(t),\, t \in \mathbb{R}^+,\, i \in I$$

where I is an index set for the different discrete switching modes of the system. Given an arbitrary switching signal, it is possible to obtain a diverging system although, independently, each linear system is exponentially stable. An interesting problem is hence to guarantee the stability of the switched system under arbitrary switching signals. In [LA09], a survey on the stability problem of switched linear systems is given.

A sufficient condition for stability is the existence of a common quadratic Lyapunov function for the set of linear systems [LA09]. In [DM99], Dayawansa and Martin prove the converse theorem: In the case of a linear switched system that is globally uniformly exponentially stable and if the set of state matrices is compact, a common Lyapunov-function for the system always exists. However, it is possible that a Lyapunov-function that is also quadratic cannot be found.
The search for a common quadratic Lyapunov-function can be done numerically as the conditions for its existence can be formulated as Linear Matrix Inequalities [LT04]. To find algebraic conditions for the existence of a common Lyapunov-function is a difficult problem. At the moment, necessary conditions for a switched system of more than 2 modes seem to only have been derived for a second order linear time-invariant switched systems [SN00].

In [RLdB11], Russo et al. investigate on the contraction of a class of switched systems.

Chapter 2

Mathematical Preliminaries

Throughout the report, we will make use of the concept of the measure of a matrix. This chapter gives an introduction to this topic and provides the link to the negative definiteness of a matrix and its use in the analysis of differential equations.

2.1 Definiteness and measure of a matrix

Hurwitz-Stability of a matrix

A matrix $A \in \mathbb{R}^{n \times n}$ is called **Hurwitz** if the real parts of all eigenvalues are strictly negative [NS10].

Definiteness of a matrix

A *symmetric* matrix $A \in \mathbb{R}^{n \times n}$ is called **negative definite** if and only if one of the following equivalent statements is true [KK00, KZB11b]:

- $x^T Ax < 0$ holds for any non-zero vector $x \in \mathbb{R}^n$.
- The eigenvalues of A are strictly negative.
- Sylvester's Criterion is satisfied.

The matrix is positive definite if $x^T Ax > 0$. If it holds that $x^T Ax \leq 0$ or $x^T Ax \geq 0$ the matrix is said to be negative (positive) semi-definite.

Sylvester's Criterion states that a symmetric matrix in $\mathbb{R}^{n \times n}$ is negative definite if the determinant of the ith leading principal minors is negative if $i = 1..n$ is odd and positive if i is even. For positive definiteness the ith leading principal minors need to be positive for all $i = 1..n$. The leading principal minors are the n quadratic upper left parts of the matrix A.

The criteria used to show definiteness of a matrix are defined for symmetric matrices. If the matrix is non-symmetric, it is negative (positive) definite if the **symmetric part of the matrix**

$$A_{sym} = \frac{1}{2}\left(A + A^T\right) \tag{2.1}$$

is negative (positive) definite.

Matrix Measure

Given is a vector norm on $\mathbb{C}^n$ and the induced matrix norm $||...||_i$ corresponding to the vector norm.

Definition 1 *[Vid93, p.22] Let $||...||_i$ be an induced matrix norm on $\mathbb{C}^{n\times n}$. Then the corresponding* ***matrix measure*** *is the function $\mu(...) : \mathbb{C}^{n\times n} \to \mathbb{R}$ defined by*

$$\mu(A) = \lim_{\epsilon\to 0^+} \frac{||I + \epsilon A||_i - 1}{\epsilon} \tag{2.2}$$

In the following, the matrix measures associated to the Euclidean norm, the l_1 -norm and the l_∞ -norm are given:

$$\mu_2(A) = \max_i \left(\lambda_i \left\{\frac{A + A^T}{2}\right\}\right) \tag{2.3}$$

$$\mu_1(A) = \max_j \left(a_{jj} + \sum_{i\neq j} |a_{ij}|\right) \tag{2.4}$$

$$\mu_\infty(A) = \max_i \left(a_{ii} + \sum_{j\neq i} |a_{ij}|\right) \tag{2.5}$$

Note that the statements $\mu_2(A) < 0$ and A *is negative definite* are hence equivalent.

Theorem 1 *[Des09, p.31] Given is an induced matrix norm $||...||_i$ and the corresponding matrix measure $\mu(.)$. Useful properties of $\mu(.)$ are:*

1. $-||A||_i \leq \mu(A) \leq ||A||_i, \forall A \in \mathbb{C}^{n\times n}$
2. $\mu(\alpha A) = \alpha\, \mu(A), \forall A \in \mathbb{C}^{n\times n}, \alpha \geq 0$
3. $\mu(A + B) \leq \mu(A) + \mu(B), \forall A, B \in \mathbb{C}^{n\times n}$
4. *If λ is an eigenvalue of $A \in \mathbb{C}^{n\times n}$, then $-\mu(-A) \leq \mathrm{Re}\{\lambda\} \leq \mu(A)$*
5. *For any nonsingular matrix N and any vector norm $||...||$, with the induced matrix measure μ, $||Nx||$ defines another vector norm and its induced matrix measure μ_N is given by $\mu_N(A) = \mu(NAN^{-1})$.*

Using a matrix measure instead of a matrix norm has the advantage, that it can obtain negative values and it is sign-sensitive.

Given a set of matrices $S = \{\mu(A^1), \mu(A^2), ..., \mu(A^K)\}$

$$\mu(S) = \max\{\mu(A^1), \mu(A^2), ..., \mu(A^K)\} \tag{2.6}$$

where μ is a matrix measure associated to a vector norm in Euclidean space and common for all A^k in S, $\mu(S)$ shall be called the induced **matrix set measure** [XS11].
According to property 4 of Theorem 1, the eigenvalues of a matrix provide a lower bound for its measure. Theorem 2 states that a measure that achieves the lower bound can always be found:

Theorem 2 *[Zah03] Let $\mathcal{N}$ be the set of all vector norms on $\mathbb{C}^n$. For any $\rho \in \mathcal{N}$, the corresponding matrix measure is denoted by μ_ρ. Then, for any matrix in $A \in \mathbb{C}^{n\times n}$, we have*

$$\max_{1\leq j\leq n} \mathrm{Re}\,\{\lambda_j(A)\} = \inf_{\rho\in\mathcal{N}} \mu_\rho(A) \tag{2.7}$$

The proof of this theorem given in [Zah03] is based on the following simplified idea: If J is the Jordan form of A, then there is an invertible matrix N such that $J = NAN^{-1}$. If A is diagonalizable, J is a matrix where the only non-zero entries are its eigenvalues on the diagonal. As J is a symmetric matrix, $\mu_2(J) = \mu_2^N(A)$ yields exactly the maximal eigenvalue of A and achieves hence the lower bound for all matrix measures of A.

By definition, the measure $\mu_2(.)$ achieves the lower bound for any symmetric matrix. In fact, this holds for the entire class of normal matrices [SÖ6]. The measures $\mu_1(.)$ and $\mu_\infty(.)$ may yield sharper bounds than $\mu_2(.)$ for a given matrix F as the following example illustrates:

$$F = \begin{bmatrix} -1 & 0.1 \\ 10 & -12 \end{bmatrix}$$

The matrix is Hurwitz-stable, but not negative definite as $\mu_2(F) \approx 1.93$. The matrix measure associated to l_∞-norm yields $\mu_\infty = -0.9$.
Note that in order to achieve the lower bound of μ_1 or μ_∞ one must search for a transformation that yields a diagonally dominant matrix with small or even negative diagonal elements. Optimally this results in a diagonalization, similar to the case of μ_2.

Differential Equations Solution Estimates

The concept of matrix measures can provide a bound for the solution of a differential equation.

Given is a differential equation $\dot{x} = A(t)x(t)$, $t \geq 0$ where $x \in \mathbb{R}^n$ and $A(t)$ is continuous. With $\mu(...)$ being the measure associated to a vector norm $||...||$ on Euclidean space, we have that [Vid93, p.47]:

$$||x(t)|| \leq ||x(t_0)||e^{\int_0^t \mu(A(\tau))\mathrm{d}\tau} \tag{2.8}$$

Chapter 3

Stability of Dynamical Systems

This chapter provides an introduction to the Theory of Lyapunov and Contraction Theory. Additionally, a comparison is given.
Throughout this chapter, consider the following differential equation:

$$\dot{x} = f(x, t) \tag{3.1}$$

where $x \in \mathbb{R}^n$ and f is a continuous, non-linear vector function.

Definition 2 *[Vid93, p.3] A vector $x^* \in \mathbb{R}$ is called an **equilibrium** of the system in 3.1 if*

$$f(x^*, t) = 0 \;\; \forall t > 0 \tag{3.2}$$

3.1 Lyapunov Theory

In this section we will assume without loss of generality that the equilibrium considered is at the origin, hence $x^* = \mathbf{0}$.

Stability in the sense of Lyapunov characterizes the equilibrium by the behavior of the dynamical system close to it. If, for instance, any trajectory of the system that starts within a ball around the equilibrium, will never leave this ball, the equilibrium is said to be stable.

Definition 3 *[Vid93, p.136] The equilibrium x^* is **stable** if for each $\epsilon > 0$ and each $t_0 \in \mathbb{R}_+$ there exists a $\delta(\epsilon, t_0)$ such that*

$$||x(t_0)|| < \delta(\epsilon, t_0) \Rightarrow ||x(t)|| < \epsilon \;\; \forall t \geq t_0 \tag{3.3}$$

*The equilibrium is said to be **uniformly stable** if for each $\epsilon > 0$ there exists a $\delta(\epsilon)$ such that*

$$||x(t_0)|| < \delta(\epsilon), t_0 \geq 0 \Rightarrow ||x(t)|| < \epsilon \;\; \forall t \geq t_0 \tag{3.4}$$

$x(t)$ is the solution of 3.1 corresponding to the initial condition $x(t_0)$.

As we will discuss time-invariant systems throughout this report, the following notions are only given in the uniform version. If the equilibrium is not only stable, but the trajectories in a close area converge to it, the equilibrium is called asymptotically stable. To classify the rate of convergence the notion of exponential stability is introduced. Exponential stability implies asymptotic stability and is a stronger property.

If the system 3.1 only has a single equilibrium, it may be possible that the stability properties do not only hold locally in an area close to x^*, but also globally, for any initial condition in $\mathbb{R}^n$.

Definition 4 *[Vid93, p.143] The equilibrium x^* is* ***globally uniformly asymptotically stable*** *if it is uniformly stable and for each pair of positive numbers M, ϵ with M arbitrarily large and ϵ arbitrarily small there exists a finite number $T(M, \epsilon)$ such that*

$$||x(t_0)|| < M, t_0 \geq 0 \Rightarrow ||x(t)|| < \epsilon, \ \ \forall t, t_0 \geq 0 \tag{3.5}$$

The equilibrium x^* is ***globally exponentially stable*** *if there exist constants $a, b > 0$ such that*

$$||x(t_0 + t)|| \leq a||x(t_0)||exp(-bt) \ \ \forall t, t_0 \geq 0 \tag{3.6}$$

Lyapunov's Direct Method

In order to characterize the stability of an equilibrium, Lyapunov's direct method can be used. Only the method to verify global asymptotic stability for a time-invariant system $\dot{x} = f(x)$ shall be presented here.

Definition 5 *A continuous function V is said to be radially unbounded if $V(x^*) = 0$ and there exists a continuous function α with $\alpha(r) \to \infty$ as $r \to \infty$ such that*

$$\alpha(||x||) \leq V(x) \ \ \forall t \geq 0, \ \ \forall x \in \mathbb{R}^n \tag{3.7}$$

see [Vid93, p.148]

Theorem 3 *The equilibrium x^* of $\dot{x} = f(x)$ is globally (uniformly) asymptotically stable if there exists a continuous, continuously differentiable and radially unbounded Lyapunov-function V such that*

1. $V(x) > 0 \ \ \forall x \in \mathbb{R}^n, \ \ \forall x \neq x^*$
2. $\dot{V}(x) < 0 \ \ \forall x \in \mathbb{R}^n, \ \ \forall x \neq x^*$
3. $V(x^*) = 0, \ \ \dot{V}(x^*) = 0$

see [Vid93, p.173] and [KZB11a]

V can be considered as a function that provides a bound for the energy of the system. If the energy of the system is continuously decreasing following any trajectory towards the equilibrium, the equilibrium is stable [Vid93, p.157].

Finding such a Lyapunov-function is not easy. Typical choices are

- $V(x) = \frac{1}{2}x^T x$
- $V(x) = x^T Px$ where P is a positive definite matrix

Such Lyapunov-functions are called *quadratic*.

In the case of a *linear* time-invariant system $\dot{x} = Ax$, global exponential stability of the equilibrium holds if one of the following statements is true:

Theorem 4 *[Vid93, p. 199] Given a matrix $A \in \mathbb{R}^{n\times n}$, the following three statements are equivalent:*

1. *A is a Hurwitz-matrix.*
2. *There exists some positive definite matrix $Q \in \mathbb{R}^{n\times n}$ such that the* ***Lyapunov Matrix Equation***
$$A^T P + PA = -Q \tag{3.8}$$
has a corresponding unique solution for P and this P is positive definite and real symmetric.
3. *For every positive definite matrix $Q \in \mathbb{R}^{n\times n}$, 3.8 has a unique solution for P and this solution is positive definite.*

The corresponding Lyapunov-function is then $V(x) = x^T Px$.

3.2 Contraction Theory

Contraction Theory proposed by Slotine et al. in [LS98] provides a different approach to the analysis of the behavior of non-linear dynamical systems. It is based on the idea that if moving along a trajectory of the contracting system, the (virtual) distance to its neighboring trajectories decreases.

Considering the system 3.1, Slotine et al. define a virtual displacement δx, which is an infinitesimal displacement at fixed time between two trajectories and derive the following relation:

$$\delta\dot{x} = \frac{\partial f(x,t)}{\partial x}\,\delta x \tag{3.9}$$

Defining the distance between two neighboring system trajectories as $||\delta x|| = \delta x^T \delta x$, its derivative is then bounded by

$$\frac{d}{dt}(\delta x^T \delta x) = 2\ \delta x^T \delta \dot{x} = 2\ \delta x^T \frac{\partial f}{\partial x} \delta x \leq 2\ \lambda_{max}\ \delta x^T \delta x \tag{3.10}$$

where λ_{max} is the largest eigenvalue of the symmetric part of the Jacobian.
If $\lambda_{max}(x,t)$ is uniformly strictly negative, one can then conclude that $||\delta x||$ converges exponentially to zero and hence the neighboring trajectories converge towards each other as

$$||\delta x|| \leq ||\delta x_0||\ e^{\int_0^t \lambda_{max}(x,t)dt} \tag{3.11}$$

Essential for Contraction Theory is the generalization by applying a differential coordinate transformation via a square matrix $\Theta(x,t)$:

$$\delta z = \Theta(x,t)\delta x \tag{3.12}$$

Definition 6 *[LS98, p.8] Given the system equations $\dot{x} = f(x,t)$, a region of the state space is called a* ***contraction region*** *with respect to a uniformly positive definite metric $M(x,t) = \Theta^T\Theta$, if equivalently the Generalized Jacobian*

$$F(x,t) = \left(\dot{\Theta} + \Theta \frac{\partial f(x,t)}{\partial x}\right)\Theta^{-1} \tag{3.13}$$

is uniformly negative definite or

$$\frac{\partial f}{\partial x}^T M + M\frac{\partial f}{\partial x} + \dot{M} \preceq -\beta_M M \tag{3.14}$$

with constant $\beta_M > 0$ in that region.

Theorem 5 *[LS98, p.8] Given the system equations $\dot{x} = f(x,t)$, any trajectory, which starts in a ball of constant radius with respect to the metric M(x,t), centered at a given trajectory and contained at all times in a contraction region with respect to M(x,t), remains in that ball and converges exponentially to this trajectory.*
Furthermore global exponential convergence to the given trajectory is guaranteed if the whole state space is a contraction region with respect to the metric M(x,t).

Negative definiteness of the Generalized Jacobian F is equivalent to $\mu_2(F) < 0$ (see chapter 2). In fact, these results hold for other matrix measures,too.
If an invertible matrix $\Theta(x,t)$ exists such that $M(x,t) = \Theta^T\Theta$ is positive definite and such that the Generalized Jacobian F satisfies

$$\exists c > 0,\ \mu(F(x,t)) \leq -c,\ \forall t \geq t_0 \tag{3.15}$$

where $\mu(.)$ is a matrix measure associated to a vector norm in Euclidean space, the system $\dot{x} = f(x,t)$ is contracting [LS98, RdBS11].

3.2.1 Partial Contraction

Slotine and Wang propose the concept of Partial Contraction as a generalization of Contraction:

Theorem 6 *[WS05] Consider a non-linear system of the form*

$$\dot{x} = f(x, x, t) \tag{3.16}$$

and assume that the auxiliary system

$$\dot{y} = f(y, x, t) \tag{3.17}$$

is contracting with respect to y. If a particular solution of the auxiliary system verifies a smooth specific property, then all trajectories of the original x-system verify this property exponentially. The original system is said to be partially contracting.

Such a smooth specific property could be for instance an equilibrium. If the equilibrium of the auxiliary system is stable in the sense of contraction, this equilibrium is exponentially stable in the original system, too.

3.3 Comparison of Contraction Theory and Lyapunov Theory

Lyapunov Theory characterizes the behavior of the system with respect to its equilibria whereas Contraction Theory does not require the explicit knowledge of an equilibrium. A contraction region in $\mathbb{R}^n$ even might not contain an equilibrium and all trajectories do in this case converge towards one trajectory, i.e. a limit cycle, instead of an equilibrium. In [JF10], Jouffroy and Fossen state that Contraction leads to an incremental form of stability which is a stronger property than stability with respect to the origin.

However, if a system is globally contracting and autonomous, all trajectories converge exponentially towards a unique equilibrium and the corresponding Lyapunov-function can be chosen as $V(x) = f(x)^T M(x,t) f(x)$ [LS98, APS08].

Note that in the linear time-invariant case where the system is defined as $\dot{x} = Ax$, the Lyapunov Matrix Equation 3.8 corresponds exactly to equation 3.14 where $\dot{M}(x,t) = 0$. Hence, it is equivalent to prove stability of the equilibrium via the contraction metric $M = \Theta^T \Theta$ or the quadratic Lyapunov-function $V = x^T M x$ [APS08].
Using the induced matrix measure shows this correspondence in yet again another light: If a matrix measure $\mu_{||...||}$ that is induced by a vector norm $||...||$ on $\mathbb{R}^n$ can be found such that

$$\mu_{||...||}(A) < 0 \tag{3.18}$$

the linear system is stable and a Lyapunov-function can be chosen as $V(x) = ||x||$ [Vid78].

Chapter 4

Dynamical Systems for Reproduction of a motion

Throughout this report, we will only consider motion learning via a Gaussian Mixture Model and reproduction via a dynamical systems approach. Hence, this chapter provides a short introduction to this topic.
Teaching a robot a point-to-point motion via a Gaussian Mixture Model consists of the steps [Cal09, p.41]:

- Produce a demonstration data set of N data points $\mathbf{x}$ with D dimensions
- Encode the data set in a Gaussian Mixture Model (GMM)
- Reproduce the motion through Gaussian Mixture Regression (GMR)

Gaussian Mixture Model

Given is the data set of N data points $\mathbf{x}$ with D dimensions. We will consider the case where a data vector $\mathbf{x}_i$ consists of the position x_i and the current velocity $\dot{x}_i$ at position x_i.

We will first assume a Gaussian Mixture Model with the number of Gaussian functions K and the set $\mathbf{S}$ of all parameters of the K Gaussian functions: mean μ^k, covariance Σ^k and the priors π^k of each Gaussian function.

The probability distribution of position x and velocity $\dot{x}$ can then be described as [KZB11a]:

$$P(x, \dot{x}; \mathbf{S}) = \sum_{k=1}^{K} P(k) P(x, \dot{x}|k) \tag{4.1}$$

where $P(x, \dot{x}|k) = \mathcal{N}(x, \dot{x}; \mu^k, \Sigma^k)$.

To learn the parameters in **S** of the Gaussian Mixture Model, it is common to use the Expectation-Maximization Algorithm [Cal09, p.47-49].
Choosing an appropriate number of Gaussian components K can be done i.e. via the BIC Criterion [Cal09, p.61]:

$$S_{BIC} = -\mathcal{L} + \frac{n_p}{2} log(N)$$

where $\mathcal{L} = \sum_{i=1}^{N} log(P(\mathbf{x}_i))$ is the log-likelihood of the model, n_p the number of parameters of the GMM and N the number of demonstration data points.

Gaussian Mixture Regression

To derive the dynamical system $\dot{x} = f(x)$, note the following notation for mean and covariance:

$$\mu^k = \begin{pmatrix} \mu_\xi^k \\ \mu_{\dot{\xi}}^k \end{pmatrix} \ \& \ \Sigma^k = \begin{pmatrix} \Sigma_\xi^k & \Sigma_{\dot{\xi}\xi}^k \\ \Sigma_{\xi\dot{\xi}}^k & \Sigma_{\dot{\xi}}^k \end{pmatrix}$$

The posterior mean estimate $P(\dot{x}|x)$ yields a non-linear dynamical system [KZB11a]:

$$\dot{x} = \sum_{k=1}^{K} h^k(x)(A^k x + b^k) \tag{4.2}$$

where

$$A^k = \Sigma_{\dot{x}x}^k (\Sigma_x^k)^{-1} \tag{4.3}$$

$$b^k = \mu_{\dot{x}}^k - A^k \mu_x^k \tag{4.4}$$

$$h^k(\xi) = \frac{P(k)P(x|k)}{\sum_{i=1}^{K} P(i)P(x|i)} \tag{4.5}$$

Equation 4.2 hence extends the motion described by the Gaussian Mixture Model from the demonstration area to the entire state space.

Chapter 5

Global Stability of Gaussian Mixture Regression

This chapter analyzes the stability in the sense of contraction of the non linear dynamical system, obtained by Gaussian Mixture Regression when reproducing a learned skill.

5.1 A sufficient condition

Recall equation 4.2, which describes the non-linear time-invariant dynamical system by:

$$\dot{\xi} = \sum_{k=1}^{K} h^k(x)(A^k x + b^k) \tag{5.1}$$

In [KZB11a], Khansari-Zadeh and Billard have proven that a Lyapunov-function $V(x) = x^T x$ exists and hence asymptotic stability in the sense of Lyapunov is guaranteed if

$$\left(A^k\right)^T + A^k \prec 0 \ \ \forall k = 1..K$$

This can be extended to the use of quadratic Lyapunov-functions of the form $V(x) = x^T P x$ where P is a positive definite, real symmetric matrix. Similar to the proof in [KZB11a], it can be shown that if a matrix P can be found such that

$$\left(A^k\right)^T P + P A^k \prec 0 \ \ \forall k = 1..K \tag{5.2}$$

holds, the equilibrium is asymptotically stable. Note that this result corresponds to the Lyapunov Matrix Equation in equation 3.8.
Hence, V represents a common Lyapunov-function for the set of linear systems $A^k x + b^k$ where the equilibrium x^* is such that $A^k x^* + b^k = 0$ for all $k = 1..K$.

The Jacobian of the system in 5.1 is given as

$$J_{GMR}(x) = \sum_{k=1}^{K} h^k(x)A^k + \frac{\partial h^k(x)}{\partial x}(A^k x + b^k) \tag{5.3}$$

The concept of Partial Contraction, see section 3.2.1, leads to the auxiliary dynamical system:

$$\dot{y} = \sum_{k=1}^{K} h^k(x)(A^k y + b^k) \tag{5.4}$$

where y is the auxiliary state variable. As stated in Theorem 6, if the auxiliary system is globally contracting and has an equilibrium, the trajectories of the original system will also *exponentially* converge to the equilibrium and hence global exponential stability is ensured.

The Jacobian of the auxiliary system is:

$$J_{Aux} = \sum_{k=1}^{K} h^k(x)A^k \tag{5.5}$$

Note that, as $0 \leq h^k(x) \leq 1$ and $\sum_{k=1}^{K} h^k(x) = 1$, the auxiliary Jacobian corresponds to the convex hull of matrices described by the vertex matrices A^k.

Note that Partial Contraction and the existence of a unique, globally stable equilibrium x^* imply Contraction of the system 5.1. The inverse is not necessarily true.

As the auxiliary system provides a simpler expression of the Jacobian, we will analyze the auxiliary system in the following sections.

To analyze the contraction of the auxiliary system, one has to show whether a contraction metric Θ and a matrix measure μ exist that guarantee that the matrix measure of the *Generalized Jacobian* F is uniformly negative:

$$\mu(F) = \mu\left((\dot{\Theta} + \Theta J_{Aux})\Theta^{-1}\right) < 0, \forall t > 0, \forall y \in \mathbb{R}^n$$

To simplify this problem we will restrict ourselves to a contraction metric that is constant, such that the Generalized Jacobian becomes $F = \Theta J_{Aux}\Theta^{-1}$. Additionally, using the matrix measure properties 2, 3 and 5 in Theorem 1, chapter 2, we obtain:

$$\mu(F) = \mu^{\Theta}(J_{Aux}) \leq \sum_{k=1}^{K} h^k(x)\mu^{\Theta}(A^k) \tag{5.6}$$

The results in equations 5.2 and 5.6 can be used equivalently as both ensure the stability of the linear subsystems in equation 5.1 (see section 3.3). Sufficient conditions for global exponential stability of the system in equation 5.1 are given in the following theorem.

Theorem 7 *Given the system in 5.1 with a unique equilibrium* x^* *such that*

$$A^k x^* + b^k = 0 \tag{5.7}$$

and considering the set of vertex matrices

$$S = \{A^1, A^2, ...A^K\},\ A^k \in \mathbb{R}^{n\times n} \tag{5.8}$$

the system's trajectories globally uniformly exponentially converge to x^* *if there is an invertible matrix* Θ *such that equivalently the Lyapunov-function* $V(x) = (x - x^*)^T P(x - x^*)$ *with a matrix* $P = \Theta^T\Theta$ *satisfies*

$$\left(A^k\right)^T P + PA^k \prec 0\ \ \forall k = 1..K \tag{5.9}$$

or the matrix set measure $\mu^\Theta(S)$ *associated to the vector norm* $||.||^\Theta$ *satisfies*

$$\mu^\Theta(S) = \max\{\mu^\Theta(A^1), \mu^\Theta(A^2), ..., \mu^\Theta(A^K)\} < 0 \tag{5.10}$$

This result also corresponds to the statement in example 2.2 in [WS05, p.4].

5.2 Discussion

The condition in equation 5.10 requires Hurwitz-stability of all vertex matrices:

Lemma 1 *Let* λ_j^k *be the* j*th eigenvalue of* $A^k \in \mathbb{R}^{n\times n}$*. If for any* $k = 1...K$ *and any* $j = 1..n$ *there is*

$$\mathrm{Re}\,\{\lambda_j^k\} \geq 0 \tag{5.11}$$

a common matrix measure as in equation 5.10 or a common Lyapunov-function as in equation 5.9 does not exist.

Proof: see matrix measure property 4 in Theorem 1, chapter 2.1

Note that Hurwitz-stability of all modes is a necessary condition for the stability under arbitrary switching in the case of linear switched systems, too [LA09].

The proof in [KZB11a] is sufficient to show global asymptotic stability of the equilibrium. The equivalence of Contraction Theory and Lyapunov Theory in Theorem 7 shows that, in fact, if the system is globally asymptotically stable, this implies the stronger property of exponential convergence.

If we consider the system in equation 5.1 as a switched linear system with infinitely many modes where the state matrices describe a convex hull, the results in [DM99] lead to the conjecture that in the case of the system 5.1 global asymptotic stability generally implies exponential stability.

Additionally, regarding the system as a set of linear systems leads to the requirement of a common unique equilibrium for all linear subsystems to guarantee global asymptotic stability.

Even when considering the system as a switched linear system, the conditions in Theorem 7 are conservative. Surprisingly, the common Lyapunov-function for a stable linear switched system might not be quadratic [DM99]. The switching rule $h^k(x)$ is not arbitrary but state dependent and might hence stabilize the system even if some vertex matrices are not Hurwitz. However, the switching rule is unknown as it depends on the parameters of the Gaussian Mixture Model.

Global stability of the system 5.1 might hence be ensured without a stabilizer and quadratic Lyapunov function or constant contraction metric respectively, but the search for such a function or metric remains an open problem.

Chapter 6

Solution Design

To ensure the stability of the system 4.2 with respect to an equilibrium, the solution approach needs to satisfy the following two conditions:

1. Existence of an equilibrium x^* at the target of the motion

2. (Global) Asymptotic stability of x^*

If $b^k = -A^k x^*$ holds in equation 4.2, then there is an equilibrium in x^* with $f(x^*) = 0$ and condition 1 is satisfied. The system's equations then become

$$\dot{x} = \sum_{k=1}^{K} h^k(x) A^k (x - x*) \tag{6.1}$$

Requiring the existence of an equilibrium is not sufficient to guarantee that the motion is asymptotically stable at the target. As can be seen in figures 6.1 and 6.2, over all stability properties may even become worse compared to the unconstrained system. The system's trajectories in figure 6.2 diverge in case of the slightest perturbation.

In order to guarantee the second condition, too, we investigate on three different approaches:

- The first approach seeks to achieve global exponential stability by adding a control matrix to the learned system. This approach did not provide satisfactory results as the control matrix strongly modifies the motion.

- The second approach proposes a modification of the Stable Estimator of Dynamical Systems (SEDS) proposed by Khansari-Zadeh and Billard in [KZB11a] and hence achieves global exponential stability, too.

- The third approach ensures local stability with respect to the equilibrium. A stabilizer becomes active in a region close to the target or after a certain time has passed.

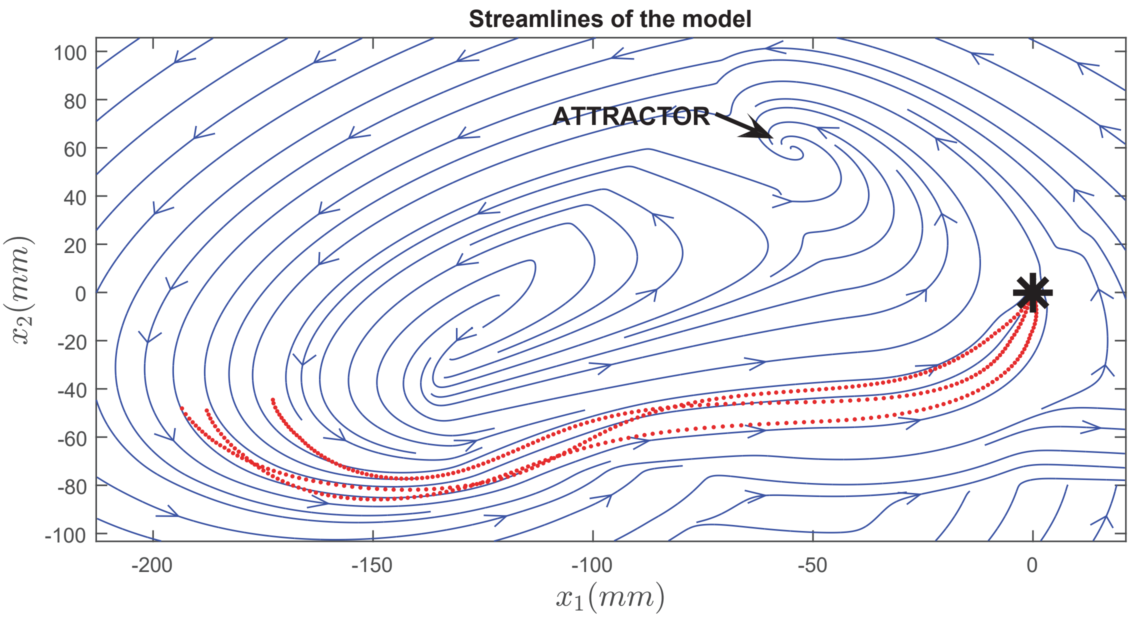


Figure 6.1: Unstable Gaussian Mixture Regression: Demonstrations of the motion (red-dashed lines) and streamlines (blue) of the dynamical system learned with Expectation-Maximization. The target (black) is not stable.

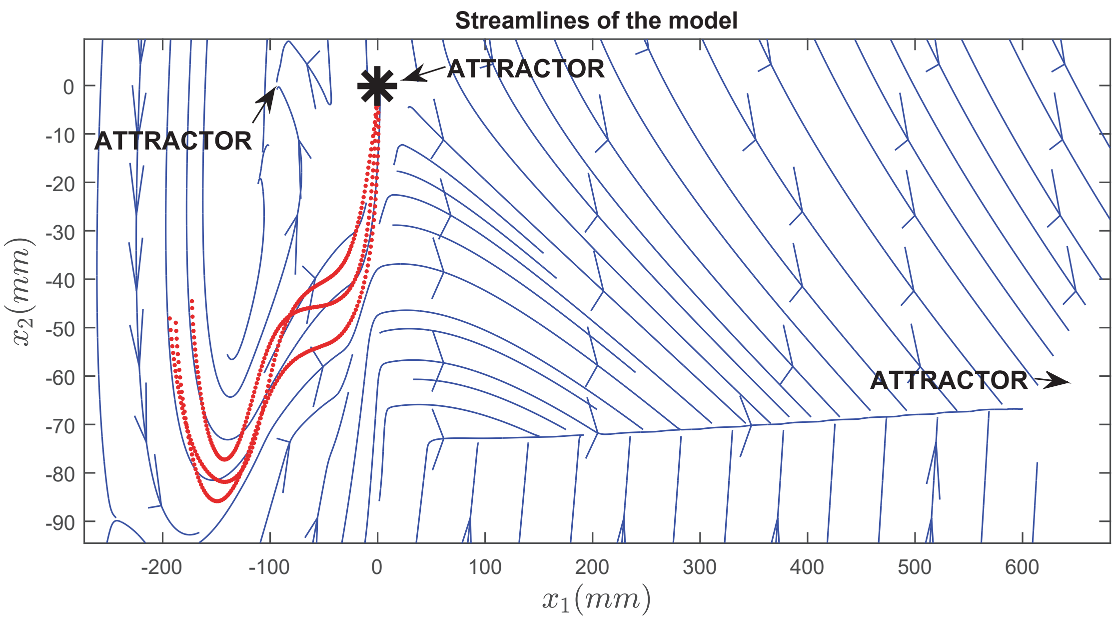


Figure 6.2: Unstable constraint Gaussian Mixture Regression: Demonstrations of the motion (red-dashed lines) and streamlines (blue) of the dynamical system learned with Expectation-Maximization and optimized under the constraint of an equilibrium at the target (black). In case of slight perturbations, the system's trajectories may converge to a spurious attractor.

6.1 Trial: Additive Control Matrix

Similar to the approach in [KZB11a], the parameters of the Gaussian Mixture Model are first learned by applying the Expectation-Maximization and then again optimized under the constraint $b^k = -A^k x^*$ to ensure $f(x^*) = 0$.
According to Theorem 7 the system is partially contracting if the set of K linear systems

$$s^k(x) = (A^k x + b^k) = A^k(x - x^*) \tag{6.2}$$

is contracting in a *common* metric $M(x,t) = \Theta^T\Theta$. As the systems are linear, the K Jacobians simply correspond to A^k. With M being a constant metric, the Generalized Jacobians are formulated as $F^k = \Theta A^k \Theta^{-1}$. If it can be shown, that there exists a matrix Θ and a measure μ for whom the measure of the set of Generalized Jacobians is negative, the system is partially contracting. In general, a common metric for all $k = 1..K$ will not exist.
Our first approach proposes therefore to choose a subset of A^k that will be modified with a constant control matrix U^k such that a common metric for *all* K subsystems can be found.

$$\dot{x} = \sum_{k=1}^{K} h^k(x)\left((A^k + U^k)(x - x^*)\right) \tag{6.3}$$

How to choose the control matrix U^k

The control matrix U^k should guarantee global stability without strongly affecting the reproduction of the motion. Choosing the elements of U^k via the matrix measure associated to the Euclidean norm is a difficult problem since this means modifying the eigenvalues of the symmetric part. Using the matrix measures associated to the norms l_1 or the l_∞ provides simpler expressions (see equations 2.3 - 2.5).
For the matrix measure associated to the l_1-norm

$$\mu_1(A) = \max_j \left(a_{jj} + \sum_{i \neq j} |a_{ij}| \right)$$

we propose a procedure in Algorithm 1 to determine U such that $\mu_1(A + U) < 0$. Note that replacing A by its transpose A^T allows the use of Algorithm 1 for $\mu_\infty(.)$ instead of $\mu_1(.)$.
Using Algorithm 1 does not yet take advantage of the generalization of Contraction Theory via a differential coordinate transformation Θ.

Algorithm 1 Choose a control matrix U such that measure $\mu_1(A+U) < 0$

Require: D // dimension of state space
Require: A // a matrix $A \in \mathbb{R}^{D \times D}$
Require: p // a parameter $p > 1$
  $U \leftarrow$ Matrix of zeros in $\mathbb{R}^{D \times D}$
  for $d = 1$ to D **do**
    $diagonal \leftarrow$ d-th diagonal element of A^k
    $c \leftarrow \sum_{i=1}^{D} |a_{ij}^k| - |diagonal|$

    if $diagonal > 0$ **and** $|diagonal| < c$ **then**
      $u \leftarrow -\ diagonal - p * c$
    else if $diagonal > 0$ **and** $|diagonal| > c$ **then**
      $u \leftarrow -\ 2 * diagonal$
    else if $diagonal < 0$ **and** $|diagonal| < c$ **then**
      $u \leftarrow -\ c$
    else
      $u \leftarrow 0$ // no need to modify the (d,d)th element of A
    end if
    $U(d,d) \leftarrow u$
  end for
  return U

Search for a suitable metric $M = \Theta^T\Theta$

To find a constant matrix Θ such that $\mu_1(\Theta A^k \Theta) < 0$ holds for all $k \in 1..K$, we first investigate on finding a metric for a single matrix A.
The Graphical Approach proposed by Russo et al. in [RdBS11] provides a procedure to check if sufficient conditions for the existence of such a diagonal constant coordinate transformation matrix Θ are satisfied.
We investigated on applying this procedure simultaneously for the set of linear systems. The conditions become too restrictive in this case.
In [APS08], a numerical approach to finding a contraction metric is proposed. We applied the approach to find a contraction metric that allows the largest possible symmetric uncertainties for a given matrix A. The resultant uncertainty intervals were very small and in general only included one of the vertex matrices in $S = \{A^1, A^2, ...A^K\}$.

Searching for a suitable contraction metric then led to the method proposed in the next section.

6.2 SEDS via a suitable Contraction Metric

As in the previous approach, $b^k = -A^k \xi^*$ is set as an optimization constraint to ensure the existence of the equilibrium ξ^*, similar to [KZB11a].
To guarantee global stability the authors in [KZB11a] set the additional optimization constraint $\mu_2(A^k) < 0$. At each iteration of the optimization, Sylvester's criterion is used to verify negative definiteness. The symmetric part of the matrices A^k is

$$B^k = A^k + (A^k)^T$$

The constraints $\mathcal{C}$ according to Sylvester's criterion correspond to the sign of the determinants of the principal minors in B^k:

$$\mathcal{C}_{(k-1)d+c} : (-1)^{c+1} det(B^k_{1:c,1:c}) < 0 \;\; \forall c \in 1..d, \;\; \forall k \in 1..K \tag{6.4}$$

Details can be found in the Technical Report in [KZB11b].

Given a coordinate transformation matrix Θ, we propose a modification of the likelihood optimization constraint such that $\mu_2(\Theta A^k \Theta^{-1}) < 0$. Thus, the symmetric part becomes

$$(B^k)_{modified} = \Theta A^k \Theta^{-1} + (\Theta A^k \Theta^{-1})^T$$

and we have to satisfy the constraints in 6.4 for $(B^k)_{modified}$. As solving the likelihood optimization problem requires the derivative of the constraints with respect to the optimization parameters, the modified derivatives are given below:

The stability constraints do not depend on the priors π^k and the means μ^k. Hence, as in the original approach, the derivative of $\mathcal{C}$ disappears:

$$\frac{\partial \mathcal{C}_{(k-1)d+c}}{\partial \pi^k} = 0 \;\; \forall c \in 1..d, \;\; \forall k \in 1..K \tag{6.5}$$

$$\frac{\partial \mathcal{C}_{(k-1)d+c}}{\partial \mu^k} = 0 \;\; \forall c \in 1..d, \;\; \forall k \in 1..K, \;\; \forall i \in 1..2d \tag{6.6}$$

As the matrix A^k is determined via $A^k = \Sigma^k_{\dot{\xi}\xi}(\Sigma^k_{\xi})^{-1}$, the derivative of $\mathcal{C}$ with respect to Σ^k does not disappear. To guarantee the positive definiteness of the covariance matrices Σ^k, a change of variables is proposed in [KZB11b] where $L^k = Chol(\Sigma^k)$ are the lower $2d \times 2d$ triangle matrices obtained by the Cholesky decomposition of Σ^k. Hence, the derivatives are formulated with respect to the $ij-$th element of L^k:

$$\frac{\partial \mathcal{C}_{(k-1)d+c}}{\partial L^k_{ij}} = (-1)^{c+1} tr \left(adj \left((B_{1:c,1:c})_{modified} \right) (\mathcal{X}_{1:c,1:c})_{modified} \right) \tag{6.7}$$

where the $2d \times 2d$ $\mathcal{X}_{modified}$ is defined by:

$$\Phi = \mathbf{0}^{ij}\left(L^k\right)^T + L^k\left(\mathbf{0}^{ij}\right)^T \tag{6.8}$$

$$\Psi = \left(-A^k\Phi_\xi + \Phi_{\dot{\xi}\xi}\right)\left(\Sigma_\xi\right)^{-1} \tag{6.9}$$

$$\mathcal{X}_{modified} = \Theta\Psi\Theta^{-1} + \left(\Theta\Psi\Theta^{-1}\right)^T \tag{6.10}$$

The matrix $\mathbf{0}^{ij}$ is a $2d \times 2d$ matrix of zeros, where only the $ij-$th component is 1.

Note that in our approach we only consider the likelihood optimization from [KZB11a].

How to learn the coordinate transformation Θ

The modification proposed above requires a coordinate transformation matrix Θ. We will use results in [Zah03], see Theorem 2, to learn Θ via a test matrix T.

This test matrix T can be the center of the state matrices A^k:

$$T_1 = \frac{1}{K}\sum_{k=1}^{K} A^k \tag{6.11}$$

Given the auxiliary Jacobian $J(x) = \sum_{k=1}^{K} h^k(x)A^k$ and N position data points $\{x_1...x_N\}$ in the demonstration area, T can also be chosen as the center of the auxiliary Jacobians obtained at position x_n:

$$T_2 = \frac{1}{N}\sum_{n=1}^{N}\left(\sum_{k=1}^{K} h^k(x_n)A^k\right) \tag{6.12}$$

The choice of the test matrix described above is inspired by representing the set of matrices by a single matrix $(T + \Delta T)$ where ΔT represents uncertainties [Zah03]. If it holds that

$$\mu(\Delta T) < -\mu(T) \;\;\&\;\; \mu(T) < 0$$

we can conclude via the matrix measure property 3 in Theorem 1 that

$$\mu(T + \Delta T) < 0$$

The coordinate transformation matrix is chosen as the invertible matrix Θ such that $V = \Theta T \Theta^{-1}$ is the Jordan form of T. The Jordan Tranformation can be defined for real matrices, too [GLR06]. To avoid complex entries in Θ we hence use the real Jordan form. Choosing Θ via the Jordan form is based on Theorem 2 and its proof in chapter 2. Hence, the measure $\mu_2^\Theta(T)$ should be close to its lower bound. If additionally, $\mu_2^\Theta(T) < 0$, $\mu(\Delta T)$ may be positive while the system is still stable.

If T has at least one positive eigenvalue, it is hence calculated once more, but by taking into account only the A^k (or $J_{aux}(x)$ respectively) that have negative eigenvalues. However, the measure $\mu_2^\Theta(T)$ might still be positive and the contraction metric $\Theta^T\Theta$ is not necessarily a good choice for the system.

6.3 Local Stabilizer

Similar to the concept of Control Lyapunov functions (see [KZB14]), it is possible to stabilize a system via a Control Contraction Metric [MS14]. To guarantee local stability of the equilibrium, we hence propose an additive feedback control law that is active only within a ball close to the target of the motion. We abandon the search for a suitable differential coordinate transformation and the contraction metric $M(x, t)$ will be the identity matrix.

If the upper condition $b^k = -A^k x^*$ for the existence of an equilibrium holds, this is equivalent to

$$\sum_{k=1}^{K} h^k(x)\ b^k = -\sum_{k=1}^{K} h^k(x) A^k x^* \ \ \forall x \in \mathbb{R}^n$$

The control law is then derived as follows:
Given is an equilibrium x^* and the system:

$$\dot{x} = \sum_{k=1}^{K} h^k(x)(A^k x + b^k) = \sum_{k=1}^{K} h^k(x) A^k x + \sum_{k=1}^{K} h^k(x) b^k$$

If at each time step, we choose a square matrix $U^i \in \mathbb{R}^{n \times n}$ such that

$$\mu_1 \left(U^i + \sum_{k=1}^{K} h^k(x_i) A^k \right) < 0$$

the auxiliary system with control in equation 6.13 is contracting and the original system is hence exponentially stable.

$$\dot{y} = \underbrace{\sum_{k=1}^{K} h^k(x) A^k y + \sum_{k=1}^{K} h^k(x) b^k}_{System} + \underbrace{\omega \left(Uy - \sum_{k=1}^{K} h^k(x) b^k \right)}_{Control} \tag{6.13}$$

We determine U at each time step via the procedure given in algorithm 1. As input we set $A \leftarrow \sum_{k=1}^{K} h^k(x_i) A^k$.

To activate the control law, we will use a sigmoid function $\omega(x_i, a, c)$.

$$\omega(x_i, a, c) = \frac{1}{1 + e^{-a(x_i + c)}}$$

ω smoothly activates the control law as soon as a predefined distance to x^*, hence a ball around the equilibrium, has been reached or a certain time limit has passed.

Chapter 7

Simulation Results

This chapter gives the results of an experimental evaluation of the proposed approaches. Simulations have been conducted with MATLAB.
The demonstration data set consists of 20 different 2-D hand writing motions provided by the LASA laboratory and was downloaded together with the sourcecode for the SEDS approaches from:

http://lasa.epfl.ch/sourcecode/

Each model is obtained by three or four demonstrations of the model and hence we simulated and compared the reproduction for three (or four respectively) starting points corresponding to each demonstration. The sampling time of the demonstrations is $0.02s$.
As the demonstration data set and the simulation data set are of different length, we use *Dynamic Time Warping* to measure the performance in terms of accuracy of the different approaches. Dynamic Time Warping allows to eliminate timing differences of two time series by a non-linear alignment of the data. The algorithm was originally proposed in the context of speech recognition (see [SC78]).
As a measure for the similarity of motion trajectories it has already been used i.e. in the context of gesture recognition to cope with uncertainties that arise when the gestures are performed by different users (see [SL13a]). The motion data is multidimensional and hence we use the MD-DTW algorithm given in [San12].

7.1 Simulation Results for Trial: Additive Control Matrix

This section gives the exemplary simulation results for three motions using the additive control matrix proposed in section 6.1 to guarantee global contraction of the system. The results indicate that while according to the streamlines in figure 7.1 the system is globally stable, the simulated motion in figure 7.2 differs from the demonstration data.

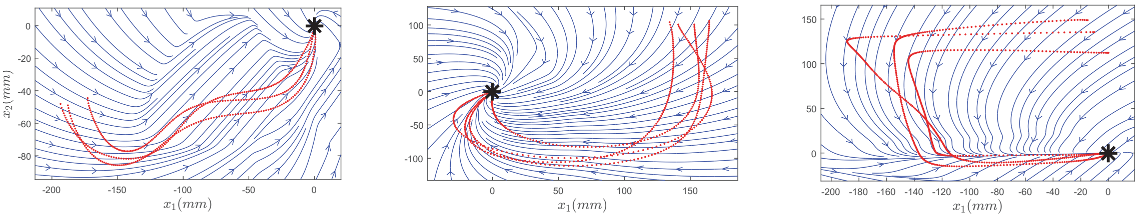


Figure 7.1: Additive Control Matrix, Trial: The streamlines (blue) of the models obtained from demonstrations (red-dashed lines) indicate that for any starting point in $\mathbf{R}^n$ the motion will converge to the target (black asterisk).

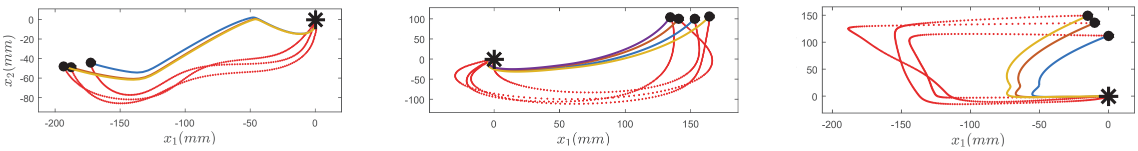


Figure 7.2: Additive Control Matrix, Trial: Although the simulated motion (continuous lines) reaches the target (black asterisk), it does not correctly reproduce the demonstration data (red-dashed lines).

7.2 Simulation Results for Modified SEDS and Local Stabilizer

This section gives a more extensive evaluation of the performance of the modified SEDS and the Local Stabilizer approach in terms of accuracy and training time. We give a comparison of existing approaches (SEDS and SEDS II [KZB11a, KZB14]) and our propositions (Local Stabilizer and Modification of SEDS by considering a contraction metric learned via the test matrices T_1 or T_2). As in general the performance of SEDS II strongly depends on the chosen form of the Control Lyapunov-function we opted for a Lyapunov-function of zero asymmetric components, hence the quadratic function $V(x) = x^T Px$ (CLF0) and a Lyapunov-function with three asymmetric components (CLF3), which should be adaptable to more complex motions.

As all the six approaches depend on the random initialization of the Expectation-Maximization Algorithm, we ran the different approaches 15 times for each motion. The number of Gaussian Components is determined via the BIC-criterion where instead of awaiting the minimal BIC we set a threshold on the decrease of the BIC as otherwise experiments yielded more Gaussian components than suitable (see Discussion, chapter 8). We hence obtained 300 dynamical systems learned with Gaussian Mixture Regression.

During our experiments, the simulated trajectories of the SEDS II approach with three asymmetric components often began to jitter. This may be due to similar numerical issues as already described in [KZB14]. We hence set the time steps of the simulation to $0.002s$ as suggested in [KZB14] and additionally eased the convergence criteria for this approach by checking the distance to the target instead of the velocity. However, we were not able to remove this error for all the motions and hence do not consider the motions "Khamesh" and "WShape" for our experiments.

Considering the globally stable SEDS approaches we could not verify the stability for two of the 300 dynamical systems as crossing trajectories and extreme accelerations were observed: one case concerned the original SEDS approach for the motion "NShape", another one concerned the modified SEDS for the motion "Line" where the metric was learned from test matrix T_1. Similar to above, this may be due to a numerical error. The approaches derived in the previous section hold under the assumption of a continuous system and hence the application of discrete time steps may cause problems.
As the Local Stabilizer only guarantees global convergence to the target after a certain time has passed, it may yield strange trajectories if the Gaussian Mixture Regression is not well parametrized. This was observed for 5 of the 15 initializations for the motion "Sshape" (see figures 7.3 and 7.4). We removed the corresponding initializations and therefore only considered simulation results based on an acceptable parametrization of Gaussian Mixture Regression.

The final simulation data set consists of 18 handwriting motions yielding ten different models each.

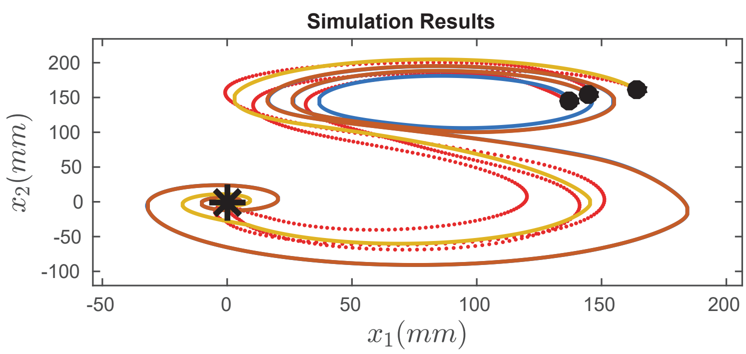


Figure 7.3: Bad initialization for Local Stabilizer: Three Gaussian components are used. Contrary to the demonstrations (red-dashed lines), two of the simulated trajectories (blue, brown continuous lines) yield cyclic behavior before the stabilizer is activated. Only the yellow simulation trajectory directly converges to the target (black asterisk).

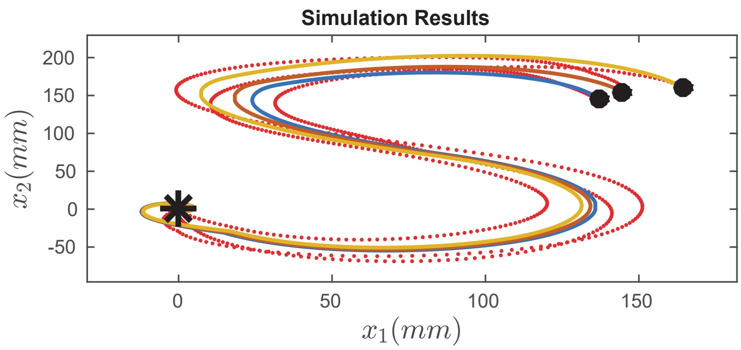


Figure 7.4: Good initialization for Local Stabilizer: Based on four Gaussian components, the simulated trajectories (continuous lines) yield a behavior similar to the demonstrated trajectories (red-dashed lines).

As both the measured Dynamic Time Warping distances and the measured training time are not normally distributed, we consider the median instead of the mean to measure the typical performance. To indicate the maximal and minimal deviation from the typical performance, we provide the location of the 10%− and the 90%−quantile via the black error bars.

According to figure 7.5, the globally stable SEDS approaches do not differ significantly from one-another. While the original SEDS approach performs slightly better regarding the median, it may sometimes produce worse results than the modified SEDS approaches as is indicated by the 90%−quantile.
The approaches ensuring stablity via a stabilizer perform better than the globally stable approaches. Yet the Local Stabilizer does not achieve the same performance as the SEDS II approaches using a Control Lyapunov-function. Using a Control Lyapunov-function with zero asymmetric quadratic components may produce worse results compared to the use of a Control Lyapunov-function with three asymmetric quadratic components. However, the median yields quite similar performance, as most of the motions we consider are suitable for the use of quadratic Lyapunov-functions.

The median and the 10%− and the 90%−quantiles of the training time are shown in figure 7.6 for the six approaches.
The training time for the stabilization approaches mainly consists of the computation time of the E-M-Algorithm. For the SEDS II approach, obtaining the Control-Lyapunov-function may be time-consuming, especially in the case of three asymmetric quadratic components. The different SEDS approaches require the computation of the constraint likelihood-optimization problem additionally to the E-M-Algorithm. Changing the contraction metric and hence the optimization constraints does not seem to influence the training time.

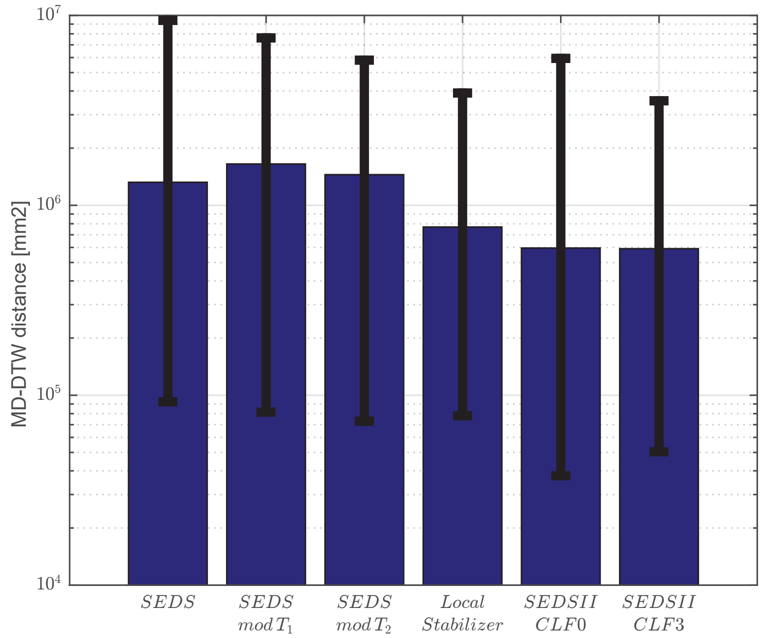


Figure 7.5: Median MD-DTW distance of the six approaches for 18 hand-writing motions. The blue bars indicate the median, the black lower and upper error bar the 10%− and the 90%−quantile respectively.

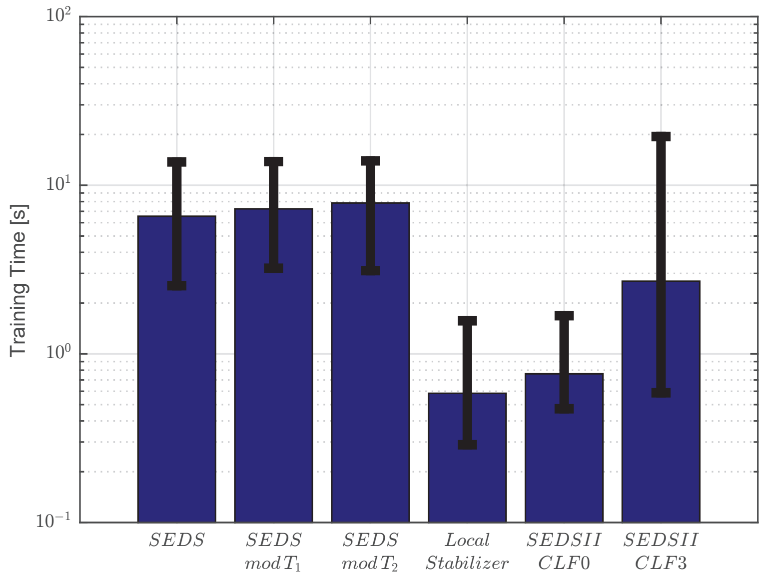


Figure 7.6: Median Training Time in seconds of the six approaches for 18 hand-writing motions. The blue bars indicate the median, the black lower and upper error bar the 10%− and the 90%−quantile respectively.

We observed that the performance of the approaches differs significantly considering different motions. We hence give the exemplary results for three motions, "JShape", "Bump" and "SharpC" in the following.

The median Dynamic Time Warping distance for the motion "JShape" is given in figure 7.7. The model is learned with two Gaussian components. In the case of this motion, the stabilization approaches, Local Stabilizer and SEDS II, perform worse than the approaches that yield global stability. As can be seen in the in figures 7.8b, d and f, one of the simulated trajectories slightly drifts away from the demonstration area. Note the different behavior of Local Stabilizer in figure 7.8b and SEDS II approach in figure 7.8d in the area close to the target.

The median Dynamic Time Warping distance for the motion "Bump" learned with four Gaussian components is given in figure 7.9. All approaches, except the modified SEDS approach where a metric is learned from the matrix T_1, equation 6.11, yield acceptable performance. The corresponding motion in figure 7.10c does not follow the curvature of the motion.

Figure 7.11 shows the performance in terms of accuracy for the motion "SharpC" learned with three Gaussian components. While using a suitable Contraction Metric yields better performance than the original SEDS approach (see figures 7.12a, c, e), the stabilization approaches Local Stabilizer and SEDS II still outperform the globally stable dynamical systems.

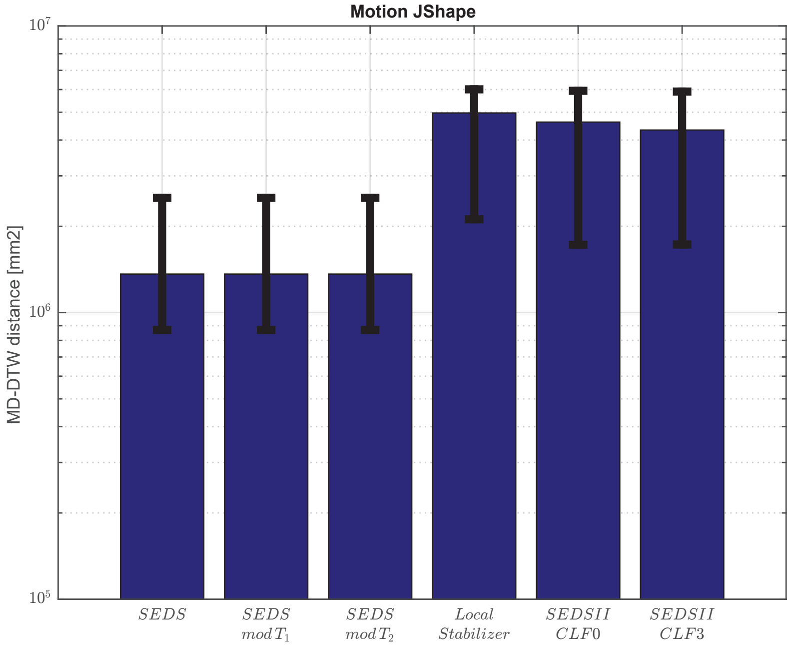
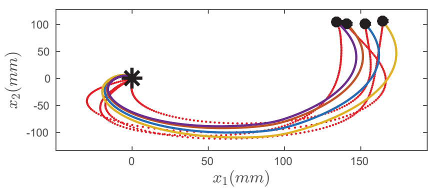
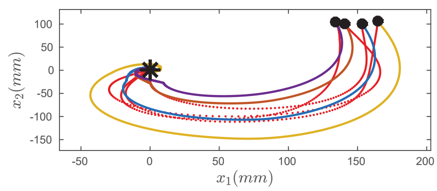


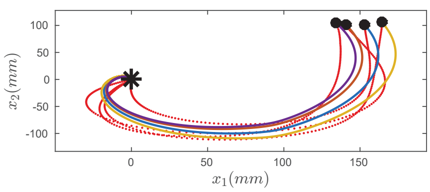
Figure 7.7: Median MD-DTW distance of the six approaches for the motion "JShape". The blue bars indicate the median, the black lower and upper error bar the 10%− and the 90%− quantile respectively.



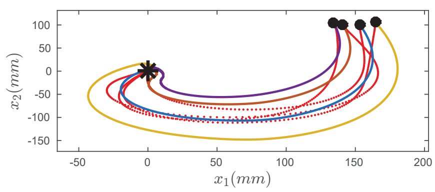
(a) SEDS



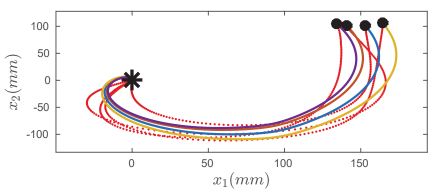
(b) Local Stabilizer



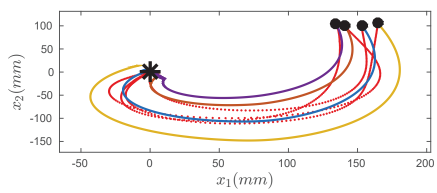
(c) SEDS mod T_1



(d) SEDS II CLF0



(e) SEDS mod T_2



(f) SEDS II CLF3

Figure 7.8: Simulation (continuous lines) and demonstrations (red-dashed lines) for the motion "JShape".

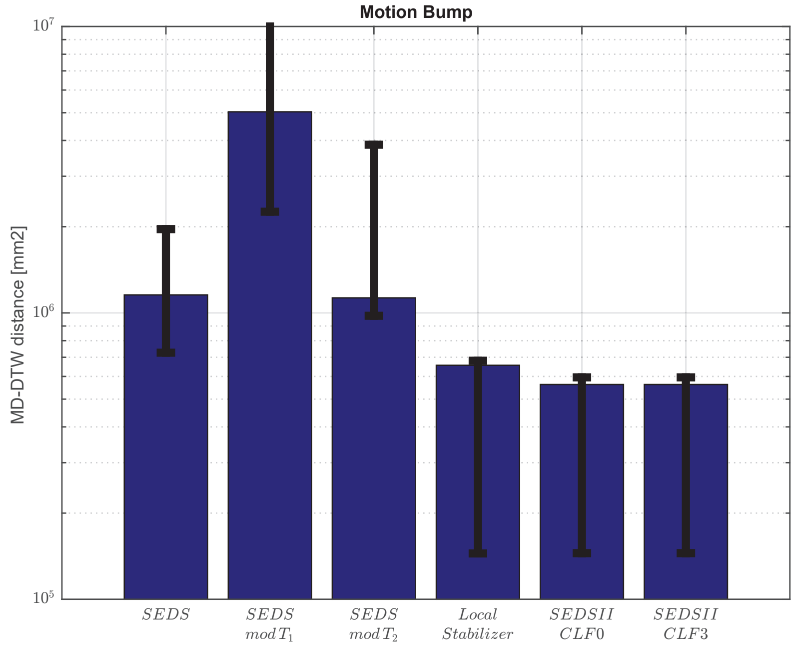
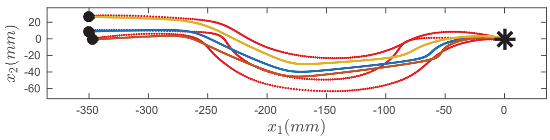
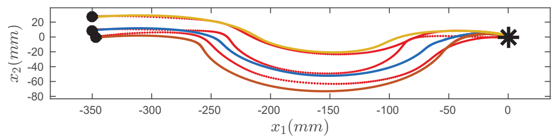


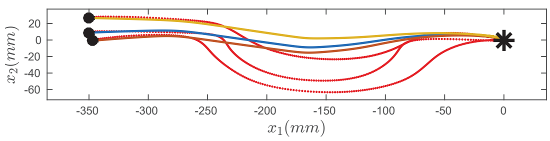
Figure 7.9: Median MD-DTW distance of the six approaches for the motion "Bump". The blue bars indicate the median, the black lower and upper error bar the 10%− and the 90%−quantile respectively.



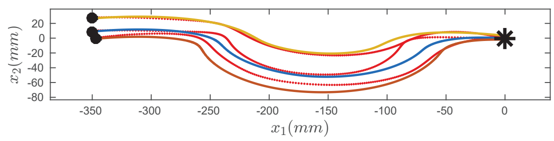
(a) SEDS



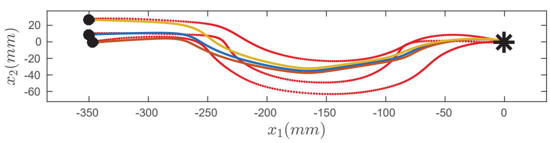
(b) Local Stabilizer



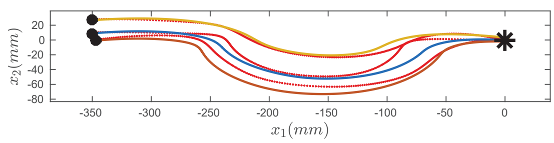
(c) SEDS mod T_1



(d) SEDS II CLF0



(e) SEDS mod T_2



(f) SEDS II CLF3

Figure 7.10: Simulation (continuous lines) and demonstrations (red-dashed lines) for the motion "Bump".

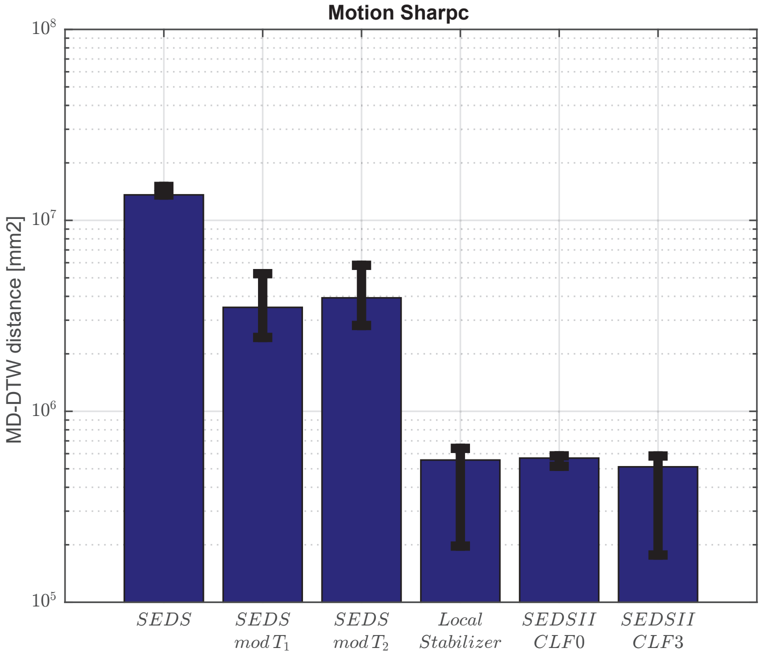
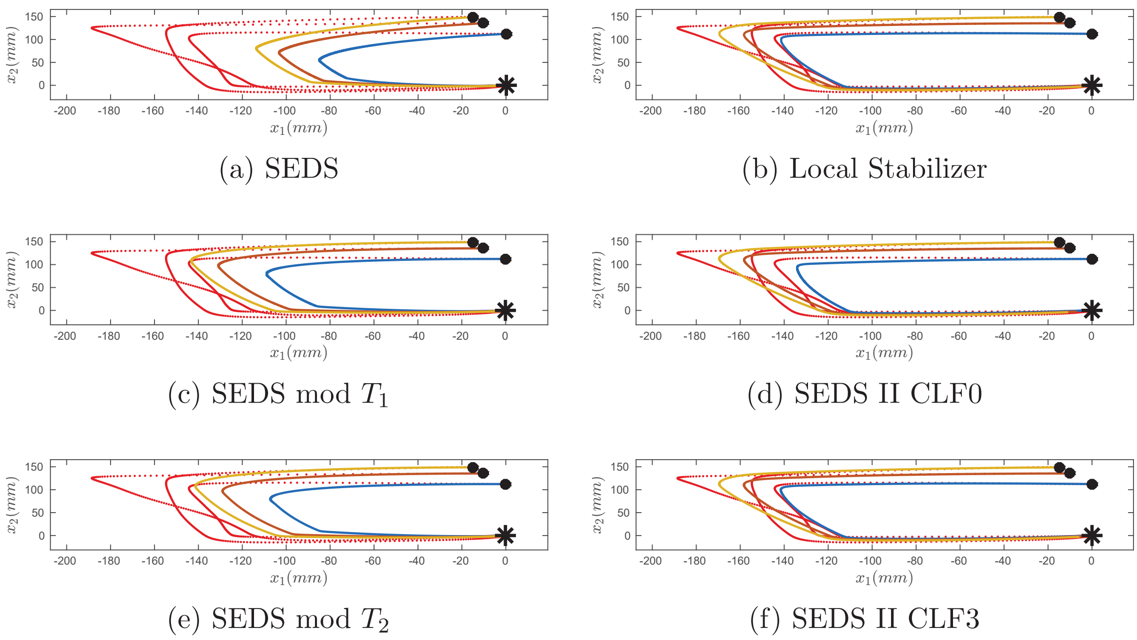


Figure 7.11: Median MD-DTW distance of the six approaches for the motion "Sharp C". The blue bars indicate the median, the black lower and upper error bar the 10%− and the 90%− quantile respectively.



(a) SEDS

(b) Local Stabilizer

(c) SEDS mod T_1

(d) SEDS II CLF0

(e) SEDS mod T_2

(f) SEDS II CLF3

Figure 7.12: Simulation (continuous lines) and demonstrations (red-dashed lines) for the motion "SharpC".

Chapter 8

Discussion

The Stabilization approaches, Local Stabilizer and SEDS II, yield better performance than SEDS and modified SEDS both in terms of accuracy and training time (see figures 7.5 and 7.6). These approaches use primarily the unstable dynamical system to reproduce the motion. As we do not consider perturbations or obstacles the stabilizer only intervenes when the motion is near the target (Local Stabilizer) or when the motion is not following the negative gradient of the Lyapunov-function (SEDS II). Note that if the Lyapunov-function is well-chosen in SEDS II, the stabilizer should only intervene very little in the demonstration area.
On the other hand, in the case of SEDS and modified SEDS, the model is inherently stable and the trade-off between accuracy and stability becomes noticeable.

The stabilization approaches are also very sensitive to the number of Gaussian components and can only be as good as the Gaussian Mixture Regression. As the results in figures 7.7 and 7.8 suggest, the SEDS approaches are able to compensate this problem. Additionally, in the case of the SEDS approaches, a smaller number of Gaussian components is desirable as it leads to a smaller number of optimization constraints.

The results for the motion "SharpC" in figure 7.11 and 7.12 show the limits of using a quadratic Lyapunov-function $V(x) = x^T P x$. While using a suitable constant contraction metric can improve the performance, the globally stable models still cannot keep up with the unconstrained approaches. Our approach proposed in section 6.2 to learn a contraction metric may yield good results as in the case of the motion "SharpC", but can also fail as in the case of the motion "Bump" when using T_1 to learn the metric (figure 7.10c).

The modification of SEDS maintains the advantageous property of inherent global stability while being in some cases more adapted to the motion. However, the property of SEDS, that the sum of two SEDS systems is globally stable, too, does not hold for the modification anymore. Additionally, the approach still faces the limitations given by the requirement of quadratic stability.

According to the results in section 7.1, adding a control matrix to the dynamical system in equation 5.1 has a negative effect on the reproduction behavior. Therefore we opted to only locally stabilize the system by the means of an additive control matrix. While the Local Stabilizer gives a simple and general expression to guarantee convergence, its comparison with the SEDS II approaches show that the influence of the control matrix is still noticeable. To face this problem, the radius of the ball around the equilibrium can be decreased. However, in this case a sufficient number of Gaussian components is required to guarantee that all trajectories determined by Gaussian Mixture Regression reach this ball.

Chapter 9

Conclusion and future work

Throughout this work, we have shown that searching to achieve global stability for the system in 4.2 yields similar results via both Contraction Theory and Lyapunov Theory. The existence of a constant contraction metric implies the existence of a quadratic Lyapunov-function.
However, Contraction Theory provides different tools to analyze the stability of a system. This allowed to state that the stability problem in 4.2 corresponds to typical problems in switched linear systems and uncertain systems and to use the results for an approach to learn a contraction metric.
Additionally, we propose a stabilizer that forces the motion to reach the target without changing significantly the over-all motion.

Further work will consider an experimental evaluation on real robots.
Approaches in the theory of uncertain or switched systems that yield the existence of a higher-order Lyapunov-Candidate or Contraction Metric and an improvement of the time clock that is used in case of perturbations for the Local Stabilizer by using a more robust approach require further investigation, too.
Other regression techniques mentioned in the introduction such as i.e. Gaussian Processes can be advantageous compared to Gaussian Mixture Regression, especially in terms of accuracy. Therefore, further work will also consider the stability of those methods by the means of Contraction Theory.

List of Figures

Bibliography

[APS08] E.M. Aylward, P.A. Parrilo, and J.J.E. Slotine. Stability and robustness of nonlinear systems via Contraction Metrics and SOS programming. *Massachusetts Institute of Technology, Technical Report 2691*, February 2008.

[Cal09] S. Calinon. *Robot Programming by Demonstration: A probabilistic approach*. EPFL Press, 2009.

[CDS+10] S. Calinon, F. D'Halluin, E. Sauser, D. Caldwell, and A. Billard. Learning and reproduction of gestures by imitation: An approach based on Hidden Markov Model and Gaussian Mixture Regression. *IEEE Robotics and Automation Magazine*, 2010.

[Des09] C. A. Desoer. *Feedback Systems: Input Output Properties*. Society for Industrial and Applied Mathematics, 2009.

[DM99] W.P. Dayawansa and C.F. Martin. A converse Lyapunov Theorem for a class of dynamical systems which undergo switching. *IEEE Transactions on automatic control*, 1999.

[GLR06] I. Gohberg, P. Lancaster, and L. Rodman. *Invariant Subspaces of Matrices with Applications*. Society for Industrial and Applied Mathematics, 2006.

[JF10] J Jouffroy and T.I. Fossen. A tutorial on incremental stability analysis using Contraction Theory. *Modeling, Identification and Control*, 2010.

[JIS02] J. Nakanishi J.A. Ijspeert and S. Schaal. Movement imitation with nonlinear dynamical systems in humanoid robots. *Proceedings of the International Conference on Robotics and Automation (ICRA)*, 2002.

[KK00] G. A. Korn and T. M. Korn. *Mathematical handbook for scientists and engineers: definitions, theorems, and formulas for reference and review.* Courier Corporation, 2000.

[KKB15] K. Kronander, M. Khansari, and A. Billard. Incremental motion learning with locally modulated dynamical systems. *Robotics and Autonomous Systems*, 2015.

[KZB10] S. M. Khansari-Zadeh and A. Billard. BM: An iterative algorithm to learn stable non-linear dynamical systems with Gaussian Mixture Models. *Proceedings of the International Conference on Robotics and Automation (ICRA)*, 2010.

[KZB11a] S. M. Khansari-Zadeh and A. Billard. Learning stable non-linear dynamical systems with Gaussian Mixture Models. *IEEE Transaction on Robotics*, 2011.

[KZB11b] S. M. Khansari-Zadeh and A. Billard. The derivatives of the SEDS optimization cost function and constraints with respect to the learning parameters. *Learning Algorithms and systems laboratory (LASA), Technical Report*, April 2011.

[KZB14] S. M. Khansari-Zadeh and A. Billard. Learning control Lyapunov function to ensure stability of dynamical system-based robot reaching motions. *Robotics And Autonomous Systems*, 2014.

[LA09] H. Lin and P.J. Antsaklis. Stability and stabilizability of switched linear systems: A survey of recent results. *IEEE Transactions on automatic control*, 2009.

[LS98] W. Lohmiller and J.J.E. Slotine. On Contraction analysis for nonlinear systems. *Automatica*, 1998.

[LT04] D. Liberzon and R. Tempo. Common Lyapunov Functions and gradient algorithms. *IEEE Transactions on automatic control*, 2004.

[MS14] I.R. Manchester and J.J.E. Slotine. Control Contraction Metrics and universal stabilizability. *Proceedings of the 2014 IFAC World Congress, preprint arXiv:1311.4625*, 2014.

[NLS13] K. Neumann, A. Lemme, and J. J. Steil. Neural learning of stable dynamical systems based on data-driven Lyapunov candidates. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* , 2013.

[NS10] K.S. Narendra and R. Shorten. Hurwitz stability of Metzler matrices. *IEEE Transactions on automatic control*, 2010.

[PHAS09] P. Pastor, H. Hoffmann, T. Asfour, and S. Schaal. Learning and generalization of motor skills by learning from demonstrations. *IEEE International Conference on Robotics and Automation (ICRA)*, 2009.

[PM09] O. Pastravanu and M.-H. Matcovschi. Matrix measures in the qualitative analysis of parametric uncertain systems. *Mathematical Problems in Engineering*, 2009.

[RdBS11] G. Russo, M. di Bernardo, and J.J.E. Slotine. A graphical approach to prove Contraction of nonlinear circuits and systems. *Transactions on Circuits and Systems*, 2011.

[RLdB11] G. Russo, Davide Liuzza, and M. di Bernardo. Contraction analysis of switched systems: the case of Caratheodory systems and networks. *arXiv preprint arXiv:1110.0855* , 2011.

[RW06] C. Rasmussen and C. Williams. *Gaussian Processes for machine learning*. MIT Press, 2006.

[S0̈6] G. Söderlind. The logarithmic norm. History and modern theory. *BIT Numerical Mathematics*, 2006.

[San12] P. Sanguansat. Multiple multidimensional sequence alignment using generalized dynamic time warping. *WSEAS Transactions on Mathematics*, 2012.

[SC78] H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1978.

[SKS00] S. Schaal, S. Kotasaka, and D. Sternad. Nonlinear dynamical systems as movement primitives. *IEEE International Conference on Humanoid Robots*, 2000.

[SL13a] M. Saveriano and D. Lee. Invariant representation for user independent motion recognition. *IEEE International Symposium on Robot and Human Interactive Communication (Ro-Man)*, 2013.

[SL13b] M. Saveriano and D. Lee. Point cloud based dynamical system modulation for reactive avoidance of convex and concave obstacles. *IEEE International Conference on Intelligent Robots and Systems (IROS)*, 2013.

[SN00] R.N. Shorten and K.S. Narendra. Necessary and sufficient conditions for the existence of a common quadratic Lyapunov function for M stable second order linear time-invariant systems. *Proceedings of the American Control Conference*, 2000.

[SS04] A. J. Smola and B. Schölkopf. A tutorial on support vector regression. *Statistics and Computing*, 2004.

[Vid78] M. Vidyasagar. On matrix measures and convex Lyapunov functions. *Journal of Mathematical Analysis and Applications*, 1978.

[Vid93] M. Vidyasagar. *Nonlinear Systems Analysis*. Prentice Hall International, Inc., 1993.

[WS05] W. Wang and J.J.E. Slotine. On Partial Contraction analysis for coupled nonlinear oscillators. *Biological Cybernetics*, 2005.

[XS11] J. Xiong and Z. Sun. Approximation of extreme measure for switched linear systems. *IEEE International Conference on Control and Automation*, 2011.

[Zah03] Z. Zahreddine. Matrix measure and application to stability of matrices and interval dynamical systems. *International Journal of Mathematics and Mathematical Sciences*, 2003.

License