

---

# Identifikation, Analyse und Regelung nichtlinearer dynamischer Systeme mittels des Koopman-Operators, künstlicher neuronaler Netze und linearer Zustandsregler

---



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

Vom Fachbereich Elektrotechnik und Informationstechnik  
der Technischen Universität Darmstadt zur Erlangung  
des Grades eines Doktor-Ingenieurs (Dr.-Ing.) genehmigte Dissertation

von

**Marcel Bonnert, M.Sc.**

geboren am 25.03.1990 in Langen

Referent: Prof. Dr.-Ing. Ulrich Konigorski  
Korreferent: Prof. Dr.-Ing. Dr. h.c. Rolf Isermann

Tag der Einreichung: 24. Juni 2021  
Tag der Prüfung: 8. November 2021

D 17 · Darmstadt 2021

---

Genehmigte Dissertation von Marcel Bonnert, M.Sc.:

Identifikation, Analyse und Regelung nichtlinearer dynamischer Systeme mittels des Koopman-Operators, künstlicher neuronaler Netze und linearer Zustandsregler

D 17 · Darmstadt 2021

Darmstadt, Technische Universität Darmstadt,  
Tag der mündlichen Prüfung: 08.11.2021

Jahr der Veröffentlichung der Dissertation auf TUprints: 2022

Bitte zitieren Sie dieses Dokument als:

URN: urn:nbn:de:tuda-tuprints-19959

Dieses Dokument wird bereitgestellt von tuprints,

E-Publishing-Service der TU Darmstadt

<http://tuprints.ulb.tu-darmstadt.de>

[tuprints@ulb.tu-darmstadt.de](mailto:tuprints@ulb.tu-darmstadt.de)



Die Veröffentlichung steht unter folgender Creative-Commons-Lizenz:  
Namensnennung – Weitergabe unter gleichen Bedingungen 4.0 International  
<https://creativecommons.org/licenses/by-sa/4.0/>

*Für Heike, Peter  
und Ricarda*

---

# Vorwort

Die vorliegende Arbeit entstand während meiner Zeit als wissenschaftlicher Mitarbeiter am Fachgebiet für Regelungstechnik und Mechatronik des Instituts für Automatisierungstechnik der Technischen Universität Darmstadt.

Zum Erfolg dieser Arbeit trugen viele Menschen bei, bei denen ich mich an dieser Stelle von Herzen bedanken möchte. Ein besonderer Dank geht dabei an meinen Doktorvater Herrn Prof. Dr.-Ing. Ulrich Konigorski für die Unterstützung bei der Themenfindung, für seine wissenschaftliche Begleitung sowie für eine Vielzahl professioneller und inspirierender Ratschläge. Nur durch die von ihm gewährten Freiheiten konnte ich meine Forschung ganz nach meinem Interesse ausrichten. Außerdem danke ich Prof. Dr.-Ing. Dr. h.c. Rolf Isermann für die Übernahme der Korreferats meiner Arbeit.

Danken möchte ich ebenso den wissenschaftlichen Mitarbeitern des Fachgebietes für Regelungstechnik und Mechatronik für viele wertvolle Diskussionen und aufschlussreiche Gespräche, die für die Entwicklung und Fertigstellung meiner Dissertation von großer Bedeutung waren. Mein besonderer Dank gilt Dr.-Ing. Eric Lenz, der mich durch seine fachliche Kompetenz in der Systemidentifikation besonders unterstützt hat. Auch gilt mein Dank für die vielen Studierenden, die mit ihren Abschluss- und Projektarbeiten meine Ideen aufgriffen und zum Teil weiterentwickelten. Insbesondere ist hier Herr Alexander Stephan Gräfe zu nennen, der mich im Rahmen seiner Masterthesis besonders unterstützen konnte.

Auch möchte ich den Eltern meiner Frau Ricarda für ihre Hilfe und Unterstützung und das mühevollen Korrekturlesen danken sowie allen Freunden und Verwandten, die mir auch in schwierigeren Phasen der Promotion immer zur Seite standen. Abschließend danke ich meiner Frau Ricarda für ihren emotionalen Rückhalt und ihre Ermutigungen sowie meinen Eltern Heike und Peter, die mich während meiner gesamten Studienzeit fürsorglich begleitet und mich in jeder Lebensphase bedingungslos unterstützt haben.

Herborn, Dezember 2021

*Marcel Bonnert*

# Inhaltsverzeichnis

|  |            |
|--|------------|
| <b>Symbole und Abkürzungen</b>   | <b>VII</b> |
| <b>Zusammenfassung/Abstract</b>  | <b>XI</b>  |
| <b>1 Einführung</b>  | <b>1</b>   |
| 1.1 Motivation . . . . .   | 1          |
| 1.2 Literaturübersicht . . . . .   | 4          |
| 1.3 Abgrenzung gegenüber vorangegangenen Arbeiten . . . . .  | 11         |
| 1.4 Struktur der Arbeit . . . . .  | 11         |
| <b>2 Grundlagen</b>  | <b>13</b>  |
| 2.1 Nomenklatur . . . . .  | 13         |
| 2.2 Mathematische Grundlagen . . . . .   | 13         |
| 2.2.1 Normierung . . . . .   | 13         |
| 2.2.2 Stellenergie . . . . .   | 14         |
| 2.2.3 Operatoren, Eigenfunktionen und Spektrum . . . . .   | 15         |
| 2.2.4 Prädiktionsinvarianz eines Unterraumes zum Koopman-Operator . . . . .                                    | 17         |
| 2.3 Identifikationsverfahren . . . . .   | 19         |
| 2.3.1 Klassische Verfahren . . . . .   | 19         |
| 2.3.2 Künstliche neuronale Netze . . . . .   | 22         |
| 2.3.3 Modellvalidierung . . . . .  | 24         |
| <b>3 Der Koopman-Operator</b>  | <b>25</b>  |
| 3.1 Koopman-Operator autonomer Zustandsraumsysteme . . . . .   | 25         |
| 3.1.1 Zusammenhang zwischen der Stabilität nichtlinearer Systeme und ihren Koopman-linearen Systemen . . . . . | 31         |
| 3.1.2 Weitere Eigenschaften Koopman-linearisierbarer Systeme . . . . .   | 36         |
| 3.2 Koopman-Operator angeregter Zustandsraumsysteme . . . . .  | 37         |
| 3.2.1 Eigenschaften angeregter Koopman-linearisierbarer Systeme . . . . .                                      | 41         |
| 3.3 Systemdynamische Eigenschaften eines Koopman-linearen Systems . . . . .                                    | 46         |
| 3.3.1 Beobachtbarkeit des Koopman-linearen Systems . . . . .   | 46         |
| 3.3.2 Steuerbarkeit des Koopman-linearen Systems . . . . .   | 47         |
| 3.4 Regelung nichtlinearer Systeme mittels Koopman-linearen Systemen . . . . .                                 | 49         |
| 3.4.1 Wahl der Gewichtungsmatrizen bei LQ-Regelung für Koopman-lineare Systeme . . . . .                       | 52         |
| 3.5 Identifizierbarkeit eines Koopman-linearen Systems . . . . .   | 54         |
| 3.5.1 Schätzung bei nicht vollständig gemessenem Zustand . . . . .   | 56         |

---

|               |  |            |
|---------------|--|------------|
| <b>4</b>      | <b>Schätzung Koopman-linearer Zustandsraumsysteme</b>  | <b>59</b>  |
| 4.1           | Deep Koopman Operator Learning with control (DKLc)   | 59         |
| 4.1.1         | Auswahl des Optimierers  | 63         |
| 4.1.2         | Trainingsphilosophien  | 64         |
| 4.1.3         | Wahl der Batch Size  | 66         |
| 4.1.4         | Umsetzung der Koopman-Zustandsrückführung mit DKLc   | 66         |
| 4.1.5         | Wahl der Gewichtungsmatrizen bei LQ-Zustandsregelung   | 68         |
| 4.1.6         | Einfluss von Mess- und Schätzungenauigkeiten auf die DKLc-Methode  | 69         |
| 4.2           | Abschätzung der Systemordnung eines Koopman-linearen Systems   | 70         |
| 4.3           | Methodik zur Identifikation, Analyse und Regelung nichtlinearer dynamischer Systeme  | 72         |
| <b>5</b>      | <b>Analyse und Regelung dynamischer Systeme</b>  | <b>73</b>  |
| 5.1           | Numerische Beispiele   | 74         |
| 5.1.1         | Van der Pol Oszillator   | 74         |
| 5.1.2         | Duffing Oszillator   | 94         |
| 5.2           | Drei-Tank Prüfstand  | 106        |
| 5.3           | Zusammenfassende Bewertung der Beispiele   | 117        |
| <b>6</b>      | <b>Zusammenfassung und Ausblick</b>  | <b>121</b> |
| <b>Anhang</b> |  | <b>124</b> |
| A.1           | Herleitungen   | 124        |
| A.1.1         | Herleitung der eigenwertabhängigen Gleichung zur Zeitlösung eines linearen Zustandsraumsystems   | 124        |
| A.1.2         | Stabilität von nichtlinearen Systemen und ihren exakten Koopman-linearen Äquivalenten  | 125        |
| A.1.3         | Herleitung der zeitdiskreten Gleichung des Koopman-Linearen Systems auf Basis von nichtlinearen Systemen mit Eingangsnichtlinearitäten | 126        |
| A.1.4         | Herleitung des Gradienten von Observables für DKLc   | 127        |
| A.1.5         | Exaktes Minimierungsmaß der Netzprädiktion   | 128        |
| A.1.6         | LMI Formulierung der zeitdiskreten Lyapunov-Gleichung  | 129        |
| A.1.7         | Existenz eines Integrators bei mittelwertfreier Anregung   | 129        |
| A.1.8         | Nachweis der schwachen Beobachtbarkeit des Drei-Tank Prüfstandes   | 130        |
| A.2           | Hardware Aufbau des Drei-Tank Prüfstandes  | 132        |
| A.3           | Weiterführende Plots und Modellkonfigurationen   | 134        |
| A.3.1         | Amplituden Verlauf der dominanten Frequenzen des mathematischen Pendels  | 134        |
| A.3.2         | Batch-Size und Bias-Varianz-Analyse von DKLc   | 134        |
| A.3.3         | Van der Pol Oszillator   | 135        |
| A.3.4         | Duffing Oszillator   | 135        |
| A.3.5         | Drei-Tank  | 136        |
|               | <b>Literaturverzeichnis</b>  | <b>141</b> |

# Symbole und Abkürzungen

## Symbole

|                    |   |
|--------------------|---|
| $A$                | Querschnittsfläche ( $\text{m}^2$ )   |
| $f$                | Frequenz ( $\text{Hz} \hat{=} 1/\text{s}$ )   |
| $g$                | Erdbeschleunigung ( $\text{m}/\text{s}^2$ )   |
| $h$                | Füllhöhe (m)  |
| $q$                | Querschnittsfläche ( $\text{m}^2$ )   |
| $t$                | Zeit (s)  |
| $T$                | Endzeit eines Intervalls (s)  |
| $T_s$              | Abtastzeit (s)  |
| $U_{C/s}$          | Steuer-/Sensorspannung (V)  |
| $\dot{V}$          | Volumenstrom ( $\text{m}^3/\text{s}$ )  |
| $a$                | Koeffizient des Nennerpolynoms einer Übertragungsmatrix / allgemeiner Koeffizient   |
| $\underline{a}$    | Vektor der Koeffizienten des Nennerpolynoms einer Übertragungsmatrix  |
| $\underline{A}$    | Allgemeine Beispielmatrix / Zustandsmatrix eines linearen Zustandsraumsystems   |
| $b$                | Koeffizient des Zählerpolynoms einer Übertragungsmatrix / Batch Size  |
| $\underline{B}$    | Matrix mit Koeffizienten der Zählerpolynome einer Übertragungsmatrix /<br>Allgemeine Beispielmatrix / Eingangsmatrix eines linearen Zustandsraumsystems |
| $\underline{b}$    | Skalierung der Datennormierung / Bias-Vektor der Schichten künstlicher neuronaler Netze   |
| $c$                | Time-Shift Operator der $\mathcal{Z}$ -Transformation   |
| $\underline{c}$    | Dynamik des Eingangs  |
| $\mathcal{C}^n$    | Raum der $n$ -fach differenzierbaren (Vektor-)Funktionen  |
| $\underline{C}$    | Ausgangsmatrix eines linearen Zustandsraumsystems / Reglermatrix  |
| $\underline{D}$    | Durchgriffsmatrix eines linearen Zustandsraumsystems  |
| $\mathcal{D}^n$    | Raum der $n$ -fach differenzierbaren Diffeomorphismen   |
| $e$                | Trainingsepochen  |
| $\underline{e}_i$  | $i$ -ter Einheitsvektor / Fehlervektor  |
| $E$                | Erwartungswert  |
| $E_u$              | Stellenergie  |
| $f$                | allgemeine Funktion   |
| $f_{\text{PBSID}}$ | Zukunftshorizont ( <i>future</i> ) der Subspace-Identifikation  |
| $\underline{f}$    | Vektor einer nichtlinearen Dynamik  |
| $\underline{F}$    | Matrix einer nichtlinearen Dynamik  |
| $g$                | Allgemeine Koopman-Observables  |
| $g_w$              | Einzelne Koopman-Observables des Eingangs   |

---

|                        |  |
|------------------------|--|
| $\underline{g}_w$      | Koopman-Observables des Eingangs   |
| $\underline{g}_z$      | Einzelne Koopman-Observables des Zustandes   |
| $\underline{g}_z$      | Koopman-Observables des Zustandes  |
| $\underline{g}_w^{-1}$ | Approximierte Inverse der Koopman-Observables des Eingangs                                 |
| $\underline{g}_z^{-1}$ | Approximierte Inverse der Koopman-Observables des Zustandes                                |
| $\underline{G}$        | Matrix einer Cholesky-Zerlegung  |
| $h$                    | Schrittweite / Ausgangsfunktion eines Zustandsraumsystems                                  |
| $\underline{h}$        | Ausgangsvektorfunktion eines Zustandsraumsystems   |
| $\underline{H}$        | Matrix von Systemausgängen zur Beschreibung einer Einbettung                               |
| $i$                    | Zähler einer Iteration (Iterator)  |
| $\underline{I}_x$      | Einheitsmatrix der Dimension $x$   |
| $j$                    | Zähler einer Iteration (Iterator) / Komplexe Zahl  |
| $J$                    | Verlustfunktion / Verlustfunktional  |
| $k$                    | Zeitschritt / Zähler einer Iteration (Iterator)  |
| $K$                    | (Mathematischer) Körper  |
| $\underline{K}$        | Kalman-Filter Matrix   |
| $\underline{K}_z$      | Koopman Eingangsmatrix   |
| $\underline{K}_z$      | Koopman Zustandsmatrix   |
| $\mathcal{K}$          | Koopman-Operator   |
| $m$                    | Eingangsordnung / Mittelwert   |
| $m_e$                  | erweiterte Eingangsordnung   |
| $n$                    | Systemordnung  |
| $n_B$                  | Anzahl von <u>B</u> atches eines Datensatzes   |
| $n_e$                  | erweiterte Systemordnung   |
| $n_r$                  | reduzierte Systemordnung   |
| $N$                    | Anzahl von Messpunkten   |
| $N_P$                  | Anzahl von <u>P</u> arametern  |
| $N_T$                  | Anzahl von <u>T</u> rainingsschritten eines künstlichen neuronalen Netzes                  |
| $\underline{O}$        | Beobachtbarkeits-/Identifizierbarkeitsmatrix   |
| $p$                    | Prädiktionshorizont  |
| $p_x$                  | Vergangenheitshorizont ( <i>past</i> ) der $x = \text{PBSID/OKID/DKLC}$ -Identifikation    |
| $\underline{P}$        | Lösung der Lyapunov-Gleichung  |
| $q$                    | Anzahl der Messgrößen / Eintrag in der Gewichtungsmatrix $\underline{Q}$ der LQ-Regelung   |
| $\underline{Q}$        | Zustandsgewichtungsmatrix der LQ-Regelung / Matrix für eine quadratische Lyapunov-Funktion |
| $r$                    | zusätzliche Ordnung / Anzahl nicht steuerbarer Eigenwerte                                  |
| $\underline{R}$        | Eingangsgewichtungsmatrix der LQ-Regelung  |
| $s$                    | Ergebnis einer Supremumsnorm / komplexe Frequenz   |
| $\underline{S}$        | Lösung einer nichtlinearen Differential-/Differenzgleichung (engl. <u>S</u> olution)       |
| $\underline{T}$        | Transformationsmatrix  |
| $\underline{u}$        | Vektor der Eingänge  |
| $\underline{U}$        | Basisvektoren der Singulärwertzerlegung  |

---

|                       |  |
|-----------------------|--|
| $\mathcal{U}$         | Untervektorraum / Eingangsraum   |
| $v$                   | Varianz  |
| $\underline{v}$       | Rechts-Eigenvektor / Additives Eingangssignal                                      |
| $\mathcal{V}$         | Vektorraum   |
| $V$                   | Lyapunov-Funktion / Rechts-Eigenvektormatrix                                       |
| $\underline{V}$       | Matrix der Singulärwertzerlegung   |
| $\underline{w}$       | Links-Eigenvektor / Eingangsvektor eines Koopman-linearen Systems                  |
| $W$                   | Links-Eigenvektormatrix / Schichtgewicht künstlicher neuronaler Netze              |
| $x$                   | Systemzustand / allgemeine Variable  |
| $\underline{x}$       | Zustandsvektor   |
| $\mathcal{X}$         | Zustandsraum   |
| $\underline{y}$       | Datenvektor / Ausgang einer verdeckten Schicht eines künstlichen neuronalen Netzes |
| $\underline{Y}$       | Datenmatrix der Ausgänge   |
| $z$                   | Koopman Zustand / Time-Shift Operator der $\mathcal{Z}$ -Transformation            |
| $\underline{z}$       | Koopman Zustandsvektor   |
| $\alpha$              | Größe zur Stabilitätsbewertung / Koeffizient einer Verlustfunktion                 |
| $\beta$               | Koeffizient  |
| $\delta$              | Vergessensfaktor / Koeffizient   |
| $\epsilon$            | Fehler   |
| $\eta$                | Regularisierungsterm der <i>Information Criteria</i> Optimierung                   |
| $\gamma$              | Größe zur Stabilitätsbewertung / Koeffizient einer Verlustfunktion                 |
| $\kappa$              | Koeffizient  |
| $\lambda$             | Eigenwert  |
| $\underline{\Lambda}$ | Eigenwertmatrix  |
| $\mu$                 | Koeffizient  |
| $\omega$              | Eigen-/Kreisfrequenz (rad/s)   |
| $\varphi$             | Eigenfunktion  |
| $\sigma$              | Standardabweichung ( $\sigma^2$ : Varianz)   |
| $\sigma(\mathcal{O})$ | Spektrum eines Operators $\mathcal{O}$   |
| $\underline{\Sigma}$  | Matrix mit Singulärwerten  |
| $\underline{\theta}$  | Vektor mit trainierbaren Koeffizienten   |
| $\tau$                | Zeit (s)   |
| $\xi$                 | Steuerbarkeitsmaß  |
| $\mathcal{Z}$         | Koopman-Zustandsraum   |

### Abkürzungen

|      |  |
|------|--|
| AICc | <i>corrected Akaike's Information Criterion</i>                |
| ADAM | <i>ADaptive Moment estimation</i>                              |
| ANN  | <i>Artificial Neural Network</i> (künstliches neuronales Netz) |
| AP   | <i>ArbeitsPunkt</i>  |
| ARX  | <i>Auto-Regressive with eXogenous inputs</i>                   |
| CPU  | Prozessor (engl. <i>Central Processing Unit</i> )              |

|                      |  |
|----------------------|--|
| DKLc                 | <i><u>D</u>ee<u>p</u> <u>K</u>oopman-<u>O</u>perator <u>L</u>earning with <u>C</u>ontrol</i>       |
| DMD                  | <i><u>D</u>ynamic <u>M</u>ode <u>D</u>ecomposition</i>   |
| eDMD                 | <i><u>e</u>xtended <u>D</u>ynamic <u>M</u>ode <u>D</u>ecomposition</i>                             |
| EHK                  | <i><u>E</u>in<u>H</u>eits<u>K</u>reis</i>  |
| ERA                  | <i><u>E</u>igensystem <u>R</u>ealization <u>A</u>lgorithm</i>                                      |
| FFT                  | <i><u>F</u>ast <u>F</u>ourier <u>T</u>ransform</i>   |
| GPU                  | Grafik-Prozessor (engl. <i><u>G</u>raphics <u>P</u>rocessing <u>U</u>nit</i> )                     |
| IC                   | Informations Kriterium (engl. <i><u>I</u>nformation <u>C</u>riteria</i> )                          |
| LASSO                | <i><u>L</u>east <u>A</u>bsolute <u>S</u>hrinkage and <u>S</u>election <u>O</u>perator</i>          |
| Lin                  | Linear   |
| LQ                   | <i><u>L</u>inear <u>Q</u>uadratisch</i>  |
| LRAN                 | <i><u>L</u>inear <u>R</u>ecurrent <u>A</u>utoencoder <u>N</u>etwork</i>                            |
| MIMO                 | <i><u>M</u>ultiple <u>I</u>nput <u>M</u>ultiple <u>O</u>utput</i>                                  |
| MLP                  | <i><u>M</u>ulti<u>L</u>ayer <u>P</u>erceptron</i>  |
| MPC                  | Modellprädiktive Regelung (engl. <i><u>M</u>odel <u>P</u>redictive <u>C</u>ontrol</i> )            |
| MRSE                 | <i><u>M</u>ean <u>R</u>elative <u>S</u>quared <u>E</u>rror</i>                                     |
| MSE                  | Mittlerer quadratischer Fehler (engl. <i><u>M</u>ean <u>S</u>quared <u>E</u>rror</i> )             |
| OKID                 | <i><u>O</u>bserver <u>K</u>alman-<u>F</u>ilter <u>S</u>ubspace <u>I</u>dentifikation</i>           |
| PBSID <sub>opt</sub> | <i><u>o</u>ptimized <u>P</u>redictor <u>B</u>ased <u>I</u>dentification</i>                        |
| VTOL                 | <i><u>V</u>ertical <u>T</u>ake-<u>O</u>ff and <u>L</u>anding</i>                                   |
| SIMO                 | <i><u>S</u>ingle <u>I</u>nput <u>M</u>ultiple <u>O</u>utput</i>                                    |
| SINDYc               | <i><u>S</u>parse <u>I</u>dentification of <u>N</u>onlinear <u>D</u>ynamics with <u>c</u>ontrol</i> |

---

# Zusammenfassung

Durch immer komplexer werdende technische Systeme und dem damit einher gehenden Bedarf an passenden Regelungsmethoden stellt sich die Frage, ob bestehende Methoden für die Identifikation, Analyse und Regelung von linearen Systemen in geeigneter Weise auch für nichtlineare Systeme anwendbar sind. Ein bekannter Ansatz hierfür ist die Linearisierung einer bekannten Systemgleichung oder die Identifikation eines nichtlinearen dynamischen Systems auf Basis eines solchen linearen Ansatzes. Eine solche Identifikation stellt dabei häufig eine große Herausforderung dar. Dies gilt insbesondere dann, wenn Daten aus dem Betrieb genutzt werden, ohne dass ein besonderes Augenmerk auf eine geeignete Anregung gelegt werden kann. Damit einher geht häufig, dass Daten aus geschlossenen Regelkreisen zu Problemen bei der Identifikation führen können. Eine eingehende Analyse des Systemverhaltens oder eine Regelung auf Basis eines Modells, welches das System nicht ausreichend gut, oder nur in einem sehr begrenzten Bereich des Zustandsraumes, wiedergibt, sind damit praktisch unmöglich.

Daher beschäftigt sich die vorliegende Arbeit mit dem sogenannten Koopman-Operator und dessen Identifikation und Nutzung zur Analyse und Regelung nichtlinearer Systeme. Dieser Operator ist ein mathematisches Konstrukt, welches es möglich macht, bestimmte Funktionen der Systemzustände eines nichtlinearen dynamischen Systems - sogenannte Observables - über diesen Operator linear mit der Zeit zu entwickeln. So ist es möglich, theoretisch fundiert lineare Konzepte wie Eigenbewegungen und lineare Analyse- und Regelungsmethoden auf nichtlineare Systeme zu übertragen. Insbesondere für Systeme mit Anregung können lineare Reglerentwurfsmethoden so auf nichtlineare Systeme angewendet werden, ohne starke Einschränkungen hinnehmen zu müssen, wie sie häufig Teil von nichtlinearen Regelungsmethoden sind. So müssen bei diesen nichtlinearen Entwurfsmethoden beispielsweise gewisse Bereiche des Zustandsraumes ausgeschlossen werden oder die Systemgleichung muss in einer speziellen Form vorliegen.

Durch den linearen Charakter des Koopman-Operators ist es auch möglich, Konzepte wie die Stabilität, welche bei linearen Modellen sehr einfach anhand der Eigenwerte der Systemmatrix bestimmt werden kann, auf nichtlineare Systeme zu übertragen. Hierfür werden in der vorliegenden Arbeit Konzepte für zeitdiskrete Modelle erarbeitet, wie sie häufig identifiziert werden. So kann anschaulich von der Stabilität des Koopman-linearen Systems auf die Stabilität des nichtlinearen Systems geschlossen werden. Damit ist es außerdem möglich, eine Stabilitätsgarantie für nichtlineare Systeme anzugeben, welche über einen Zustandsregler auf Basis dieses Koopman-linearen Systems berechnet werden. Die in der vorliegenden Arbeit betrachteten Nichtlinearitäten sind dabei ausschließlich kontinuierlich. So sind beispielsweise Systeme mit Hysteresen nicht Teil der Betrachtungen.

Für die Regelung ist die Steuerbarkeit des Prozesses von besonderer Bedeutung. Auch für diese wird in dieser Arbeit eine Eigenschaft von Koopman-linearen Systemen erarbeitet und für die

Regelung genutzt. Bei der Regelung wird sich in dieser Arbeit auf den linear quadratischen Ansatz beschränkt. Hierfür müssen gewisse Gewichtungen gewählt werden, deren anschauliche Wahl durch die Observables stark erschwert wird. Um die direkte Interpretierbarkeit dieser Gewichtungen auf den Koopman-Operator und seine Observables zu übertragen, wird eine neue Methode vorgeschlagen, wie diese Gewichte im Koopman-Raum gewählt werden können.

Die nun mehrfach genannten Observables sind jedoch meist vollständig unbekannte Funktionen und sind häufig nur für hinführende Beispiele ohne konkrete praktische Relevanz herzuleiten oder bekannt. Um dennoch solche Systeme zu schätzen, werden in dieser Arbeit künstliche neuronale Netze verwendet. Diese sind in ihrer Gestalt als allgemeine Funktionsapproximatoren für den vorliegenden Anwendungsfall prädestiniert. Es wird somit ein Netz erstellt, welches zunächst Observables berechnet und damit den Zustand kodiert. Diese Observables werden im Anschluss mittels eines linearen Zustandsraumsystems mit der Zeit entwickelt und abschließend dekodiert.

Diese Konzepte werden dann dafür genutzt, zwei mathematische Systeme, den Van der Pol und den Duffing Oszillator, sowie einen realen Drei-Tank Prüfstand zunächst zu schätzen, im Anschluss anhand der identifizierten Modelle zu analysieren und die Ergebnisse zu plausibilisieren und im Nachgang zu regeln. Als Erweiterung der erstellten Konzepte wird dieselbe Methodik für den Fall, dass nicht alle Zustände gemessen werden können, auf die die genannten Systeme angewendet.

# Abstract

Since technical systems are getting more and more complex, suitable control methods are required. To analyze such nonlinear systems, one may ask the question how already existing methods for identification, analysis and control can be applied to them. A well-known approach is the linearization of a known differential equation or an identification of the nonlinear dynamical system referring to the assumption, that a linear or linearized system fits the data. Such an identification is a great challenge when using data from a real-world plant or from closed loop environments. Therefore, a deep analysis or the computation of a controller, based on a model that cannot replicate the systems behavior or a model that is highly restricted to small and certain parts of the state-space, is practically impossible.

Hence, the present work uses the so called Koopman Operator and its identification to analyze and control nonlinear systems. With this mathematical operator it is possible to describe the behavior of nonlinear systems in terms functions of the state vector - so called observables - but linearly. Thus, linear concepts like modes and methods for linear analysis can be used for nonlinear systems. Furthermore, linear control methods can be used for excited nonlinear dynamical systems. In addition, this can be important since many nonlinear control methods rely on partially harsh restrictions. For example, some areas of the state space must be excluded, or the systems equation needs to be in a very specific structure.

Speaking of the operator's linear character, concepts like the stability, which is in the linear case easily characterized by the eigenvalues of the linear dynamical system, can be transferred to nonlinear systems. To do so, the present work gives the reader theorems for time discrete systems, since most of the identification methods rely on such systems, to guarantee stability as long as a certain error bound is adhered. With those theorems it is further possible to guarantee stability if the nonlinear system is controlled with a state controller based on an identified Koopman linear system. It is important to mention that only continuous nonlinearities are treated in the present work. That means that for example hysteresis are excluded throughout the study.

For the control of (nonlinear) systems the controllability is of major importance. Additionally, this work gives a theorem of the controllability of Koopman linear systems and uses it to acquired linear state controllers based on identified Koopman linear systems. For the linear quadratic controllers, used in the present work, certain weightings must be chosen, which is difficult due to the aforementioned observables. To transfer the interpretability of the weights from the statespace to the Koopman space a method is proposed.

The observable-functions are mostly completely unknown and only known for educational examples. To still estimate such systems, artificial neural networks are used in the presented work. Since those neural networks are universal function approximators, they are an excellent choice to

obtain Koopman linear systems. Thus, a network is created that estimates observables in parallel to its inverse and a Koopman linear statespace system.

All those concepts are used to estimate, analyze and control two theoretical examples, the Van der Pol and the Duffing Oscillator, and additionally a real world three-tank hydraulic plant. In addition, a vast part of the methods and concepts are extended and tested, using the aforementioned examples, to the case when only a part of the states can be measured.

# 1 Einführung

Zur Einführung in diese Arbeit wird zunächst die Fragestellung motiviert und eine umfassende Literaturübersicht über das noch junge Thema des Koopman-Operators gegeben.

## 1.1 Motivation

Moderne und immer komplexer werdende Systeme verlangen im gleichen Maße nach immer komplexeren Regelungsstrategien. Jedoch ist der aktuelle Stand der Regelungstheorie der, dass häufig nur für sehr spezielle Systeme beziehungsweise Systemstrukturen Regelungsalgorithmen existieren. Für Methoden die unabhängig von diesen speziellen Eigenschaften funktionieren, müssen andere, teilweise erhebliche, Einschränkungen hingenommen werden [2].

Ein anschauliches Beispiel hierfür ist die exakte Linearisierung. Sie kompensiert die Nichtlinearität des Systems und prägt ihm ein lineares Verhalten auf. Dabei ist es zum einen notwendig, dass das System exakt in Form einer Differentialgleichung vorliegt, da die Nichtlinearität mathematisch kompensiert wird. Zum anderen müssen häufig gewisse Teile des Zustandsraumes ausgeschlossen werden, damit die Regelungsstrategie mathematisch anwendbar ist [2].

Ein anderer bekannter Ansatz ist die sogenannte Modellprädiktive Regelung (MPC). Hierbei wird das Stellsignal sukzessive anhand eines Optimierungsproblems ermittelt. Dies hat jedoch insbesondere im Onlinebetrieb den Nachteil, dass die ausführende Hardware diese Optimierung durchführen können muss. Dadurch ist der Einsatz teurerer Computerkomponenten unabdingbar, was an vielen Stellen jedoch nicht möglich oder nicht wirtschaftlich ist.

Die Theorie von Regelungen linearer Systeme besitzt auf der anderen Seite keinerlei Einschränkungen, solange das System in linearer Form vorliegt. Dies ist jedoch in praktischer Hinsicht eine äußerst große Einschränkung, da in realen Systemen in den meisten Fällen nichtlineare Dynamikanteile, wie beispielsweise Sättigungen oder andere nichtlineare Abhängigkeiten von Systemgrößen, vorhanden sind.

Aufgrund des nichtlinearen Verhaltens, aber den attraktiven Vorteilen der linearen Regelungstheorie, verfolgt man in der Anwendung daher häufig den Weg über die Linearisierung nichtlinearer Systeme. Dazu muss das System wiederum als Differential- oder Differenzgleichung vorliegen, wie sie über eine theoretische Modellbildung ermittelt werden kann. Hierbei linearisiert man die Systemgleichung um verschiedene Arbeitspunkte, die nachfolgend, beispielsweise mittels einer Gain-Scheduling-Strategie, verkoppelt werden [2]. Alternativ kann ein lineares Ersatzsystem in Form einer Zustandsraumdarstellung, einer Übertragungsfunktion oder in Form von nichtparametrischen Modellen identifiziert werden, die dann aber nur für einen kleinen Bereich im Zustandsraum gelten. Dieser Bereich hängt stark davon ab, in wie weit die Nichtlinearitäten ausgeprägt

sind. Diese Eigenschaft führt jedoch bereits bei einfachen niederdimensionalen Systemen, wie beispielsweise einem direkt aktuierten einfachen Pendel, dazu, dass man eine erhebliche Anzahl an Arbeitspunkten benötigt, um das System adäquat im gesamten Zustandsraum zu regeln. Dies hängt unmittelbar mit dem Spektrum nichtlinearer Systeme zusammen, welches im Zusammenhang mit dem sogenannten Koopman-Operator, dem Thema der vorliegenden Arbeit, definiert ist [57, 58]. Hiermit ist es möglich, lineare Konzepte, wie beispielsweise ein lineares Spektrum (welches mit linearen Operatoren wie Matrizen verbunden ist [62]), nichtlinearen Systemen zuzuordnen. Ist dieses kontinuierlich, sodass die Dynamik des Systems sehr stark vom aktuellen Zustand beziehungsweise dem Anfangszustand abhängt, werden mehr Arbeitspunkte benötigt als es bei einem Punktspektrum, wie es zum Beispiel bei einem linearen System vorliegt, notwendig wäre [69].

Eine alternative Annäherung an diese Problematik stellt die unmittelbare Nutzung der bereits oben genannten Theorie des Koopman-Operators dar. Sie besagt, dass zu jedem nichtlinearen System ein linearer Operator existiert, der das Systemverhalten linear in einem unendlichdimensionalen Hilbertraum beschreibt [57, 58]. Die Basis dieses Hilbertraums sind die Eigenfunktionen des Koopman-Operators, die wiederum Funktionen des Zustandsvektors sind. Über Linearkombinationen der Eigenfunktionen lassen sich sogenannte *Observables* bilden. Diese Eigenfunktionen oder Observables zu finden, ist jedoch häufig keine einfache Aufgabe. Dieser unendlichdimensionale Operator ist anschaulicher Weise nicht direkt für die Regelung anwendbar. Zum einen durch die unendliche Dimension und zum anderen aufgrund der Unbekanntheit der Eigenfunktionen oder Observables. Man sucht bei der Anwendung dieser Theorie daher nach einem zum Koopman-Operator (näherungsweise) invarianten Unterraum in Form einer Systemmatrix verbunden mit Observables, die als neue Zustände aufgefasst werden.

Moderne Ansätze eine Schätzung für den Koopman-Operator zu ermitteln verfolgen daher häufig den Weg, unter einer Vielzahl von Observable-Kandidaten die besten auszuwählen. Dies birgt jedoch den großen Nachteil, dass wenn wichtige Eigenfunktionen oder Observables nicht in dieser Bibliothek von Kandidaten vorhanden sind, das gesamte Identifikationsergebnis unbrauchbar wird. Im besten Falle sollten diejenigen Parameter, welche zu den nicht geeigneten Observables gehören, zu Null geschätzt werden. Durch Nichtlinearitäten und Rauschen ist dies jedoch in der Anwendung häufig nicht der Fall. Ein Vorteil dieser Herangehensweise ist jedoch, dass die Systemordnung im vornherein nicht feststeht und diese sich im Laufe des Algorithmus ergibt.

Eine weitere Möglichkeit, näherungsweise invariante Unterräume des Koopman-Operators nichtlinearer dynamischer Systeme zu finden, ist der Einsatz künstlicher neuronaler Netze, beziehungsweise, im Falle tiefer Netze, *Deep-Learning*. Hierbei setzt man eine Systemordnung fest und schätzt Observables parallel zu einer Systemmatrix, indem man spezielle Netzstrukturen einsetzt. Da (tiefe) künstliche neuronale Netze mit nichtkonstanter Aktivierungsfunktion jede messbare Funktion beliebig genau annähern können ([48]), ist ihr Einsatz an dieser Stelle äußerst sinnvoll. Insbesondere deshalb, da es sich bei den Observables meist um unbekannt Funktionen handelt. Der Nachteil hierbei ist allerdings, dass die Systemordnung festgelegt werden muss.

Für eine Schätzung einer (linearen) Systemordnung stehen eine Vielzahl von Verfahren zur Verfügung. Die am weitesten verbreitete Methode sind die sogenannten *Information Criteria*, wobei das *Akaike Information Criterion* dabei das bekannteste darstellt. Diese basieren auf der Bewertung eines Modellfehlers in Form einer Abweichung zwischen den Messdaten und dem Modell und regularisieren das Bewertungsmaß über einen additiven Term, welcher die Modellkomplexität bewertet [29]. Es werden also eine Vielzahl von Modellen ermittelt und das Modell ausgewählt, welches den besten Kompromiss zwischen Prädiktionsqualität und Komplexität gewährleistet. Diese *Information Criteria* werden auch in dieser Arbeit betrachtet. Um sich einem Initialwert der Systemordnung fundierter zu nähern als sie nur zu raten, werden Subspace-Identifikationsverfahren eingesetzt. Diese liefern, neben einem Identifikationsergebnis, einen Hinweis auf die Systemordnung des hinter den Messdaten liegenden linearen Zustandsraumsystems [117]. Da in dieser Arbeit für eine solche Untersuchung zunächst alle Zustände als messbar angenommen werden und nach einem höherdimensionalen linearen System gesucht wird, bietet sich der Einsatz dieser Verfahren an. Grundsätzlich ist die Anwendung dieser Methoden jedoch auch dann denkbar, wenn nicht der gesamte Zustandsvektor gemessen werden kann oder latente Zustände im System vorhanden sind.

Im Allgemeinen ist es bei der Schätzung des Koopman-Operators nicht garantiert, dass mit einer größerer Anzahl von Observables die Qualität des Modells (quasi) monoton verbessert wird. Es ist viel mehr wichtig, die Eigenfunktionen oder Observables zu finden, die die Dynamik maßgeblich beeinflussen. Genau deshalb ist der Einsatz künstlicher neuronaler Netze, wie oben beschrieben, ein attraktives Werkzeug, da man, bis auf die Wahl der Systemordnung, strukturell keinen Einschränkungen unterliegt. Die Observables ergeben sich somit simultan mit dem linearen System und führen zu einer zu den Messdaten passenden Komposition in Form von Observables, ihren Inversen sowie dem linearen Zustandsraumsystem. Insbesondere die Inverse ist ein weiteres Problem bei Ansätzen, welche aus einer Anzahl an Ansatzfunktionen Observables auswählen, da so im Voraus nur Funktionen ausgewählt werden können, die auch über eine Umkehrfunktion verfügen.

Die Betrachtungen und der Algorithmus zur Ermittlung von näherungsweise invarianten Unterräumen des Koopman-Operators werden direkt zur Berechnung von linearen Zustandsreglern für nichtlineare Systeme eingesetzt. Die große Hoffnung bei der Anwendung der Theorie des Koopman-Operators ist, dass im besten Fall global gültige lineare Ersatzsysteme für ihre nichtlinearen Basissysteme ermittelt werden können. Bei Systemen, bei denen das nicht möglich ist, sollen zumindest weniger lineare Ersatzsysteme notwendig sein, als bei der Anwendung der klassischen Linearisierung durch Taylorreihenentwicklung oder einfachen linearen Identifikationsansätzen.

Der Beitrag dieser Arbeit soll daher sein, eine Methode vorzustellen mit der Koopman-lineare Systeme geschätzt werden können. Des Weiteren sollen verschiedene Eigenschaften des Koopman-Operators angeregter nichtlinearer dynamischer Systeme herausgearbeitet werden. Auf Basis dieser Ergebnisse soll dann untersucht werden, in wie weit nichtlineare dynamische Systeme mit einfachen linearen Reglerentwurfverfahren geregelt werden können.

## 1.2 Literaturübersicht

Die wohl bekannteste Möglichkeit, nichtlineare Systeme linear zu beschreiben, bietet die Linearisierung. Hierzu wird die Systemgleichung mittels einer Taylorreihe um eine Ruhelage herum entwickelt und nach dem ersten, linearen Glied abgebrochen. Dieser Ansatz ist auch heute der am weitesten verbreitete, was sicherlich in der intuitiven und anschaulichen Herangehensweise begründet ist.

Auf ähnliche Weise wird bei einer Identifikation von (nichtlinearen) dynamischen Systemen vorgegangen. Auch hier werden Messdaten aus einer Umgebung eines Arbeitspunktes verwendet und es wird für diesen begrenzten Bereich ein lineares Modell ermittelt. Erste Anstrengungen im Bereich der Identifikation begründete Karl-Johan Åström and Torsten Bohlin in [9] Mitte der 1960er Jahre. Die Ergebnisse innerhalb des Themenbereichs der Systemidentifikation der darauf folgenden Jahre fasst Åström mit Pieter Eykhoff in [10] zusammen. Im Jahr 1974 verfasst Eykhoff mit [41] ein Buch zum Thema Systemidentifikation, welches die bis dahin erlangten Erkenntnisse übersichtlich aufbereitet.

In einer Vielzahl an Veröffentlichungen, unter anderem auch in denen von Eykhoff, wird die Systemidentifikation wie folgt definiert.

**Definition 1.1** (Systemidentifikation [40, 52, 125]). Die Systemidentifikation ist die experimentelle Bestimmung von zeitlich oder räumlich (im Zustandsraum) begrenztem Verhalten eines Prozesses oder eines Systems. Zur Bestimmung dieses Verhaltens werden Messungen von Signalen des Prozesses oder des Systems verwendet, um mathematische Modelle abzuleiten. Der Fehler, respektive der Unterschied, zwischen den Messungen und dem mathematischen Modell soll dabei so klein wie möglich sein.

Ein wichtiger Punkt in dieser Definition ist die (optionale) Einschränkung auf begrenzt gültige Modelle. Dies wird in dieser Arbeit von besonderer Wichtigkeit sein.

Mit [65] veröffentlicht Lennart Ljung einige Jahrzehnte nach Eykhoff im Jahr 1999 ein Standardwerk, welches eine Vielzahl von parametrischen und nichtparametrischen Verfahren beschreibt, welche bis zu diesem Zeitpunkt erarbeitet werden konnten. Auch Heuristiken zur Identifikation von Systemen sind hierin zu finden. Rolf Isermann beschreibt in [49] und [50] den gesamten Ablauf einer Systemidentifikation von der Datenerfassung bis hin zu unterschiedlichen Identifikationsmethoden für dynamische Prozesse und fasst diese in [51, S. 333-372] und in [52] zusammen. Dabei liegt der Fokus von [52] insbesondere auf der Anwendung der Konzepte an realen Anlagen und Systemen und beschreibt die Methoden in einer anschaulichen und einfach zu verstehenden Weise.

Eines der am weitesten verbreiteten Identifikationsverfahren ist das kleinste-Quadrate-Verfahren (engl. *Least-Squares*) in all seinen Ausprägungen. In der Anwendung zur Schätzung linearer, zeitinvarianter Übertragungsfunktionen führt dies zum sogenannten Gleichungsfehler und in der Folge für eine biasfreie Schätzung zum sogenannten ARX-Modell. Gemeinsam werden Least-Squares und das ARX-Modell weitläufig als ARX-Identifikation bezeichnet. Dieses Modell ist

allerdings aufgrund eines unrealistischen Rauschmodells nur begrenzt in der Realität einsetzbar. Dennoch soll sie in dieser Arbeit als Referenz eingesetzt werden, da es trotz seiner Restriktionen durch die Einfachheit vielerorts Anwendung findet.

Weniger restriktive modellgebende lineare Identifikationsverfahren, welche ebenfalls auf dem Least-Squares-Ansatz basieren, identifizieren meist lineare Zustandsraumssysteme. In den Jahren von etwa 1990 bis circa 2010 entstehen einige solcher Methoden. Die Anfänge bildeten die ersten Subspace-Verfahren, welche auf einer Projektion der Messdaten basieren [94, 116]. Alessandro Chiuso entwickelte in [33] mit seiner *Optimized Predictor Based Subspace Identification* (PBSID<sub>opt</sub>) eine Methode, welche auch in dieser Arbeit zur Ermittlung einer linearen Modellordnung verwendet wird. Unter anderem konnte durch Marcel Bonnert in [17] die Leistungsfähigkeit zur Ordnungsschätzung auch bei realen Prozessen mittels PBSID<sub>opt</sub> experimentell für eine Vielzahl an Systemen gezeigt werden.

Die *Observer Kalman Filter Identification* (OKID) ist algorithmisch ähnlich zu den Subspace-Verfahren, vereinfacht diese jedoch, da weniger Hyperparameter benötigt werden, welche durch den Anwender festzulegen sind [119]. Genau wie die PBSID<sub>opt</sub>- und ARX-Methode, wird auch die OKID in dieser Arbeit als Vergleich beziehungsweise zur Ordnungsschätzung herangezogen. Diese Ergänzung soll die Ergebnisse der Ordnungsschätzung der PBSID<sub>opt</sub>-Methode absichern und die Identifikationsergebnisse der in dieser Arbeit erarbeiteten Methode einordnen. Des Weiteren zeigt sich beispielsweise in [104], dass eine Identifikation mittels der OKID bereits bei rein mathematischen Modellen zu besseren Schätzergebnissen des Systemverhaltens führen kann, wie sie mit der PBSID<sub>opt</sub> erreicht werden können. Dies plausibilisiert die Nutzung der OKID an dieser Stelle, da in vorangegangenen Arbeiten kein direkter Vergleich der Fähigkeit zur Ordnungsschätzung zwischen OKID und PBSID<sub>opt</sub> durchgeführt wurde. Es ist somit zum jetzigen Zeitpunkt nicht klar, ob die OKID gegenüber der PBSID<sub>opt</sub> nicht nur potenziell bessere Schätzergebnisse bezüglich der Messdaten sondern gegebenenfalls auch bessere Schätzungen für die Ordnung liefern kann.

Der Einsatz nichtlinearer Subspace-Verfahren wird unter anderem in [99] durch Amirali Sadeqi et al. besprochen. Hier zeigt sich, dass die angegebenen Methoden grundsätzlich für den Einsatz bei nichtlinearen System mit einer bestimmten Struktur geeignet sind. Dies gilt insbesondere für den Teil der Ordnungsschätzung dieser Methoden, da dies der wesentliche Teil dieser Subspace-Verfahren ist, welcher in dieser Arbeit genutzt wird. Santos Borjas und Claudio Garcia analysieren in [19] den Einsatz einiger Subspace-Methoden im Kontext realer Prozesse. Auch die darin getroffenen Aussagen zur Ordnungsschätzung stellen hilfreiche Informationen für die vorliegende Arbeit dar.

Bei der Identifikation von nichtlinearen dynamischen Systemen mittels linearer Ansätze ist jedoch zu beachten, dass hierbei lediglich lokal gültige Modelle geschätzt werden können, die auf einer Linearisierung des Modells beruhen. Einen Ansatz, diese lokale Linearität auszunutzen, stellen Oliver Nelles und Rolf Isermann bereits 1996 und 1997 unter anderem in [86] und [89] vor. Dabei werden lokal gültige lineare Modelle geschätzt, die dann mittels einer gewichteten Überlagerung zusammengefasst werden, um das nichtlineare Systemverhalten zu beschreiben. Für diese Metho-

dik existiert des Weiteren eine umfassende MATLAB-Toolbox, welche in [88] beschrieben wird und kostenlos abrufbar ist. Weiter stellt Oliver Nelles mit [87] ein umfangreiches Standardwerk zur Verfügung, welches einen umfangreichen Überblick über verschiedene Identifikationsmethoden verschafft. Dazu gehört die dynamische Identifikation, wie sie in der vorliegenden Arbeit verwendet wird, aber auch eine Vielzahl weiterer Methoden zur experimentellen Modellbildung auch von nicht dynamischem Verhalten wie beispielsweise der mathematischen Beschreibung von Abhängigkeiten zwischen verschiedenen Messgrößen.

Eine völlig andere Herangehensweise nichtlineares dynamisches Verhalten zu beschreiben wurde bereits im Jahr 1931 durch Bernard Koopman und John von Neumann in [57] und [58] begründet. Sie beschreiben einen linearen Operator, welcher das Verhalten nichtlinearer dynamischer System linear beschreibt. Eine (stückweise) Linearisierung der Dynamik, wie sie oben erläutert und in vielen Algorithmen genutzt wird, ist somit nicht mehr notwendig. Der Nachteil an dieser Methodik ist unter anderem, dass dieser Operator von unendlicher Dimension ist.

Der Vollständigkeit halber sei der zum Koopman-Operator adjungierte Perron-Frobenius Operator erwähnt. Dieser beschreibt die Dynamik eines nichtlinearen Systems nicht in den genannten Observables, sondern die dynamische Entwicklung von Dichten [28]. Marko Budisić et al. beschreibt in [28] weiter eine Analogie zwischen Euler'scher und Lagrange'scher Betrachtungsweise und dem Perron-Frobenius- und dem Koopman-Operator, wobei letzterer der Lagrange'schen Betrachtungsweise ähnelt. Christof Schutte et al. beschäftigen sich in [101] mit der Schätzung beider Operatoren und vergleicht diese. In den letzten Jahren hat der Koopman-Operator durch seine anschaulichere Gestalt und die Darstellbarkeit als lineares Zustandsraumsystem mehr an Bedeutung gewonnen als der Perron-Frobenius-Operator.

Nach einer langen Pause seit der Formulierung des Koopman-Operators wächst das Interesse an ihm seit 2009 wieder. Einer der Vorreiter dabei ist die Forschungsgruppe um Igor Mezić. Im Jahr 2009 wird die Verbindung zwischen der *Dynamic Mode Decomposition* (DMD) und dem Koopman-Operator hergestellt. Die DMD wird 2008 von Peter Schmid und Jörn Sesterhenn [100] unabhängig vom Koopman-Operator in Verbindung mit der Strömungsanalyse und einer Ableitung linearer Modelle dieser Strömungsvorgänge entwickelt. Hierbei ist die Anzahl der Zustände so enorm, dass solche Systeme durch die schiere Anzahl an Datenpunkten mit üblichen Verfahren nicht analysiert werden können. Schmid entwickelte daher die DMD und nutzt darin implizit den Koopman-Operator, indem er die Vortizität der Strömung in jedem Raumpunkt als Observable annimmt. Clarence Rowley et al. stellen dann in [97] 2009, wie eingangs erwähnt, die Verbindung mit dem Koopman-Operator her und setzen dies erneut in den Kontext von Strömungen. Matthew Williams et al. begründen in [122] erstmals den Begriff der *extended Dynamic Mode Decomposition* (eDMD), welche dem Koopman-Operator eine eigene Form der DMD zuteilt, indem ein erweiterter Zustand (engl. *extended*) als Basis der DMD-Schätzung dient. Seit dem setzte sich der Begriff der eDMD immer weiter durch und ist die Basis einer Vielzahl von Untersuchungen und Schätzungen im Kontext des Koopman-Operators. Auch in dieser Arbeit wird eine randomisierte Version dieses Algorithmus verwendet.

Im Jahr 2010 untersucht dieselbe Forschungsgruppe um Igor Mezić viele weitere Themengebiete mit dem Koopman-Operator. Beispielsweise wird die Wärmeentwicklung in Räumen verteilt über den Tag mittels des Koopman-Operators analysiert, indem sie einzelne Moden betrachten und diese verschiedenen Observables zuweisen [39]. Im selben Jahr entstehen weitere Analysen von Strömungen und Energiesystemen [96, 109]. In den Jahren danach folgen mit Yoshihiko Susuki weitere Arbeiten im Bereich von (verkoppelten) Energie Systemen, die den Koopman-Operator wieder dazu nutzen, lineare Moden dieser nichtlinearen Dynamik zu schätzen [107, 108, 110]. Neben eher praktisch orientierten Themen veröffentlichen Alexandre Mauroy und Igor Mezić in [77] 2016 eine Möglichkeit, die Stabilität nichtlinearer Systeme mittels des Koopman-Operators zu untersuchen. Auch dies wird in der vorliegenden Arbeit aufgegriffen, erweitert und zur Untersuchung der Beispiele verwendet.

In den Jahren 2015 und 2016 kommen zwei weitere Forschungsgruppen um Steven Brunton und J. Nathan Kutz von der University of Washington als weitere Vorantreiber des Koopman-Operators hinzu. Sie beschäftigen sich mit einem breiten Spektrum an Unterthemen des Koopman-Operators, wobei anfangs die Schätzung finiter Approximationen des Koopman-Operators in Form linearer Zustandsraumsysteme im Vordergrund steht [27]. Als eines der ersten Teams verwenden sie ein Koopman-lineares System, um ein nichtlineares System mit linearen Zustandsreglern zu regeln. Dabei beschränken sie sich jedoch im Wesentlichen auf das später vorgestellte sehr einfache System aus Gleichung (3.5), für das eine endlichdimensionale Koopman-lineare Darstellung existiert.

In 2018 wird die Idee dann durch Eurika Kaiser auf komplexere Systeme ohne eine endlichdimensionale Darstellung erweitert, indem sie die autonomen Teile gesondert analysiert und die Eingangsmatrix dazu passend identifiziert beziehungsweise hinzufügt [54]. Im selben Jahr, jedoch erst nach der Analyse dieser komplexeren Systeme, legt Joshua Proctor, ein weiterer Mitarbeiter von Brunton und Kutz, in [93] den allgemein gültigen mathematischen Grundstein für die Erweiterung des Koopman-Operators für Systeme mit Eingängen.

Ein weiteres Teilgebiet der Koopman-Theorie beschäftigt sich mit sogenannten Time-Delay-Koordinaten. Hierbei werden nicht nur die Zustände eines Zustandsraumsystems sondern mehrerer Zeitschritte dieser Zustände als ein Zustandsvektor aufgefasst. Eine erste Idee, wie dies mit dem Koopman-Operator zusammenhängt, wird in [26] beleuchtet. Dabei ergeben sich auf Basis des Embedding-Theorems von Floris Takens ([112]) interessante Einblicke in die Systemdynamik mit der Messung von nur einem Zustand. Durch die Anordnung der Daten in Form von Hankel-Matrizen und der Singulärwert-Zerlegung kann in einem Unterraum ein lineares Modell mit einer „intrinsischen Anregung“ ermittelt werden, welches nicht nur die Systemdynamik außerordentlich gut beschreibt. Mit dieser intrinsischen Anregung können auch Ausschläge in den Daten oder Orbitwechsel von Attraktoren, wie beispielsweise beim Lorenz Attraktor, zuverlässig vorhergesagt werden.

Die Idee dieser Time-Delay-Koordinaten wird in [55] durch Mason Kamb et al. weiter verfolgt und für den gesamten Zustandsvektor formuliert. Allerdings beschränkt sich diese Betrachtung nur auf Systeme ohne externe Anregung. Auch dieser Ansatz wird in dieser Arbeit für Systeme mit Anregung aufgegriffen und analysiert.

Die meisten der bis 2019 erlangten Erkenntnisse der Forschungsgruppe um Brunton und Kutz sind in dem Buch *Data-Driven Science and Engineering* zusammengetragen [23].

Ab dem Jahr 2017 beschäftigen sich dann immer mehr Forschungsgruppen auch mit der Einbindung eines Eingangs in das Koopman-Framework zum Zweck der Regelung. Ian Abraham et al. setzten in [1] als eine der ersten neben Brunton in [27] die Koopman-Methodik zur Regelung ein. Sie regeln dabei die gefahrenen Trajektorien eines Kugelroboters, ein Schlitten-Pendel-System sowie eines auf einer Drohne montierten Pendels (VTOL-Pendel). Xu Ma et al. verwenden einen linear-quadratischen Regelungsansatz mit Nebenbedingung, um ein nichtlineares System, für das ein Koopman-lineares System ermittelt wird, zu regeln [70]. Dabei wird versucht, die Wahl der Gewichtungsmatrizen für diesen Regelungsansatz über die Nebenbedingung im Sinne der Zustände sinnvoll zu wählen. Im Zuge der Popularität der Modell-Prädiktiven-Regelung (MPC) in der Regelungs-Community hielt diese Methodik auch beim Koopman-Operator Einzug. Der Vorreiter auf diesem Gebiet ist Milan Korda, der mit Mezić eine Vielzahl an Beiträgen zu diesem Thema veröffentlicht hat [60, 61]. Yiqiang Han et al. [47] verwendet unter der Nutzung von künstlichen neuronalen Netzen, die weiter unten noch tiefer recherchiert werden, ebenso eine MPC und regelt so simulativ eine Mondlandefähre sowie ein inverses Pendel.

In einer der jüngsten Veröffentlichungen von Korda und Mezić beschäftigen sie sich weiter mit der optimalen Konstruktion von Koopman-Eigenfunktionen speziell zum Zwecke der Regelung [59]. Diese Veröffentlichung dient speziell für den Van der Pol Oszillator auch in dieser Arbeit als Benchmark und findet an geeigneter Stelle in dieser Arbeit Erwähnung.

Durch die Dualität linearer Zustandsraumsysteme ist man jedoch nicht ausschließlich auf die Regelung beschränkt. Insbesondere da in den ersten Jahren der Neuentdeckung des Koopman-Operators zunächst nur autonome Systeme betrachtet wurden, war auch der Einsatz von Beobachtern ein untersuchtes Teilgebiet der Koopman-Operator-Theorie. Während Amit Surana und Andrzej Banaszuk in [105] und [106] den theoretischen Grundstein für Koopman-Operator basierte Beobachter legen, wendet Mouhacine Benosman et al. in [16] diese Idee auf die Bewegung von Menschenmengen und deren Vorhersage an.

Zur Schätzung eines Koopman-linearen Systems stehen im Wesentlichen drei Verfahren zur Verfügung. Durch die nichtlineare Transformation in Form der Observables sowie der Anhebung der Dimension des Zustandsraumes müssen neue Koordinaten gefunden werden. Diese können zum einen anhand der einzelnen Terme der Systemgleichung abgelesen werden. Dieser Ansatz wird auch bei dem später vorgestellten Beispiel in Gleichung (3.5) verfolgt. Eine weitere Möglichkeit ist das Raten von Ansatzfunktionen. Dies kann entweder durch Auswahl und anschließender Schätzung oder durch Lösen eines LASSO-Problems ( $L_1$  regularisiertes kleinste-Quadrat-Problem) mit jeweils einer Vielzahl von Ansatzfunktionen bewerkstelligt werden. Dieser Ansatz hat eine gewisse Ähnlichkeit zur sogenannten Carleman-Linearisierung nach Torsten Carleman, wobei hierbei allerdings ausschließlich die Auswahl von Monomen verfolgt wird [30]. Die beschriebene Methodik ist damit eine Erweiterung der Carleman-Linearisierung, da als Ansatzfunktionen einer Koopman-Schätzung beliebige Funktionen oder Funktionskompositionen verwendet werden können.

Der dritte, auch in dieser Arbeit verfolgte, Ansatz Koopman-lineare Systeme zu schätzen, nutzt künstliche neuronale Netze. Diese Vorgehensweise hat innerhalb der Koopman-Operator Community eine Vielzahl an Unterstützern. In den vergangenen Jahren befassten sich daher eine große Anzahl an Veröffentlichungen mit dem Einsatz künstlicher neuronaler Netze im Kontext des Koopman-Operators. Wie in der Motivation bereits einleitend erläutert, ist dieses Interesse in der Forschung vor allem in den zu schätzenden Observable-Funktionen begründet, die zumeist vollständig unbekannt sind. Nur für wenige Beispielsysteme lassen sich adäquate Funktionen im Vorhinein festlegen. Nahezu alle Ansätze verfolgen eine Netzstruktur, die ähnlich zu einem so genannten Autoencoder ist.

Bei einem Autoencoder ist der Eingang und der Ausgang eines Netzes derselbe und es wird versucht, durch Reduktion der Schichtbreiten eine Minimalrealisierung der Daten zu erreichen. In den Beiträgen [12, 20] konnte durch Pierre Baldi, Kurt Hornik, Hervé Boullard und Yves Kamp gezeigt werden, dass diese Struktur bei Verwendung von linearen Aktivierungsfunktionen exakt auf die Singulärwert Zerlegung zurück zu führen ist. Elad Plaut erweitert diese Methodik in [91], sodass auch die sogenannten *Loadings* ( $\underline{V}$ -Matrix der Singulärwertzerlegung  $\underline{Y} = \underline{U}\underline{\Sigma}\underline{V}^T$ ) aus dieser Darstellung rekonstruiert werden können. Ein solcher Autoencoder ist exemplarisch in Abbildung 1.1 dargestellt. Dabei bezeichnet  $\underline{x} \in \mathbb{R}^n$  die Messdaten der zu reduzierenden Größe und  $\underline{\hat{z}} \in \mathbb{R}^{n_r}$  mit  $n_r < n$  eine reduzierte Darstellung der Daten, ohne dabei einen wesentlichen Anteil der in den Messdaten enthaltenen Informationen zu verlieren. Weiter beschreibt  $\underline{\hat{x}} \in \mathbb{R}^n$  die Rekonstruktion der Messgröße, anhand der reduzierten Darstellung. Diese Struktur wird unter anderem von Masti et al. [76] auch zur direkten Schätzung nichtlinearer Systeme verwendet.

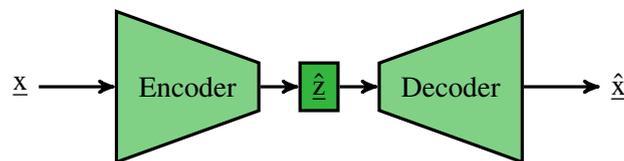


Abbildung 1.1: Exemplarischer Autoencoder mit Zustandsreduktion

Die oben angesprochene Ähnlichkeit zu einem Autoencoder bei der Schätzung von Koopman-linearen Systemen mittels künstlicher neuronaler Netze äußert sich dabei dadurch, dass der Eingang und der Ausgang des Netzes der Systemzustand und somit zunächst identisch ist. Anders als ein klassischer Autoencoder verjüngen sich diese speziellen Netze jedoch nicht, sondern erweitern sich. Dies liegt darin, dass mit der Koopman-Linearisierung meist eine Ordnungserhöhung einher geht, wie später in Kapitel 3 näher erläutert wird. Somit ergibt sich eine Struktur, die exemplarisch in Abbildung 1.2 dargestellt ist. Die Dynamik des kodierten Zustandes  $\underline{z} \in \mathbb{R}^{n_e}$  mit  $n_e \geq n$  und damit eine Schätzung des Koopman-Operators wird über die Verlustfunktion des Netzes implementiert.

Auch wenn alle Ansätze im Kontext des Koopman-Operators in dieser Richtung die Grundstruktur aus Abbildung 1.2 aufweisen, in der die Zustände zuerst kodiert und wieder dekodiert werden, so unterscheiden sie sich in den Details und vor allem in der Rekonstruktion des Systemzustandes  $\underline{x}$ . Die Veröffentlichungen von Marcel Bonnert und Ulrich Konigorski [18], Yunzhu Li et al. [64], Bethany Lusch et al. [69], Samuel Otto und Clarence Rowley [90], Morton et al. [83] und Naiya

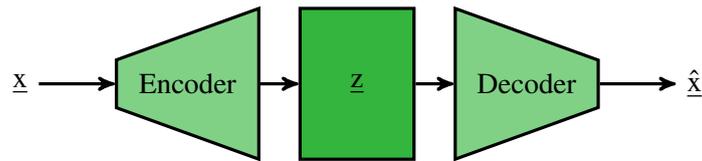


Abbildung 1.2: Exemplarischer Autoencoder mit Zustandserweiterung

Takeishi et al. [111] rekonstruieren den Systemzustand  $\underline{x}$  direkt anhand eines Encoders und eines Decoders, wie in Abbildung 1.2 gezeigt. Yiqiang Han et al. [47] gehen dabei einen noch weiter reduzierten Weg, indem der Decoder komplett ausgelassen wird. Auf der Basis eines solchen Modells wird in dieser Veröffentlichung eine MPC zur Regelung nichtlinearer Systeme verwendet, wie oben bereits erwähnt. Der Zustand wird in allen genannten Veröffentlichungen in Gestalt von  $\underline{z}$  vollständig kodiert. Das heißt der Systemzustand selbst ist nicht Teil des Observable-Vektors. Somit hat dieser kodierte Koopman-Zustand  $\underline{z}$  nur selten einen anschaulichen Bezug zum Systemzustand  $\underline{x}$ . Die eigentliche Dynamik des geschätzten linearen Systems wird daher in einem mathematisch konstruierten Hyperraum beschrieben.

Der zweite Weg bezieht den Systemzustand direkt als Observable mit ein. Der Decoder als Funktion entfällt somit, da der Zustand einfach aus dem Observable-Vektor abgelesen werden kann. Dies wurde bisher nur wenig in der Literatur untersucht. Lediglich Bonnert und Konigorski [18] (zusätzlich zum vorher erwähnten Ansatz) sowie Enoch Yeung et al. [124] befassten sich bisher unter der Nutzung künstlicher neuronaler Netze mit einer solchen Modellannahme. Dies hat den wesentlichen Grund, dass diese Forderung sehr restriktiv ist [73]. Allerdings hat sie insbesondere für die Regelung attraktive Vorteile, da ein anschaulicher Bezug zum Zustandsraum erhalten bleibt. Bei der vollständigen Transformation des Zustandes geht dieser Bezug meistens verloren. Daher werden beide Herangehensweisen in dieser Arbeit untersucht und gegenübergestellt.

Ein weiteres Unterscheidungsmerkmal zwischen den Verfahren, welche auf künstlichen neuronalen Netzen basieren, ist die Schätzung des eigentlichen linearen Zustandsraummodells. Die meisten Arbeiten schätzen das Zustandsraummodell parallel zum übrigen Netz und damit den Observables. Dies ist nicht nur anschaulich sondern besitzt auch eine gewisse Konvergenzgarantie, solange das Optimierungsproblem gut konditioniert ist. Ein Erreichen eines globalen Optimums ist dabei allerdings nicht garantiert. Qianxiao Li et al. [63] und Yunzhu Li et al. [64] gehen dagegen den Weg, die Observables schrittweise durch ein künstliches neuronales Netz zu schätzen und das Zustandsraummodell im Nachgang durch eine eDMD zu approximieren. Das Training ist somit ein alternierender Prozess. Die Konvergenz solcher Verfahren kann bisher nicht nachgewiesen werden. Für einige Beispiele in den genannten Veröffentlichungen konvergierte der Algorithmus jedoch.

Ein weiterer Punkt der für den Einsatz künstlicher neuronaler Netze in diesem Kontext spricht, ist dass Deep-Learning Algorithmen in den letzten zehn Jahren stets weiter entwickelt wurden und nach wie vor werden [34]. Denn auch wenn in den meisten Fällen durch den Einsatz von künstlichen neuronalen Netzen Modelle mit wesentlich besseren Prädiktionseigenschaften ermittelt werden können, ist deren Einsatz mit einem enormen Maß an Rechenkapazität verbunden, was

auch in dieser Arbeit später beleuchtet wird. Damit ist dies der größte Nachteil dieser Methodik im Kontext des Koopman-Operators. Durch die voranschreitende Entwicklung der Algorithmen ist es jedoch möglich, dass sich dieser Nachteil mit den kommenden Jahren relativiert.

Damit bettet sich die vorliegende Arbeit in das Feld der Schätzung nichtlinearer, angeregter, dynamischer Prozesse ein und beleuchtet insbesondere die Leistungsfähigkeit der Koopman-Theorie beim Einsatz linearer Regelungsansätze.

## 1.3 Abgrenzung gegenüber vorangegangenen Arbeiten

Diese Arbeit befasst sich mit der Schätzung des Koopman-Operators angeregter, nichtlinearer Systeme sowie deren Regelung. Mit dieser Thematik beschäftigen sich auch eine Vielzahl anderer Arbeiten. Die vorliegende Arbeit unterscheidet sich von diesen in vier wesentlichen Punkten.

1. Bei der Schätzung der Koopman-linearen Systeme mittels künstlicher neuronaler Netze wird ein **Systemeingang berücksichtigt**.
2. **Die Wahl der Systemordnung** wird durch die Verwendung von Subspace-Verfahren, in Kombination mit den *Information Criteria*, methodisiert. Die so ermittelte potenzielle Systemordnung ist somit mathematisch fundiert und basiert nicht rein auf einem Monte-Carlo Experiment mit einer theoretisch notwendigen unendlichen Anzahl an Identifikationen.
3. Speziell bei der Regelung mittels eines linear-quadratischen Regelungsansatzes stößt man auf das Problem, dass die darin verwendeten Gewichtungen bei einem vollständig kodierten Observable-Vektor nicht anschaulich wählbar sind. Diese Arbeit unterscheidet sich daher von den oben genannten Arbeiten im Sinne der Regelung in der Hinsicht, dass auch ohne das Einbeziehen des Zustandes in den Observable-Vektor die **Nutzung von klassischen LQ-Regelungen** möglich ist. Eine Nebenbedingung, wie in [70], wird damit vermieden.
4. Die verwendete Methodik ist auch für den Fall anwendbar, wenn **nicht alle Zustände messbar** sind. Durch das Zusammenfassen von Zuständen mehrerer Zeitschritte zu einem Zustandsvektor ist es somit möglich, das Systemverhalten umfassend zu beschreiben. Die Basis dafür bildet das Embedding-Theorem von Takens [112].

## 1.4 Struktur der Arbeit

Die vorliegende Arbeit teilt sich in vier Teile. Der erste Teil legt in Kapitel 2 die mathematische Basis, die zum Verständnis der Arbeit essenziell ist und die über allgemein bekannte Grundlagen hinausgeht. Der zweite Teil behandelt in Kapitel 3 und 4 das Kernthema der vorliegenden Arbeit. Hier wird der Koopman-Operator eingeführt, eine Vielzahl von Eigenschaften formuliert und eine Methode vorgestellt, wie ein Unterraum dieses Operators für angeregte Systeme in Form einer

linearen Zustandsraumdarstellung mittels künstlicher neuronaler Netze geschätzt werden kann. Der dritte Teil greift die erarbeiteten Konzepte aus den Kapiteln 3 und 4 auf und wendet diese in Kapitel 5 an, um eine Reihe von nichtlinearen Modellen zu identifizieren, zu analysieren und zu regeln. Ein Fazit der vorliegenden Arbeit und ihrer Ergebnisse sowie ein Ausblick auf mögliche nachfolgende Forschung wird in Kapitel 6, dem vierten und letzten Teil dieser Arbeit, gegeben.

## 2 Grundlagen

Dieses Kapitel dient der Zusammenstellung der mathematischen Grundlagen, auf die in dieser Arbeit zurückgegriffen wird. Umfassende Kenntnisse der linearen System- und Regelungstheorie werden vorausgesetzt. Grundlagen hierzu finden sich beispielsweise in [67] und [68]. Zu diesen Themen gehört auch ein tiefes Verständnis der linearen Algebra, welche in dieser Arbeit ebenso nicht gesondert beschrieben wird. Für eine Einführung in dieses Thema sei auf [7] und [8, S. 511 ff.] verwiesen. Ein Grundverständnis künstlicher neuronaler Netze, der darin auftretenden Gleichungen sowie ihre Optimierung wird ebenso vorausgesetzt, da in diesem Kapitel und dieser Arbeit im Allgemeinen lediglich auf markante Eckpunkte zu diesem Thema eingegangen wird. Für eine Einführung in den Themenkomplex künstlicher neuronaler Netze bietet beispielsweise [3] einen guten Überblick.

### 2.1 Nomenklatur

Skalare Variablen werden als Klein- sowie Großbuchstaben angegeben. Handelt es sich bei den Variablen um Vektoren oder Matrizen, werden diese als unterstrichene Klein- beziehungsweise Großbuchstaben geschrieben. *Liegende* Vektoren, also Matrizen im  $\mathbb{R}^{1 \times n}$ , werden mit einem T im Exponenten notiert, welches ebenso also Indikator für die Transponierte einer Matrix verwendet wird. Operatoren, Verteilungen und Räume werden in Form von kalligrafischen Buchstaben (beispielsweise  $\mathcal{G}$ ) angegeben. Der Index beim Nabla-Operator, beispielsweise  $\nabla_{\underline{x}} \underline{f}(\underline{x}, \underline{u}) = [\partial \underline{f} / \partial x_1, \dots, \partial \underline{f} / \partial x_n]$  mit  $\underline{x} \in \mathbb{R}^n$  und  $\underline{u} \in \mathbb{R}^m$ , dient der Übersicht bei komplexen Ausdrücken, um die Variable zu kennzeichnen nach der abgeleitet wird. Des Weiteren wird die Lie-Ableitung von  $\underline{h}(\underline{x}, \underline{u})$  bezüglich  $\underline{f}$  nach  $\underline{x}$  zu  $L_{\underline{f}, \underline{x}} \underline{h} = \nabla_{\underline{x}} \underline{h} \cdot \underline{f}$  definiert. Diese Ableitung ist ein  $\mathbb{R}$ -linearer Operator, was bedeutet, dass mit  $a_1, a_2 \in \mathbb{R}$  die Bedingung der Additivität und der Homogenität für den Operator  $L_{\underline{f}, \underline{x}}(a_1 \underline{h}_1 + a_2 \underline{h}_2) = a_1 L_{\underline{f}, \underline{x}} \underline{h}_1 + a_2 L_{\underline{f}, \underline{x}} \underline{h}_2$  gilt, wie sich mit obiger Definition leicht überprüfen lässt. Normalverteilte Signale werden über  $\mathcal{N}(m, \sigma^2)$  angegeben. Dabei ist  $m$  der Mittelwert und  $\sigma^2$  die Varianz des Signals.

### 2.2 Mathematische Grundlagen

#### 2.2.1 Normierung

Um anhand von Messdaten Systeme zu identifizieren, werden verschiedene Messgrößen eines Systems meist vor der eigentlichen Identifikation auf eine ähnliche numerische Größenordnung transformiert. Dies ist ein wichtiger Schritt, um numerischen Problemen, ausgelöst durch zahlenmäßig signifikant unterschiedliche Messgrößen, entgegen zu wirken. Ein anschauliches Beispiel

ist ein Prozess, der einen Druck  $p$  in Pa sowie eine Füllhöhe  $h$  in m als Messgrößen liefert. Der Prozessdruck sei dabei im Bereich  $5 \cdot 10^5 \pm 10^5$  Pa und die Füllhöhe im Bereich  $0 \leq h \leq 10$ . Die Größenordnungen unterscheiden sich somit erheblich, was die in der Identifikation verwendeten Datenmatrizen schlecht konditioniert und im Falle künstlicher neuronaler Netze zu Problemen bei der Optimierung führen kann.

Ein gängiger Bereich ist dabei  $[0, 1]$  beziehungsweise  $[-1, 1]$ . Um unabhängiger von einer physikalischen Anschauung wie im obigen Beispiel zu sein, erfolgt die Normierung eines Messdatenvektors  $\underline{x}$ , unter der Annahme dass das Signal aus einem statistischen Prozess stammt, in dieser Arbeit allgemein gemäß

$$\underline{x}_{\text{norm}} = \frac{1}{\sigma} (\underline{x} - \bar{\underline{x}}) ./ \underline{\sigma}_x$$

mit  $\bar{\underline{x}}$  als den Vektor der Mittelwerte und  $\underline{\sigma}$  als den Vektor der Standardabweichungen der einzelnen Messgrößen beziehungsweise Zuständen. Dabei steht der Operator „./“ für die elementweise Division und ist an die MATLAB-Notation angelehnt. Diese Normierung führt für normalverteilte Signale auf einen Wertebereich für den sich ein Großteil der Daten im Bereich  $[-1, 1]$  aufhalten. Bei deterministischen Signalen existiert definitionsgemäß keine Standardabweichung. Dennoch wird für solche Signale in dieser Arbeit gemäß der Gleichung  $\sigma_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$  dieser Parameter berechnet, um mit den übrigen Analysen vergleichbar zu sein.

### 2.2.2 Stellenergie

Bei der Untersuchung der Beispiele in Kapitel 5 wird häufig die Stellenergie bewertet, welche ein Regler benötigt. Dazu wird die Gleichung

$$E_{\underline{u}}(t) = \sum_{i=1}^m \int_0^t \Delta u_i(\tau)^2 d\tau$$

verwendet, um ein Maß für die Stellenergie zu erhalten. Die Größe  $m$  gibt hierin die Anzahl der Eingänge an. Weiter ist  $\Delta u_i = u_i - u_{i,\text{AP}}$  der  $i$ -te Systemeingang beziehungsweise Regler Ausgang, um den Arbeitspunkt  $u_{i,\text{AP}}$  herum. Damit gibt  $E_{\underline{u}}(t)$  die bis zu einem Zeitpunkt  $t$  benötigte Stellenergie an. Diese Definition wird gewählt, da beispielsweise in der Elektrotechnik die über einen Widerstand dissipierte Energie proportional zum Integral über das Quadrat der daran abfallenden Spannung ist. Ähnliche quadratische Abhängigkeiten der Leistungen und Energie finden sich in einer Vielzahl weiterer Teildisziplinen der Physik.

### 2.2.3 Operatoren, Eigenfunktionen und Spektrum

Eine zentrale Eigenschaft des in dieser Arbeit besprochenen Koopman-Operators ist seine Linearität. Viele lineare Operatoren  $\mathcal{K}$ , welche auf einem (unendlichdimensionalen) Vektorraum  $\mathcal{V}$  über einen Körper  $K$  definiert sind, besitzen nach

$$\mathcal{K}\varphi_i = \lambda_i\varphi_i$$

Eigenwert/Eigenfunktion Paare  $(\lambda_i, \varphi_i)$ . Die Menge aller Werte  $\lambda_i$  wird auch als Spektrum des Operators  $\mathcal{K}$  bezeichnet. Im Gegensatz zu linearen Vektorabbildungen  $\underline{y} = \underline{A}\underline{x}$  mit einer Matrix  $\underline{A} \in \mathbb{R}^{n \times n}$  als Operator, welche ein diskretes Spektrum aus genau  $n$  Eigenwert/Eigenfunktion Paaren besitzt (mit einem Eigenvektor als Eigenfunktion), kann das Spektrum allgemeiner linearer Operatoren  $\mathcal{K}$  aus unendlich vielen Eigenwerten  $\lambda_i$  mit zugehörigen Eigenfunktionen  $\varphi_i$  bestehen. Dies wird anhand des Ableitungsoperators  $\frac{\partial}{\partial x}$  deutlich. Er besitzt die Eigenwert/Eigenfunktion Paare  $(\lambda_i, e^{\lambda_i x})$  mit unendlich vielen Eigenwerten  $\lambda_i$ .

Dadurch, dass es unendlich viele Eigenwert/Eigenfunktionen Paare gibt, kann es zu verschiedenen Ausprägungen des Spektrums kommen. Die im Kontext dieser Arbeit wichtigsten sind

- Punktspektren und
- kontinuierliche Spektren.

Das Spektrum  $\sigma(\underline{K})$  endlichdimensionaler linearer Operatoren  $\underline{K}$  in Form von Matrizen ist, wie oben erwähnt, als Punktspektrum zu verstehen. Aber auch bei unendlich dimensionalen Operatoren kann es zu solchen Punktspektren kommen. Es ist dadurch definiert, dass für einen Wert  $\lambda \in \mathbb{C}$ , der Operator  $(\mathcal{K} - \lambda \text{id})$ , mit dem Identitätsoperator  $\text{id}$  (im endlichdimensionalen Fall die Einheitsmatrix  $\underline{I}$ ), nicht injektiv ist, es also keine Inverse gibt, beziehungsweise der Kern dieser Abbildung nicht leer ist. Dann ist  $\lambda \in \sigma_P(\mathcal{K})$  Teil des Punktspektrums  $\sigma_P$  des Operators  $\mathcal{K}$  und man nennt ihn Eigenwert [62]. Damit ist die Menge dieser Eigenwerte abzählbar unendlich.

Besonders problematisch bei der Schätzung endlichdimensionaler linearer Systeme auf Basis des Koopman-Operators sind die sogenannten kontinuierlichen Spektren. Sie sind streng mathematisch definiert als Werte  $\lambda \in \mathbb{C}$  für die  $(\mathcal{K} - \lambda \text{id})$  injektiv, aber nicht surjektiv ist und der selbe Operator  $(\mathcal{K} - \lambda \text{id})$  ein dichtes Bild in  $\mathcal{V}$  besitzt [62]. Die Menge von Eigenwerten eines Operators mit einem kontinuierlichem Spektrum ist somit überabzählbar unendlich.

Besitzt der Koopman-Operator zu einem nichtlinearen System ein kontinuierliches Spektrum  $\sigma_C(\mathcal{K})$ , so ist dieser praktisch nicht als lineares Zustandsraumsystem mit einer endlichen Dimension darstellbar [69]. Beispielsweise schätzten Lusch et al. in [69] solche Systeme, indem eine zusätzliche Hilfsfunktion geschätzt wird. Diese parametrisiert das linear Zustandsraumsystem in Abhängigkeit seiner aktuellen Lage im Zustandsraum und umgeht so das Problem, die Systemordnung erhöhen zu müssen. Das in [69] geschätzte Zustandsraumsystem ist allerdings durch die parametrisierende Funktion nicht mehr linear.

Die beiden besprochenen Klassen von Spektren lassen sich nach den obigen Ausführungen anschaulich wie folgt interpretieren. Systeme, welche isolierte, potenziell gedämpfte, sinusförmige Eigenbewegungen durchführen, besitzen ein Punktspektrum. Systeme, welche über eine kontinuierliche Menge an beobachteten Frequenzen verfügen, besitzen ein kontinuierliches Spektrum [69, 80]. Diese Menge an kontinuierlich verlaufenden Frequenzen kann beispielsweise durch die kontinuierliche Veränderung von Initialzuständen auftreten [80]. Um diese Problematik zu veranschaulichen, wird im Folgenden das ungedämpfte, autonome mathematische Pendel

$$\dot{\underline{x}} = \begin{bmatrix} x_2 \\ -\frac{g}{l} \sin(x_1) \end{bmatrix}$$

betrachtet. Hierbei beschreibt  $x_1$  den Winkel des Pendels zur Aufhängung und  $x_2 = \dot{x}_1$  die zugehörige Winkelgeschwindigkeit. Für das Systemmodell wird eine Punktmasse an einem masselosen, starren Stab der Länge  $l = 1$  m befestigt. Auf das Pendel wirkt die Erdbeschleunigung  $g = 9,81$  m/s<sup>2</sup>. Dieses Pendel besitzt im ungedämpften Fall die Ruhelagen in  $\underline{x} = [k\pi, 0]$ , wobei diese für ungerade  $k$  instabil und für gerade  $k$  grenzstabil sind<sup>1</sup>.

Abbildung 2.1 zeigt den Verlauf der ersten zwei dominanten Frequenzen der Fast Fourier Transformation (FFT) einer Messung des mathematischen Pendels in Abhängigkeit des Anfangszustandes  $x_{1,0}$  bei  $x_{2,0} = 0$ . Es ist deutlich die kontinuierliche Veränderung der dominanten Frequenzen zu erkennen, wonach sich kontinuierliche Spektren charakterisieren lassen [69, 80].<sup>2</sup>

Um ein solches kontinuierliches Spektrum ausreichend gut linear darzustellen, muss die Dimension des Systems daher sehr groß sein [69]. Ausgelöst durch diese Problematik wird im weiteren Verlauf der Arbeit angenommen, dass die zu beschreibenden Systeme reine Punktspektren besitzen.

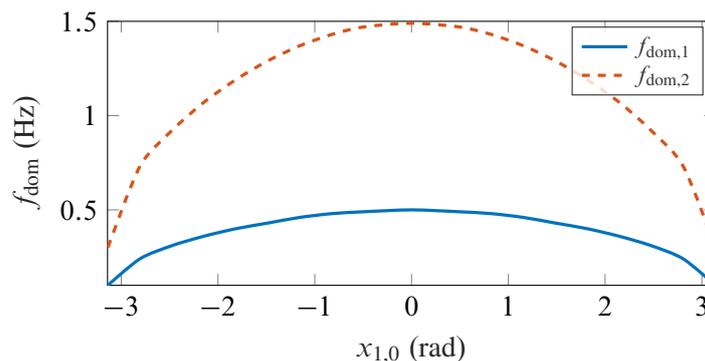


Abbildung 2.1: Visualisierung der ersten beiden dominanten Frequenzen des mathematischen Pendels in Abhängigkeit des Anfangszustandes zur Veranschaulichung seines Spektrums bei  $x_{2,0} = 0$

<sup>1</sup>Grenzstabil bedeutet hierbei, dass eine Dauerschwingung ausgeführt wird.

<sup>2</sup>Im Bereich um  $x_{1,0} = 0$  führt das Pendel nahezu reine Sinusschwingungen aus, wie Abbildung A.3 im Anhang zeigt. Der zweite Schwingungsanteil in Abbildung 2.1 im Bereich um  $x_{1,0} = 0$  spielt daher nur eine untergeordnete Rolle und soll lediglich das Gesamtbild ergänzen.

### 2.2.4 Prädiktionsinvarianz eines Unterraumes zum Koopman-Operator

Im strengen mathematischen Sinne ist ein Unterraum  $\mathcal{U} \subseteq \mathcal{V}$  invariant unter einer Abbildung  $A$  auf einem Vektorraum  $\mathcal{V}$  über einem Körper  $K$ , wenn es einen Endomorphismus  $A$  gibt, für den  $A(\mathcal{U}) \subseteq \mathcal{U}$  gilt. Dieser Unterraum wird dann auch  $A$ -Invariant genannt.

Im Rahmen dieser Arbeit wird jedoch die näherungsweise Invarianz besprochen und verwendet. Sie wurde erstmals in [121] eingeführt und in [114] durch Betrachtungen der Steuerbarkeit und Beobachtbarkeit linearer Systeme charakterisiert.

Da im Rahmen dieser Arbeit nach einem endlichdimensionalen invarianten Unterraum mit dem Operator  $\underline{K}_z$  zu einem unendlichdimensionalen linearen Operator  $\mathcal{K}$  gesucht wird, soll

$$\lim_{T \rightarrow \infty} \frac{1}{T - t_0} \int_{t_0}^T \left\| \mathcal{K} \underline{g}_z(\underline{x}(t)) - \underline{K}_z \underline{g}_z(\underline{x}(t)) \right\|^2 dt = 0$$

gelten. Beide Operatoren entwickeln die ausgewählten Funktionen  $\underline{g}_z$  mit der Zeit. Zur besseren Verständlichkeit wird zunächst der autonome Fall betrachtet. Die endlichdimensionale Approximation  $\underline{K}_z$  des Operators in Form einer Matrix auf dem ausgewählten Unterraum muss also die Funktionen  $\underline{g}_z$  im quadratischen Mittel exakt präzisieren. Da der wahre Operator  $\mathcal{K}$  unbekannt ist, wird die näherungsweise Invarianz des Unterraumes im Rahmen dieser Arbeit durch die Auswertung des obigen Integrals nach der Rücktransformation betrachtet. Es gilt also folgende Definition.

**Definition 2.1** (näherungsweise Invarianz). Eine endlichdimensionale lineare Approximation des Operators  $\mathcal{K}$  in Form einer Matrix  $\underline{K}_z$  wird als näherungsweise invariant zum Operator  $\mathcal{K}$  bezeichnet, wenn der Fehler zwischen gemessenem und geschätztem Zustand  $\underline{x}$  im gesamten betrachteten Zustandsraum  $\mathcal{X}_r$  im quadratischen Mittel verschwindet, sodass

$$\lim_{T \rightarrow \infty} \frac{1}{T - t_0} \int_{t_0}^T \left\| \underline{x}(t) - \hat{\underline{x}}(t) \right\|^2 dt = 0, \quad \forall \underline{x}_0 \in \mathcal{X}_r$$

gilt.

Dies kann unter Berücksichtigung von Abbildung 2.2 als Einhaltung eines Abstandes  $\Delta$ , beziehungsweise für höher dimensionale Probleme, als Einhaltung einer Grenze der euklidischen Norm des Prädiktionsfehlers im Zeitintervall  $\lim_{T \rightarrow \infty} [t_0, T]$ , verstanden werden.

Für komplexe nichtlineare Systeme ist eine unendlich lange Prädiktion mit linearen Operatoren besonders schwierig beziehungsweise in der Realität nicht umsetzbar. Des Weiteren verkompliziert sich die Suche nach einem System mittels Optimierung für sehr lange Horizonte noch weiter. Daher wird auf Basis von Definition 2.1 eine näherungsweise Prädiktionsinvarianz definiert.

**Definition 2.2** (näherungsweise Prädiktionsinvarianz). Eine endlichdimensionale lineare Approximation des Operators  $\mathcal{K}$  in Form einer Matrix  $\underline{K}_z$  wird als näherungsweise prädiktionsinvariant

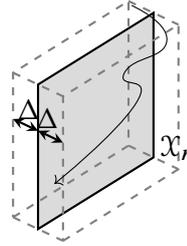


Abbildung 2.2: Visualisierung der näherungsweisen Invarianz [114]

zum Operator  $\mathcal{K}$  bezeichnet, wenn der Fehler zwischen gemessenem und geschätztem Zustand  $\underline{x}$  über einen Prädiktionshorizont  $t_p$  im gesamten betrachteten Zustandsraum  $\mathcal{X}_r$  im quadratischen Mittel kleiner als eine obere Schranke  $\epsilon$  ist. Es gilt also

$$\frac{1}{t_p - t_0} \int_{t_0}^{t_p} \|\underline{x}(t, \underline{u}(t)) - \hat{\underline{x}}(t, \underline{u}(t) | \underline{x}(t_0))\|^2 dt < \epsilon, \quad \forall \underline{x}_0 \in \mathcal{X}_r,$$

wobei  $\hat{\underline{x}}(t, \underline{u}(t) | \underline{x}(t_0))$  den geschätzten Zustand zum Zeitpunkt  $t$  angibt, der nur mit dem Wissen des gemessenen Zustandes  $\underline{x}$  zum Zeitpunkt  $t_0$ , dem linearen Zustandsraumsystem und mit der im Intervall  $[t_0, t_p]$  bekannten Anregung  $\underline{u}(t)$  ermittelt wird.

Dies stellt eine Einschränkung beziehungsweise eine Aufweichung, der näherungsweisen Invarianz dar und ist ohne die rechte Seite „ $< \epsilon$ “ die Definition des mittleren quadratischen Fehlers (MSE) für kontinuierliche Signale. Für diskrete Systeme und Signale wird das Integral durch eine Summe ersetzt. Diese Definition kann direkt als Optimierungsgröße zum Training neuronaler Netze oder allgemeiner Optimierungsprobleme verwendet werden, wie weiter unten in Gleichung (4.1) angegeben. Eine graphische Darstellung der näherungsweisen Prädiktionsinvarianz ist in Abbildung 2.3 gegeben. Definitionsgemäß ist die Abweichung beim Startzustand bei  $t = t_0$  null.

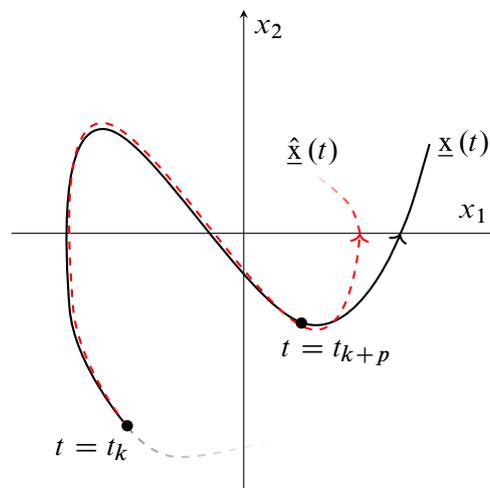


Abbildung 2.3: Illustration der näherungsweisen Prädiktionsinvarianz für eine Trajektorie im Phasenraum für ein diskrete Schätzung eines nichtlinearen Systems zweiter Ordnung

## 2.3 Identifikationsverfahren

### 2.3.1 Klassische Verfahren

Unter klassischen Verfahren werden in der vorliegenden Arbeit Identifikationsverfahren bezeichnet, die lineare Systemmodelle ohne weitere nichtlineare Transformation der Messgrößen schätzen. Sie werden in Kapitel 5 unter anderem als Gegenüberstellung zu der in Kapitel 4 vorgestellten Systematik verwendet und werden daher in Kürze erklärt.

#### Systemidentifikation mittels Ausgangsfehler

Unter den ersten Ansätzen zur Identifikation dynamischer Systeme waren die Ausgangsfehlermethoden. Sie schätzen ein lineares Modell in Form einer Übertragungsfunktion [65]. Die Betrachtung dieses Ausgangsfehlers führt in Verbindung mit einem quadratischen Fehlermaß auf ein in der Regel nicht-konvexes Optimierungsproblem, auch wenn die ARX-Modellannahme erfüllt ist. Erst durch das Einführen einer Prädiktion (durch die Zähler- und Nennerpolynome einer Übertragungsfunktion) ergibt sich der sogenannte Gleichungsfehler, der dann die in Gleichung (2.2) angegebene einfach zu lösende Form ergibt.

In Kapitel 5 werden häufig Systeme mit nur einem Eingang, aber mehreren Ausgängen betrachtet. Hierzu wird der bekannte ARX-Ansatz auf solche SIMO-Systeme (Single Input Multiple Output Systeme) erweitert. Es wird dabei der Fakt genutzt, dass zeitdiskrete reale Systeme mit Halteglied nullter Ordnung immer eine Differenzordnung von 1 aufweisen [11]. Somit liegt ein System der Form

$$\begin{bmatrix} \frac{Y_1}{U}(c) \\ \vdots \\ \frac{Y_q}{U}(c) \end{bmatrix} = \begin{bmatrix} \frac{b_{1,n-1}c^{-1} + \dots + b_{1,0}c^{-n}}{1 + a_{n-1}c^{-1} + \dots + a_{1,0}c^{-n}} \\ \vdots \\ \frac{b_{q,n-1}c^{-1} + \dots + b_{q,0}c^{-n}}{1 + a_{n-1}c^{-1} + \dots + a_{1,0}c^{-n}} \end{bmatrix} \quad (2.1)$$

vor. Dieses kann über den Verschiebungsoperator  $c$  (in der Literatur häufig auch  $z$  oder  $q$ ) in Messkoordinaten

$$\begin{aligned} \underbrace{\begin{bmatrix} y_{1,k} & \dots & y_{q,k} \end{bmatrix}}_{\underline{y}_k^T} &= - \underbrace{\begin{bmatrix} a_{n-1} & \dots & a_0 \end{bmatrix}}_{\underline{a}^T} \underbrace{\begin{bmatrix} y_{1,k-1} & \dots & y_{q,k-1} \\ \vdots & \ddots & \vdots \\ y_{1,k-n} & \dots & y_{q,k-n} \end{bmatrix}}_{\underline{Y}_k} \\ &+ \underbrace{\begin{bmatrix} u_{k-1} & \dots & u_{k-n} \end{bmatrix}}_{\underline{u}_k^T} \underbrace{\begin{bmatrix} b_{1,n-1} & \dots & b_{q,n-1} \\ \vdots & \ddots & \vdots \\ b_{1,0} & \dots & b_{q,0} \end{bmatrix}}_{\underline{B}} \end{aligned}$$

überführt werden. Mittels der Vektorisierungsregeln für Matrizen kann diese Gleichung zu

$$\underbrace{\underline{y}_k}_{\eta_k} = \underbrace{\begin{bmatrix} -\mathbf{Y}_k^T & \mathbf{I}_q \otimes \underline{u}_k \end{bmatrix}}_{\underline{\Psi}_k} \begin{bmatrix} \underline{a} \\ \text{vec}(\mathbf{B}) \end{bmatrix} \quad (2.2)$$

umgeschrieben werden. Bei rauschbehafteter Ausgangsgröße wird Gleichung (2.2) ein additiver Rauschanteil  $\underline{e}_k$  hinzugefügt. Der große Vorteil dieser Methode ist, dass man Daten verschiedener Durchläufe verwenden kann, um dasselbe System zu identifizieren. Hierzu „stapelt“ man die Daten der verschiedenen Durchläufe in  $\underline{\eta}$  und  $\underline{\Psi}$  übereinander. Der Nachteil dieser Methode ist, dass es ein sehr restriktives Rauschmodell voraussetzt. Dabei handelt es sich um die sogenannte *ARX-Modellstruktur (AutoRegressive with eXogenous Input)*. Wird Gleichung (2.2) wieder in die Form aus Gleichung (2.1) gebracht, so verbleibt ein Rauschterm, welcher das Rauschen  $\underline{e}_k$  über das Nennerpolynom aus Gleichung (2.1) filtert. Für eine biasfreie Schätzung muss das Messrauschen diesem gefilterten Rauschsignal entsprechen. Dies wird in der Realität meist nicht der Fall sein, sodass praktisch immer eine biasbehaftete Schätzung vorliegen wird.

Diese Schätzung lässt sich leicht auf den MIMO-Fall (*Multiple Input Multiple Output*) übertragen, indem man im Eingangsvektor einfach weitere Einträge für die weiteren Eingänge vorsieht und diese, wie angegeben, mit der Zeit verschiebt.

Dies ist mit den im Folgenden besprochenen Subspace-Verfahren nicht möglich. Hier muss das System anhand von einem zusammenhängenden Datensatz identifiziert werden. Die Subspace-Methoden sind dafür jedoch meist unabhängig von einem restriktiven Rauschmodell, wie es auch bei der eben vorgestellten Methode der Fall ist.

### Schätzung von Zustandsraummodellen

Zur Schätzung linearer Zustandsraumsysteme stehen eine Vielzahl von Methoden zur Verfügung. Auf der einen Seite existieren sogenannte Subspace-Verfahren, die die Messdaten in Unterräume transformieren, um so die auftretenden Gleichungen zu vereinfachen und lineare Zustandsraummodelle zu schätzen. In der vorliegenden Arbeit wird der sogenannte *optimized Predictor Based Subspace Identification* Algorithmus (PBSID<sub>opt</sub>) aus [33] verwendet. Diese Wahl ergibt sich aus vorangegangenen Untersuchungen, in denen diese Methode, im Vergleich mit einer Vielzahl anderer Subspace-Verfahren, die besten Ergebnisse auch für nichtlineare Systeme liefert [17]. Die zu wählenden Hyperparameter dieses Algorithmus sind der Vergangenheitshorizont  $p_{\text{PBSID}}$  (engl. *past horizon*) und der Zukunftshorizont  $f_{\text{PBSID}}$  (engl. *future horizon*). Der Parameter  $p_{\text{PBSID}}$  ist für die Biasfreiheit der Schätzung,  $f_{\text{PBSID}}$  für die Prädiktionsfähigkeit verantwortlich.

Eine andere Herangehensweise lineare Zustandsraummodelle zu schätzen, bietet die *Observer Kalman Filter Identification* (OKID). Bei diesem Verfahren entfällt gegenüber den PBSID<sub>opt</sub> der Zukunftshorizont, was es leichter einstellbar macht. Die Grundmethodik der OKID ist die Ermittlung der sogenannten Markov-Parameter und ist über alle ihre Ausprägungen identisch. Die

verschiedenen Abkömmlinge unterscheiden sich in der Ermittlung der eigentlichen Systemmatrizen anhand der genannten Markov-Parameter. In dieser Arbeit kommt der sogenannte *Eigensystem Realization Algorithm* (ERA) zum Einsatz und führt damit zur OKID-ERA Methode. Der Einfachheit halber bleibt die Erwähnung von „ERA“ allerdings in den nachfolgenden Untersuchungen und Beschreibungen aus und die Methode wird kurz OKID geschrieben. Eine umfassende Zusammenfassung und Analyse der verschiedenen Ansätze zur Ermittlung eines linearen Zustandsraumsystems mittels OKID ist in [119] gegeben.

Bei beiden Ansätzen, Subspace und OKID, wird ein Zustandsraumsystem in der sogenannten Innovationsform

$$\begin{aligned}\underline{x}_{k+1} &= \underline{A}\underline{x}_k + \underline{B}\underline{u}_k + \underline{K}\underline{e}_k \\ \underline{y}_k &= \underline{C}\underline{x}_k + \underline{D}\underline{u}_k + \underline{e}_k\end{aligned}$$

unterstellt, welches im Wesentlichen die Zustandsraumdarstellung eines Kalman-Filters mit einer modifizierten Ausgangsgleichung darstellt. Dabei ist  $\underline{x}_k \in \mathbb{R}^n$  der Systemzustand,  $\underline{u}_k \in \mathbb{R}^m$  der Systemeingang und  $\underline{e}_k \in \mathbb{R}^q$  die sogenannte *innovation sequence* welche ein mit Varianz  $\sigma^2$  normalverteiltes, mittelwertfreies Signal  $\mathcal{N}(0, \sigma)$  darstellt. Die Matrizen haben jeweils eine, zu den Vektoren  $\underline{x}$ ,  $\underline{u}$  und  $\underline{e}$  passende, Dimension. Die Modifikation der Ausgangsgleichung ist notwendig, sodass der Ausgang dieses Systems dem des identifizierten Systems entspricht. Dies ist zudem neben dem linearen Systemansatz die größte Einschränkung, da das Rauschen in der Zustands- sowie Ausgangsgleichung gleich sein muss. Ein speziell gefiltertes Rauschmodell, wie beim ARX-Modell, bleibt jedoch aus.

Subspace-Schätzverfahren, sowie die OKID eignen sich zusätzlich zur Schätzung einer linearen Systemordnung anhand einer darin geschätzten erweiterten Kalman-Beobachtbarkeitsmatrix oder anderer für die Systemordnung charakteristischer Matrizen. Diese Eigenschaft kommt im Rahmen dieser Arbeit zum Einsatz, um einen Startwert für die Hyperparametersuche hinsichtlich einer linearen Systemordnung zu erhalten. Näher wird dies weiter unten in Abschnitt 4.2 erläutert. Die Auswahl der oben genannten Algorithmen wurde anhand von Voruntersuchungen hinsichtlich ihrer Fähigkeit für die besprochene Ordnungsschätzung, sowie ihrer Schätzfähigkeit getroffen [17, 104].

### **(Extended) Dynamic Mode Decomposition**

Ein weiteres Verfahren, welches insbesondere im Kontext der Identifikation von Koopman-Operatoren zum Einsatz kommt, ist die (*extended*) *Dynamic Mode Decomposition* ((e)DMD). Die DMD geht auf Schmidt [100] zurück, welcher das Verfahren unabhängig von der Theorie des Koopman-Operators entwickelte. Zeitnah zur Entwicklung der DMD konnte ein Zusammenhang zum Koopman-Operator hergestellt werden [97, 122].

Die Grundidee der DMD ist es, anhand einer Vielzahl von Messgrößen die (autonome) Zustandsgleichung im Sinne der kleinsten Quadrate zu lösen. Dies geschieht meist ausschließlich für den

zeitdiskreten Fall, da Messdaten meist in Form von (äquidistanten) Messpunkten vorliegen. Im Verlauf der DMD werden einzelne Moden identifiziert, die anschaulich zur Ordnungsreduktion verwendet werden können, indem nur die signifikantesten Moden im Sinne der Singulärwertzerlegung berücksichtigt bleiben. Im Kontext einer Systemidentifikation einer Vielzahl von realen Systemen wird diese Ordnungsreduktion jedoch häufig nicht benötigt, wenn Zustandsraumssysteme mit niedriger Ordnung identifiziert werden.

Die eDMD bringt die Koopman-Theorie weiter mit der DMD zusammen [122]. Die Idee ist hier, ausgewählte Funktionen der Zustände (Observables) als Zustände zu verwenden, so wie es die gängige Vorgehensweise bei der Identifikation eines Koopman-linearen Systems ist. Das identifizierte System kann dann beispielsweise zur Berechnung von linearen Zustandsreglern auf Basis dieser Koopman-linearen Modelle verwendet werden. Für eine eingehende mathematische Erläuterung der DMD und eDMD sowie der Erweiterung der DMD um einen Eingang sei auf [92, 100, 122] verwiesen.

Die Modellqualität ist jedoch sehr stark abhängig von den gewählten Observables, sodass bei einer ungeeigneten Wahl sehr schlechte Ergebnisse, bei einer geeigneten Wahl aber auch sehr gute Ergebnisse erreicht werden können. Um dieser Problematik zu begegnen, wird in dieser Arbeit eine randomisierte Version der eDMD verwendet. Aus einer Vielzahl von Ansatzfunktionen werden zufällig  $n_e \geq n$  Funktionen ausgewählt, und damit eine eDMD durchgeführt. Dies geschieht mehrfach, sodass sich mehrere Systeme mit jeweils anderen Observables ergeben. Die Auswahl des besten Modells geschieht dann über den kleinsten Prädiktionsfehler gemäß Definition 2.2. Diese Ansatzfunktionen umfassen im Rahmen dieser Arbeit Polynome bis zur Ordnung fünf, trigonometrische Funktionen von Zuständen oder Produkte von Zuständen, sowie Produkte dieser Ansatzfunktionen. So kann beispielsweise eine Observable  $g_1(\underline{x}) = x_1^4 x_2 \sin(x_1 x_2)$  ein Teil dieser Observables sein.

### 2.3.2 Künstliche neuronale Netze

Da in der Literatur bereits eingehende Erläuterungen zur Funktionsweise künstlicher neuronaler Netze vorhanden ist, soll an dieser Stelle nur auf einen kleinen Teilbereich des Trainings eingegangen werden, der für die vorliegende Arbeit von besonderem Interesse ist. Für eine Erklärung der Grundlagen sei auf [3] verwiesen.

Einer der Hyperparameter, also der Parameter, die vom Benutzer zu wählen sind, ist die sogenannte Batch Size  $b$ . Diese gibt die Anzahl an Messpunkten an, die für einen Trainingsschritt verwendet werden. Die Anzahl der Trainingsschritte  $N_T$  unterscheidet sich von den Trainingsepochen  $e$  genau um die Anzahl an Batches  $n_B$  und es gilt somit  $N_T = n_B \cdot e$ . Im Allgemeinen ist die Optimierung eines künstlichen neuronalen Netzes stets ein nicht konvexes Optimierungsproblem [3]. Die Verringerung der Batch Size kann dabei helfen, aus flachen Gebieten und lokalen Minima des Optimierungsgebietes herauszulaufen [123].

Die Relevanz der „korrekten“ Wahl der Batch Size zeigt ein einfaches Beispiel für die Approximation der Funktion  $f(x) = x^2$ . Für ein einfaches *Multilayer Perceptron* (MLP) mit zwei verdeck-

ten Schichten mit je drei Neuronen, einer  $\text{elu}(x)$  Aktivierungsfunktion und einer Ausgangsschicht mit linearer Aktivierungsfunktion kann bereits eine gute Näherung für diese einfache Funktion mit  $N = 2.000$  gleichverteilten Messpunkten im Bereich von  $[-2, 2]$  gefunden werden. Für die Optimierung werden zehn Durchläufe mit jeweils unterschiedlichen Startgewichten gewählt. Als Verlustfunktional wird der quadrierte Ausgangsfehler definiert.

Abbildung 2.4 zeigt das Ergebnis der gemittelten Verlustfunktionale über alle Trainings hinsichtlich der Trainingssschritte  $N_T$ , sowie der Trainingszeit. Der Verlauf über die Epochen ist nicht dargestellt, da dies ein verzerrtes Bild abgibt. Hinsichtlich der Epochen konvergiert dieses Beispiel tatsächlich schneller zu gleichen Ergebnissen. Relevanter ist jedoch die eigentliche Trainingszeit beziehungsweise die echten Trainingssschritte, da mit einer geringeren Batch Size mehr Trainingssschritte in der gleichen Anzahl an Epochen durchgeführt werden.

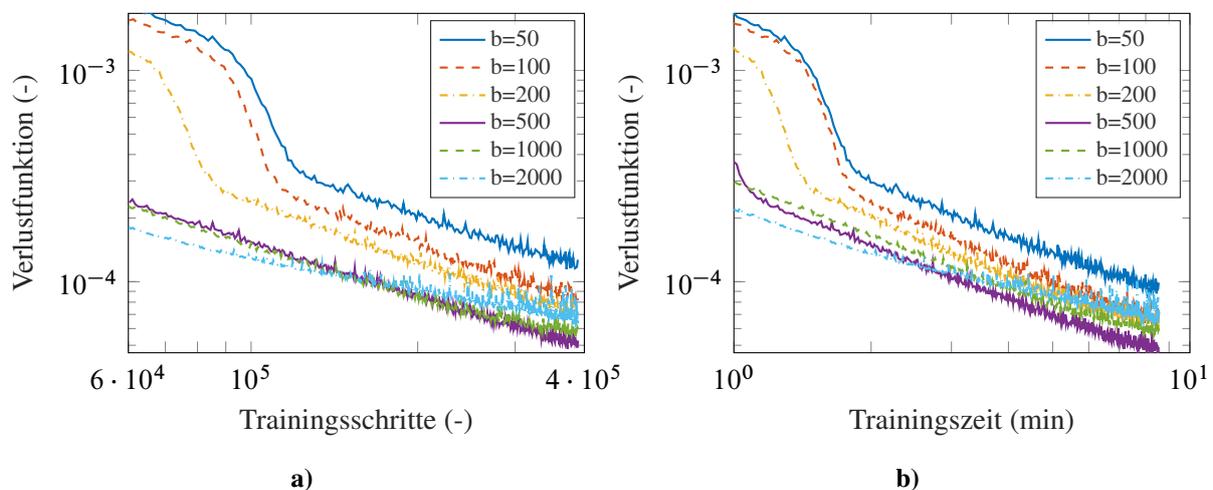


Abbildung 2.4: Veranschaulichung des Effektes von verschiedenen Batch Sizes

Es ist klar ersichtlich, dass hinsichtlich der Trainingszeit sowie der effektiven Trainingssschritte für die meisten kleinen Batch Sizes  $b < N$  ein umgekehrtes Verhalten eintritt, da mit kleiner werdender Batch Size eine Konvergenz zu ähnlichen Ergebnissen erst später eintritt. Für diese Untersuchung tritt daher meist keine schnellere Konvergenz ein, wie sie in [46, 56] festgehalten wird. In den genannten Veröffentlichungen wird allerdings lediglich der Verlauf über die Epochen betrachtet. Es kann für dieses einfache Beispiel jedoch tatsächlich eine leichte Verbesserung und schnellere Konvergenz des Modells erreicht werden, wenn sich die Batch Size in einem mittleren Bereich befindet.

Bei der gleichen Anzahl an Epochen erhöht sich jedoch die Optimierungsdauer mit der Verringerung der Batch Size. Daher muss insbesondere in praktischer Hinsicht darauf geachtet werden, diesen Parameter nicht zu klein zu wählen beziehungsweise die Anzahl an Epochen anzupassen. Für den in dieser Arbeit verfolgten Ansatz kann die Batch size jedoch nicht nur aufgrund der Optimierungszeit, sondern auch aus systemdynamischen Gründen nicht beliebig klein gewählt werden. Ist sie zu klein, geht die allgemeine Gültigkeit über alle Trajektorien beziehungsweise über einen längeren Zeithorizont, verloren. Wie sich dieser Parameter auf eine in der vorliegenden Arbeit durchgeführte Schätzung bei einem anderen Modell auswirkt wird weiter unten beleuchtet.

### 2.3.3 Modellvalidierung

Grundsätzlich werden ermittelte Modelle mittels der in Definition 2.2 angegebenen Fehlergröße validiert. Wichtig ist hier anzumerken, dass dies immer mit in der eigentlichen Schätzung nicht verwendeten Daten durchgeführt wird. Für Schätzungen mit „klassischen“ Verfahren reicht eine Aufteilung in Identifikations- und Testdaten aus. Bei der Schätzung mittels neuronaler Netze ist jedoch eine Dreiteilung erforderlich. Die Netze werden während des Trainings mit einem Trainingsdatensatz trainiert und mit einem Validierungsdatensatz validiert, da die Optimierung anhand der Bewertung der Validierungsdaten abgebrochen wird. Für die eigentliche Validierung des Modells werden im Anschluss sogenannte Testdaten verwendet, die nicht Teil der Optimierung sind. Dies ergibt abschließend ein aufschlussreiches Bild, ob das Modell das Systemverhalten ausreichend gut widerspiegeln kann und das Modell nicht übermäßig auf die Trainings- und Validierungsdaten angepasst wurde. Eine solche starke Anpassung auf die im Training involvierten Daten wird allgemein als Überanpassung beziehungsweise *overfitting* bezeichnet.

## 3 Der Koopman-Operator

Das Ziel des nun folgenden Kapitels soll es sein, eine theoretische sowie verständliche Basis für den Koopman-Operator zu schaffen und insbesondere Konzepte für angeregte Systeme zu erarbeiten. Um einen anschaulichen Einstieg in die Theorie des Koopman-Operators zu ermöglichen, wird sie im Folgenden zunächst für autonome Systeme erläutert.

### 3.1 Koopman-Operator autonomer Zustandsraumssysteme

Der nun folgende Abschnitt beschreibt den Koopman-Operator für autonome Systeme. Im Zuge dessen werden Sätze hergeleitet, die auch für die Anwendung bei Systemen mit einem Eingang beziehungsweise für geregelte Systeme von großer Wichtigkeit sind. Dazu gehören unter anderem die Übertragbarkeit von Eigenschaften wie der Stabilität eines nichtlinearen Systems und seines Koopman-linearen Modells sowie weiterer Analogien aus dem Kontext linearer Zustandsraum-systeme.

Für nichtlineare Systeme der Art

$$\dot{\underline{x}} = \underline{f}(\underline{x}) \quad (3.1)$$

mit  $\underline{x} \in \mathcal{X} \subseteq \mathbb{R}^n$  und  $\underline{f} : \mathcal{X} \rightarrow \mathcal{X}$  oder ihrem zeitdiskreten Pendant

$$\underline{x}_{k+1} = \underline{f}(\underline{x}_k) \quad (3.2)$$

existiert ein linearer Operator  $\mathcal{K}$ , welcher gemäß der Gleichung

$$\mathcal{K}^t g(\underline{x}(t)) = g \circ \underline{S}(t, \underline{x}(t)) \quad (3.3)$$

für zeitkontinuierliche beziehungsweise

$$\mathcal{K}^k g(\underline{x}) = g \circ \underline{S}(k, \underline{x}_k) \quad (3.4)$$

für zeitdiskrete dynamische Systeme sogenannte Observables  $g(\underline{x}) : \mathcal{X}_r \rightarrow \tilde{\mathcal{Z}}$  mit  $\mathcal{X}_r \subset \mathcal{X}$  und  $\tilde{\mathcal{Z}} \subseteq \mathbb{R}$ , auch Messfunktionen genannt, linear mit der Zeit entwickelt<sup>3</sup>. Die mit  $r$  indizierten Räume kennzeichnen einen reduzierten beziehungsweise eingeschränkten Raum des Zustandsraums

<sup>3</sup>Der Operator „ $\circ$ “ beschreibt eine Funktionskomposition und ist definiert über  $g \circ f = g(f)$ .

$\mathcal{X}$ , für den ein näherungsweise invarianter Unterraum zum Koopman-Operator gefunden werden kann. Die Größe  $\underline{S}$  in den Gleichungen (3.3) und (3.4) gibt die Lösung der Systemgleichung (3.1) respektive (3.2) des dynamischen Systems an (engl. *Solution*) und wird in der Literatur auch häufig mit dem Begriff *Flow* des Systems bezeichnet<sup>4</sup>. Diese Observables sind dabei Teil eines Hilbertraums<sup>5</sup>.

**Definition 3.1** (Hilbertraum [8, S. 1177]). Ein Hilbertraum ist ein Vektorraum auf dem ein Skalarprodukt definiert ist und der bezüglich der vom Skalarprodukt induzierten Norm vollständig ist.

Die nichtlineare Dynamik des Systems wird somit in einen unendlichdimensionalen Hilbertraum mittels der Observables  $g(\underline{x})$  transformiert, sodass die Dynamik in diesen neuen Koordinaten mittels des Koopman-Operators  $\mathcal{K}$  linear beschrieben werden kann.

Um die grundsätzliche Idee des Koopman-Operators zu veranschaulichen, sei das folgende Beispiel angegeben, welches aus [23] entnommen ist. Das System

$$\dot{\underline{x}} = \begin{bmatrix} \mu x_1 \\ \lambda(x_2 - x_1^2) \end{bmatrix} \quad (3.5)$$

kann mit einer endlichen Anzahl an Observables  $\underline{g}_{\underline{z}}(\underline{x}) = \underline{z} = [x_1, x_2, x_1^2]^T$  durch Substitution, linear zu

$$\dot{\underline{z}} = \begin{bmatrix} \mu & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & 2\mu \end{bmatrix} \underline{z} \quad (3.6)$$

geschrieben werden. Die dritte Gleichung ergibt sich durch das Ableiten des dritten, künstlichen Zustands (der Observable)  $z_3 = x_1^2$ , denn es gilt  $\dot{z}_3 = \partial(x_1^2)/\partial t = 2x_1\dot{x}_1 = 2\mu x_1^2 = 2\mu z_3$ . Die Systemmatrix aus Gleichung (3.6) wird dabei Koopman-Matrix genannt.

Für eine visuelle Darstellung der Anhebung der Ordnung sei Abbildung 3.1 angegeben. Man sieht, wie sich die eigentlich nichtlineare Dynamik durch die Erweiterung des Zustandes nun dynamisch linear beschreiben lässt, wenn die künstliche Koordinate  $z_3 = x_1^2$  hinzugenommen wird. Dies gilt jedoch nur, solange die Anfangszustände konsistent mit den Observables sind. Beispielsweise ist ein Startzustand  $\underline{z}_0 = [2, 1, 3] \neq [2, 1, 2^2]$  nicht zulässig, da der künstliche Zustand  $z_3 = x_1^2$  einen Startwert von  $z_{3,0} = x_{1,0}^2 = 4$  bei  $x_{1,0} = 2$  vorgibt.

Der Koopman-Operator lässt allerdings nur in sehr seltenen, und meist akademischen Beispielen wie im obigen Fall analytisch, allgemein gültig und in Form eines linearen Zustandsraumsystems schreiben. Er kann jedoch im Rahmen einer Systemidentifikation anhand von Messdaten in

<sup>4</sup>Für zeitdiskrete Systeme gibt die Systemgleichung (3.2) bereits die Zeitlösung an, sodass  $\underline{S}(k, \underline{x}_k) = \underline{f}(\underline{x}_k)$  gilt

<sup>5</sup>Für die Eigenfunktionen des Koopman-Operators, aus denen die Observables gebildet werden können, gilt selbstverständlich das gleiche. Die Eigenfunktionen sind Gegenstand des nachfolgenden Abschnittes.

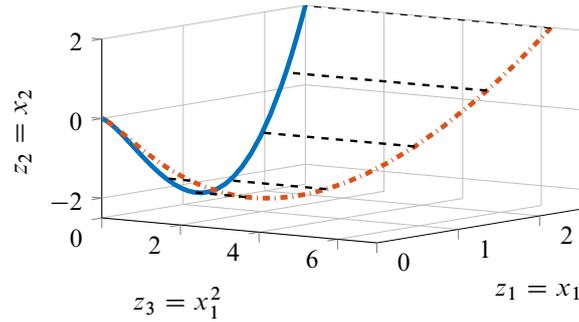


Abbildung 3.1: Autonome Beispieltrajektorie des Systems (3.5) mit  $\mu = -1$ ,  $\lambda = -0,1$ ,  $\beta = 1$  sowie einem Anfangszustand von  $\underline{z}_0 = [2,5, 2, 2,5^2] = [2,5, 2, 6,25]$

Form einer Matrix geschätzt werden. Dies wird weiter unten auf Seite 28 in diesem Abschnitt genauer beschrieben. Der wesentliche Unterschied zu anderen Verfahren zur beziehungsweise Identifikation von dynamischen Systemen ist somit die Formulierung der gemessenen Zustände in anderen Koordinaten und einer möglicherweise damit einhergehende Ordnungserhöhung, wie das obige Beispiel zeigt. Die nichtlineare Dynamik wird somit bei der Verwendung des Koopman-Operators durch die nichtlineare Transformation in Form der Observables und eine lineare dynamischen Entwicklung dieser Observables in einem unendlichdimensionalen Raum „eingetauscht“. Ausgangsmethoden oder Subspace-Verfahren für die Identifikation von beschränkt geltenden linearen Modellen im Zustandsraum, oder beispielsweise die *Sparse Identification of Nonlinear Dynamics* (SINDY) ([24, 25, 72, 95]) für die Identifikation nichtlinearer Modelle, verbleiben im Kontrast dazu in den Messkoordinaten beziehungsweise (linear transformierten) Zustandskoordinaten.

Wie viele andere, bereits in Abschnitt 2.2.3 beschriebene, lineare Operatoren besitzt auch der Koopman-Operator Eigenfunktionen  $\varphi_i(\underline{x}) : \mathcal{X}_r \rightarrow \mathcal{P} \subseteq \mathbb{C}$  in Form von Funktionalen, für die

$$\dot{\varphi}_i(\underline{x}) = \lambda_i \varphi_i(\underline{x}) \quad (3.7)$$

in zeitkontinuierlichen beziehungsweise

$$\varphi_{i,k+1}(\underline{x}_{k+1}) = \lambda_{i,d} \varphi_{i,k}(\underline{x}_k) \quad (3.8)$$

im zeitdiskreten Fall gilt. Durch die Einschränkung auf Systeme mit reinen Punktspektren existieren von diesen Eigenfunktionen abzählbar unendlich viele, was der Index  $i$  verdeutlichen soll. Diese Eigenfunktionen spannen den Hilbertraum

$$\mathcal{H} = \text{span} \{ \varphi_1, \varphi_2, \dots \}$$

auf, in dem der Koopman-Operator  $\mathcal{K}$  agiert. Für alle diese Eigenfunktionen muss für die Invarianz  $\mathcal{K}\varphi_i \in \mathcal{P}$  gelten [27].

Anhand der Gleichungen (3.7) und (3.8) ist direkt die Eigenschaft von Eigenfunktionen ersichtlich. Jede Eigenfunktion  $\varphi_i(\underline{x})$  kann über einen exponentiellen oder einen sinusförmigen Verlauf oder durch eine Mischung beider Verläufe dargestellt werden [80, 97]. Hat man des Weiteren mindestens eine Eigenfunktion gefunden, so können hieraus unendlich viele weitere abgeleitet werden. Das Produkt zweier Eigenfunktionen  $\varphi_1$  und  $\varphi_2$  zeitkontinuierlicher Koopman-Operatoren mit zugehörigen Eigenwerten  $\lambda_1$  und  $\lambda_2$  ergibt eine neue Eigenfunktion  $\varphi_1 \cdot \varphi_2$  mit Eigenwert  $\lambda_1 + \lambda_2$ , wie sich durch Einsetzen in Gleichung (3.7) überprüfen lässt. Für zeitdiskrete Koopman-Operatoren lässt sich analog der Eigenwert  $\lambda_1 \cdot \lambda_2$  zur Eigenfunktion  $\varphi_1 \cdot \varphi_2$  berechnen, wie eine kurze Zwischenrechnung mittels Gleichung (3.8) zeigt.

Da  $\varphi_i(\underline{x})$  Eigenfunktionen des Koopman-Operators sind und sich in dieser Arbeit auf Systeme mit reinen Punktspektren beschränkt wird, können die Observables  $g$  aus Gleichung (3.3) als Summe von Eigenfunktionen

$$g(\underline{x}) = \sum_{i=1}^{\infty} \tilde{g}_i \varphi_i(\underline{x}) \quad (3.9)$$

geschrieben werden. Dabei sind  $\tilde{g}_i \in \mathbb{R}$  Gewichte, welche die Aufteilung der Eigenfunktionen auf die unterschiedlichen Observables beschreiben. Durch die lineare Natur des Koopman-Operators ist es daher möglich, nichtlineare Systeme als lineare Systeme zu behandeln. Somit existiert ein Konzept, mit dem man nichtlinearen Systemen lineare Moden zuordnen kann. Der große Vorteil dabei ist, dass dadurch eine Methode erarbeitet werden kann, mit der es möglich ist, nichtlineare Systeme mit linearen Konzepten zu analysieren und, wie später gezeigt wird, mit linearen Regelungsansätzen zu regeln.

Die positive Eigenschaft der Linearität bringt jedoch den Nachteil der nichtlinearen Transformation der Eigenfunktionen beziehungsweise der daraus ableitbaren Observables und einer unendlichen Dimension des Koopman-Operators mit sich. Die nichtlineare Transformation ist, solange sie bekannt oder zumindest identifizierbar ist, jedoch höchstens in praktischer Hinsicht ein Nachteil. Die unendliche Dimension hingegen macht die praktische Anwendung dieses Konzeptes zunächst unmöglich. Daher wird nach einem  $n_e \geq n$  dimensionalen, (näherungsweise) invarianten Unterraum  $\mathcal{Z} \subseteq \mathbb{R}^{n_e}$  zum Koopman-Operator  $\mathcal{K}$  gesucht, sodass dieser Operator als Matrix  $\underline{K}_{\mathcal{Z}}$  dargestellt werden kann. Die Dimension  $n_e$  beschreibt dabei die erweiterte Systemordnung. Damit ergibt sich die Gleichung

$$\dot{\underline{g}}_{\mathcal{Z}} = \underline{K}_{\mathcal{Z},c} \underline{g}_{\mathcal{Z}} \quad (3.10)$$

mit dem Observable-Vektor  $\underline{g}_{\mathcal{Z}} : \mathcal{X}_r \rightarrow \mathcal{Z}$  und der Koopman-Matrix  $\underline{K}_{\mathcal{Z},c} \in \mathbb{R}^{n_e \times n_e} : \mathcal{Z} \rightarrow \mathcal{Z}$ . Diese zeitkontinuierliche Zustandsraumdarstellung ergibt sich anhand von Gleichung (3.3) mittels eines sogenannten infinitesimalen Generators  $\tilde{\mathcal{K}}$  der Einparametergruppe zum Koopman-Operator  $\mathcal{K}$ , welcher über eine skalare Observable mit

$$\dot{g}(\underline{x}) = \tilde{\mathcal{K}}g(\underline{x}) = \lim_{t \rightarrow 0} \frac{\mathcal{K}^t g - g}{t} = \lim_{t \rightarrow 0} \frac{g \circ \underline{S} - g}{t}$$

definiert ist [74]. Für den zeitdiskreten Fall gilt analog

$$\underline{g}_{z,k+1} = \underline{K}_z \underline{g}_{z,k},$$

was sich direkt aus Gleichung (3.4) ableiten lässt. Gleichung (3.9) lässt sich dann mit einer endlichen Anzahl an Eigenfunktionen zu

$$\underline{g}_z(\underline{x}) \approx \sum_{i=1}^{n_e} \tilde{g}_i \varphi_i(\underline{x}) \quad (3.11)$$

mit dem Gewichtungsvektor  $\tilde{g}_i \in \mathbb{R}^{n_e}$  schreiben. Durch die Limitierung auf  $n_e$  Eigenfunktionen kann eine Observable, welche Gleichung (3.3) beziehungsweise (3.4) erfüllt, nur noch approximativ ermittelt werden<sup>6</sup>. Da der Koopman-Operator, anschaulich gesprochen, nach Gleichung (3.3) auch die Zeitlösung einer Eigenfunktion angibt, gilt

$$\mathcal{K}^t \varphi_i(\underline{x}) = e^{\lambda_i t} \varphi_i(\underline{x}). \quad (3.12)$$

für den zeitkontinuierlichen, respektive

$$\mathcal{K}^k \varphi_i(\underline{x}) = \lambda_i^k \varphi_i(\underline{x}).$$

für den zeitdiskreten Fall. Wird der Koopman-Operator auf Gleichung (3.11) angewendet und Gleichung (3.12) beachtet, so ergibt sich

$$\mathcal{K}^t \underline{g}_z(\underline{x}(t_0)) = \underline{g}_z(\underline{x}(t)) = \sum_{i=1}^{n_e} \tilde{g}_i \varphi_i(\underline{x}(t_0)) e^{\lambda_i(t-t_0)}. \quad (3.13)$$

Diese Darstellung wird auch *Koopman Decomposition* mit dem Koopman-Tripel  $\left\{ \lambda_i, \varphi_i, \tilde{g}_i \right\}_{i=1}^{n_e}$  genannt<sup>7</sup> [79]. Dabei wird  $\tilde{g}_i$  häufig auch als Koopman-Mode bezeichnet.

Die Zeitlösung von Gleichung (3.10) ergibt sich nach Gleichung (A.1) zu

<sup>6</sup>Dies geht mit der Definition eines näherungsweise invarianten Unterraumes (zum Koopman-Operator) einher, wie er in Abschnitt 2.2.4 beschrieben wird.

<sup>7</sup>Im Allgemeinen schreibt man das Koopman-Tripel ohne Einschränkung der Dimension zu  $\left\{ \lambda_i, \varphi_i, \tilde{g}_i \right\}_{i=1}^{\infty}$ . Da an dieser Stelle die Projektion auf einen endlichdimensionalen Raum bereits vorgenommen wurde, wird das Tripel in seiner Anzahl limitiert.

$$\underline{g}_z(\underline{x}(t)) = \sum_{i=1}^{n_e} \underline{v}_i \underline{w}_i^T \underline{g}_z(\underline{x}(t_0)) e^{\lambda_i(t-t_0)}. \quad (3.14)$$

Beim Vergleich der Gleichungen (3.13) und (3.14) lässt sich feststellen, dass sich mittels des Produktes aus Linkseigenvektoren  $\underline{w}_i^T$  und Observable-Vektor  $\underline{g}_z$  eine Abschätzungen für die (skalierte) Eigenfunktionen  $\varphi_i(\underline{x}(t_0))$  und mit  $\underline{v}_i$  eine Approximation der Koopman-Moden angeben lässt, wenn die Realisierung des Koopman-Operators in Form der Koopman-Systemmatrix  $\underline{K}_z$  zusammen mit seinen Observables vorliegt [78]. Formell lässt sich dies zu

$$\tilde{g}_i \varphi_i(\underline{x}(t_0)) = \underbrace{\underline{v}_i}_{\tilde{g}_i/a} \underbrace{\underline{w}_i^T \underline{g}_z(\underline{x}(t_0))}_{a \cdot \varphi_i(\underline{x}(t_0))} \quad (3.15)$$

mit  $a \in \mathbb{C}$  darstellen. Die Ableitung von Eigenfunktionen aus Observables gilt analog für zeitdiskrete Systeme, wie

$$\underline{g}_z(\underline{x}_k) = \sum_{i=1}^{n_e} \underline{v}_i \underline{w}_i^T \underline{g}_z(\underline{x}(t_0)) \lambda_i^k.$$

und

$$\mathcal{K}_d^k \underline{g}_z(\underline{x}_0) = \sum_{j=1}^{n_e} \tilde{g}_j \varphi_j \lambda_j^k.$$

zeigen.

Für das Beispiel aus Gleichung (3.6) lassen sich mit der in Gleichung (3.15) angegebenen Schätzung für die Eigenfunktionen sogar die exakten Eigenfunktionen anhand der ermittelten Koopman-Matrix schätzen. Zwei der Eigenfunktionen des Koopman-Operators zum System aus Gleichung (3.5) sind mit  $\varphi_1 = x_1$  mit Eigenwert  $\mu$  und  $\varphi_2 = x_2 - \frac{\lambda}{\lambda-2\mu} x_1^2$  mit Eigenwert  $\lambda$  aus [54] bekannt und lassen sich leicht anhand von Gleichung (3.7) validieren. Dass  $z_1 = x_1$  eine Eigenfunktion ist, lässt sich direkt anhand von Gleichung (3.6) durch den Diagonaleintrag in der ersten Zeile erkennen. Die zweite Eigenfunktion kann mit Gleichung (3.15) berechnet werden<sup>8</sup>. Ermittelt man den Linkseigenvektor zum Eigenwert  $\lambda$  der obigen Systemmatrix, so erhält man  $\underline{w}_\lambda = a \cdot \left[ 0, 1, \frac{-\lambda}{\lambda-2\mu} \right]^T$  mit einer beliebigen Zahl  $a \in \mathbb{R}$ . Für  $a = 1$  gilt für die Abschätzung der Eigenfunktion  $\hat{\varphi}_\lambda = \underline{w}_\lambda^T \underline{g}_z = x_2 - \frac{\lambda}{\lambda-2\mu} x_1^2 = \varphi_2$ . Dies entspricht exakt der bekannten Eigenfunktion zum gegebenen Eigenwert  $\lambda$ . Dass bei diesem Beispiel die exakten Eigenfunktionen

<sup>8</sup>Alternativ können die Eigenfunktion auch mittels des Modellansatzes  $\varphi = a_1 x_1 + a_2 x_2 + a_3 x_1^2$  über die partielle Differenzialgleichung  $\nabla_{\underline{x}} \varphi(\underline{x}) \underline{f}(\underline{x}) = \lambda \varphi(\underline{x})$ , die sich mittels der Kettenregel aus Gleichung (3.7) ableiten lässt, herleiten.

geschätzt werden liegt maßgeblich daran, dass für das gegebene System ein endlichdimensionaler, invarianter Unterraum zum Koopman-Operator existiert. Dieses Beispiel zeigt, dass der Schritt der Ordnungserhöhung gegebenenfalls umgangen werden kann, wenn die Eigenfunktionen selbst anstatt der Observables verwendet werden.

In [78, S. 21] findet sich ein weiteres Beispiel mit einem approximierten Unterraum für ein gegebenes nichtlineares System, zu dem ein solcher invarianter Unterraum gefunden werden soll. Für die darin gegebene Approximation findet sich eine Annäherung, welche allerdings nur eng um 0 herum gültig ist. Führt man für das dort gegebene Beispiel bei gegebenen Parametern die Approximation weiter als die dargestellten vier Basisfunktionen, so wird die Schätzung schlechter. Dies ist ein anschauliches Beispiel dafür, dass es bei der Schätzung eines näherungsweise invarianten Unterraumes zum Koopman-Operator nicht ausreichend ist, schlicht möglichst viele Basisfunktionen zu verwenden. Es müssen vielmehr die richtigen sein, die zur Approximation angesetzt werden, da es sonst gar zu einer schlechteren Abschätzung kommen kann, als mit weniger Ansatzfunktionen. Für den Fall  $n_e \rightarrow \infty$  ist die Darstellung jedoch exakt, wie es die Theorie des Koopman-Operators vorgibt [58, 73].

Durch die Anhebung der Dimension des Zustandsraumes kann der Koopman-Operator auch als umgekehrte Operation zur Ordnungsreduktion (beispielsweise nach Lohmann [66]) aufgefasst werden, da bei der Ordnungsreduktion Zustände weggelassen werden, ohne einen gewissen Fehler zu überschreiten. Der Koopman-Operator hingegen versucht die Prädiktionsfähigkeit als lineares Zustandsraummodell zu erhöhen, indem er die Modellordnung künstlich anhebt, beziehungsweise nichtlinear transformiert.

### 3.1.1 Zusammenhang zwischen der Stabilität nichtlinearer Systeme und ihren Koopman-linearen Systemen

Die Koopman-Linearisierung eines nichtlinearen Systems beschreibt exakt das nichtlineare Systemverhalten. Somit müssen auch systemdynamische Eigenschaften, wie beispielsweise Stabilität, beider Darstellungen ineinander überführbar sein. Dass dies für zeitkontinuierliche Systeme erfüllt ist, wurde bereits in diversen Veröffentlichungen gezeigt [71, 78]. Dieser Abschnitt beschäftigt sich mit der Fragestellung, wie diese Konzepte auf zeitdiskrete Systeme übertragen werden können. Dies ist insbesondere deshalb wichtig, weil in der Identifikation von Koopman-Linearisierungen nahezu ausschließlich zeitdiskrete Systeme geschätzt werden.

Liegt kein Schätzfehler in der Dynamik im Koopman-Raum, sowie in der Transformation der Zustände durch die Observables vor, so lässt sich eine sehr simple Stabilitätsaussage anhand der Lyapunov-Stabilitätstheorie ableiten, wie in Anhang A.1.2 aufgeführt ist. Meist wird bei der Schätzung von Koopman-Linearisierungen jedoch ein Fehler existieren. Für die weiteren Betrachtungen wird Gleichung (3.2) mit einer Koopman-Approximation zu

$$\underline{x}_{k+1} = \underline{\bar{g}}_z \left( \underline{K}_z \underline{g}_z(\underline{x}_k) \right) + \underline{\epsilon}_{nl}(\underline{x}_k) \quad (3.16)$$

definiert. Dies ist keine Einschränkung, da der Fehlerterm  $\epsilon_{\text{nl}}(\underline{x}_k)$  sämtliche Fehler zusammenfasst. Des Weiteren wird die Schätzung der Umkehrfunktion  $\underline{\bar{g}}_z \in \mathcal{Z} \rightarrow \tilde{\mathcal{X}}_r$  fehlerbehaftet sein, sodass

$$\left(\underline{g}_z \circ \underline{\bar{g}}_z\right)(\underline{z}_k) = \underline{z}_k + \epsilon_{z,\text{trans}}(\underline{z}_k). \quad (3.17)$$

gilt. Der Raum  $\tilde{\mathcal{X}}_r$  stellt dabei einen Unterraum zu  $\mathcal{X}_r$  dar, für den die Koopman-Linearisierung näherungsweise (Prädiktions-)Invariant ist<sup>9</sup>. Gleichung (3.16) wird mittels der Zustandsobservable zu

$$\underline{g}_z(\underline{x}_{k+1}) = \underline{g}_z\left(\underline{\bar{g}}_z\left(\underline{K}_z \underline{g}_z(\underline{x}_k)\right) + \epsilon_{\text{nl}}(\underline{x}_k)\right) \quad (3.18)$$

geschrieben. Unter Zuhilfenahme der Abkürzung  $\tilde{\underline{x}}_{k+1} = \underline{\bar{g}}_z\left(\underline{K}_z \underline{g}_z(\underline{x}_k)\right)$  sowie der Taylorreihen-Approximation von  $\underline{g}_z(\underline{x}_{k+1})$  in Gleichung (3.18) um  $\tilde{\underline{x}}_{k+1}$  ergibt sich der Ausdruck<sup>10</sup>

$$\begin{aligned} \underline{g}_z(\underline{x}_{k+1}) &= \underline{g}_z(\tilde{\underline{x}}_{k+1}) + \nabla_{\underline{x}} \underline{g}_z(\underline{x}_k)(\underline{x}_{k+1} - \tilde{\underline{x}}_{k+1}) + \underline{R}(\underline{x}_k) \\ &= \underline{K}_z \underline{g}_z(\underline{x}_k) + \epsilon_{z,\text{trans}}(\underline{x}_k) + \nabla_{\underline{x}} \underline{g}_z(\underline{x}_k) \epsilon_{\text{nl}}(\underline{x}_k) + \underline{R}(\underline{x}_k) \\ &= \underline{K}_z \underline{g}_z(\underline{x}_k) + \underline{\epsilon}(\underline{x}_k). \end{aligned} \quad (3.19)$$

Somit können sämtliche, im nichtlinearen Modell vorhandene, Fehler zu einem Gesamtfehler  $\underline{\epsilon}(\underline{x}_k)$  im Koopman-Raum zusammengefasst werden. Für die Stabilität von Koopman-Linearisierungen mit einem Fehler gilt mit den obigen Ausführungen der folgende Satz.

**Lemma 3.2** (Stabilität zeitdiskreter Koopman-Linearisierungen I, [5]). Sei  $\underline{g}_z(\underline{x}) : \mathcal{X}_r \rightarrow \mathcal{Z}$  mit  $\mathcal{X}_r \subset \mathbb{R}^n$  und  $\mathcal{Z} \subset \mathbb{R}^{n_e}$  mit  $n_e \geq n$  ein Vektor von Observables des Zustandes mit  $\underline{g}_z(\underline{0}) = \underline{0}$ . Sei

$$\underline{g}_z(\underline{x}_{k+1}) = \underline{K}_z \underline{g}_z(\underline{x}_k) + \underline{\epsilon}(\underline{x}_k)$$

eine fehlerbehaftete Koopman-Linearisierung zu einem diskreten, nichtlinearen dynamischen System nach Gleichung (3.2). Besitzt  $\underline{K}_z$  nur Eigenwerte  $\left|\lambda\left(\underline{K}_z\right)\right| < 1$ , so gilt die Lyapunov-Gleichung

$$\underline{K}_z^T \underline{Q} \underline{K}_z - \underline{Q} = -\underline{P}.$$

Weiter sei  $V(\underline{x}_k) = \underline{g}_z(\underline{x}_k)^T \underline{Q} \underline{g}_z(\underline{x}_k)$  eine Lyapunov-Funktion für das Koopman-lineare Zustandsraumsystem. Gilt mit  $\|\underline{\epsilon}(\underline{0})\|_2 = 0$  die Ungleichung

<sup>9</sup>Die praktische Einhaltung eines kleinen Transformationsfehlers ist Teil von Kapitel 4.

<sup>10</sup>Der in Gleichung (3.19) zusammengefasste Fehler  $\underline{\epsilon}(\underline{x}_k)$  hängt unter anderem vom Gradienten  $\nabla_{\underline{x}} \underline{g}_z$  von  $\underline{g}_z$  ab. Dies kann gegebenenfalls zu großen Fehlern in bestimmten Bereichen des Zustandsraumes führen.

$$\sup_{\forall \underline{x}_k \in \mathcal{X} \setminus \{0\}} \frac{\|\underline{\epsilon}(\underline{x}_k)\|}{\|\underline{g}_z(\underline{x}_k)\|} \leq \alpha,$$

mit  $0 \leq \alpha < -\|\underline{K}_z\|_2 + \sqrt{\|\underline{K}_z\|_2^2 + \frac{\lambda_{\min}(\underline{P})}{\lambda_{\max}(\underline{Q})}}$ , dann ist das nichtlineare zeitdiskrete System stabil im Sinne von Lyapunov.

*Beweis.* Es gilt offensichtlich  $V(\underline{0}) = 0$  und  $V(\underline{x}_k) > 0 \forall \underline{x}_k \neq \underline{0}$ . Um eine zeitdiskrete Lyapunov-Funktion zu sein, muss weiter

$$V(\underline{x}_{k+1}) - V(\underline{x}_k) < 0$$

gelten. Mit Gleichung (3.19) gilt

$$\begin{aligned} V(\underline{x}_{k+1}) - V(\underline{x}_k) &= \left( \underline{K}_z \underline{g}_z(\underline{x}_k) + \underline{\epsilon}(\underline{x}_k) \right)^T \underline{Q} \left( \underline{K}_z \underline{g}_z(\underline{x}_k) + \underline{\epsilon}(\underline{x}_k) \right) - \underline{g}_z(\underline{x}_k)^T \underline{Q} \underline{g}_z(\underline{x}_k) \\ &= \underline{g}_z(\underline{x}_k)^T \underline{K}_z^T \underline{Q} \underline{K}_z \underline{g}_z(\underline{x}_k) + 2 \underline{g}_z(\underline{x}_k)^T \underline{K}_z^T \underline{Q} \underline{\epsilon}(\underline{x}_k) + \underline{\epsilon}(\underline{x}_k)^T \underline{Q} \underline{\epsilon}(\underline{x}_k) \\ &\quad - \underline{g}_z(\underline{x}_k)^T \underline{Q} \underline{g}_z(\underline{x}_k) \\ &= \underline{g}_z(\underline{x}_k)^T \left( \underline{K}_z^T \underline{Q} \underline{K}_z - \underline{Q} \right) \underline{g}_z(\underline{x}_k) + 2 \underline{g}_z(\underline{x}_k)^T \underline{K}_z^T \underline{Q} \underline{\epsilon}(\underline{x}_k) + \underline{\epsilon}(\underline{x}_k)^T \underline{Q} \underline{\epsilon}(\underline{x}_k) \\ &= - \underline{g}_z(\underline{x}_k)^T \underline{P} \underline{g}_z(\underline{x}_k) + 2 \underline{g}_z(\underline{x}_k)^T \underline{K}_z^T \underline{Q} \underline{\epsilon}(\underline{x}_k) + \underline{\epsilon}(\underline{x}_k)^T \underline{Q} \underline{\epsilon}(\underline{x}_k) \\ &\leq - \lambda_{\min}(\underline{P}) \|\underline{g}_z(\underline{x}_k)\|_2^2 + \lambda_{\max}(\underline{Q}) \|\underline{\epsilon}(\underline{x}_k)\|_2^2 + 2 \underline{g}_z(\underline{x}_k)^T \underline{K}_z^T \underline{Q} \underline{\epsilon}(\underline{x}_k). \end{aligned}$$

Der letzte Term lässt sich mit der Cauchy-Schwarz Ungleichung

$$\begin{aligned} 2 \underline{g}_z(\underline{x}_k)^T \underline{K}_z^T \underline{Q} \underline{\epsilon}(\underline{x}_k) &\leq 2 \|\underline{K}_z \underline{g}_z(\underline{x}_k)\|_2 \|\underline{Q} \underline{\epsilon}(\underline{x}_k)\|_2 \\ &\leq 2 \lambda_{\max}(\underline{Q}) \|\underline{\epsilon}(\underline{x}_k)\|_2 \|\underline{K}_z\|_2 \|\underline{g}_z(\underline{x}_k)\|_2 \end{aligned}$$

nach oben abschätzen, wobei  $\|\underline{K}_z\|_2$  die Spektralnorm der Matrix  $\underline{K}_z$  ist. Mit

$\|\underline{\epsilon}(\underline{x}_k)\|_2 \leq \alpha \|\underline{g}_z(\underline{x}_k)\|_2$  gilt dann

$$\begin{aligned} V(\underline{x}_{k+1}) - V(\underline{x}_k) &\leq - \lambda_{\min}(\underline{P}) \|\underline{g}_z(\underline{x}_k)\|_2^2 + \lambda_{\max}(\underline{Q}) \|\underline{\epsilon}(\underline{x}_k)\|_2^2 \\ &\quad + 2 \lambda_{\max}(\underline{Q}) \|\underline{\epsilon}(\underline{x}_k)\|_2 \|\underline{K}_z\|_2 \|\underline{g}_z(\underline{x}_k)\|_2 \\ &\leq \left( - \lambda_{\min}(\underline{P}) + \lambda_{\max}(\underline{Q}) \alpha^2 + 2 \lambda_{\max}(\underline{Q}) \|\underline{K}_z\|_2 \alpha \right) \|\underline{g}_z(\underline{x}_k)\|_2^2 \stackrel{!}{<} 0. \end{aligned}$$

Diese Ungleichung ist genau dann erfüllt, wenn die additive Lösung der quadratischen Gleichung für  $\alpha$  herangezogen wird<sup>11</sup>. Diese ergibt sich zu

$$\alpha < -\|\underline{\mathbf{K}}_z\|_2 + \sqrt{\|\underline{\mathbf{K}}_z\|_2^2 + \frac{\lambda_{\min}(\mathbf{P})}{\lambda_{\max}(\mathbf{Q})}}.$$

Hält der Dynamikfehler  $\underline{\epsilon}$  gegenüber den Observables diese obere Fehlerschranke ein, dann ist das nichtlineare System stabil im Sinne von Lyapunov mit der Lyapunov-Funktion  $V(\underline{\mathbf{x}}_k) = \underline{\mathbf{g}}_z(\underline{\mathbf{x}}_k)^T \mathbf{Q} \underline{\mathbf{g}}_z(\underline{\mathbf{x}}_k)$ .  $\square$

Der im obigen Lemma angegebene Fehler  $\underline{\epsilon}(\underline{\mathbf{x}}_k)$  kann über Messungen nach Gleichung (3.19) zu

$$\underline{\epsilon}(\underline{\mathbf{x}}_k) = \underline{\mathbf{g}}_z(\underline{\mathbf{x}}_{k+1}) - \underline{\mathbf{K}}_z \underline{\mathbf{g}}_z(\underline{\mathbf{x}}_k)$$

ermittelt werden. Es handelt sich damit um den Fehler, welchen das geschätzte System in einem Zeitschritt macht.

Eine weitere Herangehensweise die Stabilität zeitdiskreter nichtlinearer Systeme mittels ihrer Koopman-Linearisierungen nachzuweisen bietet das folgende Lemma, welches das zeitdiskrete Pendant zu [78, S. 53, Proposition 2.4] darstellt.

**Lemma 3.3** (Stabilität zeitdiskreter Koopman-Linearisierungen II). Sei  $\underline{\mathbf{g}}_z(\underline{\mathbf{x}}) : \mathcal{X}_r \rightarrow \mathcal{Z}$  mit  $\mathcal{X}_r \subset \mathbb{R}^n$  und  $\mathcal{Z} \subset \mathbb{R}^{n_e}$  mit  $n_e \geq n$  ein Vektor von Observables des Zustandes mit  $\underline{\mathbf{g}}_z(\underline{\mathbf{0}}) = \underline{\mathbf{0}}$ . Sei

$$\underline{\mathbf{g}}_z(\underline{\mathbf{x}}_{k+1}) = \underline{\mathbf{K}}_z \underline{\mathbf{g}}_z(\underline{\mathbf{x}}_k) + \underline{\epsilon}(\underline{\mathbf{x}}_k) \quad (3.20)$$

eine zeitdiskrete Koopman-Linearisierung eines Systems nach (3.2) mit einem Dynamikfehler  $\underline{\epsilon}(\underline{\mathbf{x}}_k)$ . Wenn ein  $\gamma > 0$  existiert, sodass

$$\left\| \left( c\mathbf{I} - \underline{\mathbf{K}}_z \right)^{-1} \right\|_{\mathcal{H}_{\infty}} < \frac{1}{\gamma} \quad (3.21)$$

und

$$\sup_{\forall \underline{\mathbf{x}}_k \in \mathcal{X} \setminus \{0\}} \frac{\|\underline{\epsilon}(\underline{\mathbf{x}}_k)\|_2}{\|\underline{\mathbf{g}}_z(\underline{\mathbf{x}}_k)\|_2} < \gamma \quad (3.22)$$

<sup>11</sup>Die subtraktive Lösung quadratischen Gleichung für  $\alpha$  führt auf ein  $\alpha < 0$ , was gemäß der Definition  $\|\underline{\epsilon}(\underline{\mathbf{x}}_k)\|_2 \leq \alpha \|\underline{\mathbf{g}}_z(\underline{\mathbf{x}}_k)\|_2$  nicht sinnvoll wäre.

gilt, wobei  $\mathcal{X}_r$  positiv invariant ist, dann besitzt das System eine Lyapunov-Funktion

$V(\underline{x}_k) = \underline{g}_z^T(\underline{x}_k) \underline{Q} \underline{g}_z(\underline{x}_k)$  mit  $V(\underline{0}) = 0$ ,  $V(\underline{x}_k) > 0 \forall \underline{x} \neq \underline{0}$ ,  $\underline{Q} = \text{diag}([q_1, \dots, q_{n_e}]) > 0$  und  $\lambda_i(\underline{Q}) \neq 1$  mit  $i \in [1, \dots, n_e]$ .

*Beweis.* Für den Nachweis der Stabilität nach Lyapunov muss

$$V(\underline{x}_{k+1}) - V(\underline{x}_k) < 0$$

gelten. Mit (3.20) ergibt sich somit

$$\begin{aligned} V(\underline{x}_{k+1}) - V(\underline{x}_k) &= \left( \underline{K}_z \underline{g}_z(\underline{x}_k) + \underline{\epsilon}(\underline{x}_k) \right)^T \underline{Q} \left( \underline{K}_z \underline{g}_z(\underline{x}_k) + \underline{\epsilon}(\underline{x}_k) \right) - \underline{g}_z^T(\underline{x}_k) \underline{Q} \underline{g}_z(\underline{x}_k) \\ &= \underline{g}_z^T(\underline{x}_k) \left( \underline{K}_z \underline{Q} \underline{K}_z - \underline{Q} \right) \underline{g}_z(\underline{x}_k) + \underline{\epsilon}_z^T(\underline{x}_k) \underline{Q} \underline{\epsilon}_z(\underline{x}_k) + 2 \underline{\epsilon}^T(\underline{x}_k) \underline{Q} \underline{K}_z \underline{g}_z(\underline{x}_k) \\ &= d_V \stackrel{!}{<} 0 \end{aligned} \tag{3.23}$$

Der letzte Term in Gleichung (3.23) wird mittels

$$\begin{aligned} &\left( (\underline{I} - \underline{Q})^{1/2} \underline{\epsilon}(\underline{x}_k) - (\underline{I} - \underline{Q})^{-1/2} \underline{Q} \underline{K}_z \underline{g}_z(\underline{x}_k) \right)^T \cdot \\ &\left( (\underline{I} - \underline{Q})^{1/2} \underline{\epsilon}(\underline{x}_k) - (\underline{I} - \underline{Q})^{-1/2} \underline{Q} \underline{K}_z \underline{g}_z(\underline{x}_k) \right) \geq 0 \end{aligned}$$

nach oben abgeschätzt. Die darin enthaltenen Wurzelmatrizen existiert immer, solange die Matrix invertierbar ist [43]. Die Lösung kann komplex sein und muss nicht eindeutig sein, was an dieser Stelle jedoch nicht wichtig ist. Somit ergibt sich jedoch die Einschränkung der Eigenwerte von  $\underline{Q}$ . Ausmultipliziert ergibt der obige Ausdruck

$$\underline{\epsilon}^T(\underline{x}_k) (\underline{I} - \underline{Q}) \underline{\epsilon}(\underline{x}_k) + \underline{g}_z^T(\underline{x}_k) \underline{K}_z^T \underline{Q} (\underline{I} - \underline{Q})^{-1} \underline{Q} \underline{K}_z \underline{g}_z(\underline{x}_k) \geq 2 \underline{\epsilon}^T(\underline{x}_k) \underline{Q} \underline{K}_z \underline{g}_z(\underline{x}_k).$$

Diese Abschätzung wird in Gleichung (3.23) eingesetzt und für die Stabilitätsuntersuchung gilt also fortan

$$\begin{aligned} d_V &< \underline{g}_z^T(\underline{x}_k) \left( \underline{K}_z \underline{Q} \underline{K}_z - \underline{Q} \right) \underline{g}_z(\underline{x}_k) + \underline{\epsilon}_z^T(\underline{x}_k) \underline{Q} \underline{\epsilon}_z(\underline{x}_k) \\ &\quad + \underline{\epsilon}^T(\underline{x}_k) (\underline{I} - \underline{Q}) \underline{\epsilon}(\underline{x}_k) + \underline{g}_z^T(\underline{x}_k) \underline{K}_z^T \underline{Q} (\underline{I} - \underline{Q})^{-1} \underline{Q} \underline{K}_z \underline{g}_z(\underline{x}_k) \\ &= \underline{g}_z^T(\underline{x}_k) \left( \underline{K}_z \underline{Q} \underline{K}_z - \underline{Q} + \underline{K}_z^T \underline{Q} (\underline{I} - \underline{Q})^{-1} \underline{Q} \underline{K}_z \right) \underline{g}_z(\underline{x}_k) + \left\| \underline{\epsilon}^T(\underline{x}_k) \right\|_2^2 \\ &= \underline{g}_z^T(\underline{x}_k) \left( \underline{K}_z \underline{Q} \underline{K}_z - \underline{Q} + \underline{K}_z^T \underline{Q} (\underline{I} - \underline{Q})^{-1} \underline{Q} \underline{K}_z + \gamma^2 \underline{I} \right) \underline{g}_z(\underline{x}_k) \\ &\quad + \left\| \underline{\epsilon}(\underline{x}_k) \right\|_2^2 - \gamma^2 \left\| \underline{g}_z(\underline{x}_k) \right\|_2^2 \end{aligned}$$

Wenn  $\underline{Q}$  die Lösung einer Lyapunov-Gleichung ist, was hier vorausgesetzt wird, dann entfällt der erste Term der letzten Gleichung wenn die Ungleichung (3.21) gilt, wie in A.1.6 im Anhang ausgeführt ist. Mit der Forderung  $\|\underline{\epsilon}(\underline{x}_k)\|_2^2 - \gamma^2 \|\underline{g}_z(\underline{x})\|_2^2 \stackrel{!}{<} 0$  ergibt sich (3.22), was den Beweis vervollständigt.  $\square$

Dieses Lemma ist zunächst scheinbar weniger restriktiv als Lemma 3.2, da die einzige Abschätzung den Mischterm betrifft. Das vorher ermittelte Lemma 3.2 nähert hingegen ganze Matrizen durch ihren größten, respektive ihren kleinsten, Eigenwert an. Ob sich diese Vermutung anhand von Beispielen validieren lässt, wird in Kapitel 5 untersucht. Eine wichtige Einschränkung beider Lemmata ist der Ausschluss von ungedämpft schwingungsfähigen Systemen. Dies schließt nach [80] nichtlineare Systeme aus, welche Grenzyklen durchlaufen.

### 3.1.2 Weitere Eigenschaften Koopman-linearisierbarer Systeme

Mit Hilfe des Koopman-Operators lassen sich noch eine Vielzahl weiterer interessanter Untersuchungen anstellen. Eine besondere Eigenschaft ist dabei, dass die Observables dazu genutzt werden können, eine Ähnlichkeitstransformation herzuleiten, wie man sie beispielsweise beim Basiswechsel linearer Zustandsraumsysteme verwendet.

Hierzu wird das Gleichungssystem

$$\begin{aligned} \dot{\underline{z}} &= \underline{K}_z \underline{z} + \underline{e}_{d,z}(\underline{x}) \\ \underline{z} &= \underline{g}_z(\underline{x}) \\ \hat{\underline{x}} &= \bar{\underline{g}}_z(\underline{z}) \\ \underline{x} &= \hat{\underline{x}} + \underline{e}_t(\underline{x}) \end{aligned} \tag{3.24}$$

mit  $\underline{x} \in \mathcal{X}_r$ , dem Dynamikfehler  $\underline{e}_{d,z}(\underline{x})$  und dem Trajektorienfehler  $\underline{e}_t(\underline{x})$  betrachtet, mithilfe dessen jedes autonome System als Koopman-lineares System dargestellt werden kann. Durch Ableitung der vierten Gleichung aus (3.24) erhält man

$$\begin{aligned} \dot{\underline{x}} &= \frac{d\bar{\underline{g}}_z(\underline{z})}{dt} + \dot{\underline{e}}_t(\underline{x}) \\ &= \nabla_z \bar{\underline{g}}_z(\underline{z}) \dot{\underline{z}} + \dot{\underline{e}}_t(\underline{x}) \\ &= \nabla_z \bar{\underline{g}}_z(\underline{z}) \underline{K}_z \underline{g}_z(\underline{x}) + \nabla_z \bar{\underline{g}}_z(\underline{z}) \underline{e}_{d,z}(\underline{x}) + \dot{\underline{e}}_t(\underline{x}). \end{aligned}$$

Umgeformt ergibt sich

$$\begin{aligned}
\dot{\underline{x}} &= \nabla_{\underline{z}\bar{\underline{g}}_z}(\underline{z}) \underline{K}_{\underline{z}\bar{\underline{g}}_z}(\underline{x}) + \nabla_{\underline{z}\bar{\underline{g}}_z}(\underline{z}) \underline{e}_{d,z}(\underline{x}) + \nabla_{\underline{x}\bar{\underline{e}}_t} \dot{\underline{x}} \\
(\underline{I} - \nabla_{\underline{x}\bar{\underline{e}}_t}) \dot{\underline{x}} &= \nabla_{\underline{z}\bar{\underline{g}}_z}(\underline{z}) \underline{K}_{\underline{z}\bar{\underline{g}}_z}(\underline{x}) + \nabla_{\underline{z}\bar{\underline{g}}_z}(\underline{z}) \underline{e}_{d,z}(\underline{x}) \\
\dot{\underline{x}} &= (\underline{I} - \nabla_{\underline{x}\bar{\underline{e}}_t})^{-1} \left( \nabla_{\underline{z}\bar{\underline{g}}_z}(\underline{z}) \underline{K}_{\underline{z}\bar{\underline{g}}_z}(\underline{x}) + \nabla_{\underline{z}\bar{\underline{g}}_z}(\underline{z}) \underline{e}_{d,z}(\underline{x}) \right).
\end{aligned}$$

Für den fehlerfreien Fall ohne Dynamik- und Trajektorienfehler gilt damit weiter

$$\begin{aligned}
\dot{\underline{x}} &= \underline{f}(\underline{x}) = \nabla_{\underline{z}\bar{\underline{g}}_z}(\underline{z}) \underline{K}_{\underline{z}\bar{\underline{g}}_z}(\underline{x}) \\
\rightarrow \Delta \dot{\underline{x}} &\approx \nabla_{\underline{z}\bar{\underline{g}}_z}(\underline{z}) \underline{K}_{\underline{z}\bar{\underline{g}}_z}(\underline{x}) \Delta \underline{x}
\end{aligned} \tag{3.25}$$

mit  $\Delta \underline{x} = \underline{x} - \underline{x}_{AP}$ , mit dem konstanten Arbeitspunkt  $\underline{x}_{AP}$ . Eine Beobachtung, die aus dieser Darstellung geschlossen werden kann, ist die oben genannte strukturelle Ähnlichkeit zu einer Ähnlichkeitstransformation aus der Theorie linearer Zustandsraumsysteme. Bei diesen kann ein Basiswechsel über die Darstellung  $\tilde{\underline{x}} = \underline{T}^{-1} \underline{x}$  erfolgen, sodass das autonome Zustandsraumsystem  $\dot{\underline{x}} = \underline{A} \underline{x}$  mittels  $\dot{\tilde{\underline{x}}} = \underline{T}^{-1} \underline{A} \underline{T} \tilde{\underline{x}} = \tilde{\underline{A}} \tilde{\underline{x}}$  in neuen Koordinaten  $\tilde{\underline{x}}$  dargestellt werden kann. Mit dem Koopman-Operator und seinen Observables ist es ebenso möglich eine ähnliche Darstellung zu finden, denn auch bei Gleichung (3.25) wird die Systemmatrix (Koopman-Matrix  $\underline{K}_{\underline{z}\bar{\underline{g}}_z}$ ) von rechts mit einer (nichtlinearen) Transformation und von links mit einem Äquivalent seiner Inversen multipliziert. Ist die Transformation und Rücktransformation über die Observables exakt, ist das Produkt der Gradienten von Observable-Vektor und seiner Inversen darüber hinaus durch den Satz über die Umkehrfunktion die Einheitsmatrix  $\nabla_{\underline{z}\bar{\underline{g}}_z}(\underline{z}) \nabla_{\underline{x}\bar{\underline{e}}_t}(\underline{x}) = \underline{I}$  [6, S. 223], was das Argument weiter verstärkt.

Da nun die Theorie autonomer Systeme eingehend erläutert und deren Eigenschaften erarbeitet wurden, folgt die Koopman-Operator-Theorie für angeregte, nichtlineare Zustandsraumsysteme.

## 3.2 Koopman-Operator angeregter Zustandsraumsysteme

Da sich für rein autonome Systeme keine Regler berechnen lassen, werden nun die vorgestellten Konzepte für Systeme mit Eingängen erweitert. Dass sich der Koopman-Operator allgemein auch auf Systeme mit Eingängen erweitern lässt, wurde bereits von Proctor et al. [93] gezeigt. Das Ziel dieses Abschnittes soll es daher sein, getroffene Annahmen zu Begründen und zu plausibilisieren und wichtige Eigenschaften des Koopman-Operators angeregter Systeme heraus zu arbeiten. Dabei steht insbesondere die Stabilität des geschlossenen Regelkreises mittels Zustandsrückführung im Vordergrund, wenn das nichtlineare System mit einem Zustandsregler auf Basis des Koopman-linearen Systems geregelt wird.

Für die in dieser Arbeit betrachteten Systeme werden zunächst allgemeine nichtlineare Systeme nach

$$\dot{\underline{x}} = \underline{f}(\underline{x}, \underline{u}) \in \mathcal{X} \quad (3.26)$$

sowie ihre zeitdiskrete Darstellung

$$\underline{x}_{k+1} = \underline{f}(\underline{x}_k, \underline{u}_k) \in \mathcal{X} \quad (3.27)$$

mit dem Zustand  $\underline{x} \in \mathcal{X} \subseteq \mathbb{R}^n$  und dem Eingang  $\underline{u} \in \mathcal{U} \subseteq \mathbb{R}^m$  betrachtet. Auch für solche Systeme kann die Systemdynamik mittels Observables  $g(\underline{x}, \underline{u}) : \mathcal{T} \rightarrow \mathcal{Z} \subset \mathbb{R}^{n_e}$  mit  $\mathcal{T} = \{(\underline{x}, \underline{u}) | \underline{x} \in \mathcal{X}_r \subset \mathcal{X}, \underline{u} \in \mathcal{U}_r \subseteq \mathcal{U}\}$  linear beschrieben werden. Die Dimension  $n_e \geq n$  beschreibt dabei die erweiterte Systemordnung. Die zu den Gleichungen (3.3) und (3.4) äquivalente Darstellung für angeregte Systeme lautet

$$\mathcal{K}_{\underline{u}(t)}^t g(\underline{x}(t)) = g \circ \underline{S}(t, \underline{u}(t), \underline{x}(t)) \quad (3.28)$$

für zeitkontinuierliche beziehungsweise

$$\mathcal{K}_{\underline{u}_k}^k g(\underline{x}_k) = g \circ \underline{S}(k, \underline{u}_k, \underline{x}_k) \quad (3.29)$$

für zeitdiskrete angeregte dynamische Systeme. Der Index  $\underline{u}(t)$  beziehungsweise  $\underline{u}_k$  beschreibt einen Koopman-Operator zu einem bestimmten Zeitpunkt  $t$  respektive  $k$ , für die  $\underline{u}$  stückweise konstant ist. Der Eingang ist somit als Parameter zu verstehen, sodass die Bewegung eines angeregten Systems durch eine Gruppe von Koopman-Operatoren beschrieben wird. Die Größe  $\underline{S}$  ist wie bereits bei den autonomen Systemen die Lösung der Systemgleichung (3.26) respektive (3.27) des dynamischen Systems, sodass im zeitdiskreten Fall  $\underline{S}$  direkt mit Gleichung (3.27) ersetzt werden kann. Alternativ kann nach [78] und [93] für zeitdiskrete Systeme im Allgemeinen auch

$$\mathcal{K}^k g(\underline{x}_k, \underline{u}_k) = g(\underline{S}(k, \underline{x}_k, \underline{u}_k), \underline{u}_{k+1}) \quad (3.30)$$

geschrieben werden. Durch Gleichung (3.30) wird insbesondere klar, dass die Observables für angeregte Systeme im allgemeinen Funktionen des Zustandes und des Eingangs sind. Mit den dargestellten Grundlagen soll nun die lineare Zustandsgleichung auf Basis der Koopman-Operator-Theorie hergeleitet werden.

Für jedes angeregte nichtlineare System (3.26) kann ohne Einschränkung der Allgemeinheit eine aufgeteilte Form

$$\begin{aligned} \dot{\underline{x}} &= \underline{f}_{\underline{x}, \underline{u}}(\underline{x}, \underline{u}) + \underline{F}_{\underline{x}}(\underline{x}) \underline{f}_{\underline{u}}(\underline{u}) \\ &= \underline{f}_{\underline{x}, \underline{u}}(\underline{x}, \underline{u}) + \sum_{i=1}^{m_{\underline{x}}} f_{\underline{u}, i}(\underline{u}) \underline{f}_{\underline{x}, i}(\underline{x}) \end{aligned} \quad (3.31)$$

hergeleitet werden, wobei  $f_{\underline{x},i}$  und  $f_{\underline{u},i}$  jeweils die  $i$ -te Spalte von  $\underline{F}_{\underline{x}}$  beziehungsweise das  $i$ -te Element von  $\underline{f}_{\underline{u}}$  ist. Die Größe  $m_{\underline{x}}$  gibt die Anzahl der Eingangsfunktionen  $f_{\underline{u},i}$  an, die im Allgemeinen von der Anzahl der Eingänge  $m$  abweichen kann. Für die zeitliche Ableitung des Observable-Vektors  $\underline{g}_{\underline{z}}(\underline{x}, \underline{u})$  gilt mit Gleichung (3.31)

$$\begin{aligned} \dot{\underline{g}}_{\underline{z}}(\underline{x}, \underline{u}) &= \nabla_{\underline{x}} \underline{g}_{\underline{z}}(\underline{x}, \underline{u}) \left( \underline{f}_{\underline{x},\underline{u}}(\underline{x}, \underline{u}) + \sum_{i=1}^{m_{\underline{x}}} f_{\underline{u},i}(\underline{u}) \underline{f}_{\underline{x},i}(\underline{x}) \right) + \nabla_{\underline{u}} \underline{g}_{\underline{z}}(\underline{x}, \underline{u}) \dot{\underline{u}} \\ &= L_{\underline{f}_{\underline{x},\underline{u}},\underline{x}} \underline{g}_{\underline{z}}(\underline{x}, \underline{u}) + \sum_{i=1}^{m_{\underline{x}}} f_{\underline{u},i}(\underline{u}) L_{\underline{f}_{\underline{x},i},\underline{x}} \underline{g}_{\underline{z}}(\underline{x}, \underline{u}) + L_{\underline{u},\underline{u}} \underline{g}_{\underline{z}}(\underline{x}, \underline{u}). \end{aligned} \quad (3.32)$$

Außerdem wird für obige Gleichung Folgendes angenommen.

**Annahme 3.4.** Der Einfluss des Terms  $L_{\underline{u},\underline{u}} \underline{g}_{\underline{z}}(\underline{x}, \underline{u})$  auf Gleichung (3.32) ist gegenüber den anderen Termen vernachlässigbar.

Wie in Abschnitt 2.1 beschrieben, handelt es sich bei der Lie-Ableitung um einen ( $\mathbb{R}$ -) linearen Operator. Dieser soll daher im Folgenden als Matrix approximiert werden, sodass eine Matrixmultiplikation der Approximation mit  $\underline{g}_{\underline{z}}(\underline{x}, \underline{u})$  möglich wird. Mit den Substitutionen  $\underline{z} = \underline{g}_{\underline{z}}(\underline{x}, \underline{u})$ , den Approximationen  $L_{\underline{f}_{\underline{x},\underline{u}},\underline{x}} \approx \underline{\tilde{K}}_{\underline{z}}$  und  $L_{\underline{f}_{\underline{x},i},\underline{x}} \approx \underline{\tilde{K}}_{\underline{w},i}$  sowie Annahme 3.4 ergibt sich somit für Gleichung (3.32)

$$\dot{\underline{z}} = \underline{\tilde{K}}_{\underline{z}} \underline{z} + \sum_{i=1}^{m_{\underline{x}}} f_{\underline{u},i}(\underline{u}) \underline{\tilde{K}}_{\underline{w},i} \underline{z}. \quad (3.33)$$

Anhand dieser Gleichung wird deutlich, dass, wenn Observables des Eingangs  $\underline{w} = \underline{g}_{\underline{w}}(\underline{u})$  verwendet werden, die wahren Nichtlinearitäten des Eingangs gute Kandidaten für die Wahl dieser Observables sind<sup>12</sup>. Die den endlichdimensionalen Unterraum zum Koopman-Operator aufspannenden Funktionen werden daher im Weiteren zu

$$\{g_{\underline{z},i}(\underline{x}, \underline{u})\}_{i=1}^{n_e} \cup \{g_{\underline{w},i}(\underline{u}) = f_{\underline{u},i}(\underline{u})\}_{i=1}^{m_{\underline{x}}} \quad (3.34)$$

gewählt. Mit dieser Basis ergibt sich dann das nicht mehr bilineare System

$$\dot{\underline{z}} = \underline{K}_{\underline{z}} \underline{z} + \underline{K}_{\underline{w}} \underline{w}. \quad (3.35)$$

<sup>12</sup>Wichtig ist an dieser Stelle zu erwähnen, dass die Observables des Einganges keine Observables im Sinne der Koopman-Theorie sind. Dies ist darin begründet, dass diese Observables keine eigene Dynamik besitzen, die linear über den Koopman-Operator ausgedrückt werden kann.

mit  $\underline{z} = \underline{g}_z(\underline{x}, \underline{u})$  und  $\underline{w} = [f_{u,1}(\underline{u}), \dots, f_{u,m_x}(\underline{u})]^T$ . Ähnliche Systemansätze für eingangsaffine Systeme sind unter anderem in [53] und [78] zu finden. Dabei kann  $\underline{K}_w$  auch als konstante Approximation zur Matrix  $[\underline{K}_{w,1}\underline{z}, \dots, \underline{K}_{w,m_x}\underline{z}]$  beziehungsweise der Lie-Ableitungen  $[L_{f_{x,1},x}\underline{g}_z(\underline{x}, \underline{u}), \dots, L_{f_{x,m_x},x}\underline{g}_z(\underline{x}, \underline{u})]$  interpretiert werden. Diese Basis ist analog für zeitdiskrete Systeme in Anhang A.1.3 hergeleitet. Für die weiteren Untersuchungen wird ein zu Gleichung (3.35) analoges diskretes System

$$\underline{z}_{k+1} = \underline{K}_z \underline{z}_k + \underline{K}_w \underline{w}_k. \quad (3.36)$$

verwendet. Um einer Überbestimmtheit des Eingangs entgegenzuwirken, wird im Folgenden angenommen, dass  $m_x = m$  gilt. Eine einzelne Eingangsgröße soll also lediglich in einer Ausprägung auftreten. Somit sind beispielsweise Systeme mit Eingangstermen  $u_1$  und  $u_1^2$  innerhalb der Systemgleichung ausgeschlossen<sup>13</sup>. Des Weiteren soll im Verlauf der Arbeit die folgende Annahme gelten.

**Annahme 3.5.** Die Observables  $\underline{z} = \underline{g}_z(\underline{x}, \underline{u})$  sind lediglich Funktionen des Zustandes und nicht des Einganges. Es gilt somit  $\underline{z} = \underline{g}_z(\underline{x}) \neq \underline{f}(\underline{u})$ .

Diese Annahme lässt sich auch mathematisch begründen. Besonders anschaulich ist dies im Zeitdiskreten, da die Observables dort über  $\underline{z}_k = \underline{g}_z(\underline{x}_k, \underline{u}_k)$  definiert sind. Schreibt man nun die Dynamik mit dieser Definition, so ergibt sich

$$\underline{z}_{k+1} = \underline{g}_z(\underline{x}_{k+1}, \underline{u}_{k+1}) \approx \underline{K}_z \underline{g}_z(\underline{x}_k, \underline{u}_k) + \underline{K}_w \underline{g}_w(\underline{x}_k, \underline{u}_k).$$

Da der Eingang keine eigene Dynamik besitzt, ist diese Gleichung nicht erfüllbar. Zusätzlich würde das bedeuten, dass der Zustand vom Eingang des nächsten Zeitschrittes abhängt, was der zugrunde gelegten Basisgleichung (3.27) widerspricht. Diese Idee lässt sich einfach auf zeitkontinuierliche Systeme übertragen, denn für ein infinitesimales Inkrement  $dt$  gilt

$$\underline{z}(t + dt) \approx \underline{z}(t) + \dot{\underline{z}}(t) dt \approx \underline{z}(t) + (\underline{K}_z \underline{z}(t) + \underline{K}_w \underline{w}(t)) dt.$$

Somit gilt hier mit  $\underline{z} = \underline{g}_z$  analog die Unabhängigkeit der Observables  $\underline{g}_z$  von  $\underline{u}$ , da  $\underline{z}(t + dt)$  unabhängig von  $\underline{u}(t + dt)$  ist.

Bei genauerer Betrachtung für den zeitdiskreten sowie den zeitkontinuierlichen Fall fällt auf, dass die Observables des Eingangs  $\underline{g}_w$  sehr wohl Funktionen des Zustandes  $\underline{x}$  und des Eingangs  $\underline{u}$  sein

<sup>13</sup>Grundsätzlich lässt die Koopman-Theorie eine Erweiterung des Eingangsraumes mit der Dimension  $m_e \geq m$  zu. Da jedoch durch eine solche Erweiterung die Reglerberechnung erheblich verkompliziert wird und keine linearen Regelungsverfahren ohne weitere Vorkehrungen verwendet werden können, werden auch solche in dieser Arbeit ausgeschlossen. Grundsätzlich kann es jedoch, insbesondere für die Prädiktion, Vorteile haben, den Eingangsraum in seiner Dimension zu erweitern, wie in [18] gezeigt wird.

können, wie die Gleichungen (3.32) und (3.33) zeigen. Da in dieser Arbeit jedoch gewöhnliche Zustandsregler verwendet werden sollen, werden auch die Eingangsobservables als Funktionen angenommen, welche ausschließlich vom Eingang abhängen.

### 3.2.1 Eigenschaften angeregter Koopman-linearisierbarer Systeme

Wie aus den voran gegangenen Abschnitten klar wird, bietet der Koopman-Operator die Möglichkeit der Anwendung linearer systemtheoretischer Methoden und, bei vorhandenem Eingang, linearen Regelungsstrategien auf nichtlineare, dynamische Systeme. Für den Fall autonomer Systeme existiert mindestens ein System (Gleichung (3.5)), welches über eine endlichdimensionale, zum Koopman-Operator invariante Darstellung verfügt. Auch mit einer additiven linearen Anregung  $u$  in der Ableitung des zweiten Zustandes existiert diese noch. Dieses System ist jedoch in dieser Konfiguration nicht vollständig steuerbar. Für vollständig steuerbare nichtlineare Systeme kann keine Koopman-invariante Darstellung nach (3.36) gefunden werden. Ein Nachweis, dass eine solche Linearisierung nicht global existieren kann, wurde bereits in [5] geführt. Für die in dieser Arbeit angegebene Herleitung, welche die Existenz auch für lokal gültige invariante Koopman-Linearisierungen nach Gleichung (3.36) ausschließt, wird zunächst die folgende Randbedingungen definiert.

**Definition 3.6** (Lokale Steuerbarkeit). Für ein nichtlineares System nach (3.27) existiert ein Raum  $\mathcal{X} \subseteq \mathbb{R}^n$  mit  $\underline{0}$  als inneren Punkt, sodass jeder Zustand  $\underline{x}_2 \in \mathcal{X}$ , ausgehend von jedem Punkt  $\underline{x}_1 \in \mathcal{X}$ , mit einer geeigneten Steuerfolge  $\underline{u}_k \in \mathcal{U}$  mit  $k = [0, \dots, T]$  in endlicher Zeit  $T$  erreicht werden kann, ohne dass der Raum  $\mathcal{X}$  verlassen wird.

Weiter gelte folgende Definition.

**Definition 3.7** (Unreduzierbare Koopman-Linearisierungen). Die Koopman-Linearisierung nach Gleichung (3.36) heißt unreduzierbar, wenn keine lineare, reguläre Transformationsmatrix  $\underline{T} \in \mathbb{R}^{n_e \times n_e}$  existiert, sodass mindestens ein Element von  $\underline{Tg}_z(\underline{x})$  für alle  $\underline{x} \in \mathcal{X}$  null ist.

Dies schließt überflüssige, nicht angeregte Observables aus, welche keinen Einfluss auf die Systemdynamik haben. Weiter ergibt sich daraus der folgende Folgesatz.

**Folgesatz 3.8.** Eine nicht-unreduzierbare Koopman-Linearisierung kann immer in einer unreduzierbare Koopman-Linearisierung überführt werden.

Dies kann einfach durch das Finden einer Matrix  $\underline{T}$  bewerkstelligt werden, für das mindestens ein Element von  $\underline{Tg}_z(\underline{x})$  auf dem gesamten  $\mathcal{X}$  null wird. Diese Elemente werden für die nachfolgenden Schritte ignoriert. Für die weiteren Ausführungen ist außerdem das folgende Lemma von besonderer Bedeutung.

**Lemma 3.9** (Satz der Invarianz der Dimension [21]). Es ist nicht möglich, zwischen einer  $n$ -dimensionalen Mannigfaltigkeit und einer  $n + r = n_e$ -dimensionalen Mannigfaltigkeit mit  $r > 0$  eine eindeutige und stetige Abbildung zu finden.

Wichtig ist dabei der Geltungsbereich über Mannigfaltigkeiten. Ein einfaches Beispiel dafür ist die Funktion  $\underline{f}(\underline{x})^T = [x, x^2]$  mit  $x \in \mathbb{R}^+$ . Diese Funktion ist nur so lange bijektiv solange man sich auf der Linie  $[x, x^2]$  im  $\mathbb{R}^2$  bewegt. Weicht man von dieser Linie ab, ist die Funktion nicht mehr bijektiv. Dies liegt am oben angegebenen Lemma und dem Sachverhalt, dass diese Linie eine eindimensionale Mannigfaltigkeit ist.

Mit den angegebenen Annahmen und Definitionen sowie Lemma 3.9 lässt sich nun beweisen, dass für vollständig steuerbare Systeme keine exakt gültige Koopman-Linearisierung mit  $n_e > n$  existieren kann.

**Satz 3.10** (Nichtexistenz von Koopman-Linearisierungen mit Ordnungsanhebung vollständig steuerbarer, nichtlinearer, dynamischer Systeme [5]). *Für ein nichtlineares System (3.27) gelte Annahme 3.6. Es sei eine stetige bijektive Transformationsvorschrift  $\underline{z}_k = \underline{g}_z(\underline{x}_k) : \mathbb{R}^n \rightarrow \mathcal{Z} \subset \mathbb{R}^{n_e}$  mit  $n_e > n$  und  $\underline{g}_z(\underline{0}) = \underline{0}$ , sowie eine weitere stetige bijektive Transformation  $\underline{w}_k = \underline{g}_w(\underline{u}_k) : \mathcal{U}_r \subseteq \mathbb{R}^m \rightarrow \mathcal{W} \subseteq \mathbb{R}^m$ , für die  $\underline{0}$  ein innerer Punkt von  $\mathcal{W}$  ist, definiert. Für diese Transformationen, existiert unter Definition 3.7 keine unreduzierbare Koopman-Linearisierung zum nichtlinearen System.*

*Beweis.* Sei  $\underline{x}_0 \neq \underline{0} \in \mathcal{X}$  ein Startwert des nichtlinearen Systems (3.27). Da das System gemäß Annahme 3.6 lokal auf  $\mathcal{X}$  steuerbar ist, kann eine Ruhelage  $\underline{x}_{AP,i} \in \mathcal{X}$  durch eine geeignete Eingangstrajektorie in endlicher Zeit erreicht werden. Angenommen es existiert eine im Satz angegebene Transformation mit linearem Zustandsraumsystem, dann muss auch im transformierten Raum der Punkt  $\underline{g}_z(\underline{x}_{AP,i})$  ausgehend von  $\underline{g}_z(\underline{x}_0)$  in endlicher Zeit erreichbar sein. Angenommen  $(\underline{K}_z, \underline{K}_w)$  sei nicht vollständig steuerbar, dann lässt sich eine Kalman-Zerlegung

$$\begin{bmatrix} \underline{z}_{1,k+1} \\ \underline{z}_{2,k+1} \end{bmatrix} = \begin{bmatrix} \underline{K}_{z,11} & \underline{K}_{z,12} \\ \underline{0} & \underline{K}_{z,22} \end{bmatrix} \begin{bmatrix} \underline{z}_{1,k} \\ \underline{z}_{2,k} \end{bmatrix} + \begin{bmatrix} \underline{K}_{w,1} \\ \underline{0} \end{bmatrix} \underline{w}_k \quad (3.37)$$

finden. Angenommen mindestens ein Eigenwert von  $\underline{K}_{z,22}$  sei ungleich null, dann kann das System die Ruhelage nicht mehr in endlicher Zeit erreichen, da nach Annahme 3.7 mindestens für ein  $\underline{x}_0$  diese Mode nur durch einen Anfangswert angeregt ist. Das lineare System muss also vollständig steuerbar sein.

Angenommen alle Eigenwerte von  $\underline{K}_{z,22}$  seien gleich null. Dann ist wäre möglich, die Ruhelage in genau  $n_e$  Zeitschritten und somit in endlicher Zeit, zu erreichen. Jedoch ist es dann nicht mehr möglich von dort aus den Anfangszustand  $\underline{x}_0$  und damit  $\underline{g}_z(\underline{x}_0)$  zurück zu kommen, was aber aufgrund der vollständigen Steuerbarkeit des nichtlinearen Systems möglich sein muss. Daher muss das lineare System wieder vollständig steuerbar sein.

Wenn die Koopman-Linearisierung aber vollständig steuerbar wäre, so wäre es möglich aus  $\underline{z} = \underline{0}$  eine  $n_e$ -dimensionale Mannigfaltigkeit zu erreichen. Die Transformation  $\underline{g}_z$  würde somit bijek-

tiv eine  $n$ - in eine  $n_e$ -dimensionale Mannigfaltigkeit transformieren. Dies würde das Lemma 3.9 widersprechen<sup>14</sup>.  $\square$

Diese Erkenntnis ist insbesondere für nichtlineare Systeme mit mehreren Ruhelagen anschaulich, da lineare Systeme nur eine Ruhelage besitzen, die stets in  $\underline{x} = \underline{0}$  angenommen werden kann. Eine Approximation von mehreren Ruhelagen über ein lineares Modell ist nur in seltenen Ausnahmefällen möglich. Einer dieser Fälle ist der in Abschnitt 5.1.2 besprochene Duffing Oszillator.

Auch wenn dieses Ergebnis zunächst scheinbar in starker Konkurrenz zu einer Vielzahl an Veröffentlichungen steht, so haben diese jedoch gezeigt, dass es trotz Satz 3.10 möglich ist zumindest näherungsweise invariante Unterräume zum Koopman-Operator mit der Koopman-Linearisierung nach Gleichung (3.36) zu approximieren. Sie kann jedoch niemals zu einer exakten Darstellung konvergieren. An dieser Stelle wird klar, warum sich auch in der grundlegenden Definition 1.1 der Systemidentifikation auf zeitlich oder räumlich begrenzte Systeme beschränkt wird. Es ist in vielen Situationen, wie auch dieser, schlicht nicht möglich, global beziehungsweise zeitlich und räumlich unbegrenzt gültige Systeme zu schätzen. Des Weiteren ist die Ermittlung von beschränkt gültigen Systemen häufig einfacher, da man sich auch ohne die Koopman-Theorie auf Bereiche festlegen kann, die in guter Näherung linear darstellbar sind.

Mit einer Schätzung eines Systems geht in den meisten Fällen ein Schätzfehler einher. Da im angeregten Fall mit dem angenommenen Systemmodell unter Nutzung der Koopman-Operator-Theorie immer ein Fehler vorhanden sein wird, sind die nachfolgenden Ergebnisse um so wichtiger für den weiteren Verlauf der Arbeit. Die in Abschnitt 3.1.1 ausgeführten Herleitungen werden im Folgenden auf Systeme mit Eingängen erweitert. Für die Zustandsobservables gilt dabei unverändert Gleichung (3.17). Für den Eingangstransformationsfehler gilt

$$\left( \underline{g}_w \circ \bar{\underline{g}}_w \right) (\underline{w}_k) = \underline{w}_k + \epsilon_{w,trans} (\underline{u}_k). \quad (3.38)$$

Des Weiteren wird ein Ersatzsystem

$$\underline{x}_{k+1} = \bar{\underline{g}}_z \left( \underline{K}_z \underline{g}_z (\underline{x}_k) + \underline{K}_w \underline{g}_w (\underline{u}_k) \right) + \epsilon_{nl} (\underline{x}_k, \underline{u}_k) \quad (3.39)$$

für die Systemdynamik aus (3.27) definiert. Wie bereits beim autonomen Fall kann jedes nichtlineare System nach (3.27) in eine solche Form gebracht werden kann, da der Fehlerterm  $\epsilon_{nl} (\underline{x}_k, \underline{u}_k)$  sämtliche Fehler beinhaltet. Gleichung (3.39) wird mittels der Zustandsobservable  $\underline{g}_z (\cdot)$  zu

$$\underline{g}_z (\underline{x}_{k+1}) = \underline{g}_z \left( \bar{\underline{g}}_z \left( \underline{K}_z \underline{g}_z (\underline{x}_k) + \underline{K}_w \underline{g}_w (\underline{u}_k) \right) + \epsilon_{nl} (\underline{x}_k, \underline{u}_k) \right) \quad (3.40)$$

<sup>14</sup>Dieser Satz gilt nicht für unendlichdimensionale Systeme, da Lemma 3.9 eine Schlussfolgerung des *invariance of domain*-Theorems ist und dieses nicht für unendlichdimensionale Räume gilt [115]. Damit steht dies nicht in Konkurrenz zur Koopman-Operator-Theorie im Allgemeinen.

geschrieben. Unter Zuhilfenahme der Abkürzung  $\tilde{\underline{x}}_{k+1} = \bar{\underline{g}}_z \left( \underline{K}_z \underline{g}_z(\underline{x}_k) + \underline{K}_w \underline{g}_w(\underline{u}_k) \right)$  sowie der Taylorreihen-Approximation von  $\underline{g}_z(\underline{x}_{k+1})$  aus Gleichung (3.40) um  $\tilde{\underline{x}}_{k+1}$  ergibt sich der Ausdruck<sup>15</sup>

$$\begin{aligned} \underline{g}_z(\underline{x}_{k+1}) &= \underline{g}_z(\tilde{\underline{x}}_{k+1}) + \nabla_{\underline{x}} \underline{g}_z(\underline{x}_k) (\underline{x}_{k+1} - \tilde{\underline{x}}_{k+1}) + \underline{R}(\underline{x}_k, \underline{u}_k) \\ &\quad \underline{K}_z \underline{g}_z(\underline{x}_k) + \underline{K}_w \underline{g}_w(\underline{u}_k) + \underline{\epsilon}_{z,\text{trans}}(\underline{x}_k, \underline{u}_k) \\ &\quad + \nabla_{\underline{x}} \underline{g}_z(\underline{x}_k) \underline{\epsilon}_{\text{nl}}(\underline{x}_k, \underline{u}_k) + \underline{R}(\underline{x}_k, \underline{u}_k) \\ &= \underline{K}_z \underline{g}_z(\underline{x}_k) + \underline{K}_w \underline{g}_w(\underline{u}_k) + \underline{\epsilon}_{\text{lin}}(\underline{x}_k, \underline{u}_k). \end{aligned} \quad (3.41)$$

Wird das System nun mit einem Koopman-Zustandsregler

$$\underline{u}_k = \bar{\underline{g}}_w \left( -\underline{C} \underline{g}_z(\underline{x}_k) \right) \quad (3.42)$$

nach [18] geregelt, ergibt sich mit den definierten Transformationsfehlern aus den Gleichungen (3.17) und (3.38)

$$\begin{aligned} \underline{x}_{k+1} &= \bar{\underline{g}}_z \left( \underline{K}_z \underline{g}_z(\underline{x}_k) + \underline{K}_w \underline{g}_w \left( \bar{\underline{g}}_w \left( -\underline{C} \underline{g}_z(\underline{x}_k) \right) \right) \right) + \underline{\epsilon}_{\text{nl}} \left( \underline{x}_k, \bar{\underline{g}}_w \left( -\underline{C} \underline{g}_z(\underline{x}_k) \right) \right) \\ &= \bar{\underline{g}}_z \left( \left( \underline{K}_z - \underline{K}_w \underline{C} \right) \underline{g}_z(\underline{x}_k) + \underline{\epsilon}_{w,\text{trans}} \left( -\underline{C} \underline{g}_z(\underline{x}_k) \right) \right) + \underline{\epsilon}_{\text{nl}} \left( \underline{x}_k, -\underline{C} \underline{g}_z(\underline{x}_k) \right) \\ &= \bar{\underline{g}}_z \left( \left( \underline{K}_z - \underline{K}_w \underline{C} \right) \underline{g}_z(\underline{x}_k) + \underline{\epsilon}_{w,\text{trans}}(\underline{x}_k) \right) + \tilde{\underline{\epsilon}}_{\text{nl}}(\underline{x}_k) \end{aligned} \quad (3.43)$$

im Zustandsraum beziehungsweise

$$\begin{aligned} \underline{g}_z(\underline{x}_{k+1}) &= \left( \underline{K}_z - \underline{K}_w \underline{C} \right) \underline{g}_z(\underline{x}_k) + \underline{\epsilon}_{w,\text{trans}}(\underline{x}_k) + \underline{\epsilon}_{\text{lin}} \left( \underline{x}_k, -\underline{C} \underline{g}_z(\underline{x}_k) \right) \\ &= \left( \underline{K}_z - \underline{K}_w \underline{C} \right) \underline{g}_z(\underline{x}_k) + \underline{\epsilon}(\underline{x}_k) \end{aligned} \quad (3.44)$$

im Koopman-Raum. Das heißt, dass selbst bei vorhandenem Koopman-Zustandsregler das System in die Form aus den Sätzen 3.2 und 3.3 gebracht werden kann. Weiter bedeutet das, dass wenn der Koopman-Zustandsregler  $\underline{C}$  so gewählt wird, dass das geregelte lineare Koopman-Zustandsraumsystem  $\underline{K}_z - \underline{K}_w \underline{C}$  nur Eigenwerte innerhalb des Einheitskreises besitzt, die Stabilität des nichtlinearen System gesichert ist, solange der Fehler in definierten Grenzen bleibt. Dies bildet eine der essenziellen Erkenntnisse der vorliegenden Arbeit.

Sicherlich gibt diese Herleitung keinen Aufschluss über die Dynamik, welche mittels des Koopman-Reglers nach Gleichung (3.42) am nichtlinearen System eingestellt wird, jedoch kann so ge-

<sup>15</sup>Der in Gleichung (3.41) zusammengefasste Fehler  $\underline{\epsilon}_{\text{lin}}(\underline{x}_k, \underline{u}_k)$  hängt unter anderem vom Gradienten  $\nabla_{\underline{x}} \underline{g}_z$  von  $\underline{g}_z$  ab. Dies kann gegebenenfalls zu großen Fehlern in bestimmten Bereichen des Zustandsraumes führen.

zeigt werden, dass ein Koopman-Regler, trotz der möglicherweise komplexen nichtlinearen Transformation in  $\underline{g}_z$  das nichtlineare System zumindest nicht destabilisiert, was eine Grundvoraussetzung für die Regelung von Systemen ist. Die besprochenen und erarbeiteten Erkenntnisse werden im folgenden Satz zusammengefasst.

**Satz 3.11** (Stabilität nichtlinearer, über lineare Koopman-Zustandsregler geregelter, dynamischer Systeme). *Sein  $\underline{g}_z(\underline{x}) : \mathcal{X}_r \rightarrow \mathcal{Z}$  mit  $\mathcal{X}_r \subset \mathbb{R}^n$  und  $\mathcal{Z} \subset \mathbb{R}^{n_e}$  mit  $n_e \geq n$  ein Vektor von Observables des Zustandes mit  $\underline{g}_z(\underline{0}) = \underline{0}$  mit einer lokal auf  $\mathcal{X}_r$  gültigen Umkehrfunktion  $\bar{\underline{g}}_z(\underline{z}) : \mathcal{Z} \subset \mathbb{R}^{n_e} \rightarrow \mathcal{X}_r \subset \mathbb{R}^n$  mit  $\bar{\underline{g}}_z(\underline{0}) = \underline{0}$ . Weiter sein  $\underline{g}_w(\underline{u}) : \mathcal{U}_r \subseteq \mathcal{U} \subseteq \mathbb{R}^m \rightarrow \mathcal{W} \subseteq \mathbb{R}^{m_e}$  Observables des Eingangs mit  $\underline{g}_w(\underline{0}) = \underline{0}$  mit einer lokal auf  $\mathcal{U}_r$  gültigen Umkehrfunktion  $\bar{\underline{g}}_w(\underline{w}) : \mathcal{W} \subseteq \mathbb{R}^{m_e} \rightarrow \mathcal{U}_r \subseteq \mathbb{R}^m$  mit  $\bar{\underline{g}}_w(\underline{0}) = \underline{0}$ .*

Sei

$$\underline{x}_{k+1} = \underline{f}(\underline{x}_k, \underline{u}_k)$$

ein nichtlineares System, welches über einen Koopman-Zustandsregler  $\underline{u}_k = \bar{\underline{g}}_w(-\underline{C}\underline{g}_z(\underline{x}_k))$  auf Basis des Koopman-linearen Systems  $\underline{z}_{k+1} = \underline{K}_z \underline{z}_k + \underline{K}_w \underline{w}_k$  geregelt wird. Weiter wird die Lyapunov-Funktion  $V(\underline{x}_k) = \underline{g}_z(\underline{x})^T \underline{Q} \underline{g}_z(\underline{x})$  definiert.

Die Reglermatrix  $\underline{C}$  wird so gewählt, dass  $\underline{K}_z - \underline{K}_w \underline{C} = \bar{\underline{K}}_z$  nur Eigenwerte  $|\lambda(\bar{\underline{K}}_z)| < 1$  besitzt. Gilt dann  $0 \leq \alpha < -\|\bar{\underline{K}}_z\|_2 + \sqrt{\|\bar{\underline{K}}_z\|_2^2 + \frac{\lambda_{\min}(\underline{P})}{\lambda_{\max}(\underline{Q})}}$  mit  $\|\underline{\epsilon}(\underline{0})\|_2 = 0$ ,  $\underline{P}$  sowie  $\underline{Q}$  aus der zeitdiskreten Lyapunov-Gleichung  $\underline{K}_z^T \underline{Q} \underline{K}_z - \underline{Q} = -\underline{P}$  und

$$\sup_{\forall \underline{x}_k \in \mathcal{X} \setminus \{0\}} \frac{\|\underline{\epsilon}(\underline{x}_k)\|}{\|\underline{g}_z(\underline{x}_k)\|} \leq \alpha,$$

beziehungsweise

$$\left\| \left( c\underline{I} - \underline{K}_z \right)^{-1} \right\|_{\mathcal{H}_\infty} < \frac{1}{\gamma}$$

mit

$$\sup_{\forall \underline{x}_k \in \mathcal{X} \setminus \{0\}} \frac{\|\underline{\epsilon}(\underline{x}_k)\|_2}{\|\underline{g}_z(\underline{x}_k)\|_2} < \gamma,$$

und  $\lambda_i(\underline{Q}) \neq 1$  mit  $i \in [1, \dots, n_e]$  wobei  $\mathcal{X}_r$  positiv invariant ist, dann ist das geregelte nichtlineare System stabil im Sinne von Lyapunov.

*Beweis.* Dies ergibt sich aus den vorherigen Ausführungen mit den Gleichungen (3.38)-(3.44) sowie den Beweisen zu den Hilfssätzen 3.2 und 3.3.  $\square$

Selbes gilt für zeitkontinuierliche Systeme mit Satz 3 aus [71].

Wichtig für die praktische Schätzung sowie Regelung nichtlinearer Systeme ist, dass obiger Satz unabhängig von einer zeitlich, über einen gewissen Zeithorizont, begrenzten Gültigkeit des Modells, also einer näherungsweise Prädiktionsinvarianz nach Definition 2.2, ist. Es ist nur ausschlaggebend, dass der Fehlerterm innerhalb eines Zeitschrittes nicht über einen maximal zulässigen Fehler hinaus geht.

Der vorangegangene Abschnitt legt die Grundlage für sämtliche weiteren Untersuchungen. Die Sätze 3.2 und 3.3 sowie alle Erweiterungen stellen, unter gewissen Voraussetzungen, sicher, dass lineare Zustandsregler auf Basis von Koopman-Linearisierungen das nichtlineare, dynamische System nicht destabilisieren, was eine Grundvoraussetzung für Regler im Allgemeinen ist. Diese Erkenntnis macht es erst gesichert möglich, lineare Zustandsregler auf Basis geschätzter Koopman-linearer Systeme aufzubauen, was zuvor auch ohne diese theoretische Basis in beispielsweise [27, 54, 70] durchgeführt wurde, um einige zu nennen.

### 3.3 Systemdynamische Eigenschaften eines Koopman-linearen Systems

Der folgende Abschnitt soll Eigenschaften wie die Beobachtbarkeit und die Steuerbarkeit des Koopman-Operators ermitteln und legt somit eine weitere Grundlage für die im späteren Verlauf dieser Arbeit angewendeten Konzepte zur Identifikation und Regelung nichtlineare Systeme mittels geschätzter Koopman-linearer Modelle.

#### 3.3.1 Beobachtbarkeit des Koopman-linearen Systems

Zur besseren Verständlichkeit sei zunächst die Beobachtbarkeit von dynamischen Systemen definiert

**Definition 3.12** ([2, S. 526]). Ein System nach Gleichung (3.26) mit einer Ausgangsgleichung  $\underline{y} = \underline{h}(\underline{x}, \underline{u})$  heißt beobachtbar, wenn jeder beliebige Anfangszustand  $\underline{x}(t_0) \in \mathcal{X}$  durch Kenntnis des Ausgangsgrößenverlaufes  $\underline{y}(t)$  und des Eingangsgrößenverlaufes  $\underline{u}(t) \in \mathcal{U} \subseteq \mathcal{C}^{n-1}$  innerhalb eines endlichen Zeitintervalls  $[t_0, T]$  eindeutig bestimmt werden kann.

Wenn nun alle Systemzustände gemessen werden, ist das System vollständig beobachtbar, denn dann kann schlicht für einen beliebigen Zeitpunkt der Zustand aus den Messungen abgelesen werden. Somit sind in diesem Fall auch alle Eigenfunktionen und damit alle Observables des Koopman-Operators beobachtbar, da sie Funktionen der Zustände sind.

Dies gilt jedoch analog auch für Systeme, bei denen nur ein Teil der Zustände gemessen wird, was man sich anhand der Definition der schwachen Beobachtbarkeit nach [2, S. 530] klar machen kann. Denn können aus der Messung einzelner Zustände und des Eingangs alle Zustände rekonstruiert werden, so ist dies auch für alle Eigenfunktionen und den daraus ableitbaren Observables möglich. Weitere Einblicke für den Fall, dass nicht alle Zustände gemessen werden, finden sich in Abschnitt 3.5.

### 3.3.2 Steuerbarkeit des Koopman-linearen Systems

Um die Ausführungen in diesem Abschnitt zu begründen, wird zunächst die Steuerbarkeit eines Systems definiert.

**Definition 3.13** ([2, S. 175]). Ein System nach Gleichung (3.26) mit wegzusammenhängenden Mengen  $\mathcal{X}$  und  $\mathcal{U}$  heißt vollständig steuerbar, wenn jeder beliebige (Anfangs-)Zustand  $\underline{x}(t_0) \in \mathcal{X}$  durch einen Eingangsgrößenverlauf  $\underline{u}(t) \in \mathcal{U}$  innerhalb eines endlichen Zeitintervalls  $[t_0, T]$  mit einer Trajektorie  $\underline{x}(t) \in \mathcal{X}$  in jeden (End-)Zustand  $\underline{x}(T) \in \mathcal{X}$  überführt werden kann.

Ein angeregtes, nichtlineares, dynamisches System  $\dot{\underline{x}} = \underline{f}(\underline{x}, \underline{u})$  sei vollständig steuerbar und kann über die Transformationsvorschrift

$$\begin{aligned}\underline{g}_{\underline{z}}(\underline{x}) &: \mathcal{X}_r \subset \mathbb{R}^n \rightarrow \mathcal{Z} \subset \mathbb{R}^{n_e} \\ \underline{g}_{\underline{w}}(\underline{u}) &: \mathcal{U}_r \subset \mathbb{R}^m \rightarrow \mathcal{W} \subset \mathbb{R}^{n_e} \\ \underline{z} &= \underline{g}_{\underline{z}}(\underline{x}), \underline{w} = \underline{g}_{\underline{w}}(\underline{u})\end{aligned}$$

in ein lineares Zustandsraumsystem (3.35) mit einer Schätzung der Systemmatrizen überführt werden. Es ist nun wichtig zu verstehen, dass grundsätzlich  $\underline{K}_{\underline{z}} : \mathcal{Z} \rightarrow \mathcal{Z}$  gilt, weiter jedoch  $\underline{K}_{\underline{z}} : \mathcal{Z}_R \subseteq \mathcal{Z} \rightarrow \mathcal{Z}_R$  gelten muss. Dieser Unterraum ist eine, für das Beispiel in Abbildung 3.2 rot dargestellte, Mannigfaltigkeit, denn hier gelten die nichtlinearen Verkopplungen, wie sie in  $\underline{g}_{\underline{z}}$  vorgegeben sind. Damit wird weiter klar, dass  $\mathcal{Z}_R$  ein geregelt invarianter Unterraum von  $\mathcal{Z}$  ist [13, 113], welcher (zunächst) die Dimension  $n_R \leq n$  haben muss, denn es können maximal  $n$  der  $n_e$  Zustände von (3.35) unabhängig gesteuert werden, ohne dabei aus  $\mathcal{Z}_R$  heraus zulaufen. Hierzu wird folgende Heuristik angegeben.

**Heuristik 3.14.** Genau  $n$  Eigenwerte des Koopman-Operators eines vollständig steuerbaren nicht-linearen Systems (3.1) beziehungsweise (3.2) sind steuerbar

*Begründung:* Wenn mehr oder weniger als  $n$  Eigenwerte steuerbar sein sollten, dann können die zugehörigen Eigenfunktionen eine Mannigfaltigkeit erreichen, welche eine höhere Dimension als  $n$  besitzt. Dadurch können die Eigenfunktionen Trajektorien ausführen, die nicht mit dem nichtlinearen System vereinbar sind. Dies würde den Grundgleichungen der Koopman-Operator-Theorie

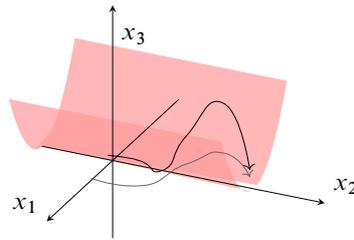


Abbildung 3.2: Visualisierung der Verkopplung zwischen den Zuständen eines Koopman-linearen Systems. Die graue Linie zeigt die Projektion der schwarzen Linie auf die  $x_1 - x_2$ -Ebene, welche sich im Unterraum  $Z_R$  (rote Mannigfaltigkeit) des  $\mathbb{R}^3$  bewegt

(3.28) beziehungsweise (3.29) widersprechen. Durch die unendliche Dimension des Koopman-Operators kann jedoch nicht Lemma 3.9 angewendet werden, um den Ordnungsunterschied in den Mannigfaltigkeiten zu falsifizieren.

Eine direkte Folge ist, dass für asymptotisch stabile endlichdimensionale Koopman-Linearisierungen nach (3.36) nur räumlich und zeitlich begrenzt gültig sein können. Denn sind nur  $n$  Eigenwerte steuerbar, verbleiben bei stabilen Systemen  $n_e - n$  Dynamikanteile in Form von abklingenden oder dauerhaft schwingenden Eigenbewegungen vorhanden. Daher kann das nichtlineare Systemverhalten nur für eine gewisse Zeit aus einem Initialzustand heraus prädiziert werden. Nach einer endlichen Zeit muss das Koopman-lineare System wieder mit dem aktuellen Zustand reinitialisiert werden. Dieser Horizont, für den das Koopman-lineare System gilt, lässt sich mit der in Kapitel 4 vorgestellten Methode mitlernen. Dieser Effekt wird zudem in den Beispielen in Kapitel 5 verdeutlicht.

Aufgrund numerischer Unsicherheiten und Schätzfehlern werden Koopman-lineare Systeme, welche anhand von Messdaten geschätzt werden, jedoch die Eigenschaft nach Heuristik 3.14 nicht eindeutig erfüllen, da der Beweis eine fehlerfreie Schätzung voraussetzt. Dies wird Gegenstand der Untersuchungen in Kapitel 5 sein.

Für die Schätzung und eine damit auszulegende Regelung ist weiterhin folgendes Lemma wichtig.

**Lemma 3.15** ([22]). Eine konstante Observable  $\underline{g}_z(\underline{x}) = \text{const.}$  ist eine Eigenfunktion des Koopman-Operators mit Eigenwert 0 bei zeitkontinuierlichen beziehungsweise 1 bei zeitdiskreten Zustandsraummodellen.

*Beweis.* Für die Eigenfunktiongleichung (3.7) beziehungsweise (3.8) ist das sofort ersichtlich.  $\square$

Eine direkte Folge aus Lemma 3.15 ist, dass eine endlichdimensionale Approximation eines nichtlinearen Systems als Koopman-lineares System, sollte es für eine Regelung verwendet werden, keine konstante Observable  $\underline{g}_z(\underline{x})$  beinhalten darf. Wenn das nichtlineare System keinen konstanten Zustand enthält, dann kann diese Eigenfunktion und der dazugehörige Eigenwert (in 0 im Kontinuierlichen oder 1 im Zeitdiskreten) nicht Teil der  $n$  steuerbaren Eigenwerte sein, da sonst die Steuerbarkeit der eigentlichen Zustände verloren gehen würde. Da es sich dabei um einen

Integrator handelt, ist dieser nicht stabilisierbar [42], was dazu führt, dass kein linear quadratischer Regler berechnet werden kann, denn nach Trentelman [113] muss das Integral des linear quadratischen Reglers (beziehungsweise die Summe im zeitdiskreten Fall) konvergieren. Ist eine instabile Mode  $x_1$  (eines diagonalisierten Systems), und somit auch ein Integrator, nicht steuerbar, das Gesamtsystem also nicht stabilisierbar, dann konvergiert das Optimierungsintegral

$$J = \int_0^{\infty} (x_1^2 + u^2) d\tau$$

nicht. Numerisch ergibt sich die optimale Ansteuerung dann zu  $u = 0$ , was in einer Reglermatrix  $\underline{C} = \underline{0}$  resultiert.

In praktischer Hinsicht ist es daher außerdem äußerst wichtig, dass bei der Normierung der Daten darauf geachtet wird, dass bei mittelwertfreier Anregung die Zustände mittelwertfrei sind, da sonst ein wahrscheinlich nicht stabilisierbarer Integrator mitgeschätzt wird. Dies ist leicht anhand der Gleichung

$$(\underline{x}_{k+1} + \bar{x}) = \underline{A}(\underline{x}_k + \bar{x}) + \underline{B}u_k$$

mit dem konstanten Anteil  $\bar{x}$  erkennbar. Dieser konstante Teil der Gleichung muss invariant unter der Matrix  $\underline{A}$  sein. Dies ist nur der Fall wenn sie mindestens einen Eigenwert in 1 besitzt. Eine Herleitung ist in Anhang A.1.7 gegeben. Kann also kein Gleichanteil in den Zuständen ohne einen Gleichanteil in der Anregung gefunden werden, so besitzt das System keinen Integrator. Wird dieser Eigenwert in 1, respektive 0, nur durch falsch normierte Daten und nicht durch eine im System nicht vorhandene Dynamik induziert, so wird dieser wahrscheinlich bei einer Schätzung nur schlecht oder gar nicht steuerbar sein.

Verbleibt also ein Gleichanteil im Zustandssignal ohne einen solchen Anteil in der Anregung, führt dies für den Fall  $n_e = n$  dazu, dass das Schätzergebnis verfälscht wird. Für den Fall  $n_e > n$  kann es nicht nur zu Verfälschungen der Dynamik, sondern wieder zu der eben beschriebenen Problematik kommen, dass ein nicht stabilisierbarer Integrator geschätzt wird und somit für das System kein linear quadratischer Regler entworfen werden kann.

Die Vermeidung der Konvergenz hin zu solchen Eigenfunktionen oder Observables ist somit zusätzlich eine wichtige Anforderung an den Schätzalgorithmus, da diese keine Dynamikinformation enthalten und somit das Schätzergebnis nur negativ beeinflussen können [73].

### 3.4 Regelung nichtlinearer Systeme mittels Koopman-linearen Systemen

Eines der Ziele des Koopman-Frameworks ist es, nichtlineare Systeme ganzheitlich als lineares Zustandsraumsystem darzustellen. Dies würde die Möglichkeit eröffnen, die lineare System- und



anschauliche Gewichtungsmatrizen gewählt werden. Der Regler ergibt sich dann in Relation zu diesen Gewichten. Hierbei wird das Gütemaß

$$J_{\underline{x},c} = \int_{t=0}^{\infty} \left( \underline{x}(\tau)^T \underline{Q} \underline{x}(\tau) + \underline{u}^T(\tau) \underline{R} \underline{u}(\tau) \right) d\tau \quad (3.45)$$

für zeitkontinuierliche beziehungsweise

$$J_{\underline{x},d} = \sum_{k=0}^{\infty} \underline{x}_k^T \underline{Q} \underline{x}_k + \underline{u}_k^T \underline{R} \underline{u}_k \quad (3.46)$$

für zeitdiskrete Systeme minimiert. Dies geschieht für lineare Zustandsraumssysteme mittels der Riccati-Gleichung. Der große Vorteil in dieser Herangehensweise ist, dass man gegebenenfalls wichtigere Zustände über die Gewichtungsmatrizen  $\underline{Q}$  und  $\underline{R}$  höher gewichten kann, sodass der Regler auf diese mehr Einfluss nimmt und im besten Fall eine höhere Güte, weniger Überschwingen oder andere vorteilhafte Eigenschaften erreichen kann.

Soll diese Methodik nun für Koopman-linearisierte Systeme verwendet werden, ergibt sich die folgende Problematik. Bei einigen Systemen führt die Verwendung der Zustände selbst als Observables zu schlechteren Ergebnissen hinsichtlich der Prädiktionsfähigkeit. Abhilfe kann hier die Transformation des gesamten Zustandes schaffen. Man bewegt sich dann allerdings mit dem Modell in einem anderen Zustandsraum, dessen Zustände nicht ohne Weiteres direkt als Zustandäquivalente interpretierbar sind. Eine sinnvolle beziehungsweise anschauliche Wahl für die Gewichtungsmatrizen ist daher stark erschwert bis unmöglich.

Für zeitdiskrete Koopman-lineare Systeme geht Gleichung (3.46) in

$$J_{\underline{z}} = \sum_{k=0}^{\infty} \underline{g}_{\underline{z}}(\underline{x}_k)^T \underline{Q}_{\underline{z}} \underline{g}_{\underline{z}}(\underline{x}_k) + \underline{g}_{\underline{w}}(\underline{u}_k)^T \underline{R}_{\underline{w}} \underline{g}_{\underline{w}}(\underline{u}_k) \quad (3.47)$$

über. Sind die Systemzustände selbst Zustände des Koopman-linearen Systems, so wählt man die Matrizen gemäß einer anschaulichen Gewichtung zwischen den Zuständen und der Eingänge und wählt den Rest zu null. Für den vollständig kodierten Zustand ist die LQ-Regelung ein bisher wenig betrachtetes Thema in der Literatur. Lediglich Ma et al. nutzt in [70] einen komplexeren LQ-Ansatz mit Nebenbedingung.

Im Folgenden soll eine analytische und anschauliche Herangehensweise vorgestellt werden, wie die Gewichtungsmatrizen  $\underline{Q}_{\underline{z}}$  und  $\underline{R}_{\underline{w}}$  gewählt werden können, wenn die Systemzustände nicht Teil der Koopman-Linearisierung sind. Dies ist insbesondere deshalb wichtig, da eine solche Transformation häufig zu besseren Schätzergebnissen führt und somit ein größerer dynamischer Rückhalt hinsichtlich der echten Dynamik des Systems existiert, wie sich in Kapitel 5 zeigen wird. Dies führt außerdem dazu, dass durch den geringeren Schätzfehler eher eine Stabilitätsgarantie auf Basis von Satz 3.11 gegeben werden kann, als bei einer potenziell schlechteren Schätzung, welche die Zustände als Observables mit einbezieht.

### 3.4.1 Wahl der Gewichtungsmatrizen bei LQ-Regelung für Koopman-lineare Systeme

Nachdem im vorangegangenen Abschnitt die Problematik erläutert wurde, wird in diesem Abschnitt eine Lösung diskutiert. Hierzu werden Linearisierungen der Observables verwendet. Weiter wird davon ausgegangen, dass nur  $n$  Eigenwerte der Koopman-Linearisierung steuerbar sind, wie es auch Heuristik 3.14 besagt. Um konsistent in der Darstellung der Eigendynamik und den  $n$  steuerbaren Eigenwerten zu bleiben, werden zunächst die Eigenfunktionen anstatt der anschaulicheren Observables verwendet. Die vollständige Steuerbarkeit des nichtlinearen Systems wird gemäß Heuristik 3.14 vorausgesetzt. Für die folgenden Ausführungen, werden somit die  $n$  steuerbaren der insgesamt  $n_e$  Eigenfunktionen ausgewählt und in einen Vektor  $\underline{\varphi}_{z,1}(\underline{x}_k) \in \mathcal{Z}_1 \subset \mathbb{C}^n$  zusammengefasst. Die für die Regelung vernachlässigten Eigenfunktionen werden im Vektor  $\underline{\varphi}_{z,2}(\underline{x}_k) \in \mathcal{Z}_2 \subset \mathbb{C}^{n_e-n}$  gelistet. Um die Regelung in dieser Arbeit nicht weiter zu verkomplizieren, wird die Dimension des Eingangsraumes nicht angehoben, da die Funktionen in  $\underline{g}_w$  sonst ein überbestimmtes Gleichungssystem ergeben würden. Eine ausführliche Erläuterung ist weiter unten angegeben.

Da eine zusätzliche Gewichtung nicht steuerbarer Dynamiken beim linear-quadratischen Regelungsentwurf keinen Effekt hat, werden nur die  $n$  steuerbaren Dynamiken gewichtet. Eine ausschließliche Gewichtung nicht steuerbarer Dynamiken führt gar dazu, dass sich für den berechneten Regler  $\underline{C} \approx \underline{0}$  ergibt. Beide Anmerkungen lassen sich leicht mit den Argumenten für die Stabilisierbarkeit aus Abschnitt 3.3.2 validieren.

Damit geht die Optimierungsgleichung in

$$\begin{aligned} J_z &= \sum_{k=0}^{\infty} \begin{bmatrix} \underline{\varphi}_{z,1}(\underline{x}_k)^T & \underline{\varphi}_{z,2}(\underline{x}_k)^T \end{bmatrix} \begin{bmatrix} \underline{Q}_{\varphi,1} & \underline{0} \\ \underline{0} & \underline{0} \end{bmatrix} \cdot \begin{bmatrix} \underline{\varphi}_{z,1}(\underline{x}_k) \\ \underline{\varphi}_{z,2}(\underline{x}_k) \end{bmatrix} + \underline{g}_w(\underline{u}_k)^T \underline{R}_w \underline{g}_w(\underline{u}_k) \\ &= \sum_{k=0}^{\infty} \underline{\varphi}_{z,1}(\underline{x}_k)^T \underline{Q}_{\varphi,1} \underline{\varphi}_{z,1}(\underline{x}_k) + \underline{g}_w(\underline{u}_k)^T \underline{R}_w \underline{g}_w(\underline{u}_k). \end{aligned}$$

über. Nun werden die Eigenfunktionen an einem Tupel  $(\underline{x}_{AP}, \underline{u}_{AP})$  einer Ruhelage linearisiert, das heißt in eine Taylorreihe entwickelt, und nach dem ersten Glied abgebrochen, sodass sich

$$\begin{aligned} J_z &\approx \sum_{k=0}^{\infty} \Delta \underline{x}_k^T \nabla_{\underline{x}} \underline{\varphi}_{z,1}(\underline{x}_{AP})^T \underline{Q}_{\varphi,1} \nabla_{\underline{x}} \underline{\varphi}_{z,1}(\underline{x}_{AP}) \Delta \underline{x}_k \\ &\quad + \Delta \underline{u}^T \nabla_{\underline{u}} \underline{g}_w(\underline{u}_{AP})^T \underline{R}_w \nabla_{\underline{u}} \underline{g}_w(\underline{u}_{AP}) \Delta \underline{u} \end{aligned} \quad (3.48)$$

ergibt.

Gleichung (3.48) hat dieselbe Form wie Gleichung (3.46) für die Linearisierungen nichtlinearer Systeme, da die Zustände direkt gewichtet werden. Lediglich die Gewichtungsmatrizen unterscheiden sich durch eine Art Ähnlichkeitstransformation. So können die Gewichtungsmatrizen

vom Zustandsraum in den Koopman-linearen Zustandsraum nach Koeffizientenvergleich über

$$\begin{aligned}\underline{Q}'_x &= \nabla_{\underline{x}} \varphi_{-z,1}(\underline{x}_{AP})^T \underline{Q}_{\varphi,1} \nabla_{\underline{x}} \varphi_{-z,1}(\underline{x}_{AP}) \\ \underline{R}'_u &= \nabla_{\underline{u}} \underline{g}_w(\underline{u}_{AP})^T \underline{R}_w \nabla_{\underline{u}} \underline{g}_w(\underline{u}_{AP}).\end{aligned}\quad (3.49)$$

transformiert werden. Die Matrizen sind gestrichen dargestellt, da sie nicht exakt den Matrizen aus (3.46) entsprechen müssen, sondern Gleichung (3.48) nur strukturell zu Gleichung (3.46) ähnlich ist. Daher werden diese gestrichenen Matrizen als Äquivalent angesehen.

Für die nachfolgenden Beschreibungen wird folgende Annahme getroffen.

**Annahme 3.17.** Die Jacobi-Matrix der  $n$  steuerbaren Eigenfunktionen  $\nabla_{\underline{x}} \varphi_{-z,1}(\underline{x})$ , beziehungsweise ihrer Abschätzung  $\nabla_{\underline{x}} \varphi_{-z,1}(\underline{x}) \approx \underline{W}_1^T \nabla_{\underline{x}} \underline{g}_z(\underline{x})$  nach Gleichung (3.15), ist invertierbar.

Mit dieser Annahme ist es möglich, die Gewichtungsmatrix  $\underline{Q}_{\varphi,1}$  in Koopman-Eigenkoordinaten anhand der anschaulich wählbaren Gewichtung in Zustandskoordinaten zu berechnen.

$$\underline{Q}_{\varphi,1} = \left( \nabla_{\underline{x}} \varphi_{-z,1}(\underline{x}_{AP})^T \right)^{-1} \underline{Q}'_x \left( \nabla_{\underline{x}} \varphi_{-z,1}(\underline{x}_{AP}) \right)^{-1} \quad (3.50)$$

$$\underline{R}_w = \left( \nabla_{\underline{u}} \underline{g}_w(\underline{u}_{AP})^T \right)^{-1} \underline{R}'_u \left( \nabla_{\underline{u}} \underline{g}_w(\underline{u}_{AP}) \right)^{-1}. \quad (3.51)$$

Hierbei ergibt sich allerdings das Problem, dass hierfür auch tatsächlich Eigenfunktionen geschätzt werden müssen. Insbesondere in dieser Arbeit werden, wie in Kapitel 4 beschrieben, lediglich die Observables geschätzt, welche eine Linearkombination dieser Eigenfunktionen darstellen. Wie bereits in Annahme 3.17 verwendet und in Abschnitt 3.1 mit Gleichung (3.15) genauer erläutert wird, können diese Eigenfunktionen mit den Linkseigenvektoren der Koopman-Matrix und den Observables abgeschätzt werden. Gleichung (3.50) gilt jedoch nur für die Darstellung in Eigenkoordinaten. Die Darstellung in Observable-Koordinaten ergibt sich mit der Transformationsmatrix  $\underline{W}_1$  zu

$$\underline{Q}_z = \underline{W}_1 \left( \nabla_{\underline{x}} \underline{g}_z^T \underline{W}_1 \right)^{-1} \underline{Q}'_x \left( \underline{W}_1^T \nabla_{\underline{x}} \underline{g}_z \right)^{-1} \underline{W}_1^T. \quad (3.52)$$

Durch die lineare Transformation ist mit Annahme 3.17 der Rang von  $\underline{Q}_z$  und  $\underline{Q}'_x$  derselbe trotz dessen, dass  $\dim(\underline{Q}_z) > \dim(\underline{Q}'_x)$  gilt. Somit wird die Gewichtung in approximierten Eigenkoordinaten auf die Observables verteilt. Dies kann durch die zweifache Approximation ( $\varphi \approx \underline{W}_1^T \underline{g}_z$  und  $\underline{g}_z \approx \nabla_{\underline{x}} \underline{g}_z \Delta \underline{x}$ ) sowie der nicht garantierten Einhaltung der nichtlinearen Verkopplungen in  $\varphi$  in der Berechnung der Reglermatrix  $\underline{C}$  dazu führen, dass die Abhängigkeit zwischen den gewählten Zustandsgewichten und den Auswirkungen bei der Regelung nicht mehr direkt übertragen

werden. Dies wird in Kapitel 5 anhand von Beispielen näher beleuchtet. Für die Anwendung des Zustandsregler ist dies jedoch zunächst nicht problematisch.

Anders gestaltet sich dies, wenn der Eingangsraum in seiner Dimension erweitert wird, was für die reine Prädiktion des Systemverhaltens ohne Regelung Vorteile haben kann [18]. Die einfache Verwendung einer klassischen LQ-Zustandsregelung, mittels den angepassten Gewichten aus den Gleichungen (3.50) und (3.51), würde eine solche Erhöhung des Eingangsraumes im Koopman-Raum jedoch durch die nichtlinearen Verkopplungen in  $\underline{g}_w$  unmöglich machen. Grundsätzlich könnte ein solcher Zustandsregler mit angepasstem Gewicht  $\underline{R}_w$  mittels der Pseudoinversen von  $\nabla_{\underline{u}} \underline{g}_w$  für Modelle mit  $m_e > m$  ähnlich zu Gleichung (3.51) berechnet werden. Beim Aufschalten des Reglers ergibt sich dann jedoch die Problematik, dass das Stellsignal  $\underline{w}_{c,k} = \underline{C} \underline{z}_k$  die nichtlinearen Verkopplungen in  $\underline{g}_w$  nicht zwingend einhält. Das sich ergebende Stellsignal ist somit gemäß  $\underline{g}_w$  und  $\bar{\underline{g}}_w$  nicht definiert<sup>16</sup> und es muss eine LQ-Zustandsregelung mit Nebenbedingungen verwendet werden. Eine solche LQ-Regelung mit Nebenbedingung wurde im Kontext des Koopman-Operators ohne Erweiterung des Eingangsraumes bereits in [70] diskutiert.

Für die Anwendung von Gleichung (3.52) ist des Weiteren folgende Annahme wichtig.

**Annahme 3.18.** Die Linearisierung am untersuchten Arbeitspunkt  $\underline{x}_{AP}$  des nichtlinearen Systems ist vollständig beobachtbar über die künstliche Ausgangsmatrix  $\bar{\underline{Q}}_z \nabla_{\underline{x}} \underline{g}_z$  mit  $\underline{Q}_z = \bar{\underline{Q}}_z^T \bar{\underline{Q}}_z$ .

Diese Annahme liegt in der Forderung begründet, dass gemäß [68, S. 306] ein LQ-geregeltes lineares System über  $\bar{\underline{Q}}_x$  beobachtbar sein muss. Dabei ist die künstliche Ausgangsmatrix  $\bar{\underline{Q}}_x$  die Zerlegung der Zustandsgewichtung gemäß  $\underline{Q}_x = \bar{\underline{Q}}_x^T \bar{\underline{Q}}_x$ . Gleichung (3.47) wird für die Berechnung der modifizierten Zustandsgewichtung aus Gleichung (3.52) linearisiert. Daraus ergibt sich mit  $\underline{Q}'_x = \nabla_{\underline{x}} \underline{g}_z^T \underline{Q}_z \nabla_{\underline{x}} \underline{g}_z$  die Forderung, dass die Linearisierung des nichtlinearen Systems über  $\bar{\underline{Q}}_z \nabla_{\underline{x}} \underline{g}_z$  am untersuchten Arbeitspunkt vollständig beobachtbar sein muss. Dabei ist  $\bar{\underline{Q}}_z$  die Matrix einer Cholesky-Zerlegung von  $\underline{Q}_z$ .

### 3.5 Identifizierbarkeit eines Koopman-linearen Systems

In Abschnitt 3.3.1 wurde bereits gezeigt, dass bei vollständig sowie unvollständig gemessenem Zustandsvektor die Koopman-Eigenfunktionen und damit der Koopman-Operator beobachtbar ist. Der folgende Abschnitt soll diese Idee weiterführen, indem zunächst die Identifizierbarkeit des Koopman-Operators erläutert wird. Sie bietet eine Erweiterung der Beobachtbarkeit nichtlinearer Systeme. Nachfolgend wird die Identifikation beziehungsweise die Schätzung von Koopman-linearen Systemen diskutiert, wenn nicht alle Zustände gemessen werden oder latente Zustände im System vorhanden sind.

<sup>16</sup>Dies gilt insbesondere für die späteren Approximation durch das DKLC-Verfahren, welches in Kapitel 4 beschrieben wird. Das Modell kann an diesen Stellen nicht trainiert werden, da sich die Größen  $\underline{w}$  aus  $\underline{g}_w = \underline{w}$  ergeben und man somit nicht frei in der Wahl dieser Koopman-Eingänge ist.

Um Koopman-lineare Systeme zu ermitteln, werden diese in der Regel identifiziert, da eine analytische Herleitung nur für sehr einfache Systeme möglich ist, wie das System aus Gleichung (3.5) beispielhaft zeigt. Dieser Abschnitt soll daher in das nachfolgende Kapitel 4, welches sich mit der Schätzung von Koopman-linearen Systemen beschäftigt, überleiten, sowie die in den vorangegangenen Abschnitten erarbeiteten Konzepte einer Koopman-Zustandsregelung auf Systeme übertragen, bei denen nicht alle Zustände des nichtlinearen Systems messbar sind.

Für die folgenden Betrachtungen gilt die vollständig steuerbare Systemgleichung (3.26) mit der Ausgangsgleichung

$$\underline{y} = \underline{h}(\underline{x}) \in \mathcal{Y}. \quad (3.53)$$

Da zur vollständigen dynamischen Beschreibung alle Zustände notwendig sind, muss das System zur Identifikation vollständig beobachtbar sein. Andernfalls würde ein Teil der Dynamik nach der Identifikation verborgen bleiben und das Systemverhalten kann nicht vollständig geschätzt werden. Weiter muss das System identifizierbar sein. Die Identifizierbarkeit gibt anschaulich an, ob zusätzlich die Parameter des Modells geschätzt (beobachtet) werden können. Zur vollständigen Identifizierbarkeit eines Koopman-linearen Systems muss folgender Satz gelten.

**Satz 3.19.** *Ein Koopman-lineares System (3.35) mit Observables  $\underline{z} = \underline{g}_z(\underline{x})$  und eindeutiger Rücktransformation  $\underline{h}(\underline{x}) = \underline{g}_z^{-1}(\underline{g}_z(\underline{h}(\underline{x}, \underline{\theta}))) + \underline{e}_h(\underline{x}, \underline{\theta}) = \bar{\underline{h}}(\underline{x}, \underline{\theta}) + \underline{e}_h(\underline{x}, \underline{\theta})$  eines nichtlinearen Systems (3.26) ist unter der Bedingung*

$$\text{rank}(\underline{O}(\tilde{\underline{x}}_0)) = \text{rank} \left( \begin{bmatrix} \frac{\partial}{\partial \tilde{\underline{x}}} (\bar{\underline{h}}(\tilde{\underline{x}})) \\ \frac{\partial}{\partial \tilde{\underline{x}}} (L_f \bar{\underline{h}}(\tilde{\underline{x}})) \\ \frac{\partial}{\partial \tilde{\underline{x}}} (L_f^2 \bar{\underline{h}}(\tilde{\underline{x}})) \\ \vdots \\ \frac{\partial}{\partial \tilde{\underline{x}}} (L_f^{n+n_p-1} \bar{\underline{h}}(\tilde{\underline{x}})) \end{bmatrix} + \begin{bmatrix} \frac{\partial}{\partial \tilde{\underline{x}}} (\underline{e}_h(\tilde{\underline{x}})) \\ \frac{\partial}{\partial \tilde{\underline{x}}} (L_f \underline{e}_h(\tilde{\underline{x}})) \\ \frac{\partial}{\partial \tilde{\underline{x}}} (L_f^2 \underline{e}_h(\tilde{\underline{x}})) \\ \vdots \\ \frac{\partial}{\partial \tilde{\underline{x}}} (L_f^{n+n_p-1} \underline{e}_h(\tilde{\underline{x}})) \end{bmatrix} \right)_{\tilde{\underline{x}}=\tilde{\underline{x}}_0} = n+n_p \quad (3.54)$$

mit  $\tilde{\underline{x}}^T = [\underline{x}^T, \underline{\theta}^T]$  (lokal) identifizierbar in einer Umgebung  $\mathcal{U}(\tilde{\underline{x}}_0)$  um  $\tilde{\underline{x}}_0$ . Dabei ist  $\underline{\theta} \in \mathbb{R}^{n_p}$  der Vektor von Parametern im System.

*Beweis.* Nach [120] gilt Gleichung (3.54) mit  $\underline{e}_h = \underline{0}$  bereits direkt für nichtlineare Systeme (3.26). Da es sich bei der Lie-Ableitung um einen linearen Operator handelt, ergibt sich mit der Definition des Ausgangs in Form eines Fehlers eine Summe zweier Matrizen.  $\square$

Besonders interessant ist hierbei der Fall, wenn eine fehlerfreie Koopman-Linearisierung mit  $\underline{e}_h$  existiert. Denn dann ist es für die Identifizierbarkeit eines solchen System nicht von belang, ob im Koopman-linearen System mehr Parameter geschätzt werden müssen, als beim nichtlinearen

System. In der praktischen Identifikation, sollten dennoch mindestens so viele Datenpunkte wie Parameter im System vorhanden sein, wie Gleichung (2.1) beispielhaft zeigt. Im Falle künstlicher neuronaler Netze, wie sie in Kapitel 4 verwendet werden, werden mindestens zwei bis drei mal so viele Datenpunkte benötigt wie Parameter im Netz vorhanden sind [3].

Für den Fall dass ein Fehler vorhanden ist, ist dieser Satz weniger von praktischer Relevanz, da der Fehler in der Regel nicht analytisch ausgedrückt werden kann. Dennoch muss er so beschaffen sein, dass er durch Summenbildung mit der linken Matrix aus Gleichung (3.54) nicht den Rang herabsetzt.

### 3.5.1 Schätzung bei nicht vollständig gemessenem Zustand

In den bisherigen Abschnitten wurde stets von einem vollständig messbaren Zustand ausgegangen. Dies ist jedoch häufig nicht der Fall, da aus Kostengründen oder physikalischen Gegebenheiten keine Messung des gesamten Zustandes möglich ist oder sogar latente Zustände im System vorhanden sind. Somit stehen nur ein Teil der Zustände, oder nur Messgrößen, welche Funktionen der Zustände gemäß Gleichung (3.53) sind, zur Verfügung. Das heißt, es können nur die Messgrößen zur Bildung von Observables  $\underline{g}_z(\underline{y})$  genutzt werden. Dies ist allerdings aus theoretischer Sicht zunächst keine Einschränkung, da Satz 3.19 genau so gilt. Ist das nichtlineare System also identifizierbar, dann ist auch das Koopman-lineare System auf Basis von Messgrößen identifizierbar.

Dieser Satz trifft jedoch keine Aussage darüber, wie das System identifiziert werden kann. Ein gängiger Ansatz, sich diesem Problem anzunehmen, wurde bereits in Abschnitt 2.3 in Form einer zeitlichen Verschiebung der Messgrößen vorgestellt. Im Kontext des Koopman-Operators wurden dazu bereits einige Arbeiten verfasst [5, 26, 55, 81, 111].

Alle haben gemein, dass sie auf eine andere Weise Takens' Embedding-Theorem mit dem Koopman-Operator autonomer Systeme in Zusammenhang bringen. Takens' Embedding-Theorem ist für zeitdiskrete Systeme wie folgt definiert:

**Satz 3.20** (Takens' Embedding-Theorem [103, 112, 118]). *Sei  $\mathcal{X}$  eine kompakte Menge der Dimension  $n$ . Wenn  $d > 2n$  gilt, dann ist die Menge an Funktionen  $(\underline{f}(\underline{x}), h(\underline{x}))$ , für die*

$$\begin{aligned} \underline{H}(\underline{x}_k) &= \left[ h(\underline{x}_k) \quad h(\underline{f}(\underline{x}_k)) \quad \cdots \quad h(\underline{f}^{d-1}(\underline{x}_k)) \right] : \mathcal{X} \rightarrow \mathbb{R}^d \\ &= \left[ h(\underline{x}_k) \quad h(\underline{x}_{k+1}) \quad \cdots \quad h(\underline{x}_{k+d-1}) \right] : \mathcal{X} \rightarrow \mathbb{R}^d \end{aligned} \quad (3.55)$$

eine Einbettung ist, offen und dicht in  $\mathcal{D}^r(\mathcal{X}) \times \mathcal{C}^r(\mathcal{X}, \mathbb{R})$  mit  $r \geq 1$ .

Dabei ist  $\mathcal{D}^r(\mathcal{X})$  die Menge an  $r$ -fach ableitbaren Diffeomorphismen,  $\mathcal{C}^r(\mathcal{X}, \mathbb{R})$  die Menge an  $r$ -fach ableitbaren skalaren Messgrößen und  $\underline{f}^i$  die  $i$ -fache Funktionskomposition von  $\underline{f}$  mit sich selbst. Eine Einbettung  $f : \mathcal{A} \rightarrow \mathcal{B}$  ist dabei eine Abbildung zwischen zwei topologischen Räumen  $\mathcal{A}$  und  $\mathcal{B}$  für die  $f$  ein Homöomorphismus ist.

Im Kontext zeitdiskreter Systeme ist es anschaulich, dass dies mit der Beobachtbarkeit zusammenhängen muss, was in [4] untersucht wird. Ist  $\underline{H}$  ein Homöomorphismus, so kann aus der Kenntnis von skalaren Messgrößen über einen gewissen Horizont auf den ursprünglichen Zustand zurückgerechnet werden. Dabei gibt Takens' Embedding-Theorem nur eine untere Schranke an, ab der garantiert ist, dass Gleichung (3.55) mit gegebenem  $d$  eine Einbettung ist. In der Realität ist diese Grenze jedoch häufig etwas niedriger, was man sich wieder leicht mit der Beobachtbarkeit linearer Systeme verdeutlichen kann. Denn dort werden lediglich  $n$  Zeitschritte zur Bewertung der Beobachtbarkeit, und damit zur Berechnung des kompletten Systemzustandes anhand von potenziell weniger Messgrößen als Zuständen, herangezogen.

Dieser Satz ist jedoch nicht direkt auf angeregte Systeme übertragbar. Gleiches gilt für Systeme, die nicht auf einen Attraktor zulaufen. In [103] wird der Satz für zeitdiskrete, stochastische (stochastisch angeregte) Systeme erweitert, für die das dynamische System über einen  $m$  dimensionalen Eingang  $\underline{u}$  der Art  $\underline{u}_{k+1} = \underline{c}(\underline{u}_k)$  angeregt wird. Dabei erhöht sich die untere Grenze auf  $d > 2(n + m)$  wenn  $\underline{u}$  unbekannt ist. Dies ist auch anschaulich, da zusätzlich zum Zustand auch der Eingang rekonstruiert werden muss. Ist das Eingangssignal bekannt, so bleibt die Grenze bei  $d \geq 2n + 1$  bestehen [102, 103]. In [118, S. 14] wird der Ansatz zusätzlich für beliebig angeregte Systeme beschrieben.

Auf Basis dieser Erkenntnisse wird ein neuer Zustandsvektor  $\underline{x}_e^T = [\underline{y}_k^T, \underline{y}_{k-1}^T, \dots, \underline{y}_{k-d}^T] \in \mathcal{Y}_e \subseteq \mathbb{R}^{q \cdot (d+1)}$  (engl. *extended*) mit einem Delay Horizont  $d$  definiert, der gegebenenfalls nach Takens' Embedding-Theorem und seinen Erweiterungen ausgelegt werden kann. Dies ist im Rahmen der Koopman-Theorie keine Einschränkung, da die Observables lediglich Funktionen von  $\underline{x}$  sein müssen, was sie mit  $\underline{y} = \underline{h}(\underline{x})$  sind. Der Einsatz dieser Herangehensweise ist auch bei vollständig gemessenem Zustandsvektor denkbar, da im Kontext dieser Arbeit nichtlineare Systeme identifiziert werden sollen, für die Takens' Embedding-Theorem im Allgemeinen gilt.

Es ist möglich, sich einen weiteren, heuristischen Ansatz für die Bildung von Observables zu überlegen. Der Eingang wirkt sich unmittelbar auf die nachfolgenden Zustände aus. Somit ist es intuitiv, auch die Eingänge bei der Bildung von Zustandsobservables zu berücksichtigen. Der neue Zustandsvektor ergibt sich dann zu  $\underline{x}_e^T = [\underline{y}_k^T, \underline{y}_{k-1}^T, \underline{u}_{k-1}^T, \dots, \underline{y}_{k-d}^T, \underline{u}_{k-d}^T] \in \mathcal{Y}_e \subseteq \mathbb{R}^{q \cdot (d+1) + d \cdot m}$ .

Genaugenommen gilt das Embedding-Theorem von Takens' wie bereits erwähnt für Systeme, die auf keinen Attraktor zulaufen, nicht. Dennoch wird im Rahmen dieser Arbeit experimentell erforscht, in wie weit Delay-Koordinaten für solche Systeme verwendet werden können. Um bei der Wahl der Delay-Dimension nicht raten zu müssen, wird auch ohne theoretischen Rückhalt diese Dimension unter anderem gemäß des Takens' Embedding-Theorems gewählt.

Bei der Zustandsregelung unter Verwendung von Delay-Koordinaten muss insbesondere die Beschaffenheit des künstlichen Zustandsvektors beachtet werden. Durch die Einführung der Delay-Koordinaten ist der Zustandsvektor ein erweiterter Vektor aus den gemessenen Ausgängen, die über einen gewissen Zeithorizont verschoben in einem Vektor zusammengefasst werden. Daraus ergibt sich das Problem, dass für die ersten  $d - 1$  Zeitschritte nach dem Anlaufen kein valider Zustandsvektor zur Verfügung steht. Somit ist hier besondere Vorsicht bei der Zuschaltung des

Stellsignals geboten. Wird der Zustandsvektor also falsch beziehungsweise zu weit weg vom eigentlichen Betriebspunkt initialisiert, so kann es zu unerwünschtem Verhalten oder gar Instabilität kommen. Daher müssen hierfür Vorkehrungen getroffen werden.

Beispielsweise kann für die Übergangszeit bis zum Zeitschritt  $d$  das System ohne Regelung betrieben werden. Dies ist jedoch ausschließlich für garantiert stabile oder gedämpft schwingende Systeme anwendbar, insoweit dieser Betrieb nicht zu Problemen führt. Für eine sehr langsame Dynamik (oder eine sehr schnelle Abtastzeit) könnte der Zustandsvektor auch mit den selben Werten für alle Zeitschritte initialisiert werden. Eine dritte Möglichkeit ist eine Art Übergangsregelung zu implementieren, die das System garantiert (semi) stabil hält solange der Koopman-Zustandsregler auf ausreichende Werte wartet.

Außerdem kann es durch die begrenzte Messung von Zuständen dazu kommen, dass bestimmte Eigenwerte nicht beobachtbar sind und somit nicht identifiziert werden. Grundsätzlich hängt die Steuerbarkeit eines System nicht von einzelnen Messgrößen und damit von der Beobachtbarkeit des Systems ab. Doch auch wenn diese Abhängigkeit systemtheoretisch nicht vorliegt, so hängt doch die Beobachtbarkeit einzelner Eigenwerte bei (Koopman-) linearen Systemen und damit die Identifizierbarkeit seiner Eigenwerte von den Messgrößen ab. Kommt es durch eine ungeeignete Messung dazu, dass steuerbare Eigenwerte nicht beobachtbar sind, so tauchen diese Eigenwerte nicht im identifizierten Modell auf. Die Anzahl nicht beobachtbarer, aber steuerbarer, Eigenwerte wird im weiteren Verlauf mit  $r$  bezeichnet. Eine Berücksichtigung von  $n$  Eigenwerten bei der Reglerberechnung nach Gleichung (3.52), obwohl nur  $n - r$  Eigenwerte einen mathematischen beziehungsweise physikalischen Bezug zu den am realen System steuerbaren Eigenwerten des Koopman-Operators haben, kann sich auf zwei Weisen auf den resultierenden Regler, und damit das Verhalten des Systems im geschlossenen Regelkreis mittels Zustandsrückführung, auswirken.

1. Aufgrund der Argumentation im ersten Absatz von Abschnitt 3.4.1 hat eine Berücksichtigung eines nicht steuerbaren Eigenwerts im LQ-Verfahren keine Auswirkung auf den Regler und die Reglermatrix ist entsprechend dieselbe wie bei ausbleibender Berücksichtigung nichtsteuerbarer Eigenwerte.
2. Durch Schätzfehler kann es dazu kommen, dass Heuristik 3.14 am identifizierten Modell nicht erfüllt ist und mehr als  $n$  Eigenwerte steuerbar sind. Werden nun Eigenwerte im LQ-Entwurf berücksichtigt, die zwar modellseitig steuerbar sind, aber nicht den „korrekt“<sup>17</sup> steuerbaren Eigenwerten des physikalischen System entsprechen, so kann es zu unerwünschtem Verhalten kommen. Dies kann sich unter anderem dadurch äußern, dass sich die beabsichtigte Beeinflussung einzelner Zustände, durch entsprechende Gewichtung im LQ-Entwurf, nicht wie vorgesehen ausprägt. Dies wird in Kapitel 5 tiefer analysiert und erläutert.

Um an dieser Stelle unerwünschtes Verhalten zu vermeiden, ist es zwingend erforderlich den berechneten Regler unter strenger Beobachtung am entsprechenden System zu testen.

<sup>17</sup>„korrekt“ meint an dieser Stelle die identifizierten steuerbaren Eigenwerten, die den  $n$  steuerbaren Eigenwerten des Koopman-Operators des nichtlinearen dynamischen Systems entsprechen.

## 4 Schätzung Koopman-linearer Zustandsraumsysteme

Das vorangegangene Kapitel befasste sich mit den theoretischen Grundlagen des Koopman-Operators. Außerdem wurden Erweiterungen in Bezug auf angeregte Systeme sowie unvollständig gemessene Zustände vorgenommen und systemtheoretische Begriffe, wie beispielsweise der Steuerbarkeit, im Kontext des Koopman-Operators betrachtet und formuliert. Diese Konzepte werden im nun folgenden Kapitel aufgegriffen und es wird hieraus ein neues Konzept zur Schätzung von Koopman-linearen Systemen entwickelt<sup>18</sup>.

### 4.1 Deep Koopman Operator Learning with control (DKLc)

Wie bereits in den vorherigen Kapiteln erläutert wurde, ist es selbst mit der Bedingung  $\dot{\varphi}(\underline{x}, \underline{u}) = \nabla_{\underline{x}}\varphi(\underline{x}, \underline{u})\underline{f}(\underline{x}, \underline{u}) \approx \lambda\varphi(\underline{x}, \underline{u})$  und bekanntem Systemmodell<sup>19</sup> eine enorme Herausforderung, Koopman-Eigenfunktionen  $\varphi(\underline{x})$ , oder die daraus ableitbaren Observables, analytisch zu bestimmen. Aufgrund dieser großen Einschränkung bieten sich Funktionsapproximationsalgorithmen an.

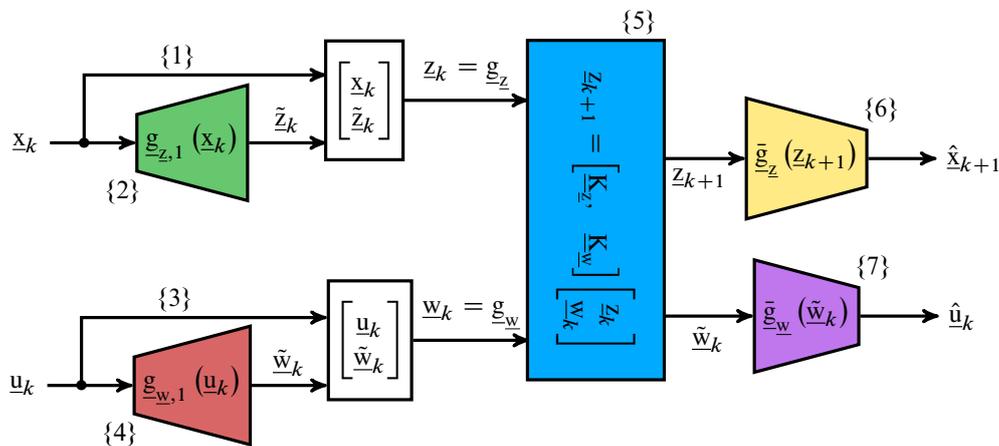


Abbildung 4.1: DKLc-Netzstruktur. Spezielle Struktur eines künstlichen neuronalen Netzes zur Approximation eines näherungsweise invarianten Unterraumes zum Koopman-Operator eines nichtlinearen Systems der Art aus Gleichung (3.36) anhand einer Reihe von von Messdaten  $[\underline{x}_{k+1}, \underline{x}_k, \underline{u}_k]$  mit  $k \in [0, N - 1]$

<sup>18</sup>Sämtlicher Programmcode, welcher zur Anfertigung dieser Arbeit verwendet und erarbeitet wurde, ist unter <https://github.com/MarcelBonnert/DKLc-Diss.git> abrufbar.

<sup>19</sup>Diese Gleichung ergibt sich durch die Annahme, dass der Term  $\nabla_{\underline{u}}\varphi(\underline{x}, \underline{u})\dot{\underline{u}}$  aus  $\dot{\varphi}(\underline{x}, \underline{u}) = \nabla_{\underline{x}}\varphi(\underline{x}, \underline{u})\underline{f}(\underline{x}, \underline{u}) + \nabla_{\underline{u}}\varphi(\underline{x}, \underline{u})\dot{\underline{u}} = \lambda\varphi(\underline{x}, \underline{u})$  gegenüber  $\nabla_{\underline{x}}\varphi(\underline{x}, \underline{u})\underline{f}(\underline{x}, \underline{u})$  nur einen geringen Anteil besitzt (ähnlich zu Annahme 3.4).

Ein ausreichend tiefes MLP ist in der Lage, jede messbare Funktion beliebig genau anzunähern [48]. Dabei bieten diese Netze im vorliegenden Fall einen weiteren erheblichen Vorteil. Es müssen keinerlei Annahmen über die Funktion der Koopman-Observables festgelegt werden. Abbildung 4.1 zeigt den, in dieser Arbeit verwendeten, speziellen Aufbau eines künstlichen neuronalen Netzes.

Im Gegensatz zum Vorgehen in [69] werden in dieser Arbeit Observables und keine Eigenfunktionen geschätzt. Dies hat den maßgeblichen Grund, dass Eigenfunktionen sehr restriktiv in Ihrer Beschaffenheit sind. Diese Eigenschaft würde das Optimierungsproblem, insbesondere beim Hinzufügen eines Eingangs, weiter erschweren. Für die Regelung, wie sie im Späteren durchgeführt wird, ist es unerheblich, wie genau das Modell aussieht und welche Basisfunktionen vorliegen. Wichtiger ist, dass der Fehler zwischen Koopman-linearem und nichtlinearem Modell klein bleibt, wie in Satz 3.11 angegeben. Hier bietet die Schätzung von Observables einen klaren Vorteil gegenüber Eigenfunktionen.

Die Grundlage bei der Anhebung der Dimension des Zustandsraumes bildet dabei eine Encoder/Decoder Netzstruktur. Diese Struktur ist durch den Aufbau ähnlich zu einem sogenannten Autoencoder, wie er im maschinellen Lernen häufig eingesetzt wird. Eine Beschreibung der Unterschiede zwischen einem Autoencoder und der hier genutzten Struktur findet sich in Abschnitt 1.2. Das künstliche neuronale Netz besteht aus sieben Teilen welche folgend aufgelistet und erläutert sind.

- {1} Linearer Zustandsencoder - Eingang:  $\underline{x}_k$ , Ausgang:  $\underline{x}_k$ ,
  - Leitet den Systemzustand direkt zur Koopman-Operator Approximationsschicht
- {2} Nichtlinearer Zustandsencoder  $\underline{g}_{z,1}(\underline{x}_k) : \mathcal{X}_r \rightarrow \mathcal{G}_z$  - Eingang:  $\underline{x}_k$ , Ausgang:  $\tilde{z}_k$ ,
  - Ist für das optionale Anheben der Zustandsraumdimension und die nichtlineare Transformation in Form der Funktionen  $\underline{g}_z$  des Zustandsraums erforderlich
- {3} Linearer Eingangsender - Eingang:  $\underline{u}_k$ , Ausgang:  $\underline{u}_k$ ,
  - Leitet die Systemanregung direkt zur Koopman-Operator Approximationsschicht
- {4} Nichtlinearer Anregungsender  $\underline{g}_{w,1}(\underline{u}_k) : \mathcal{U}_r \rightarrow \mathcal{G}_w$  - Eingang:  $\underline{u}_k$ , Ausgang:  $\tilde{w}_k$ ,
  - Ist für das optionale Anheben der Dimension des Eingangsraumes und die nichtlineare Transformation in Form der Funktion  $\underline{g}_w$  des Anregungsraums erforderlich
- {5} Koopman-Operator Schicht - Eingang:  $[\underline{z}_k^T, \underline{w}_k^T]^T$ , Ausgang  $\underline{z}_{k+1}$ ,
  - Approximiert eine lineare Zustandsraumrealisierung zu einem zum Koopman-Operator näherungsweise prädiktionsinvarianten Unterraum
- {6} Zustandsdecoder - Eingang:  $\underline{z}_{k+1}$ , Ausgang:  $\hat{\underline{x}}_{k+1}$ ,
  - Approximiert eine Inverse zu  $\underline{g}_z$
- {7} Anregungsdecoder - Eingang:  $\underline{w}_k$ , Ausgang:  $\hat{\underline{u}}_k$ .
  - Approximiert eine Inverse zu  $\underline{g}_w$ , zur Umrechnung des Stellsignals vom Koopman- in den Zustandsraum

Die Dimensionen der Räume  $\mathcal{G}_z$  und  $\mathcal{G}_w$  sind dabei abhängig von der Wahl der einzelnen Encoderstrukturen. Wie sich in Kapitel 5 zeigt, können die linearen Encoder entfallen. Somit gilt bei vorhandenen linearen Encodern  $\mathcal{G}_z \subset \mathbb{R}^{n_e-n}$  beziehungsweise  $\mathcal{G}_w \subset \mathbb{R}^{m_e-m}$  und bei eine vollständigen Kodierung des Zustandes ohne linearen Encoder  $\mathcal{G}_z \subset \mathbb{R}^{n_e}$  beziehungsweise  $\mathcal{G}_w \subset \mathbb{R}^{m_e}$ .

Das zu lösende Optimierungsproblem des künstlichen neuronalen Netzes lautet

$$\min J(\underline{\theta}).$$

Das Kostenfunktional  $J(\underline{\theta})$  wird zu

$$J = \alpha_1 J_{\text{pred},\underline{x}} + \beta_1 J_{\text{pred},\underline{x},\infty} + \alpha_2 J_{\text{pred},\underline{z}} + \beta_2 J_{\text{pred},\underline{z},\infty} + \alpha_3 J_{\text{auto}} + \beta_3 J_{\text{in},\infty} + \alpha_4 J_{\text{in}} + \alpha_5 J_{\text{DMD}} + \gamma_1 \|\underline{\theta}\|_1 + \gamma_2 \|\underline{\theta}\|_2 \quad (4.1)$$

mit

$$\begin{aligned} J_{\text{pred},\underline{x}} &= \frac{1}{p} \sum_{k=0}^{p-1} \frac{\delta_1^k}{\delta_{s,1}} \|\hat{\underline{x}}_k - \underline{x}_k\|_2^2 \\ J_{\text{pred},\underline{x},\infty} &= \max_{k \in \{1, \dots, p\}} \|\hat{\underline{x}}_k - \underline{x}_k\|_\infty \\ J_{\text{pred},\underline{z}} &= \frac{1}{p-1} \sum_{k=1}^{p-1} \frac{\delta_2^k}{\delta_{s,2}} \|\hat{\underline{z}}_k - \underline{g}_{\underline{z}}(\underline{x}_k)\|_2^2 \\ J_{\text{pred},\underline{z},\infty} &= \max_{k \in \{1, \dots, p\}} \|\hat{\underline{z}}_k - \underline{g}_{\underline{z}}(\underline{x}_k)\|_\infty \\ J_{\text{auto}} &= \frac{1}{p} \sum_{k=0}^{p-1} \|\bar{\underline{g}}_{\underline{z}}(\underline{g}_{\underline{z}}(\underline{x}_k)) - \underline{x}_k\|_2^2 \\ J_{\text{in}} &= \frac{1}{p} \sum_{k=0}^{p-1} \|\bar{\underline{g}}_{\underline{w}}(\underline{g}_{\underline{w}}(\underline{u}_k)) - \underline{u}_k\|_2^2 \\ J_{\text{in},\infty} &= \max_{k \in \{1, \dots, p\}} \|\bar{\underline{g}}_{\underline{w}}(\underline{w}_k) - \underline{u}_k\|_\infty \\ J_{\text{DMD}} &= \frac{1}{p} \sum_{k=0}^{p-1} \|\underline{g}_{\underline{z}}(\underline{x}_{k+1}) - \underline{K}_{\underline{z}} \underline{g}_{\underline{z}}(\underline{x}_k) - \underline{K}_{\underline{w}} \underline{g}_{\underline{w}}(\underline{u}_k)\|_2^2 \end{aligned}$$

definiert. Dabei gibt  $\underline{\theta}$  den Vektor an, welcher sämtliche zu optimierenden Parameter im angegebenen Netz beinhaltet. Dazu gehören die Gewichtungsmatrizen und Biasterme der einzelnen Schichten. Die Abhängigkeit von  $\underline{g}_{\underline{z}}$ ,  $\underline{g}_{\underline{w}}$ ,  $\bar{\underline{g}}_{\underline{z}}$ ,  $\bar{\underline{g}}_{\underline{w}}$ ,  $\underline{K}_{\underline{z}}$  und  $\underline{K}_{\underline{w}}$  und damit implizit  $\hat{\underline{x}}$  von  $\underline{\theta}$  wurde der Übersichtlichkeit halber weggelassen. Zur anschaulicheren Darstellung sind die Ausdrücke für  $J_{\text{pred},\underline{x}}$  und  $J_{\text{pred},\underline{z}}$  an dieser Stelle vereinfacht. Eine ausführliche Formulierung ist in Anhang A.1.5 angegeben.

Die Größen  $J_{\text{auto}}$  und  $J_{\text{in}}$  bewerten, in wie weit das Netz die Zustands- und Eingangsdaten rekonstruieren kann. Unter der Rekonstruktion der Daten wird verstanden, dass die Kodierung und Dekodierung die Daten mit einem möglichst kleinen Fehler verläuft. Sind die gemessenen Zustände oder Eingänge jeweils Teil der Observable-Vektoren, entfällt die Bewertung dieser Anteile

und es wird kein Decoder geschätzt, da die Zustände aus den ersten  $n$  Elementen des Observable-Vektors geschlossen werden können. Die zu minimierenden Funktionale  $J_{\text{pred},x}$  und  $J_{\text{pred},z}$  geben an, wie gut das Netz die Daten im Zustands- sowie im Koopman-Raum  $p$  Zeitschritte in die Zukunft präzidieren kann. Hierfür werden die Anfangszustände jeder, in den Messdaten vorhandenen, Trajektorie allein aus den Systemparametern über die lineare Zustandsgleichung in {5} präzidiert und zu jedem Schritt mit den Messdaten verglichen. Um das Optimierungsproblem besser zu konditionieren und eine Überanpassung zu vermeiden, werden nach dem Beispiel von [90] Vergessensfaktoren  $0 < \delta_j < 1$  und ihre Summe  $\delta_{s,j} = \sum_{k=0}^{p-1} \delta_j^k$  hinzugefügt. Diese gewichten anschaulich die kurzfristige Prädiktion höher als die langfristige und erleichtern so die Konvergenz des Netzes hin zu einer validen Lösung. Der Term  $J_{\text{DMD}}$  bewertet die Einschnittprädiktion nach dem Vorbild der DMD und wird mit dem Gedanken integriert, die Einschnittprädiktion im Hinblick auf Satz 3.11 und damit die Regelungseigenschaften zu verbessern. Die Anteile, welche eine Supremumsnorm  $\|\cdot\|_\infty$  integrieren, sollen Ausreißer verhindern.

Das Einbeziehen eines Prädiktionshorizontes, speziell in der Koopman-Schicht {5}, ist besonders wichtig, da im Allgemeinen mehrere Netze zum gleichen Fitergebnis führen [35]. Diese Mehrdeutigkeit wird durch die erzwungene Gültigkeit des Modells über eine gewisse Anzahl an Zeitschritten eingeschränkt. Die sich ergebenden Observables sind jedoch weiterhin nicht unbedingt eindeutig, was auch in [18] erkennbar ist. Dort wird die in diesem Abschnitt beschriebene Methode zum ersten Mal mit einem Prädiktionshorizont von  $p = 1$  verwendet. Auch wenn die allgemeinen Fitergebnisse gut sind, so sind die geschätzten Eigenwerte durch den sehr kurzen Prädiktionshorizont nicht die, die man im Hinblick auf die Dynamik erwarten würde. Beispielsweise ergibt sich darin für das Pendelsystem kein Pol, welcher mit der dominanten Frequenz korrespondiert. Zudem variieren die Ergebnisse sehr stark mit unterschiedlichen Initialgewichten des Netzes.

Die in dieser Arbeit weiterentwickelte und verwendete DKLC-Methode nutzt hingegen einen erweiterten Prädiktionshorizont und eine komplexere Kostenfunktion, sodass sich ein System ergibt, welches, bei geeigneter Wahl der Abtastzeit, robust und unabhängig von Startgewichten die dominanten Eigenwerte schätzt, wie später in Kapitel 5 dargestellt. Die dominanten, der dort ermittelten Eigenwerte der Koopman-linearen Systemäquivalente sind in einer Umgebung der Linearisierung des nichtlinearen Systems und sind somit plausibel hinsichtlich der realen Systemdynamik. Die Einbindung einer Prädiktion ist in der Literatur auch als *Linear Recurrent Autoencoder Network*-Struktur (LRAN) bekannt und wurde für autonome Systeme durch Otto et al. [90] untersucht.

Weiter verfügt die vorgestellte DKLC-Methode über die Möglichkeit, die Zustände oder manuell gewählte Observables direkt in den Observable-Vektor einzubinden. Dabei ist das folgende Lemma zunächst wichtig.

**Lemma 4.1.** Die (linear transformierten) Zustände eines nichtlinearen Systems  $\dot{\underline{x}} = \underline{f}(\underline{x}) \in \mathcal{X}$  können nicht alle Eigenfunktionen des Koopman-Operators auf dem gesamten  $\mathcal{X}$  sein.

*Beweis.* Wenn alle Zustände (oder eine lineare Transformation  $\tilde{\underline{x}} = \underline{T}^{-1} \underline{x}$  dieser) Eigenfunktionen des Koopman-Operators wären, dann könnten sie linear in der Form  $\tilde{\dot{x}}_i = \lambda_i \tilde{x}_i$  dargestellt werden. Damit wären sie Zustände eines linearen und nicht eines nichtlinearen Systems.  $\square$

Auch wenn diese Zustände somit keine Eigenfunktionen des Koopman-Operators sein können, so können sie dennoch Observables sein. Dies zeigt sich auch anhand von Gleichung (3.9), wonach sich Observables aus der Summe der Eigenfunktionen ergeben. Für das Beispiel aus Gleichung (3.5) kann der dort verwendete Observable-Vektor, welcher unter anderem aus den Zuständen besteht, direkt aus den Eigenfunktionen gebildet werden<sup>20</sup>. Grundsätzlich handelt es sich bei der Hinzunahme der Zustände als Observables allerdings um eine Einschränkung bezüglich des potenziell erreichbaren Fitergebnisses, wie sich auch in Kapitel 5 anhand einiger Beispiele zeigen wird. Diese Darstellung besitzt jedoch durch seine Anschaulichkeit bei der Wahl der Gewichtungsmatrizen der LQ-Regelung sowie durch die verringerte Komplexität über den entfallenden Zustandsdecoder Vorteile gegenüber dem vollständig kodierten Zustand.

Die Software zur Schätzung von Koopman-linearen Systemen unter Nutzung von Deep-Learning-Methoden nutzt dabei das Tensorflow Paket in der Version 1.14 für die Programmiersprache Python [75]. Verwendet wird genauer Python in der Version 3.6. Wie der Name bereits verrät, basiert Tensorflow auf mehrdimensionalen Tensoren, sowie sogenannte *Dataflow Graphs*. Hierzu werden anschaulich die Rechenprozesse zunächst als eine Art Flowchart beschrieben und erst zur Ausführungszeit mit Zahlenwerten versehen. So ist es möglich, komplexe Modelle, wie beispielsweise künstliche neuronale Netze, aufzubauen und diese zu optimieren.

Für die Optimierung solcher Modelle müssen verschiedene Randbedingungen definiert werden, was das Thema der folgenden Abschnitte sein soll.

### 4.1.1 Auswahl des Optimierers

Im nachfolgenden Abschnitt werden die vielversprechendsten Optimierer für künstliche neuronale Netze aufgelistet, in Kurzform beschrieben und anhand ihrer Eigenschaften eine Auswahl getroffen. Der im Training künstlicher neuronaler Netze am häufigsten eingesetzte Optimierer verwendet die sogenannte ADAM-Gleichung. Die allgemeine Updategleichung für viele Optimierer vom ADAM-Typ künstlicher neuronaler Netze lautet

$$\underline{\theta}_k = \underline{\theta}_{k-1} - h \cdot \underline{\hat{m}}_k ./ \sqrt{\underline{\hat{v}}_k}.$$

Dabei ist  $\underline{\theta}$  der zu schätzende Parametervektor,  $\underline{\hat{m}}_k$  und  $\underline{\hat{v}}_k$  jeweils eine Schätzung des Gradienten und der Varianz des Gradienten der Optimierungsgröße und  $h$  die Schrittweite oder, wie im Kontext künstlicher neuronaler Netze häufig genannt, die Lernrate. Der Operator „./“ und die Wurzel geben elementweise Operationen an. Weiter wird der Vektor  $h ./ \sqrt{\underline{\hat{v}}_k}$  als effektive Schrittweite bezeichnet und bringt den Effekt mit sich, dass das Verfahren weniger abhängig von der gewählten Schrittweite  $h$  ist [31].

<sup>20</sup>Anhand dieses Beispiels zeigt sich jedoch, dass einzelne Zustände (dort  $x_1$ ) Eigenfunktionen des Koopman-Operators sein können. Dies ist genau dann der Fall, wenn, wie auch im genannten Beispiel, einzelne Zustände vollständig linear beschrieben werden, wie sich leicht durch den Beweis in Lemma 4.1 zeigen lässt.

Der ADAM-Typ von Optimierern basiert daher auf einer Schätzung der ersten beiden statistischen Momente des Gradienten der zu schätzenden Variablen [31]. Daraus leitet sich sein Name *ADaptive Moment estimation* ab. Die Updategleichung lässt sich intuitiver mit

$$\underline{\theta}_k = \underline{\theta}_{k-1} - h \hat{E}(\underline{\nabla J}_k) ./ \sqrt{\hat{\text{Var}}(\underline{\nabla J}_k)} \quad (4.2)$$

angeben. Ist die Varianz des Gradienten groß, so ist die Unsicherheit groß und der Parameter wird nur numerisch wenig aktualisiert, da der Einfluss des hinteren Terms aus Gleichung (4.2) mit höherer Varianz abnimmt. Diese Eigenschaft ist vor allem dann sinnvoll, wenn die Lerndaten stark verrauscht sind, was für die Anwendung in realen Problemen von großem Vorteil ist. Der Optimierer wird dann also anschaulich gesprochen „vorsichtiger“.

Eine Variante des ADAM-Algorithmus ist der ADAMax-Algorithmus, welcher die Varianz über die Maximum-/Unendlichkeitsnorm aktualisiert. Komplexere Optimierer wie Adagrad ([37]) oder ADADELTA ([126]) sind für das in dieser Arbeit verwendete und vergleichsweise simple Netz nicht notwendig, können aber für komplexe Aufgaben Vorteile mit sich bringen [98].

Da die in dieser Arbeit beschriebene Methode einen enormen Rechenaufwand benötigt, bleibt eine detaillierte Analyse einzelner Optimierer anhand verschiedener Modelle und Ordnungen an dieser Stelle aus. Tests in einem kleinen Umfang liefern keine auffälligen Unterschiede zwischen den Optimierern. Des Weiteren liefert ein Optimierer vom ADAM-Typ in den meisten Studien ein besseres Ergebnis hinsichtlich der dabei konkret verwendeten Verlustfunktion als andere. Er wird folglich in dieser Arbeit als Optimierungsalgorithmus verwendet.

Zu jedem Optimierer für künstliche neuronale Netze muss eine Schrittweite gewählt werden. Anhand zahlreicher Tests konnte eine Schrittweite von  $h = 10^{-4}$  als guter Kompromiss zwischen Konvergenzgeschwindigkeit und dem robusten Auffinden niedriger Minima der Verlustfunktion gefunden werden. Jedoch ist dies nicht der einzige Parameter für die Optimierung abseits der Gewichte der einzelnen Terme der Verlustfunktion. Ebenso muss eine sogenannte Batch Size und die Art des Batchings für den Datensatz gewählt werden, mit dem sich die beiden folgenden Abschnitte beschäftigen.

### 4.1.2 Trainingsphilosophien

Beim Training eines DKLc-Netzes können im Wesentlichen zwei Trainingsansätze verfolgt werden. Das *Sample Batching* und das *Trajectory Batching*. Durch die Namensgebung wird beschrieben, in welcher Dimension der Datensatz in Batches aufgeteilt wird. Eine visuelle Gegenüberstellung beider Konzepte ist in Abbildung 4.2 dargestellt.

#### Sample Batching

Das Sample Batching stellt den allgemeineren Ansatz dar. Dabei wird das Netz in vier Dimensionen trainiert, das heißt die Daten werden in einem vier-dimensionalen Array gespeichert und

verarbeitet. Diese sind die Sample-Dimension, die dem Ansatz den Namen verleiht, der Prädiktionshorizont, die Anzahl der Trajektorien und die Systemordnung beziehungsweise die Anzahl der Messgrößen. Gegebenenfalls kann die Trajektorien Dimension weggelassen und durch eine niedrigere Batch Size ersetzt werden, um, je nach Hardware, Rechenzeit einzusparen.

Der Vorteil bei diesem Ansatz liegt darin, dass für mehrere gemessene Trajektorien über viele verschiedene Messzeitpunkte ein Modell ermittelt wird. Dies macht die Optimierung allerdings auch erheblich schwieriger, da für ein nichtlineares Modell ein lineares Modell über den kompletten Datensatz ermittelt werden muss. Die Daten müssen entsprechend aufbereitet und geeignet sein. Wenn dieser Ansatz allerdings zu einer sinnvollen Lösung kommt, ist das Modell über die komplette Zeitdauer des Datensatzes und somit allgemeiner gültig. Der Prädiktionshorizont dient hierbei im Wesentlichen dazu, die Dynamik über kurze Zeitabstände besser zu präzisieren. Im Gegensatz dazu sind die Modelle des *Trajectory Batchings* nur über den Prädiktionshorizont (tatsächlich) gültig. Im Grenzfall, wenn die Anzahl der Samples dem Prädiktionshorizont entspricht, geht der Sample Batching Ansatz direkt in das Trajectory Batching über.

Nach der Validierung einer Vielzahl von Versuchen und anhand der oben genannten Ausführungen wird dieser Ansatz nicht weiter verfolgt, da er das Optimierungsproblem zu sehr einschränkt, was sich in einer schlechten Konvergenz hin zu sinnvollen Ergebnissen äußert.

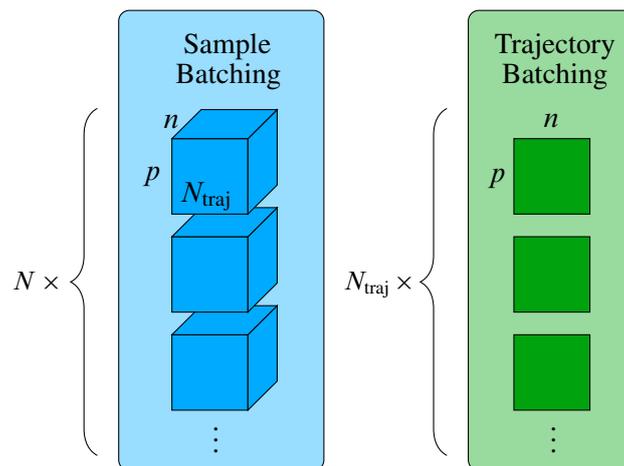


Abbildung 4.2: Visualisierung der Datenanordnung für unterschiedlichen Trainingsphilosophien. Die geschwungene Klammer gibt dabei die Dimension an, über die die Daten aufgeteilt werden

### Trajectory Batching

Das Trajectory Batching stellt einen reduzierten Ansatz dar. Die Sample-Dimension entfällt bei diesem Ansatz, sodass nun lediglich in drei Dimensionen trainiert wird. Dadurch erhöht sich allerdings die Anzahl an Trajektorien  $N_{\text{traj}}$  der Länge  $p$ . Das Modell muss somit einen Datenpunkt nur  $p$  Zeitschritte in die Zukunft präzisieren können, da jede Messreihe beziehungsweise Trajektorie nur  $p$  Samples lang ist. Somit ist es für das Optimierungsproblem wesentlich einfacher ein Modell zu finden, welches unter dem Verlustfunktional (4.1) minimal ist. Diese Modelle sind allerdings nur für diesen trainierten Prädiktionshorizont, und je nach Fit gegebenenfalls sogar über einen

kürzeren Horizont, gültig. Durch eine größere Anzahl an Trajektorien wird versucht, eine allgemeinere Gültigkeit des Modells zu schaffen, was jedoch insbesondere für einzelne Trajektorien nicht immer der Fall sein muss, wie sich in Kapitel 5 zeigen wird.

Dennoch wird diese Trainingsphilosophie ausgewählt, um die beschriebenen Netze zu trainieren. Heuristisch ist insbesondere für die Regelung eine kurzzeitige Präzidierbarkeit des Systems ausreichend, um das nichtlineare Modell adäquat zu regeln. Weiter ist das Optimierungsproblem deutlich besser konditioniert und es ist wahrscheinlicher ein Modell zu ermitteln, welches das Systemverhalten zumindest über den trainierten Zeithorizont gut reproduzieren kann.

Auch wenn die allgemeine Trainingsphilosophie ausgewählt ist, ist es wichtig zu ermitteln, wie groß die Batch Size gewählt werden sollte. Mit der eigentlichen Auswahl dieser beschäftigt sich der nachfolgende Abschnitt.

### 4.1.3 Wahl der Batch Size

Eine allgemeine Beschreibung der Batch Size findet sich in Abschnitt 2.3.2. Darin wurde bereits verdeutlicht, dass die Batch Size ein sensibler Parameter ist, der dabei helfen kann, die Approximationsfähigkeit des Netzes zu erhöhen. Bei ungeeigneter Wahl kann er jedoch auch zu einer Verminderung der Schätzqualität führen.

Um den Einfluss dieses Parameters im Rahmen einer DKLc-Schätzung zu beurteilen, wird erneut das Referenzsystem aus Gleichung (3.5) mit unterschiedlichen Batch Sizes im Bereich von 10 bis 5000 geschätzt. Die Modellparameter sind in Abschnitt A.3.2 im Anhang dargestellt. Für diese Schätzung ergibt sich jedoch ein anderes Bild als in Abschnitt 2.3.2, denn der Parameter hat für eine DKLc-Schätzung nur einen vernachlässigbaren Effekt. Die Schätzung mit einer sehr kleinen Batch Size liefert zwar ein etwas besseres Ergebnis hinsichtlich des Verlustfunktional, führt jedoch, wie erläutert, zu einer wesentlich längeren Optimierungsdauer.

Rückblickend auf das vergleichsweise einfache Beispiel im Grundlagenkapitel in Abschnitt 2.3.2 wird im Verlauf der Arbeit dieser Parameter auf einen mittleren Wert von 512 festgelegt, was etwa einem Zehntel bis Zwanzigstel der Menge an Trainingstrajektorien entspricht, abhängig von der Größe der Datensätze. So wird die Optimierungsdauer bei gleichbleibender Anzahl an Epochen nicht unnötig verlängert und es kann der potentiell positive Effekt einer kleineren Batch Size genutzt werden.

### 4.1.4 Umsetzung der Koopman-Zustandsrückführung mit DKLc

Abbildung 3.3 zeigt die Koopman-Zustandsrückführung nach [18]. In Verbindung mit der DKLc-Methode muss dieser jedoch leicht angepasst werden. Durch die Biasterme im Netz führt ein  $\underline{x} = \underline{0}$  beziehungsweise  $\underline{x} = \underline{x}_{AP}$  im Allgemeinen zu einem  $\underline{z} \neq \underline{0}$ . Das gleiche gilt für den Eingang. Die um diese stationären Werte angepasste Zustandsrückführung ist in Abbildung 4.3 dargestellt.

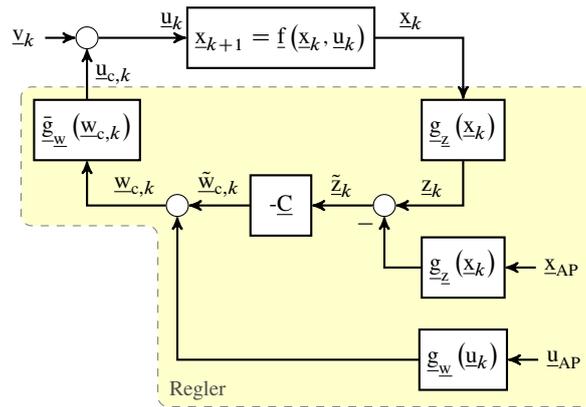


Abbildung 4.3: Verallgemeinerte Koopman-Zustandsrückführung für DKLc

Die Struktur ist ähnlich zu einem linearen Zustandsregler, welcher auf ein nichtlineares System angewendet wird. Auch hier muss der Arbeitspunkt korrigiert werden, bevor und nachdem die Signale auf die Reglermatrix beziehungsweise das Stellsignal auf das System gegeben werden. Der maßgebliche Unterschied ist die Rücktransformation des Stellsignals von  $\mathcal{W}$  nach  $\mathcal{U}$ . Formell ergibt sich das Regelgesetz zu

$$\underline{u}_{c,k} = \bar{\underline{g}}_{\underline{w}} \left( \underline{g}_{\underline{w}}(\underline{u}_{AP}) - \underline{C} \left( \underline{g}_{\underline{z}}(\underline{x}_k) - \underline{g}_{\underline{z}}(\underline{x}_{AP}) \right) \right). \quad (4.3)$$

Anhand dieser Gleichung ist zu erkennen, dass die Transformation und Rücktransformation des Eingangs mindestens am Arbeitspunkt fehlerfrei sein muss, um den Arbeitspunkt ohne Abweichung einzustellen. Ähnlich zu Gleichung (3.17) gilt für die Eingangstransformation

$$\underline{u}_k = \bar{\underline{g}}_{\underline{w}} \left( \underline{g}_{\underline{w}}(\underline{u}_k) \right) + \underline{\epsilon}(\underline{u}_k). \quad (4.4)$$

Angenommen der Arbeitspunkt  $\underline{x}_{AP}$  wird durch den Regler in Gleichung (4.3) zu einem bestimmten Zeitschritt  $\tilde{k}$  eingestellt. Dann gilt für das Stellsignal  $\underline{u}_{c,\tilde{k}} = \bar{\underline{g}}_{\underline{w}} \left( \underline{g}_{\underline{w}}(\underline{u}_{AP}) \right)$ . Wenn diese Transformation am Arbeitspunkt nun gemäß Gleichung (4.4) einen Fehler aufweist, wird das System durch den Transformationsfehler  $\underline{\epsilon}$  angeregt, was zu einem Verlassen des Arbeitspunktes führt. Durch den fehlenden Integrator im Regler kann dann nicht garantiert werden, dass diese Abweichung zum Arbeitspunkt asymptotisch verschwindet.

Weiter ist der Satz 3.11 über die Stabilität von Koopman geregelten Systemen additiv abhängig von einem Fehler in der Stellgrößentransformation, sodass ein solcher Fehler die Stabilitätsgarantie beeinflusst. Dies ist insbesondere bei der Wahl der Gewichtungen  $\alpha_3$  und  $\beta_3$  im Gütefunktional (4.1) zu berücksichtigen.

### 4.1.5 Wahl der Gewichtungsmatrizen bei LQ-Zustandsregelung

Grundsätzlich werden die Gewichte gemäß dem Vorgehen aus Abschnitt 3.4.1 gewählt und berechnet. Für ein DKLC-Modell kann der Gradient der Observables anhand der Schicht-Gewichte des künstlichen neuronalen Netzes explizit symbolisch angegeben werden. Eine Herleitung dafür ist in Anhang A.1.4 angegeben.

Auch wenn es heuristisch anschaulich ist, dass die Eigenwerte, die in der Nähe derer des linearisierten Systems liegen, die nach Heuristik 3.14  $n$  steuerbaren Eigenwerte sind, so ist es nicht immer sichergestellt, dass eine solche Linearisierung bekannt ist. Daher müssen diese zunächst identifiziert, beziehungsweise in Ihrer Steuerbarkeit bewertet, werden. Des Weiteren kann es durch die fehlerbehaftete Schätzung der Systeme dazu kommen, dass eine binäre Entscheidung über die Steuerbarkeit der einzelnen Eigenwerte nicht möglich ist. Durch die in der Approximation auftretenden Fehler kann es passieren, dass sich weitere schlechter steuerbare Eigenwerte ergeben.

Für die Quantifizierung der Steuerbarkeit einzelner Eigenwerte stehen in der Literatur einige Verfahren zur Verfügung. Dazu gehört unter anderem das Steuerbarkeitskriterium nach Gilbert [44], sowie modale Steuerbarkeitsmaße nach Müller und Lückel [84], die beide auf einer Diagonalisierung des untersuchten linearen Modells basieren.

Das Steuerbarkeitskriterium nach Gilbert ist grundsätzlich zur Evaluierung der binären Eigenschaft der Steuerbarkeit einzelner Eigenwerte konzipiert. Das lineare Zustandsraummodell wird über eine Ähnlichkeitstransformation mittels der Eigenvektoren der Systemmatrix diagonalisiert. Dabei handelt es sich um einen Basiswechsel in Eigenkoordinaten, welcher ebenso auf die Eingangsmatrix angewendet wird. So ist es möglich, durch Betrachtung der einzelnen, zu den Eigenwerten korrespondierenden Zeilen der Eingangsmatrix, darauf rückzuschließen, ob ein einzelner Eigenwert steuerbar ist oder nicht. Wenn die Koordinaten jedoch zusätzlich normiert sind, ist ebenso eine Beurteilung des Grades der Steuerbarkeit oder der Empfindlichkeit einzelner Eingänge auf spezifische Eigenwerte anhand der Einträge in der Eingangsmatrix möglich. Die Evaluierung des Grades der Steuerbarkeit nach Gilbert kann allerdings für Systeme mit mehreren Eingängen unter Umständen problematisch sein, da ein einfacher numerischer Vergleich der Einträge in der Eingangsmatrix nicht mehr möglich ist. Außerdem ist bei der Evaluierung der Steuerbarkeit bei konjugiert komplexen Eigenwerten zu beachten, dass, wenn einer dieser Eigenwerte steuerbar ist, dies auch für den anderen gilt, wie sich leicht durch eine Modaldarstellung eines Zustandsraumsystems nachweisen lässt. Die Steuerbarkeit jedes einzelnen Elements dieses Eigenwertpaares wird dann als gleichwertig angesehen.

Für die Bewertung der Steuerbarkeit einzelner Eigenwerte nach Müller und Lückel gilt im zeitdiskreten Fall

$$\xi_{i,\text{ctrb}} = \frac{\mathbf{w}_i^T \mathbf{K}_w \mathbf{K}_w^T \mathbf{w}_i}{\mathbf{w}_i^T \mathbf{w}_i} e^{1-|\lambda_i|}. \quad (4.5)$$

Dieses Maß beurteilt anschaulich gesprochen die Verschiebbarkeit der Eigenwerte und setzt zusätzlich die Stabilität beziehungsweise Dynamik über den multiplikativen Anteil der Exponentialfunktion ins Verhältnis. Somit werden schnellere Dynamikanteile über diesen Term stärker gewichtet. Gegebenenfalls kann es sinnvoll sein diesen multiplikativen Anteil für die reine Beurteilung der Steuerbarkeit, ohne Einbeziehung der Dynamik, zu vernachlässigen. Eine Normierung der Basis ist bei diesem Maß nicht notwendig, da das gesamte Maß bereits durch den Nenner  $\underline{w}_i^T \underline{w}_i$  in Gleichung (4.5) normiert wird.

#### 4.1.6 Einfluss von Mess- und Schätzungenauigkeiten auf die DKLc-Methode

Da es sich beim Deep Koopman-Operator Learning um eine Methode handelt, welche auf künstlichen neuronalen Netzen beruht, ist eine einfache Aussage für den Zusammenhang zwischen Signalrauschen und Identifikationsungenauigkeiten, wie beispielsweise bei Subspace-Verfahren ([19, 32]) oder autoregressiven Ansätzen ([65]), bisher nicht möglich. Das Gleiche gilt für eine Abschätzung bei der Variation verschiedener Anfangsgewichte der Schichten des künstlichen neuronalen Netzes, was nachfolgend experimentell untersucht wird.

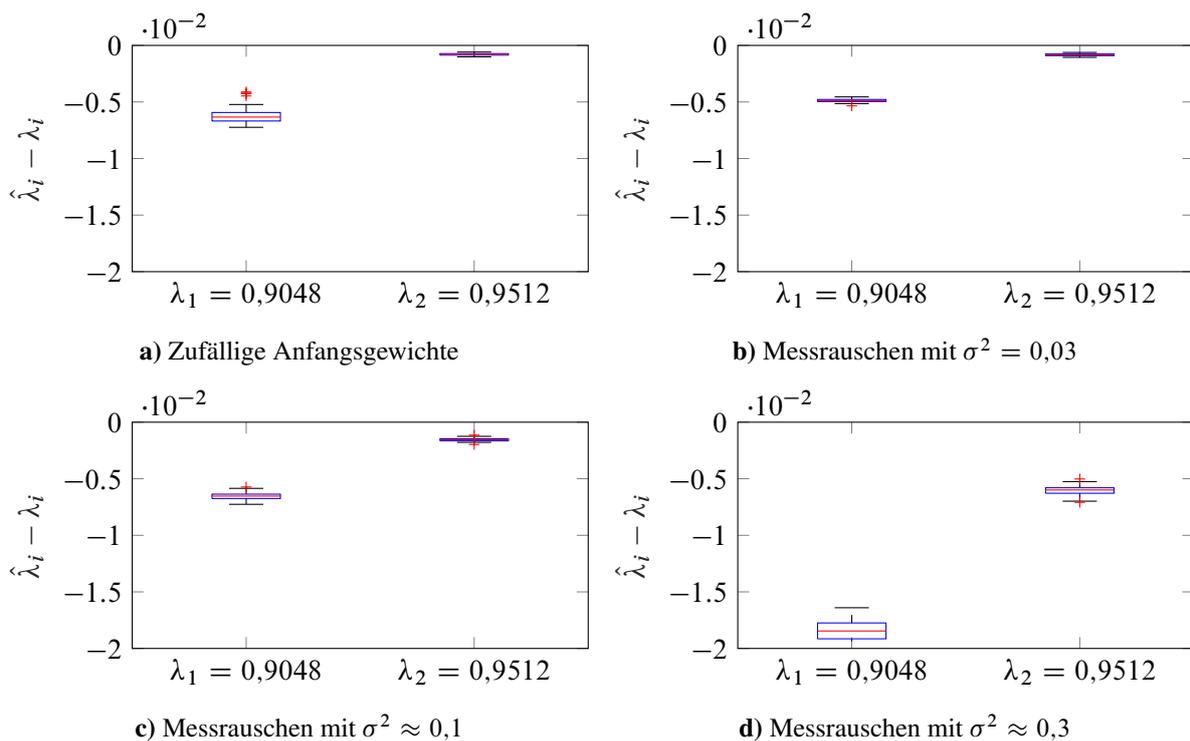


Abbildung 4.4: Zusammenhang von Schätzungenauigkeiten in den Eigenwerten über variierende initiale Schichtgewichte sowie Rauschamplituden

Da mit dem System aus (3.5) ein Modell zur Verfügung steht, für das eine exakte Lösung mit einer endlichen Anzahl an Observable- und Eigenkoordinaten bekannt ist, eignet sich dieses zur Untersuchung der Auswirkung von Rauschen und unterschiedlichen Initialisierungsgewichten. In [5] wurde die DKLc-Methode bereits in soweit evaluiert, dass sie gegen die analytische Lösung

konvergiert. In Abbildung 4.4 wird der Zusammenhang von Bias und Varianz bei einer DKLc-Schätzung über verschiedene Rauschamplituden und Initialgewichte aufgezeigt. Zu diesem Zweck werden jeweils 100 Systeme geschätzt und ausgewertet. Dabei wird das System mit einer geringeren Anzahl an Epochen trainiert, als die in Kapitel 5 betrachteten Systeme. Dies liegt maßgeblich an der Masse der zu schätzenden Systeme, um eine valide statistische Aussage treffen zu können. Hinsichtlich der praktischen Anwendung des DKLc-Verfahrens können die Modelle nicht über die Epochenanzahl aus Kapitel 5 (10.000) trainiert werden, da die Schätzung sonst mehrere Wochen in Anspruch nehmen würde. Vorangegangene Testidentifikationen lieferten jedoch die Erkenntnis, dass sich für das betrachtete System bereits nach etwa 2.000 Epochen eine Lösung einstellt, die hinsichtlich des Verlustfunktional kaum noch verbessert werden kann. Die Struktur der einzelnen Schichten des Netzes sowie die Parameter des Verlustfunktional (4.1) ist in Anhang A.3.2 angegeben.

Für die Schätzung des Rauscheinflusses wird das Netz mit 100 Datensätzen mit unterschiedlichen Rauschrealisierungen bei gleichen Initialgewichten trainiert. Bei der Untersuchung der Startgewichte wird lediglich ein fester Datensatz ohne Rauschen verwendet. Es ist ein deutlicher Zusammenhang zwischen der Rauschamplitude und der Streuung sowie einem Bias erkennbar. Bei der Erhöhung des Messrauschens auf  $\sigma^2 \approx 0,3$  kommt es zu einer vergleichsweise erheblichen Vergrößerung des Bias, sodass dieser etwa quadratisch mit der Varianz ansteigt. Der Einfluss unterschiedlicher Initialgewichte ist bezüglich des Bias ähnlich zum Fall mit Rauschen bei  $\sigma^2 \approx 0,1$ . Die Varianz ist jedoch etwas größer als im Rauschfall.

Des Weiteren ist auffällig, dass die Schätzung mittels DKLc die Modelle immer mit einer schnelleren Dynamik versieht, als es das Referenzsystem vorsieht. Ein konkreter Grund hierfür konnte nicht ermittelt werden. Setzt man voraus, dass das nichtlineare Modell des später betrachteten Drei-Tanks korrekt ist, so stellt sich für den langsamsten Eigenwert der umgekehrte Effekt ein. Daher scheint dieses Verhalten systemabhängig zu sein.

Für eine bessere Einordnung der Rauschamplitude ist in Abbildung 5.10 eine Beispieltrajektorie des später besprochenen Van der Pol Oszillators mit verschiedenen Rauschamplituden und -realisierungen gezeigt. In diesem Plot ist die Amplitude des Systemsignals gegenüber dem besprochene Beispielsystem (3.5) etwas geringer, sodass das Signal-Rausch-Verhältnis in dieser Darstellung etwas geringer ist.

## 4.2 Abschätzung der Systemordnung eines Koopman-linearen Systems

Um einen theoretisch fundierten Ausgangspunkt der zu verwendenden Systemordnung zu erhalten, und somit die Anzahl an Observables abzuschätzen, werden in dieser Arbeit die sogenannten Subspace-Verfahren verwendet. Im darauf folgenden Schritt werden dann die sogenannten *Information Criteria* angewendet, um die Systemordnung weiter zu optimieren.

Die Subspace-Identifikationsverfahren liefern neben einem linearen Zustandsraummodell auch eine Abschätzung der Systemordnung eines linearen Modells ausgehend von Messungen des Ausgangs. Da im Rahmen dieser Arbeit, ausgehend von einer Messung des gesamten Zustandsvektors, nach linearen Systemen höherer Ordnung gesucht wird, erscheint die Anwendung dieser Methoden sinnvoll. Abbildung 5.2a zeigt beispielhaft vorgreifend für das später betrachtete Van der Pol System den Verlauf der Singulärwerte der charakteristischen Matrix, welche eine Systemordnung abschätzt.

Es ist klar erkennbar, dass eine eindeutige Aussage über die Systemordnung nur bedingt möglich ist. Aus [17] sowie den Veröffentlichungen von Sadeqi et al. [99] und Borjas und Garcia [19] kann jedoch diejenige Zahl als Systemordnung ausgewählt werden, nach der sich eine Art Plateau in den Singulärwerten ausbildet. Für das betrachtete Beispiel kann somit eine Ordnung von vier beziehungsweise sieben oder acht ermittelt werden, je nachdem welchen Subspace-Algorithmus zur Ordnungsbestimmung man zugrunde legt. Da dies jedoch keine allgemein gültige Aussage ist, muss die Systemordnung weiter optimiert werden.

Dafür werden die *Information Criteria* verwendet. Sie bewerten anhand des logarithmierten mittleren quadratischen Fehlers  $\epsilon^2 = 1/N \sum_{i=1}^N (\underline{x}_i - \hat{\underline{x}}_i)^2$  zwischen Modell und Messdaten und einem Term  $\eta(N, N_p)$

$$\text{IC} = N \ln(\epsilon^2) + \eta(N, N_p)$$

die Qualität des Modells gegenüber seiner Komplexität und sollen somit ein Overfitting vermeiden. Die Komplexität wird über die Anzahl der Modellparameter  $N_p$ , welche in  $\eta$  normiert werden, bewertet. Das in der Literatur am häufigsten verwendete Kriterium ist das korrigierte *Akaike Information Criterion* (AICc), welches in Gleichung (4.6) angegeben ist [29].

$$\text{AIC}_c = 2N \ln(\epsilon) + \frac{2N N_p}{N - N_p - 1} \quad (4.6)$$

Die Korrektur des AIC ist durch Schätzung mit Messdaten und der damit verbundenen endlichen Anzahl an Messpunkten notwendig [29, S. 63]. Für  $N \rightarrow \infty$  geht das  $\text{AIC}_c$  in das AIC über.

Ursprünglich ist  $N_p$  definiert als die Anzahl an Parametern im Modell. In der Literatur ist dies jedoch nicht einheitlich. Häufig, insbesondere bei der Schätzung von Zustandsraummodellen, wird anstatt der Anzahl der freien Parameter die Ordnung  $N_p = n$  verwendet. Dies hat den Grund, dass die Anzahl an effektiven freien Parametern bei Zustandsraummodellen mit  $\mathcal{O}(n \cdot (1 + m))$  zunimmt. Dies würde bei wachsender Ordnung und Erhöhung des Eingangsraumes beziehungsweise einer großen Anzahl an Eingängen sehr schnell dazu führen, dass  $\eta$  sehr groß wird und der Einfluss der Schätzqualität  $\epsilon$  abnimmt, sodass praktisch immer Modelle mit niedriger Ordnung den Vergleich gewinnen, auch wenn die Schätzqualität der Modelle mit nur leicht erhöhter Komplexität gegebenenfalls deutlich bessere Prädiktionsergebnisse liefern.

Grundsätzlich ist es außerdem denkbar, auch die Komplexität der geschätzten künstlichen neuronalen Netze mit  $N_P$  zu bewerten. Dafür kann beispielsweise die Schichtdicke der einzelnen Encoder und Decoder oder die Anzahl an Neuronen pro Schicht herangezogen werden. Da die Netze im nachfolgenden Kapitel jedoch nicht, beziehungsweise nur äußerst geringfügig variiert werden, ist eine Bewertung der Netzkomplexität in dieser Arbeit nicht notwendig.

Wichtig ist bei der Anwendung der *Information Criteria* vor allem zu verstehen, dass es nicht auf den absoluten Wert des  $AIC_c$  ankommt. Vielmehr ist das Verhältnis zwischen den einzelnen Werten der verschiedenen Modelle von Belang. Daher wird häufig auch eine Differenz untereinander bewertet, worauf in dieser Arbeit durch die visuelle Darstellung verzichtet wird.

### 4.3 Methodik zur Identifikation, Analyse und Regelung nichtlinearer dynamischer Systeme

Das gesamte Framework, von Voranalyse des Systems und der Daten, über die Identifikation bis hin zur Regelung des nichtlinearen Systems, aus diesem und dem vorangegangenen Kapitel 3, ist in Abbildung 4.5 noch einmal übersichtlich zusammengefasst. Da die Vorverarbeitung der Daten optional ist, ist dieser Prozess gestrichelt dargestellt. Der graue gestrichelte Pfeil zeigt an, dass eine umfassende, strukturelle Anpassung des Modells notwendig sein kann, sollte das Training nicht konvergieren oder der Prädiktionsfehler zu groß sein.

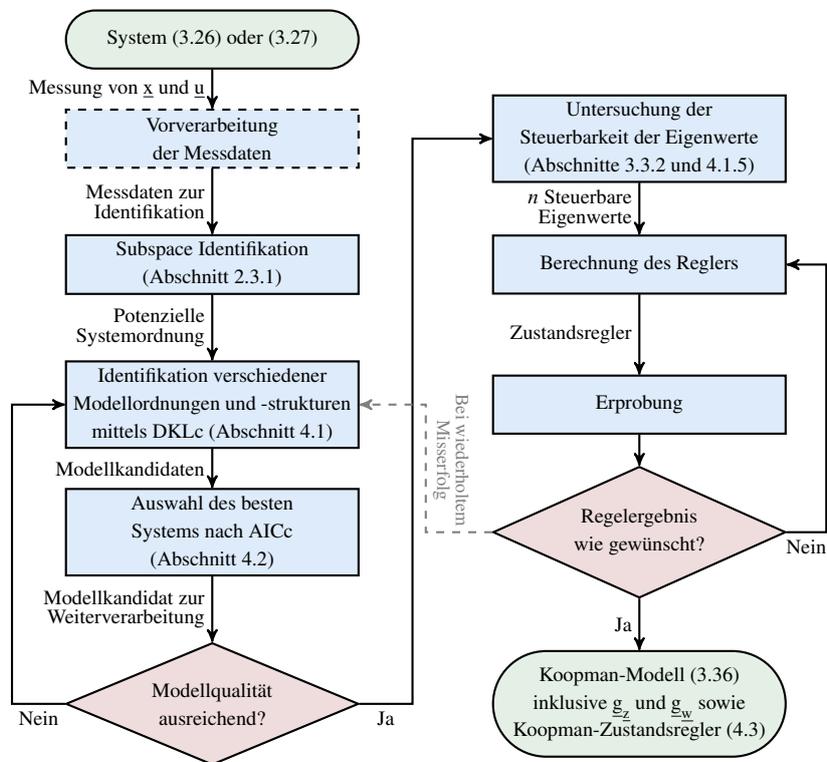


Abbildung 4.5: Grafische Darstellung des in dieser Arbeit entwickelten Identifikations-, Analyse- und Regelungs-Frameworks

## 5 Analyse und Regelung dynamischer Systeme

Dieses Kapitel greift die in den vorangegangenen Kapiteln erarbeiteten Konzepte und Methoden auf und wendet sie auf zwei simulierte Systeme und einen realen Prüfstand an. Dabei werden gemäß dem Schema aus Abbildung 4.5 zunächst Daten des jeweiligen Systems gemessen und gegebenenfalls durch Filterung und Unterabtastung vorverarbeitet. Im Anschluss wird das System auf Basis dieser Daten identifiziert, analysiert und geregelt. Die Ergebnisse dieses Vorgehens werden zusätzlich mit anderen, in Kapitel 2.3 beschriebenen, Methoden verglichen. Die künstlichen neuronalen Netze zur Ermittlung der Modelle werden auf einem Rechner mit AMD Ryzen 3700X Prozessor und Nvidia GeForce<sup>®</sup> RTX 2080 Grafikkarte trainiert.

Um die in dieser Arbeit vorgestellte DKLc-Methode praktisch nutzbar zu machen, ist eine leistungsfähige Grafikkarte von höchster Bedeutung. Durch den physischen Aufbau einer GPU kann die Zeit pro Optimierungsschritt gegenüber einer CPU deutlich reduziert werden [15]. Je nach Netzkonfiguration und Datenmenge ist eine Optimierung auf einer GPU für das DKLc-Verfahren zwischen drei- bis zehnfach schneller als eine Optimierung auf einer CPU.

In diesem Kapitel werden neben dem eigentlichen Verfahren auch verschiedene Modellkonfigurationen untersucht. Diese sind dabei durch eine Zahl mit vorangestelltem „S“ (für Struktur) gekennzeichnet. Die Benennung sowie die entsprechende Konfiguration ist in Tabelle 5.1 aufgetragen. Genauere Angaben zur Konfiguration der Observables sind im entsprechenden Abschnitt beschrieben.

Tabelle 5.1: Erläuterung der untersuchten Modellstrukturen

| Bezeichnung | Observable-Vektoren   | Erläuterung   |
|-------------|---|---|
| S1          | $\underline{g}_z(\underline{x}) = \left[ \underline{x}^T, \underline{g}_{z,1}(\underline{x})^T \right]^T$ $\underline{g}_w(\underline{u}) = \underline{u}$                      | Der Zustand ist Teil des Observable-Vektors.<br>Der Eingang ist der Eingangsobservable-Vektor |
| S2          | $\underline{g}_z(\underline{x}) = \left[ \underline{x}^T, \underline{g}_{z,1}(\underline{x})^T \right]^T$ $\underline{g}_w(\underline{u}) = \underline{g}_{w,1}(\underline{u})$ | Der Zustand ist Teil des Observable-Vektors.<br>Der Eingang wird vollständig kodiert.         |
| S3          | $\underline{g}_z(\underline{x}) = \underline{g}_{z,1}(\underline{x})$ $\underline{g}_w(\underline{u}) = \underline{u}$  | Der Zustand wird vollständig kodiert.<br>Der Eingang ist der Eingangsobservable-Vektor        |
| S4          | $\underline{g}_z(\underline{x}) = \underline{g}_{z,1}(\underline{x})$ $\underline{g}_w(\underline{u}) = \underline{g}_{w,1}(\underline{u})$                                     | Zustand und Eingang werden vollständig kodiert  |

Grundsätzlich lässt es das DKLc-Verfahren zu, dass zusätzliche Eingangsobservable generiert werden, sodass  $m_e > m$  gilt. Dies wird jedoch aufgrund der Ausführungen in Abschnitt 3.4.1 in dieser Arbeit nicht betrachtet. Daher ist der Eingangsobservable-Vektor für die Konfigurationen S1 und S3 der gemessene Eingangsvektor und für die Modellstrukturen S2 und S4 eine vollständige Transformation des Eingangs ohne eine Ordnungsanhebung.

Die Funktionsfähigkeit der vorgestellten Methodik wird in [5] analysiert und für das System aus Gleichung (3.5) gezeigt. Hierin wird eine konsistente Konvergenz hin zu den Eigenfunktionen des Systems experimentell gezeigt. Daher wird sich in diesem Abschnitt darauf beschränkt, komplexere Systeme zu analysieren und diese zu regeln.

Es sei darauf hingewiesen, dass sich die in dieser Arbeit gezeigten Ergebnisse durch die randomisierte Initialisierung der Schichtgewichte bei erneuter Identifikation leicht verändern können. Dadurch kann es im Einzelfall dazu kommen, dass eine andere Ordnung ausgewählt wird oder sich das Prädiktions- und/oder Regelverhalten geringfügig verändert. Durch die begrenzte Rechenkapazität ist eine umfassende und mehrfache Identifikation von einzelnen Konfigurationen mit der gegebenen Hardware nicht realistisch möglich, da eine komplette Identifikation inklusive einer Hyperparameteroptimierung sowie Ordnungssuche, wie sie in den folgenden Abschnitten gezeigt werden, für ein einzelnes System bis zu einem Monat reine Rechenzeit in Anspruch nehmen kann.

Um die in der vorliegenden Arbeit entworfenen Verfahren und theoretischen Grundlagen anhand von Beispielen zu validieren, werden diese einleitend anhand von zwei rein numerischen Beispielen erprobt. Danach folgt die Anwendung auf einen realen Prüfstand<sup>21</sup>.

## 5.1 Numerische Beispiele

Die mathematisch motivierten dynamischen Systeme werden ausgewählt, weil sie nicht nur über ausgeprägte Nichtlinearitäten, sondern auch über mehrere Arbeitspunkte verfügen. Dies gilt insbesondere für das zweite Beispiel, den Duffing Oszillator. Zur besseren Verständlichkeit beginnt die Validierung der Methoden anhand des etwas weniger komplizierten Van der Pol Systems.

### 5.1.1 Van der Pol Oszillator

Der Van der Pol Oszillator ist ein nichtlineares System, welches eine Dauerschwingung, beziehungsweise einen Grenzzyklus aufweist. Das hier verwendete zugehörige nichtlineare und angelegte Zustandsraum-Modell lautet

$$\dot{\mathbf{x}} = \begin{bmatrix} x_2 \\ \mu (1 - x_1^2) x_2 - x_1 \end{bmatrix} + \begin{bmatrix} 0 \\ \beta \end{bmatrix} u.$$

<sup>21</sup>Das auf Seite 59 angegebene Github-Repository beinhaltet neben der DKLc-Software sämtliche, in diesem Kapitel besprochene, Beispiele mit Skripten und Abbildungen.

Für die Parameter gilt  $[\mu, \beta] = [1, 1]$ . Beim Van der Pol Oszillator handelt es sich um ein selbsterregtes System. Das autonome System besitzt lediglich eine Ruhelage in  $\underline{x}_R = \underline{0}$ . Über den Eingang kann jeder beliebige stationäre Zustand der Form  $\underline{x}_{R,\mu} = [u_R, 0]^T$  eingestellt werden. Abbildung 5.1 zeigt die autonomen Trajektorien des Van der Pol Oszillators für verschiedene Anfangswerte. Es ist deutlich der charakteristische Grenzzyklus des Oszillators zu erkennen, auf den alle Trajektorien bei unterschiedlichen Anfangswerten zulaufen. Dieses Verhalten ist unabhängig davon, ob die Trajektorien außerhalb oder innerhalb des Attraktors starten.

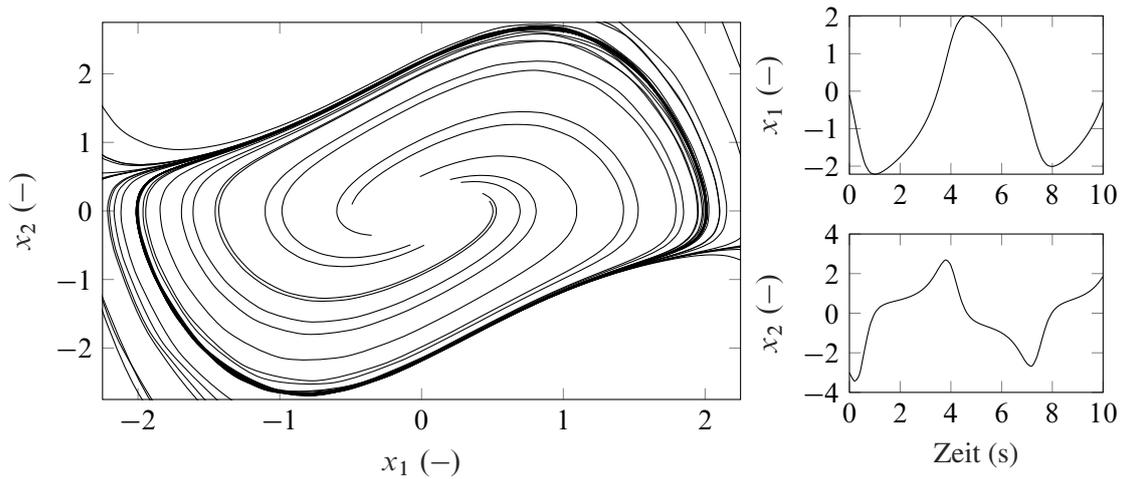


Abbildung 5.1: Autonome Trajektorien des Van der Pol Oszillators

Das System wird für die Identifikation mit einem normalverteilten Signal mit  $\mathcal{N}(0, 1)$  und einem gleichverteilt zufälligen Anfangszustand vom Rand einer der beiden Räume  $\mathcal{X}_0 = \{\underline{x} \mid x_1 \in [-2.5, \dots, 2.5], x_2 \in [-3, \dots, 3]\}$  und  $\mathcal{X}_1 = \{\underline{x} \mid x_1^2 + x_2^2 = 0,5^2\}$  angeregt. Welcher der beiden Räume für einen einzelnen Startzustand ausgewählt wird ist ebenso gleichverteilt zufällig. Simuliert wird das System ausgehend von einer Vielzahl von Anfangszuständen mit einer Abtastzeit von  $T_S = 0,1$  s über  $T = 30$  s. So ergeben sich mehrere einzelne Trajektorien, welche auch das transiente Verhalten mit einbeziehen, sodass dieses ebenso mit geschätzt werden kann.

### Identifikation

Das beschriebene System wird bei gegebenen Daten mit dem in Kapitel 4 beschriebenen Vorgehen identifiziert, wobei zunächst rauschfreie Daten analysiert werden. Die zum Training der Netze verwendeten Parameter sind in Abschnitt A.3.3 im Anhang aufgeführt. Zur ersten Abschätzung der Ordnung kommt eine Subspace-Identifikation auf Basis einer einzelnen Trajektorie zum Einsatz. Wie Abbildung 5.2a zeigt, kann je nach Herangehensweise und Methode eine Systemordnung von vier, respektive sieben oder acht abgeschätzt werden. In der Literatur wird häufig die Ordnung als signifikant angesehen, für die sich der zugehörige Singulärwert oberhalb des Mittelwertes aus größtem und kleinsten logarithmierten Singulärwert befindet [14]. Wie unter anderem durch Bonnert [17], Sadeqi et al. [99] und Borjas und Garcia [19] beschrieben, liefert der Wert nach dem sich im Singulärwertplot eine Art Plateau ausbildet eine bessere Abschätzung einer linearen Systemordnung. Dadurch ergibt sich die erwähnte Ordnung von sieben respektive acht.

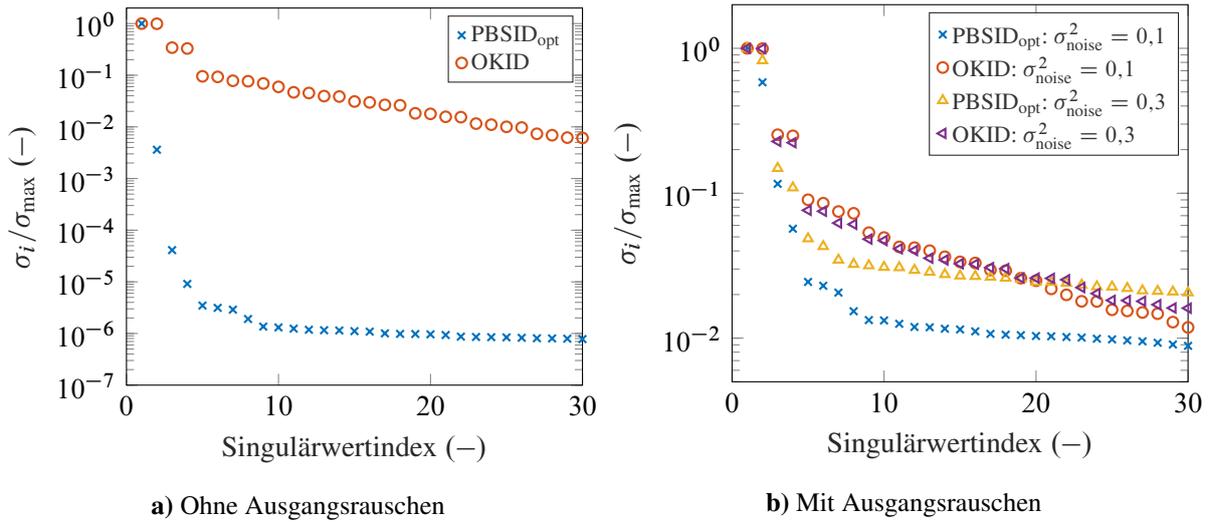


Abbildung 5.2: Singulärwerte einer PBSID<sub>opt</sub> Schätzung mit  $p_{\text{PBSID}} = 10$  und  $f_{\text{PBSID}} = 40$  und OKID-ERA Schätzung mit  $p_{\text{OKID}} = 50$

Diese lineare Systemordnung dient als Ausgangspunkt für die Schätzung mittels DKLc. Für die Schätzung werden Ordnungen zwischen zwei und zehn, sowie eine vergleichsweise hohe Systemordnung von 20 verwendet, um einen Vergleichswert für große Ordnungen zu generieren. Abbildung 5.3 zeigt den Verlauf des AIC<sub>c</sub> für verschiedene Systemkonfigurationen nach Tabelle 5.1 sowie Systemordnungen. Die auf der Abszisse aufgetragene Systemordnung schließt die im Observable-Vektor direkt verwendeten Zustände mit ein, sodass beispielsweise bei den Strukturen S1 und S2 bei einer Systemordnung von vier, zwei zusätzliche Observables geschätzt werden.

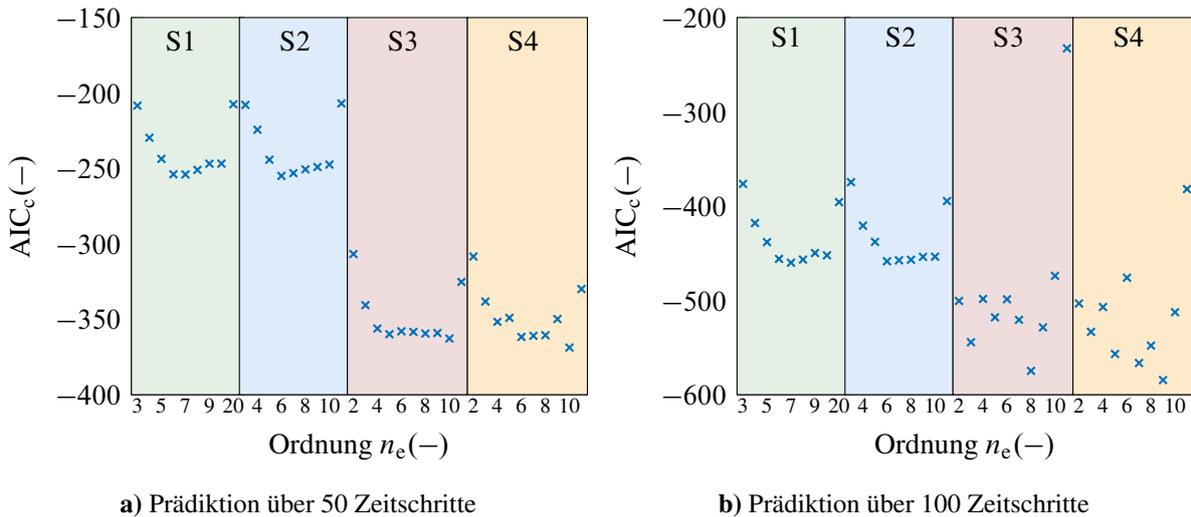


Abbildung 5.3: AIC<sub>c</sub> bei verschiedenen Systemordnungen für die DKLc-Schätzung. Durch die eng beieinander liegenden Datenpunkte ist nur jede zweite Ordnung auf der Abszisse aufgetragen.

Durch die Logarithmierung des Fehlers geht ein numerisch kleiner Fehler mit einem großen negativer Wert im AIC<sub>c</sub> und somit mit einer guten Schätzung einher. Bei gleicher Ordnung ist dies

nach Gleichung (4.6) also mit einem kleineren Prädiktionsfehler gleichzusetzen. Besonders im Bereich kleiner Ordnungen dominiert der Fehler, was zu einem numerisch größeren  $AIC_c$  führt. Für die Einbeziehung der Zustände direkt als Observable ist deutlich erkennbar, dass für den Van der Pol Oszillator dies eine erhebliche Einschränkung zu sein scheint, da für ausschließlich über Observables geschätzte Systeme (Modellstrukturen S3 und S4) deutlich niedrigere Werte für den  $AIC_c$  erreicht werden.

Dieses Bild bestätigt sich auch in Abbildung 5.5a, welche direkt den *Mean Relative Squared Error* (MRSE) in Prozent ohne Einbezug der Systemkomplexität zeigt. Der Fehler nimmt nahezu monoton für alle Modellstrukturen mit einer Erhöhung der Systemordnung ab. Für die Strukturen S1 und S2 verringert sich ab einer Ordnung von etwa  $n_e = 6$  der Fehler nicht mehr maßgeblich, weshalb der  $AIC_c$  dies als beste Systemordnung auswählt.

Es ist festzuhalten, dass für den in der Schätzung gewählten Prädiktionshorizont von  $p_{DKLc} = 50$  Zeitschritten die Systemordnung von vier, welche einer Interpretation der Subspace-Schätzung nach als Basiswert dient, zumindest zu den niedrigsten  $AIC_c$  Werten gehört, aber nie das Minimum bildet, wie Abbildung 5.3 zeigt. Für einen Prädiktionshorizont von  $p = 50$  nimmt der  $AIC_c$  für alle Konfigurationen bis zu einer Ordnung von fünf bis sechs über fast alle Modellstrukturen monoton ab und steigt für die Strukturen mit Zustandsdurchgriff dann wieder an. Dies gibt Aufschluss darüber, dass die erhöhte Modellkomplexität nicht oder nur wenig zu einer verbesserten Schätzqualität führt. Für komplett zustandstransformierte Systeme ist dies weniger bis gar nicht ausgeprägt, sodass der  $AIC_c$  mit Erhöhung der linearen Modellordnung häufig weiter abnimmt.

Für die folgende Betrachtung wird das Systemmodell, welches nur für  $p_{DKLc} = 50$  Zeitschritte trainiert wurde, über  $p = 100$  Zeitschritte prädiziert. Für diese erweiterte Prädiktion bleibt die nach dem  $AIC_c$  beste Systemordnung über alle Strukturen nahezu erhalten, wie in Abbildung 5.3b dargestellt. Für die Systemmodelle S3 und S4 ändert sich das allgemeine Bild jedoch maßgeblich. In den Fällen in denen die Prädiktion über 50 Zeitschritte eine Systemordnung von zehn als beste ausgewählt werden würde, ist es bei der längeren Prädiktion eine Systemordnung von acht beziehungsweise neun. Rein numerisch unterscheidet sich diese Ordnung nicht zu sehr von der mit  $p = 50$  ermittelten Ordnung. Allerdings besitzt sie durch den nun weniger glatten Verlauf des  $AIC_c$ , und den damit größeren Werten um diese Ordnung herum, einen sehr viel größeren *Support* [29, S. 70]. Der Begriff *Support* beschreibt dabei die Eigenschaft, wie gut das Modell das Systemverhalten abbilden kann.

Wichtig bei der Verwendung des  $AIC_c$  ist außerdem, insbesondere bei diesem Beispiel, dass nicht zwingend auf den niedrigsten Wert geachtet werden muss. Wenn sich verschiedene Modelle im  $AIC_c$  numerisch nur geringfügig unterscheiden, so können alle als geeignet und gleichwertig angesehen werden. Auch dies fällt unter den im vorherigen Absatz erwähnten Begriff *Support* [29, S. 70]. Das heißt, wenn verschiedene Modelle mit ähnlichem  $AIC_c$  hinsichtlich anderen Kriterien, wie beispielsweise einer Regelung, unterschiedliche Eigenschaften aufweisen, so ist das Modell zu bevorzugen, welches hinsichtlich des ausgewählten Kriteriums bessere Eigenschaften besitzt. Dies gilt somit auch dann, wenn das ausgewählte Modell nicht den niedrigsten  $AIC_c$  aufweist. Ein Beispiel dafür sind die Systemordnungen für S1 aus Abbildung 5.3a. Hier ist der  $AIC_c$  für  $n_e = 7$

im Vergleich zu  $n_e = 6$  zwar niedriger, aber ihre numerischen Werte liegen sehr dicht beieinander. Die beiden Modelle besitzen somit einen ähnlichen *Support*. Sollte in einem bestimmten Anwendungsfall die Systemordnung ein kritischer Faktor sein, so kann die niedrigere Systemordnung durch den ähnlichen *Support* ohne maßgebliche Qualitätseinbußen ausgewählt werden.

Abbildung 5.4 zeigt beispielhaft den Verlauf einer Trajektorie für alle vier Konfigurationen bei einer Systemordnung von  $n_e = 7$ . Die Auswahl dieser Systemordnung wird auf Basis des besten  $AIC_c$  Scores der Struktur S1 getroffen. Es bestätigt sich das Bild aus Abbildung 5.3a. Über den Verlauf von 50 Zeitschritten ist der MRSE der Modellstrukturen S1 und S2 deutlich größer als bei den anderen beiden. Beim Zustand  $x_2$  ist dieses Verhalten besonders ausgeprägt. Der MRSE der einzelnen Modelle ist dabei in der Legende aufgetragen.

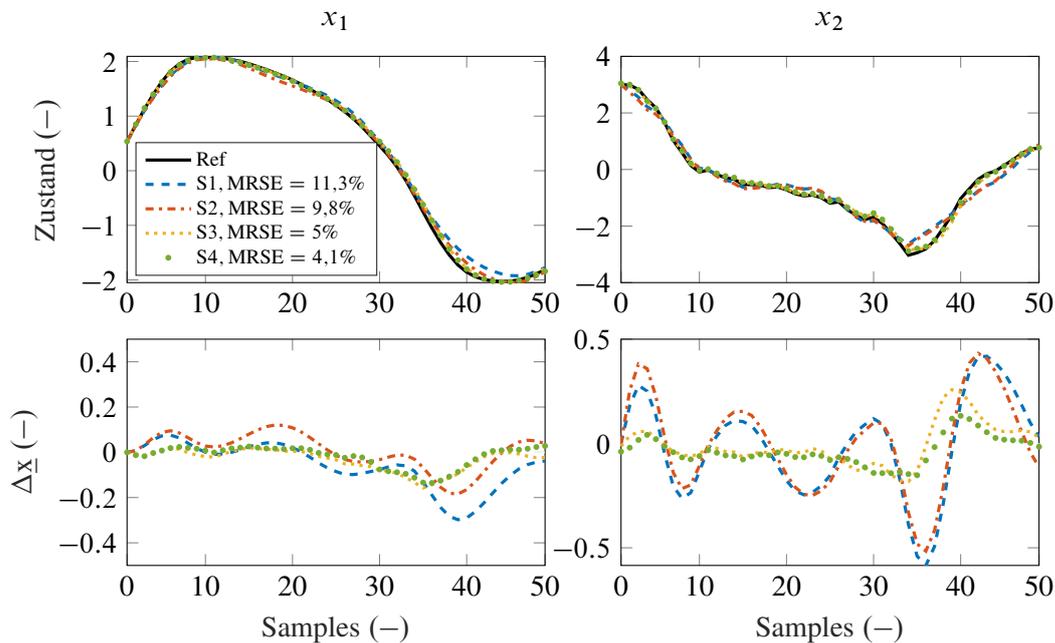


Abbildung 5.4: Modellprädiktion und Fehlerverlauf bei  $n_e = 7$

Eine Vergleichsschätzung mit  $PBSID_{opt}$  und einer randomisierten eDMD-Schätzung ist in Abbildung 5.5 für verschiedene Ordnungen dargestellt. Da die Schätzung mit OKID sowie ARX keine zufriedenstellenden Ergebnisse erreicht, sind diese aus der Bewertung ausgenommen. Es ist zu beachten, dass für die  $PBSID_{opt}$  und OKID Schätzung nicht, wie bei den anderen Verfahren, eine Vielzahl von einzelnen Trajektorien geschätzt wird. Wie in Abschnitt 2.3 erläutert, sind beim  $PBSID_{opt}$  Verfahren zwei Horizonte zu wählen, die für unterschiedliche Teile der Schätzung verantwortlich sind. Um vergleichbar mit den DKLC-Schätzungen zu bleiben, werden die Horizonte zu  $p_{PBSID} = 10$  und  $f_{PBSID} = 40$  gewählt, um so auf den gleichen Prädiktionshorizont von  $p_{DKLC} = 50$  zu kommen. Der Parameter  $f_{PBSID}$  bestimmt die Prädiktion und wird deshalb etwas größer gewählt. Für die OKID Schätzung wird passend ein Horizont von  $p_{OKID} = 50$  gewählt. Für die randomisierte eDMD wird stets das beste Modell aus 100 geschätzten zufälligen Modellen ausgewählt. Die eDMD lässt für die Schätzung keinen Horizont zu, sodass für die 100 Modelle zusätzlich 100 unterschiedliche Messreihen mit einer Länge von jeweils 50 Samples verwendet werden. Das anhand von Testdaten ermittelte beste Modell wird dann in die Darstellung als Re-

ferenz aufgenommen. Die tendenziell großen Fehler im Bereich um 20 % ergeben sich durch die Vielzahl von getesteten Trajektorien. Für einzelne Trajektorien ist der Fehler meist im einstelligen Prozentbereich, wie Abbildung 5.4 anhand einer Beispielordnung zeigt.

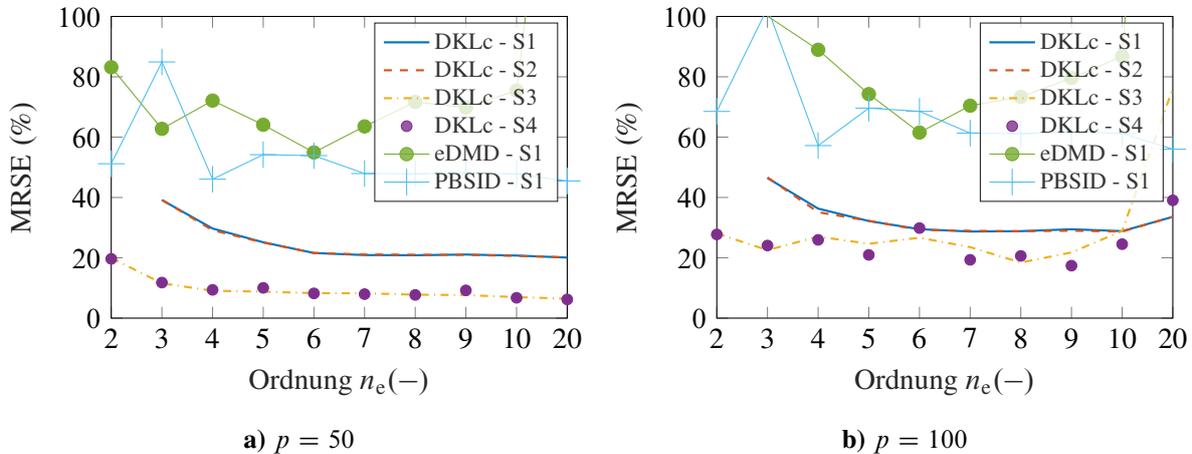


Abbildung 5.5: MRSE verschiedener Schätzverfahren gegenüber Koopman-Operator basierten Verfahren über einen Prädiktionshorizont von  $p = 50$  und  $p = 100$  Samples

Abbildung 5.5 zeigt anschaulich die reine Schätzqualität der verschiedenen Verfahren ohne Bewertung der Modellkomplexität. Es wird schnell deutlich, dass das DKLc-Verfahren über fast alle Systemordnungen die anderen Methoden hinsichtlich des MRSE übertrifft. Lediglich für eine sehr hohe Ordnung von  $n_e = 20$  sowie Prädiktionshorizonte  $p > 50$  kann das Verfahren nicht die selbe Qualität wie der PBSID<sub>opt</sub>-Algorithmus erreichen. Dies könnte jedoch auch durch die erhöhte Komplexität des Modells, sowie dem starken Fokus auf die ersten  $p_{\text{DKLc}} = 50$  Zeitschritte während des Trainings ausgelöst werden und lässt sich gegebenenfalls mit einer veränderten Trainingszeit oder leicht veränderten Gewichten im Verlustfunktional (4.1) verbessern. Zudem weist diese Abbildung auf einen enormen Performancevorteil gegenüber anderen Methoden hin. Das Verfahren von Korda und Mezić aus [59], welches auf Eigenfunktionen basiert, benötigt beispielsweise für eine ähnliche Schätzqualität eine Systemordnung von  $n_e = 20$ . Von weiterem Interesse ist außerdem die Bewertung der linearen Abbildbarkeit der Dynamik in Form des Koopman-linearen Systems.

Während Abbildung 5.5 den Fehler im Zustandsraum zeigt, gibt Tabelle 5.2 den Fehler im Koopman-Raum an. Dabei wird Struktur S1 verwendet, da diese derer entspricht, die bei der eDMD-Schätzung verwendet wird. Ohne genauere Analyse verdeutlicht sich das Bild, dass der eDMD-Ansatz mit den gegebenen Ansatzfunktionen nicht in der Lage ist, das Systemverhalten adäquat zu beschreiben. Der DKLc-Ansatz hingegen bietet einen geringen Fehler auch in der linearen Prädiktion im Koopman-Raum, was zeigt, dass das Modell auch dann gute Ergebnisse liefert wenn das nichtlineare Verhalten linear prädiziert wird. Das Ergebnis wird also nicht durch die Rücktransformation im Decoder künstlich geschönt. Auch hier wird klar, dass der DKLc-Ansatz sehr viel besser in der Lage ist, eine tatsächlich lineare Dynamik in den transformierten Koordinaten darzustellen als der vielerorts verwendete eDMD-Ansatz. Dies ist eine Grundvoraussetzung für den linearen Reglerentwurf, denn selbst wenn der eDMD-Ansatz unter Umständen in den Basisoko-

ordinaten zu einem befriedigenden Ergebnis führt, die Dynamik in transformierten Koordinaten aber mit sehr großen Fehlern behaftet ist, so ist eine Stabilität im Regelungsbetrieb nicht mehr gewährleistet, wie Satz 3.11 besagt.

Tabelle 5.2: Ergebnisse linearer Schätzungen im Koopman-Raum über einen Horizont von  $p = 100$  Samples

| MRSE (%)    |      | Ordnung $n_e$ |       |       |       |        |        |       |
|-------------|------|---------------|-------|-------|-------|--------|--------|-------|
|             |      | 3             | 4     | 6     | 8     | 9      | 10     | 20    |
| Algorithmus | DKLc | 20,86         | 12,99 | 15,06 | 21,6  | 19,58  | 23,14  | 86,29 |
|             | eDMD | 55,46         | 35,55 | 59,66 | 84,14 | 177,16 | 203,03 | -     |

Für ein umfassendes und detailreicheres Bild der Prädiktionsfähigkeit der einzelnen Modelle sind die Fehlerhistogramme für die jeweils nach  $AIC_c$  besten Modelle der einzelnen Modellstrukturen in Abbildung 5.6 aufgetragen. Hier zeigt sich ein grundlegender Unterschied zwischen den Modellstrukturen mit und ohne Zustandsdurchführung. Während die Strukturen S1 und S2 nach kurzer Zeit eine größere Varianz aufweisen, diese danach jedoch etwa gleichbleibend ist, ist die Varianz des Fehlers für Strukturen S3 und S4 bis kurz vor Ende des trainierten Prädiktionshorizonts  $p_{DKLc}$  um etwa 70 % kleiner, steigt dann aber an. Des Weiteren scheint der zweite Zustand etwas schwieriger zu präzidieren zu sein, da er bei allen Modellansätzen die größte Streuung aufweist. Jedoch ist auch hier die Prädiktion bei vollständig transformiertem Zustand um einiges besser. Der Fehler nimmt jedoch mit der Zeit immer weiter zu. Diese mehr oder weniger schlagartige Vergrößerung des Fehlers ist auf den Vergessensfaktor  $\delta = 0,99$  zurückzuführen, welcher bei der Optimierung verwendet wird. Der Fehler ab einem Zeithorizont von circa  $p = 40$  fließt nur noch mit etwa  $\delta^{p_i} < 0,67$  ein und hat damit bereits nur noch weniger als 70% des Einflusses gegenüber den vorangegangenen Zeitschritten. Dies verschiebt den Fokus der Optimierung auf die kürzeren Zeithorizonte, was das Optimierungsproblem besser konditioniert und auch für eine Regelung heuristisch gesehen ausreichend ist, da für die Zustandsregelung ohnehin nur der aktuelle Zustand (ohne Verzögerungsglieder) verwendet wird. Allerdings weist dieses Verhalten zunächst auch auf eine gewisse Überanpassung hin. Der weitläufig geringere Fehler für die später präzidierten Bereiche  $k > 50$ , im Vergleich zu den Modellstrukturen S1 und S2, widerspricht dieser Überanpassung jedoch.

Abbildung 5.7 zeigt des Weiteren das Fehlerhistogramm für die beste eDMD-Schätzung nach Abbildung 5.5. Aufgrund der starken Ähnlichkeit entfällt die Darstellung der  $PBSID_{opt}$ -Schätzung. Es ist klar ersichtlich, dass, obwohl der MRSE im Mittel für die niedrigeren Systemordnungen „nur“ etwa doppelt so groß ist wie der der DKLc-Modelle, die Prädiktionsfähigkeit für viele der Testtrajektorien unbrauchbar ist. Dies ist insbesondere deswegen der Fall, da sich der Wertebereich der Daten, welche dargestellt sind, etwa im Bereich um  $\pm 1$  befindet und der dargestellte absolute Fehler damit den gesamten Wertebereich ausnutzt.

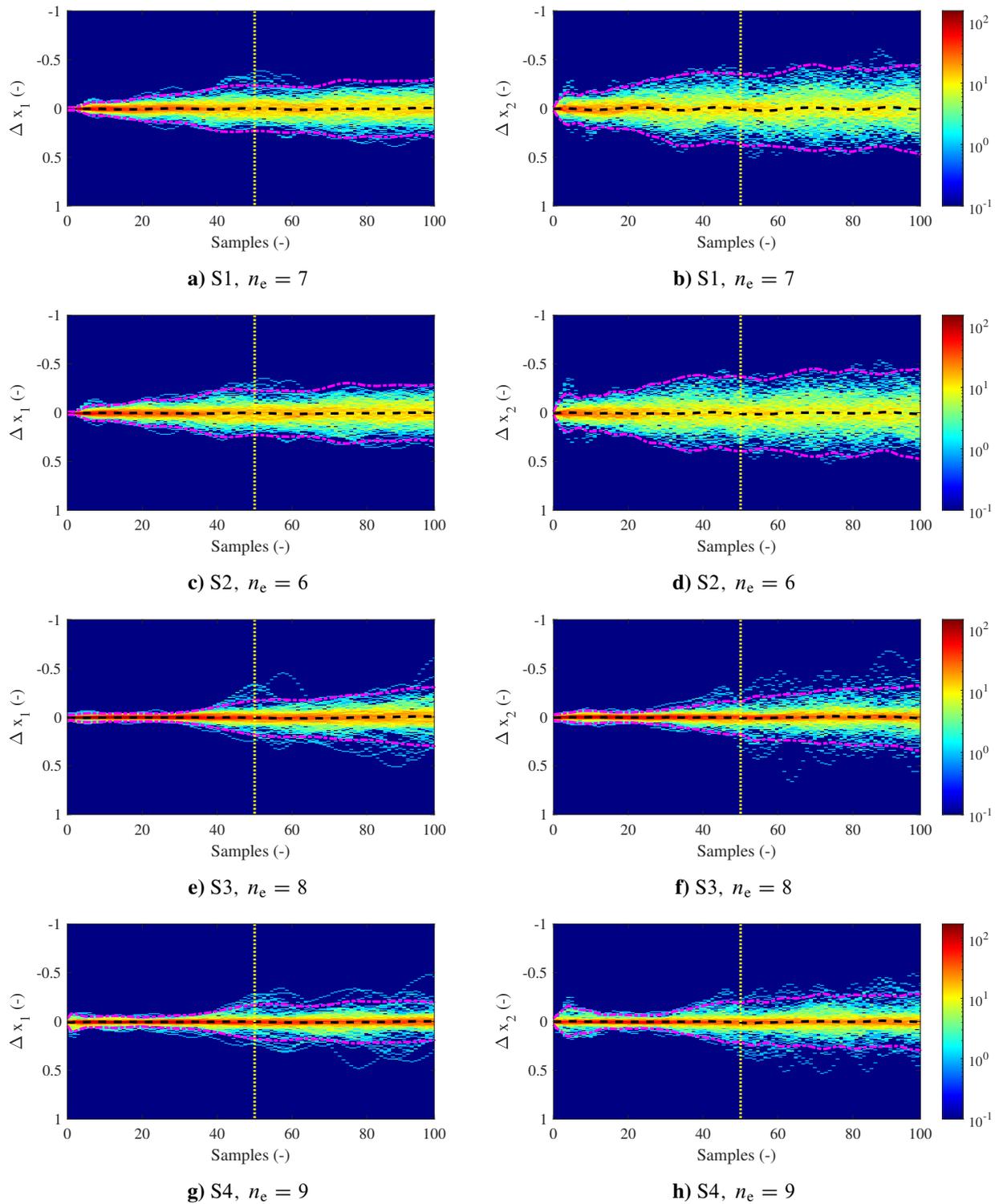


Abbildung 5.6: Fehler Histogramme der nach  $AIC_c$  mit  $p = 100$  besten Modelle des Van der Pol Systems für verschiedene Konfigurationen

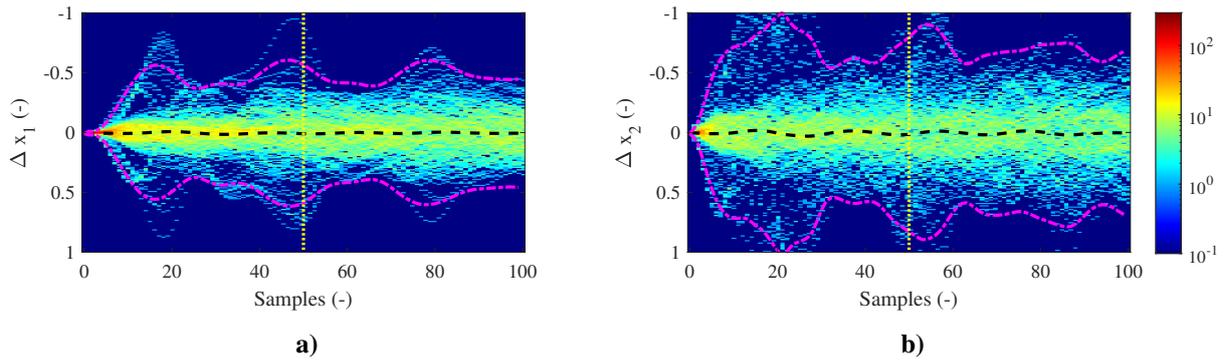


Abbildung 5.7: eDMD-Schätzung des Van der Pol Oszillators mit  $n_e = 4$

Das Einbinden einer Eingangsnichtlinearität beziehungsweise eines Eingangscoders, führt zu keiner signifikanten Verbesserung. Je nach gewählter Systemordnung wird das System etwas besser oder schlechter geschätzt, wie Abbildung 5.5 und die visuelle Betrachtung der Fehlerhistogramme aus Abbildung 5.6 zeigen. Im Großen und Ganzen deckt sich dies mit den Ausführungen aus Abschnitt 3.2, der darüber Aufschluss gibt, dass die Beschaffenheit des Eingangs auch bei der Übertragung in den Koopman-Raum erhalten bleibt<sup>22</sup>. Da das System in eingangsaffiner Form vorliegt, bringt somit die Einbindung einer Nichtlinearität keinen Vorteil. Bei genauerer Betrachtung lässt sich jedoch nicht sagen, dass der Eingangscoder eine lineare Funktion schätzt, wie Abbildung 5.8 anhand verschiedener Ordnungen zeigt. In den meisten Fällen geben die Encoder jedoch eine näherungsweise lineare Funktionen an, welche teilweise über einen Offset verfügen.

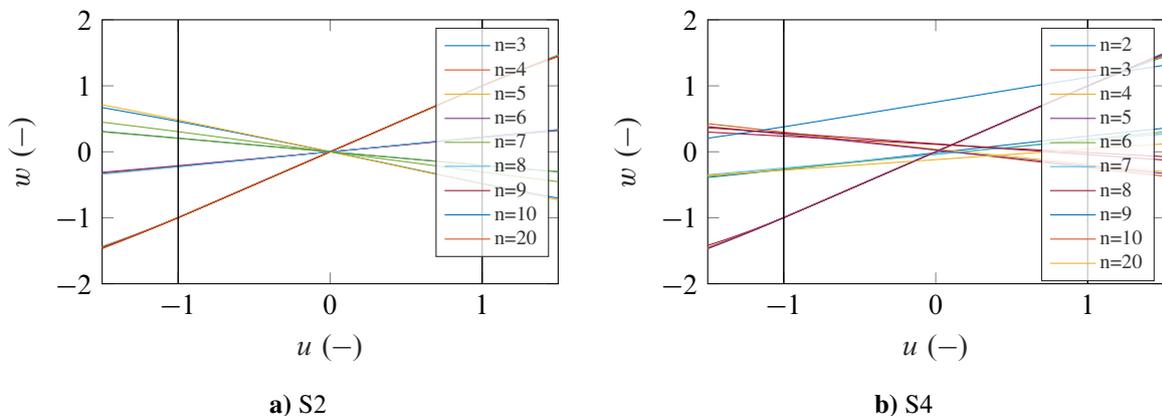


Abbildung 5.8: Geschätzte Eingangsobservable für verschiedene Ordnungen. Die vertikalen, schwarzen Linien indizieren den trainierten Wertebereich.

Bei diesem Beispielsystem bringt der Eingangscoder keinen konkreten Nutzen. In den Arbeiten von Gräfe [5, S. 56] sowie Bonnert und Konigorski [18] zeigt sich jedoch, dass bei ausgeprägter Eingangsnichtlinearität die Einbindung eines Eingangscoders notwendig sein kann, um eine adäquate Schätzung zu ermöglichen.

<sup>22</sup>Wie im erwähnten Abschnitt ebenfalls erläutert, muss genau genommen ein multiplikativer Term, welcher eine Funktion des Zustandes  $\underline{x}$  ist, eingebunden werden. Diese Beschaffenheit wird jedoch vernachlässigt, um die Regelung einfach zu gestalten.

Die reine Prädiktionsfähigkeit der Modelle kann in den Fehlerhistogrammen sehr gut wiedergegeben werden. Für tiefere dynamische Einblicke ist es jedoch von Vorteil, die Eigenwerte der Koopman-Matrix  $\underline{K}_z$  zu betrachten. Abbildung 5.9 zeigt die Eigenwertkonfiguration von verschiedenen DKLc-Modellen gegenüber den Eigenwerten der Linearisierung des nichtlinearen Systems. Zuerst fällt auf, dass alle vier Modellstrukturen ein konjugiert komplexes Eigenwertpaar bei etwa  $\lambda_{1,2} \approx 0,995 \pm 0,1j$  platzieren. Dies ist dahingehend wichtig, da dieser Eigenwert etwa mit der dominanten Frequenz von 1 rad/s des Van der Pol Oszillators bei den gegebenen Parametern korrespondiert. Des Weiteren liegt dieser in der Nähe des Eigenwertpaares der Linearisierung. Somit erscheinen die Systeme, zunächst unabhängig vom Modellansatz, nicht nur hinsichtlich ihrer Prädiktionseigenschaft sondern auch bezüglich ihrer Dynamik valide. Ein dynamisch ähnliches Modell kann durch den eDMD-Ansatz ermittelt werden. Dieses ähnelt in seiner Eigenwertkonfiguration sehr stark den Modellansätzen mit Zustandsdurchgriff S1 und S2.

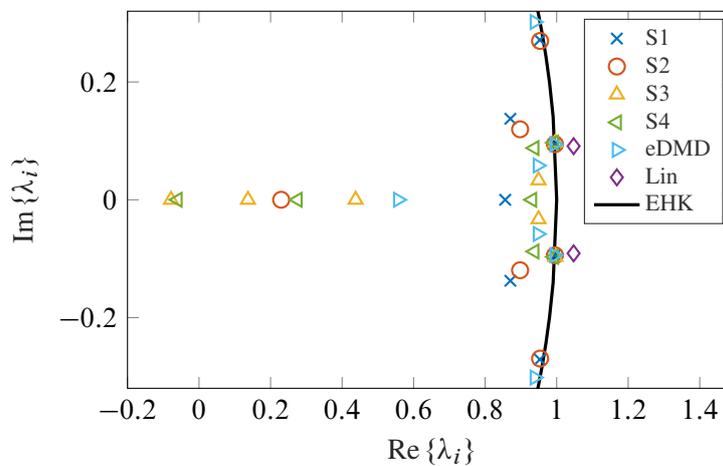


Abbildung 5.9: Zeitdiskrete Eigenwerte der Koopman-Matrix verschiedener Schätzmodelle des Van der Pol Oszillators für  $n_e = 7$

Bei genauerer Betrachtung ergeben sich jedoch signifikante Unterschiede zwischen den Eigenwertkonfigurationen der einzelnen Modellansätze. Zunächst ist es auffällig, dass die beiden Ansätze S1 und S2, welche den Systemzustand im Observable-Vektor berücksichtigen, alle Eigenwerte in die Nähe des Einheitskreises (EHK), beziehungsweise in die Nähe der dominanten Eigenwerte platzieren. Es wird ein weiteres nahezu ungedämpftes konjugiert komplexes Polpaar mit charakteristischer Frequenz von circa  $\omega_3 \approx 2,77$  rad/s geschätzt, welches in der Nähe der dritten Harmonischen der Grundschwingung liegt. Die Ansätze S3 und S4 platzieren hingegen die weiteren Eigenwerte auf die Realachse anstatt in der Nähe des Einheitskreises. Durch die deutlich schlechteren Fitergebnisse scheint es allerdings so, dass zusätzliche Eigenwerte der dritten Harmonischen der Grundschwingung der Modelle S1 und S2 eher schädlich für die Prädiktionseigenschaft sind.

Um diese Vermutung gegebenenfalls zu falsifizieren, werden die einzelnen Eigenwerte der DKLc-Modelle auf Ihre Beobachtbarkeit mittels der Verfahren nach Gilbert [44] und Müller und Lückel [84] untersucht. In Abschnitt 4.1.5 wurde bereits die Steuerbarkeit nach Müller und Lückel erläutert. Die Bewertung der Beobachtbarkeit

$$\xi_{i,\text{obsv}} = \frac{\underline{v}_i^T \underline{C}^T \underline{C} \underline{v}_i}{\underline{v}_i^T \underline{v}_i} e^{1-|\lambda_i|}.$$

unterscheidet sich dabei lediglich durch die Verwendung der transponierten Ausgangsmatrix und der Rechtseigenvektoren anstatt der Eingangsmatrix und der Linkseigenvektoren gegenüber Gleichung (4.5). Für beide Verfahren zeigt sich, dass das besprochene konjugiert komplexe Eigenwertpaar einen ähnlichen Einfluss auf die Ausgänge  $z_1 = x_1$  und  $z_2 = x_2$  wie die übrigen fünf Eigenwerte besitzt. Ein negativer Einfluss des konjugiert komplexen Eigenwertpaars  $\lambda_{3,4} = 0,9529 \pm 0,2711j$  kann somit nicht ausgeschlossen werden, da es einen maßgeblichen Anteil am Ausgang der Modelle besitzt. Dieser potenziell negative Einfluss kann daher lediglich aus der besseren Approximation mittels der anderen Modellansätze S3 und S4 geschlossen werden. Durch das Auftreten weiterer Eigenwerte ist eine isolierte Bewertung dieses Eigenwertpaars jedoch nicht vollständig valide möglich und es verbleibt eine gewisse Unsicherheit bei der Aussage über den negativen Einfluss des besprochenen Eigenwertpaars.

Die Modellansätze S3 und S4 mit kompletter Zustandstransformation scheinen hingegen bei der Platzierung der Eigenwerte völlig anders vorzugehen. Die dominante Eigenfrequenz ist, wie bereits beschrieben, ebenfalls Teil des Spektrums des Koopman-linearen Systems. Wie die Modelle mit Zustandsdurchgriff schätzen diese Ansätze auch ein weiteres gedämpftes konjugiert komplexes Polpaar in der Nähe der dominanten Eigenwerte. Die übrigen Eigenwerte werden in die Nähe von Null auf die Realachse gelegt. Dies ist ein Indiz dafür, dass diese eher unwichtig sind. Tatsächlich erreichen die Ansätze mit einer niedrigeren Modellordnung, welche keine dieser reellen Dynamiken aufweisen, bereits ein gutes Fitergebnis, wie Abbildung 5.5 zeigt. Eine Dominanzuntersuchung der einzelnen Eigenwerte ist aufgrund der oben genannten Ausführungen bei diesen Modellstrukturen nicht möglich. Da das Schätzergebnis der Modellstrukturen S3 und S4 jedoch auch für niedrige Modellordnungen bereits deutlich besser ist, als das der Modelle mit Zustandsdurchgriff, scheint primär die Transformation der Zustände und nur sekundär die Verteilung der Eigenwerte für die Schätzqualität verantwortlich zu sein.

Ein weiteres wichtigen Unterthema ist die Untersuchung der Leistungsfähigkeit des Schätzalgorithmus bei vorhandenem Rauschen. Insbesondere in der Praxis stehen meist keine unverrauschte Daten zur Verfügung. Für diese Untersuchung wird dem Zustandssignal ein mittelwertfreies normalverteiltes Rauschsignal mit Standardabweichungen von  $\sigma = 0,01$ ,  $\sigma = 0,1$  und  $\sigma = 0,3$  aufaddiert. Um einen Eindruck vom Rauscheinfluss zu erhalten, ist in Abbildung 5.10 eine Periode des Van der Pol Systems mit und ohne Rauschen aufgetragen.

Eine Subspace-Schätzung zur Abschätzung der Ordnung ist in Abbildung 5.2b gezeigt. Dabei bleibt die abschätzbare Ordnung im Vergleich zum unverrauschten Fall in etwa gleich. Dadurch und durch den Sachverhalt, dass ein kompletter Trainingslauf über alle Ordnungen mehrere Wochen in Anspruch nehmen kann, wird sich auf die weiter oben verwendete Vergleichsordnung von  $n_e = 7$  beschränkt. Bei einem Rauschen mit  $\sigma = 0,01$  sind die Ergebnisse in den Fehlerhistogrammen praktisch nicht zu unterscheiden und daher nicht dargestellt, was auch im Hinblick auf das sehr hohe Signal-Rauschverhältnis einleuchtend ist (vergleiche Abbildung 5.10). Auch die

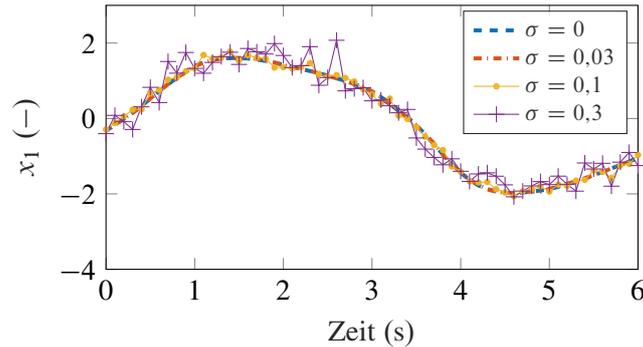


Abbildung 5.10: Beispieltrajektorien mit verschiedenen Rauschamplituden

Schätzung der Eigenwerte liefert sehr ähnliche, dominante Dynamiken. Bei einer Schätzung mit größeren Rauschamplituden von  $\sigma = 0,1$  und  $\sigma = 0,3$  kann mit DKLC kein Modell geschätzt werden, welches die Systemdynamik adäquat wiedergeben kann. Dies führt so weit, dass selbst die dominante Eigenschwingung, welche mit einem Eigenwert bei  $\lambda_{1,2} \approx 0,995 \pm 0,1j$  korrespondiert, nicht erkannt wird. Diese Problematik ist insbesondere im Hinblick auf den späteren Abschnitt 5.2, welcher sich mit Messdaten eines realen Prüfstandes beschäftigt, von enormer Relevanz.

Die Untersuchung der Stabilität des nichtlinearen Systems anhand der Koopman-Identifikation nach Satz 3.11 betrifft ebenso das geregelte System und ist daher Teil des nachfolgenden Abschnitts, welcher sich mit der Regelung des Van der Pol Oszillators unter Verwendung der in Abschnitt 3.4 beschriebenen Methoden befasst. Durch die Eigenwerte mit  $|\lambda| \approx 1$  wäre eine Untersuchung der Stabilität mit den Sätzen 3.2 und 3.3 durch den Ausschluss ungedämpft schwingungsfähiger Systeme ohnehin nicht möglich.

## Regelung

Der Regler soll anhand eines validen Systemmodells ausgelegt werden. Daher werden nur die Regelungen dargestellt, welche auf Modellen mit den besten  $AIC_c$  Score basieren. Ergebnisse mit anderen Ordnungen werden lediglich in Textform kommentiert. Eine Ausnahme bildet dabei das Modell mit Struktur S4 mit Ordnung  $n_e = 9$ . Dieses spezielle Modell ist durch das Training instabil geschätzt worden<sup>23</sup>. Um vergleichbar mit den anderen Strukturen zu bleiben, wird auf das zweitbeste Modell, hinsichtlich des  $AIC_c$ , mit Ordnung  $n_e = 7$  zurückgegriffen. So soll potenziell schlecht vergleichbarem Verhalten durch die Instabilität des Modells vorgebeugt werden.

Der nächste Schritt ist nun die Regelung und Stabilisierung des Systems in seine Ruhelage in  $\underline{x}_R = \underline{0}$  unter Nutzung der in Abschnitt 3.4 beschriebenen Methodik. Einer der großen Vorteile einer LQ-Regelung ist die anschauliche Gewichtung einzelner Zustände gemäß der Gleichungen (3.45) und (3.46). So kann gezielt Einfluss auf bestimmte Zustandsgrößen durch die Regelung genommen werden. Hierzu muss das System in der Regel erst normiert werden, um eine sinnvolle

<sup>23</sup>Ein konjugiert komplexes Eigenwertepaar liegt bei  $|\lambda_{1/2}| = |0,99 \pm 0,1j| \approx 1,001$  und ist damit nur sehr knapp instabil.

und anschauliche Gewichtung zu ermöglichen. Als Beispiel dient ein fiktives Zustandsraummodell mit zwei Zuständen, welches den Verlauf einer Temperatur  $x_1$  im Bereich von  $350 \pm 50$  K und eines Druckes  $x_2$  im Bereich von  $(10 \pm 1) \cdot 10^5$  Pa beschreibt. Wird nun eine Gewichtung von  $\underline{Q} = \text{diag}([10, 1])$  gewählt, so besitzt der Anteil der Temperatur  $x_1$  durch die wesentlich kleinere Größenordnung trotz höherer Gewichtung in  $\underline{Q}$  kaum Einfluss auf das Gütemaß in (3.45) beziehungsweise in (3.46). Dies wird in der Regel dazu führen, dass der Regler nicht die Temperatur sondern den Druck maßgeblich beeinflusst. Dies zeigt anschaulich die Notwendigkeit der Normierung der Zustandsgrößen auf eine einheitliche Größenordnung. Da die Daten, mit denen das Modell trainiert wird, für das Training bereits normiert sind und somit ein normiertes Modell identifiziert wird, entfällt die Notwendigkeit der Normierung an dieser Stelle.

Modelle mit den besten Fit-Ergebnissen, und damit auch der potenziell besseren Systembeschreibung, liefern nicht zwangsläufig die besseren Regelungsergebnisse. Als Vergleich sei hier die bereits ausgeschlossene Systemordnung  $n_e = 9$  für S4 gegeben. Diese besitzt hinsichtlich des Fits und des  $\text{AIC}_c$  zwar die potenziell beste Systembeschreibung, liefert für die Regelung aber wie der Van der Pol Oszillator selbst einen Grenzzyklus, welcher sich jedoch in Frequenz und Amplitude von der des ursprünglichen Systems unterscheidet. Sie schafft es also nicht die Ruhelage im Nullpunkt zu stabilisieren<sup>24</sup>.

Wie eingangs beschrieben, sollen die jeweils besten Modelle, mit der Ausnahme von Modell S4, geregelt werden. Hier müssen nach Heuristik 3.14 zunächst die  $n = 2$  („am besten“) steuerbaren Eigenfunktionen ausgewählt werden. Hierzu wurden in Abschnitt 4.1.5 zwei mögliche Methoden vorgestellt. Da das behandelte System über nur einen Eingang verfügt, ist die Beurteilung mittels Gilbert sehr anschaulich und es muss keine Rücksicht auf die Verteilung der Eingänge auf die einzelnen Eigenwerte oder potenzielle Aufhebungseffekte genommen werden.

In Abbildung 5.11 ist der Grad der Steuerbarkeit nach Gilbert und dem Modalmaß nach Müller und Lückel aus Gleichung (4.5) für die Modellstrukturen S3 und S4 bei der bisher als Referenz herangezogenen Modellordnung  $n_e = 7$  dargestellt. Hier zeigt sich der in Abschnitt 3.3.2 angesprochene Effekt, dass mehr als genau  $n$  Eigenwerte steuerbar sind. Für Modellstruktur S3 ist jedoch eine wesentlich ausgeprägtere Steuerbarkeit nach dem Modalmaß für ein konjugiert komplexes Eigenwertpaar gegenüber den übrigen Eigenwerten erkennbar. Auch beim Maß nach Gilbert ist eine bessere Steuerbarkeit des selben Eigenwertpaares gegeben, allerdings weit weniger ausgeprägt. Für die Modellstruktur S4 ist die Aussage zur Steuerbarkeit einzelner Eigenwerte beim Modalmaß für dasselbe Paar wie bei Struktur S3 mit geringerer Deutlichkeit nach wie vor gegeben. Für das Gilbert-Kriterium wird das nächst dominante Eigenwertpaar als dominant steuerbar erkannt. Eine wichtige Folge daraus ist, dass bei der Wahl des Eigenwertpaares  $\lambda_{4,5}$  bei der Modellstruktur S4 die Regelung wieder in einem Grenzzyklus resultiert. Erst die Wahl beziehungsweise das Einbeziehen des „korrekten“ Paares führt mit einem Koopman-Regler zu einer asymptotisch stabilen Regelung für das Van der Pol System.

<sup>24</sup>Ein Grund dafür kann die Instabilität des identifizierten Systems sein, da sie nicht mit dem Systemverhalten vereinbar ist. Der entstehende Regler agiert somit vermutlich nicht dem Systemverhalten entsprechend.

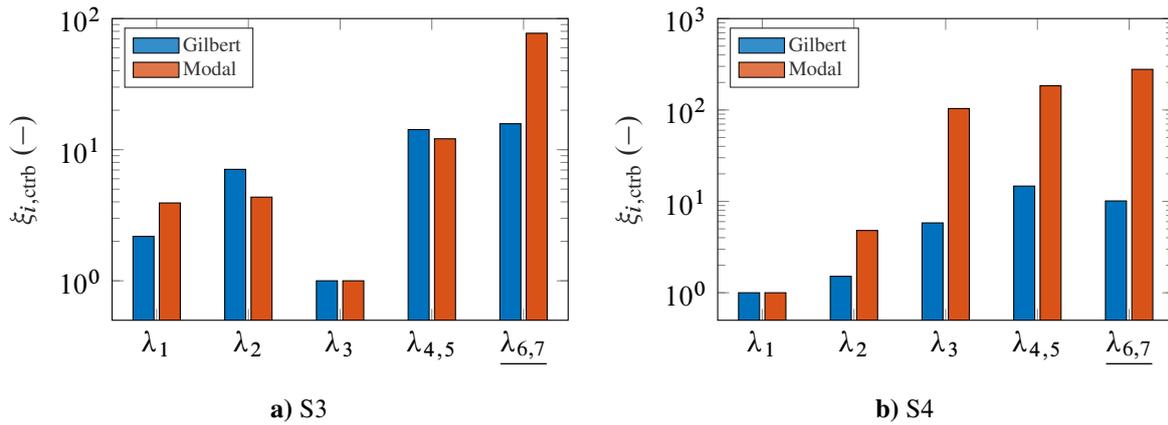


Abbildung 5.11: Steuerbarkeitsmaße  $\xi_{i,\text{ctrl}}$  geschätzter Van der Pol Systeme bei  $n_e = 7$  der einzelnen Eigenwerte. Die Maße sind für jedes Verfahren auf das kleinste Maß über alle Eigenwerte normiert. Die Eigenwerte sind aufsteigend nach Realteil sortiert. Die zur dominanten Eigenfrequenz passenden Eigenwerte sind unterstrichen dargestellt.

Um die Regler vergleichbarer zu machen, werden die Gewichte der LQ-Regelung zusätzlich über das Minimierungsproblem

$$\underset{\underline{Q}_z, \underline{R}_w}{\operatorname{argmin}} J(\underline{Q}_z, \underline{R}_w) = \alpha_0 E_{\underline{u}}(\underline{Q}_z, \underline{R}_w, T) + \frac{1}{N} \sum_{k=1}^N \sum_{i=1}^n \alpha_{1,i} \left\| \Delta x_{i,k}(\underline{Q}_z, \underline{R}_w) \right\|_2^2 \quad (5.1)$$

optimiert. Dabei sind die Größen  $\Delta x_{i,k}$  die Abweichung des  $i$ -ten Zustandes zum Zielzustand zu einem Zeitschritt  $k$ . Die Variable  $N$  gibt die Anzahl an Messzeitpunkten und  $T$  den letzten Zeitschritt an. Dieses Maß wird gewählt, um an der Idee der Einsparung von notwendiger Regelenergie festzuhalten und diese implizit mit zu optimieren. Als Startwert wird für alle ausgewählten Systeme  $\underline{Q}'_{\underline{x}} = 10 \cdot \underline{I}$  sowie  $R_u = 1$  gewählt. Für die Optimierungshyperparameter wird  $\alpha_0 = 1$ ,  $\alpha_{1,1} = 10$  und  $\alpha_{1,2} = 10$  gewählt, um eine zum Startwert ähnliche Aufteilung zu erreichen. Da die Matrizen  $\underline{Q}$  und  $\underline{R}$  positiv definite Matrizen sind<sup>25</sup>, werden bei der Optimierung nicht direkt die Gewichte, sondern ihre Cholesky-Zerlegung  $\underline{Q} = \underline{G}_Q \underline{G}_Q^T$  beziehungsweise  $\underline{R} = \underline{G}_R \underline{G}_R^T$  optimiert, um aus dem Produkt der beiden Matrizen eine positiv (semi-)definite Matrix zu erhalten. Das Regelergebnis der optimierten Koopman-Regler ist in Abbildung 5.12 dargestellt.

Es ist deutlich zu sehen, dass praktisch alle Koopman-Regler einen sehr ähnlichen Weg im Zustandsraum verfolgen, um die Ruhelage in  $\underline{x}_R = \underline{0}$  zu erreichen. Das gleiche Verhalten zeigt sich in den rücktransformierten Gewichtungsmatrizen  $\underline{Q}'_{\underline{x}}$  und  $R'_u$ , die stets eine ähnliche Größenordnung zueinander sowie eine ähnliche Verteilung der Gewichtungen auf die einzelnen Zustände beziehungsweise den Eingang besitzen.

Des Weiteren ist an dieser Stelle zu erwähnen, dass insbesondere bei der Verwendung verschiedener Modellansätze S1 bis S4 oder einer Linearisierung beziehungsweise der einer linearen Iden-

<sup>25</sup>Die Matrix  $\underline{Q}$  kann auch positiv semidefinit sein.

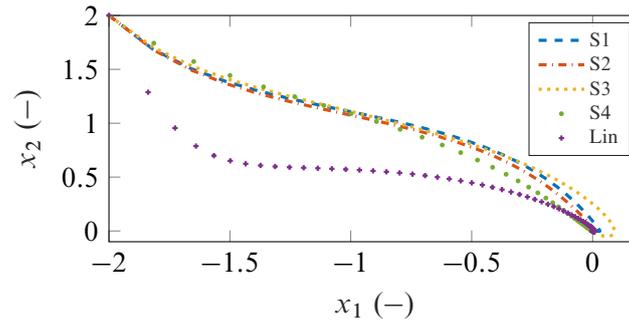


Abbildung 5.12: Ergebnis der Regelung nach Optimierung der LQ-Gewichte

tifikation nicht garantiert ist, dass das Optimierungsmaß (5.1) zu vergleichbaren Ergebnissen im geschlossenen Regelkreis mittels Zustandsrückführung führen muss. Dies ist insbesondere mit dem Blick auf den Regler auf Basis der Linearisierung in Abbildung 5.12 erkennbar. Weder mit der Optimierung nach Gleichung (5.1) noch nach weiteren händischen Einstellversuchen konnte ein Verlauf der Zustände erzeugt werden, welcher mit denen der Koopman-Regler hinsichtlich Energieverbrauch, Einschwingzeit und Einschwingverhalten vergleichbar ist.

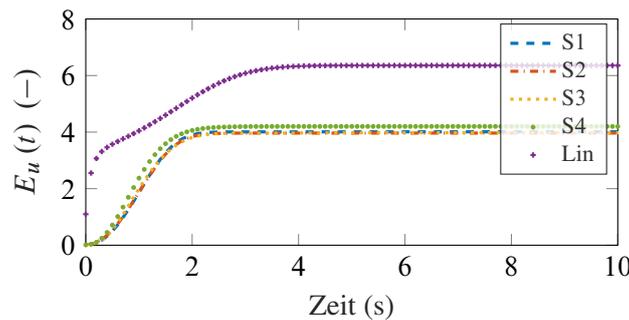


Abbildung 5.13: Stellaufwand der linearen Zustandsregler verschiedener Regelungskonzepte

Betrachtet man die von den verschiedenen Reglern aufgewendete Stellenergie in Abbildung 5.13 im Vergleich zur erreichten Dynamik, dargestellt in Tabelle 5.3 in Form des Eintritts in ein 5%-Band vom Startwert bei  $\underline{x}_0 = [-2, 2]$  aus, kann ein interessantes Verhalten beobachtet werden. Zum einen bringen die Koopman-Regler das System fast doppelt so schnell in die Ruhelage wie der Regler auf Basis der Linearisierung. Zum anderen benötigen sie hier etwa 37% weniger Stellenergie. Dies entspricht dem Idealzustand, den man an realen Anlagen erreichen möchte und zeigt das große Potenzial von Regelungen mittels des Koopman-Operators.

Tabelle 5.3:  $T_{5\%}$  der Zustände  $x_1$  und  $x_2$  für alle Regler

| Model         | S1    | S2    | S3    | S4    | Linear |
|---------------|-------|-------|-------|-------|--------|
| $T_{5\%} x_1$ | 2,0 s | 2,1 s | 1,9 s | 2,4 s | 3,7 s  |
| $T_{5\%} x_2$ | 2,5 s | 2,5 s | 2,6 s | 2,6 s | 4,1 s  |

Die Evaluierung, in wie weit die Transformation der Reglergewichte nach den Gleichungen (3.50) und (3.51) den beabsichtigten Zusammenhang zwischen den Zuständen und den gewählten Ma-

trizen erhält, ist in Abbildung 5.14 aufgetragen. Hierbei ist klar ersichtlich, dass bei Verwendung von Struktur S3 mit einer Modellordnung von  $n_e = 7$  bei den gewählten Gewichten tatsächlich der maßgeblich gewichtete Zustand die Regelung bestimmt. Besonders auffällig ist dies beim Beispiel mit  $\underline{Q}'_x = \text{diag}([1, 100])$ . Der Zustand  $x_2$  wird gemäß der Gewichtung sehr schnell in Richtung null geregelt. Durch die Systemdynamik  $\dot{x}_1 = x_2$  kann das System hinsichtlich des Zustandes  $x_1$  dann jedoch kaum noch reagieren und der Zustand „kriecht“ langsam in Richtung seiner Ruhelage.

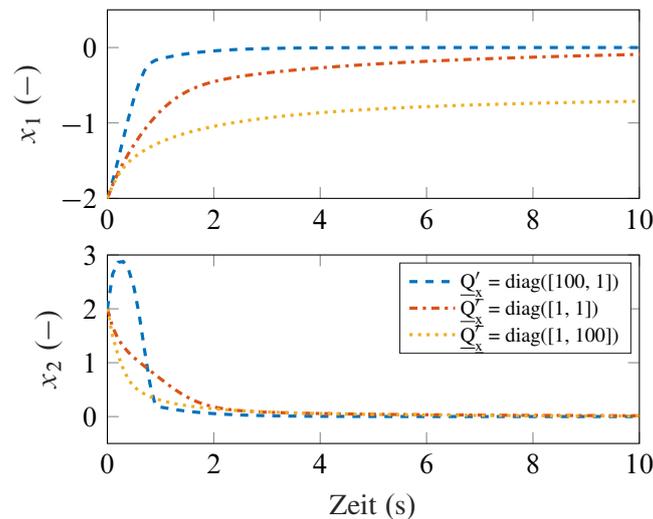


Abbildung 5.14: Zeitplot für verschiedene Zustandsgewichte mit  $\underline{R} = 0,1$  für Modellstruktur S3 und  $n_e = 7$

Der letzte Teil der Analyse, bei vollständig gemessenem Zustandsvektor, soll die Einordnung der Stabilität im geschlossenen Regelkreis mittels Zustandsrückführung sein. Satz 3.11 liefert bei  $\underline{Q}'_x = \underline{I}$ ,  $R'_u = 1$  sowie  $\underline{P}_{\text{lyap}} = \underline{I}$  für die Modellstrukturen S3 und S4 die Stabilitätsparameter  $\alpha_{\max} \approx 0,00085$  sowie  $\gamma_{\max} \approx 0,0033$ . Für S1 und S2 sind diese etwa um den Faktor zehn kleiner. Auch für  $\gamma_{\max}$  als den größeren der beiden Werte, kann kein Bereich garantierter Stabilität angegeben werden. Dies liegt in diesem Beispiel maßgeblich daran, dass durch die eingeschränkte Steuerbarkeit einzelner Eigenwerte einige Eigenwerte der geregelten Koopman-Matrix in der Nähe des Einheitskreises verbleiben. Bei solchen Konfigurationen ergeben sich für den Parameter  $\gamma$  durch die Supremumsnorm sehr kleine Werte. Die starken Einschränkungen in der Herleitung des Parameters  $\alpha$  und die damit einhergehenden potenziell negativen Folgen wurden bereits in Abschnitt 3.1.1 erwähnt und bestätigen sich anhand dieses Beispiels.

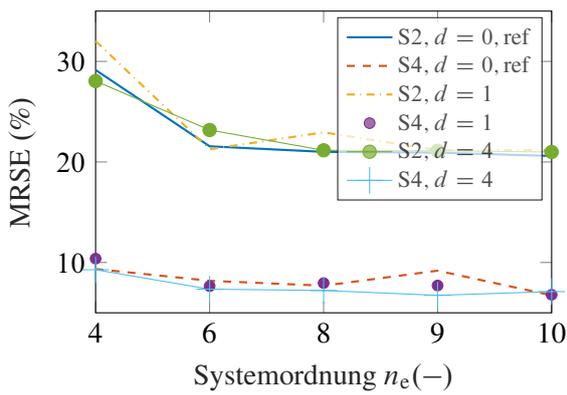
### Time-Delay Koordinaten

Mit Messung des gesamten Zustandsvektors kann nicht nur eine sehr gute Approximation des Van der Pol Oszillators, sondern auch eine hoch performante Regelung erreicht werden. Häufig ist die Situation vor allem in der praktischen Anwendung der Systemidentifikation diese, dass nur ein Teil der Zustände gemessen werden kann. Daher beschäftigt sich dieser Abschnitt damit, wie

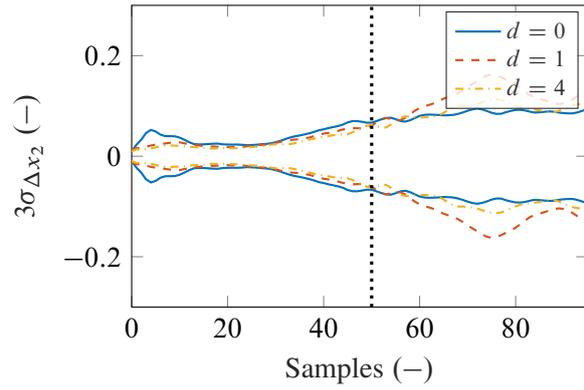
sich sogenannte Time-Delay Koordinaten auf die Schätzung und Regelung mittels des Koopman-Frameworks auswirken. Die verwendete Anzahl an Delay-Schritten  $d = 1$  und  $d = 4$  ergibt sich anhand der beiden folgenden Gründe. Zum einen soll sich anhand der Delay-Schritte wieder die Systemordnung ergeben. Aus der Messung einer Variable muss also ein Delayschritt ( $d = 1$ ) hinzugezogen werden, sodass sich ein Zustandsvektor der Länge  $n = 2$  ergibt. Der Fall  $d = 4$  ergibt sich anhand von Takens' Embedding-Theorem, welches in Abschnitt 3.5.1 besprochen wird. Hierin wird eine Einbettungsdimension von  $2n + 1$  oder größer vorausgesetzt. Mit dem aktuellen Zeitschritt ergibt sich somit eine notwendige Dimension der Einbettung von fünf, was zu  $d = 4$  notwendigen Delay-Schritten führt. Um den Trainingsaufwand so gering wie möglich zu halten, wird sich auf die Modellstrukturen S2 und S4 beschränkt. Die Wahl der Strukturen mit Eingangscodierung folgt aus der etwas geringeren Streuung der Modelle gegenüber den Strukturen S1 und S3, wie sie in Abbildung 5.6 erkennbar ist.

Zunächst soll der Einfluss von Time-Delay-Koordinaten bei der Messung des gesamten Zustandes untersucht werden. Abbildung 5.15 zeigt beispielhaft den mittleren MRSE über die verschiedene Ordnungen bei einem Prädiktionshorizont von  $p = 50$  sowie die Entwicklung der Varianz über die Prädiktion hinweg. Die Entwicklung der Varianz ist für die Modellstruktur S4 bei einer Ordnung von  $n_e = 9$  angegeben. Dies entspricht der Konfiguration, die in der Voruntersuchung nach  $AIC_c$  am besten abgeschnitten haben. Es ist deutlich zu sehen, dass das Einführen von Time-Delay-Koordinaten keinen relevanten beziehungsweise verbessernden Einfluss auf die Schätzung nimmt. Lediglich beim dargestellten Fall für Modellstruktur S4 kann eine leichte Verbesserung der Varianz innerhalb der Kurzzeitprädiktion erreicht werden. Die Langzeitprädiktion wird hingegen sogar schlechter. Derselbe positiven Einfluss ist auch im ersten Zustand etwas weniger ausgeprägt vorhanden. Durch die starke Ähnlichkeit entfällt die explizite Darstellung der Varianz des Zustandes  $x_1$  jedoch. Für die Modellstruktur S2 ist keine Verbesserung durch das Einführen von Time-Delay-Koordinaten zu erreichen, weshalb die Darstellung dieses Falls ausbleibt. Dadurch, dass keine Verbesserung durch das Einführen von Delay-Koordinaten erreicht werden kann, wird die Messung des gesamten Zustandes in Kombination mit Delay-Koordinaten nicht weiter verfolgt. Dies hat insbesondere den Grund, da das Einführen von Delays die Regelung im späteren Schritt verkompliziert, wie in es bereits Abschnitt 3.5.1 beschrieben wurde.

Die wesentlich aussagekräftigere Untersuchung stellt der Fall dar, dass nur einer der Zustände gemessen werden kann. Dies ist insbesondere in praktischer Hinsicht interessant, da hier häufig aus Kostengründen oder physikalischen Gegebenheiten keine Messung aller Zustände möglich ist oder sogar latente Zustände im System vorhanden sind. Zu diesem Zweck ist in Abbildung 5.16a eine Gegenüberstellung der MRSE über die verschiedenen Ordnungen und Delay-Schritte für  $p = 50$  für die Schätzung des Zustandes  $x_1$  gezeigt. Für eine Einordnung ist außerdem der MRSE von  $x_1$  aufgetragen, der sich aus der Messung des gesamten Zustandes, ohne Einbeziehung eines Delays, ergibt. Eine Darstellung der Schätzung anhand der isolierten Messung von  $x_1$  ohne ein Delay bleibt an dieser Stelle aus, da sich der MRSE nahezu unabhängig von der Modellordnung im Bereich von etwa 62 % befindet. Dies deckt sich mit den Erkenntnissen aus Abschnitt 3.5.1 und Takens' Embedding-Theorem. Bei alleiniger Messung von  $x_1$  erhöht sich die Schätzqualität mit der Anzahl an Delays. Ein anderes Bild ergibt sich für den vollständig kodierten Zustand bei der



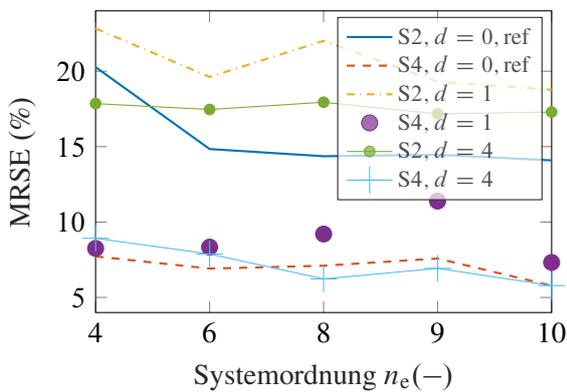
a) MRSE über verschiedenen Systemordnungen



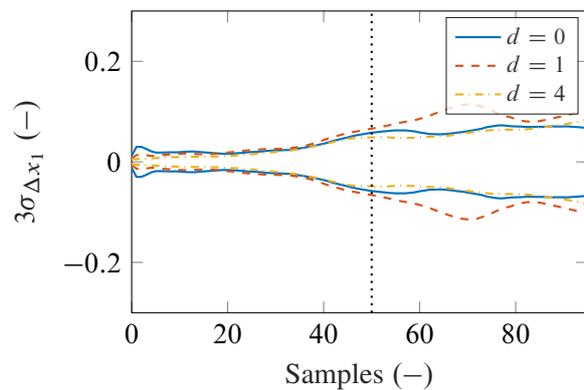
b) Varianz des zweiten Zustandes in Form der  $3\sigma$ -Grenzen für S4

Abbildung 5.15: MRSE über verschiedenen Systemordnungen sowie Varianz der DKLc-Schätzung nach Einführung von Time-Delay-Koordinaten bei vollständig gemessenem Zustand. Die Referenzmodelle geben den Fit bei  $d = 0$  und der Schätzung anhand des vollständig gemessenen Zustandsvektors an.

Struktur S4. Hier tritt ebenso eine Verbesserung der Schätzqualität mit Erhöhung der Anzahl an Delays auf. Allerdings kann für diese Struktur bei der isolierten Messung von  $x_1$  und  $d = 4$  eine ähnliche Schätzqualität erreicht werden wie bei der Schätzung mit dem gesamten Zustandsvektor. Die Unterschiede im von teilweise unter einem Prozent sind dabei eher als Schwankung in den Schätzergebnissen zu verstehen, die durch unterschiedliche Initialgewichte auftreten, und nicht als bessere Schätzung. Hinsichtlich der Varianz der Schätzung ergibt sich in Abbildung 5.16b ein ähnliches Bild, wie bei der Messung der vollständigen Zustandsvektors. Die Varianz im Bereich der Kurzzeitprädiktion nimmt mit der Anzahl an Delay-Schritten ab, wohingegen sie im weiteren Verlauf zunimmt. Für  $d = 4$  ergibt sich eine weniger streuende Schätzung zu Beginn der Prädiktion, sowie eine ähnliche Varianz nach Erreichen des trainierten Prädiktionshorizontes  $p_{DKLc} = 50$  im Vergleich mit der Referenzschätzung.



a) MRSE über verschiedenen Systemordnungen



b) Varianz in Form der  $3\sigma$ -Grenzen für S4,  $n_e = 9$

Abbildung 5.16: Fit und Varianz der DKLc-Schätzung nach Einführung von Time-Delay-Koordinaten bei Messung des Zustandes  $x_1$

Eine Darstellung der Eigenwerte bleibt an dieser Stelle aus, da sich eine zum vollständig gemessenen Zustand kaum unterscheidbare Eigenwertkonfiguration ergibt. Das gleiche gilt für die Bewertung der Steuerbarkeit der einzelnen Eigenwerte, da sich auch hier das gleiche Verhalten wie in Abbildung 5.11 beim vollständig gemessenen Zustand zeigt. Es sei allerdings angemerkt, dass Heuristik 3.14 eine Transformation  $\underline{g}_z(\underline{x})$  mit allen Zuständen als Argument voraussetzt. Da die Steuerbarkeit des Gesamtsystems jedoch nicht von der gewählten Ausgangsgröße abhängt, wird an dieser Stelle und für alle weiteren Beispiele, welche Delay-Koordinaten verwenden, angenommen, dass dies auch für Koopman-lineare Systeme gilt, die auf Basis von Delay-Koordinaten ermittelt werden und somit nicht der gesamte Zustand unmittelbar vorliegt. Für die Wahl der Matrix  $\underline{Q}'_x$  ist weiter wichtig, dass lediglich die Zustände zum Zeitpunkt  $k$ , also in der Regel die ersten  $q$  Einträge, gewichtet werden. Dies ist darin begründet, dass die übrigen Einträge im Delay-Koordinaten-Vektor dieselben Zustände, jedoch zeitversetzt, darstellen. Eine mehrfache Gewichtung des gleichen Zustandes ist daher nicht sinnvoll.

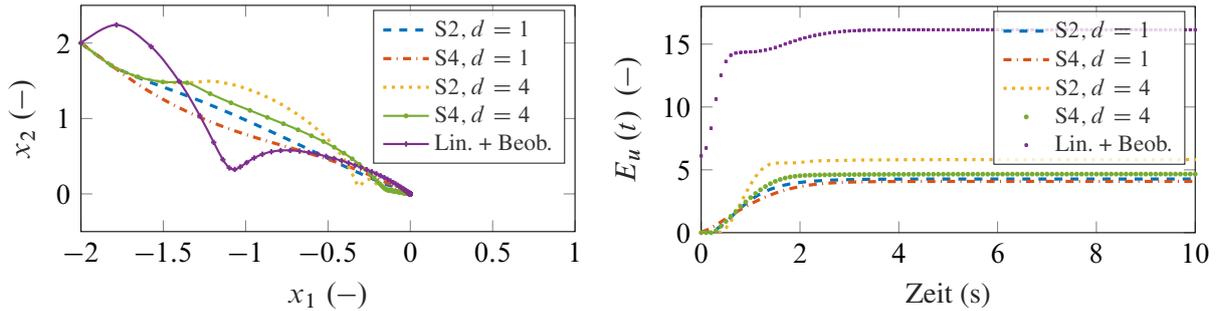
Abbildung 5.17a zeigt den geregelten Van der Pol Oszillator mittels verschiedener Regler auf Basis der Time-Delay-Koordinaten. Die Gewichtung für  $x_1$  wurde so gewählt, dass sich für alle Regler eine Einschwingzeit von etwa  $T_{5\%} \approx 3,2$  s nach Einsetzen des Reglers ergibt. Tabelle 5.4 zeigt die Konfiguration der DKLc-Regler sowie die Einschwingzeit. Die teilweise unterschiedlichen Größenordnungen der Einträge in der  $\underline{Q}'_x$  Matrix sind darauf zurück zu führen, dass es sich bei den zu wählenden  $\underline{Q}'_x$  Matrizen, wie in Abschnitt 3.4.1 erwähnt, nur um strukturell ähnliche Matrizen zu  $\underline{Q}_x$  handelt, diese jedoch nicht exakt denen entsprechen müssen.

Tabelle 5.4: Reglerkonfiguration (bei  $R = 1$ ) und Einschwingzeit der DKLc-Regler

|                 | S2, $d = 1$ | S4, $d = 1$ | S2, $d = 4$ | S4, $d = 4$ | Beobachter   |
|-----------------|-------------|-------------|-------------|-------------|--|
| $q'_{x_1}$      | 10          | 7           | 80          | 203         | 150, ( $\underline{Q}_L = \text{diag}([30, 30])$ ) |
| $T_{5\%}$ $x_1$ | 3,3 s       | 3,3 s       | 3,6 s       | 3,6 s       | 3,2 s  |
| $x_2$           | 3,1 s       | 3,3 s       | 2,5 s       | 2,5 s       | 2,3 s  |

Auch hier erfolgt die Auswahl der Modelle wieder anhand des  $AIC_c$ , wobei an dieser Stelle auf eine Darstellung verzichtet wird. Für die Regelung und die Wahl der  $n$  steuerbaren Eigenwerte muss die Systemordnung des nichtlinearen Systems bekannt sein. Um hier Abhilfe zu schaffen, kann gegebenenfalls anhand der Steuerbarkeitsmaße eine Entscheidung getroffen werden. Für bestimmte Modelle ist dies jedoch nicht immer aussagekräftig, wie Abbildung 5.11b zeigt.

Ein Vergleich einer Zustandsregelung auf Basis der Linearisierung kann nur über einen Beobachter basierten Entwurf erfolgen, da eine ausschließliche Messung von  $x_1$  betrachtet wird. Andernfalls müsste ein unterschiedliches Regelungskonzept verfolgt werden. Damit besitzt der Vergleichsregler eine eigene Dynamik, was durch die Verzögerungsglieder durch das Zusammenführen der verzögerten Zustände als ein Vektor implizit auch für den Koopman-Regler in Time-Delay-Koordinaten gilt. Hier zeigen sich jedoch erneute Vorteile gegenüber häufig verwendeten Verfahren wie einem beobachterbasierten Entwurf auf Basis einer Linearisierung des nichtlinearen Systems, wie ebenfalls in 5.17 dargestellt. Zum einen muss ein solches Systemmodell bekannt



a) Phasenplot des geregelten Van der Pol Oszillators      b) Stellenergie der verschiedenenene Regler

Abbildung 5.17: Ergebnis der Koopman-Regelung mittels Time-Delay-Koordinaten

sein, um diesen Entwurf durchführen zu können. Zum anderen führt bei ähnlicher Einschwingzeit für  $x_1$  ein solcher Regler zu erheblich größeren Stellamplituden, obwohl der Beobachterzustand dem des echten Systems entspricht. Dies ist womöglich auf die Nichtlinearität des Van der Pol Systems beziehungsweise auf die Instabilität seiner Linearisierung zurück zu führen.

Ein weiterer Regler in Form der Übertragungsfunktion

$$G_{R,alt}(s) = 45 \cdot \frac{s^2 + 4s + 3}{s^2 + 15s + 50},$$

welcher heuristisch anhand der Wurzelortskurve ausgelegt wird, erreicht dabei zwar ein dynamisches Verhalten mit ähnlicher Einschwingzeit, provoziert jedoch ein starkes Überschwingen. Die Darstellung der Stellenergie entfällt an dieser Stelle, da dies durch ihren Endwert von etwa  $E_{u,R,alt}(T \rightarrow \infty) \approx 1.251,4$  die Interpretierbarkeit der übrigen Regler stark erschweren würde. Bezüglich der Stellenergie unterscheiden sich die DKLc-Regler mit Time-Delay-Koordinaten kaum voneinander. Weiter sind sie sehr ähnlich zu den Ergebnissen ohne Delay aus Abbildung 5.13.

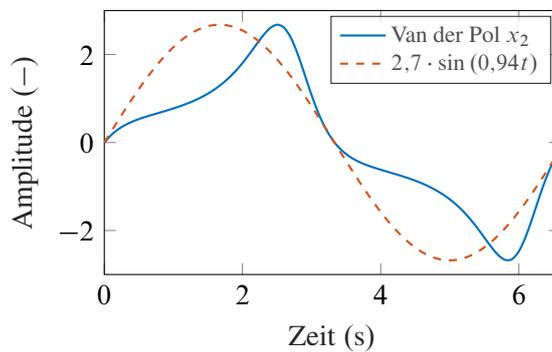


Abbildung 5.18: Darstellung einer Periode des zweiten Zustandes des Van der Pol Systems gegenüber einem skalierten Sinus

Die reine Identifikation und Regelung anhand der Messung von Zustand  $x_2$  ist in Bezug auf das Systemmodell ( $\dot{x}_1 = x_2$ ) für eine Regelung nur wenig sinnvoll, weshalb eine solche Betrachtung

tung an dieser Stelle ausbleibt. Für die Prädiktion stellt sich jedoch ebenso eine Verbesserung der Prädiktionsfähigkeit mit zunehmender Anzahl an Delay-Schritten ein. Die allgemeine Qualität der Prädiktion von  $x_2$  ist allerdings sehr viel weniger exakt wie die Schätzung des Zustandes  $x_1$  und bewegt sich im Bereich um einen Fehler von circa 60 %. Dies ist möglicher Weise durch die Observables selbst ausgelöst. Diese stellen Transformationen vom Zustandsraum in einen anderen, potenziell höherdimensionalen, Raum dar. Kann diese Transformation den Zustand in einen reinen Sinus übertragen, so wäre dies eine geeignete Möglichkeit, diese Observable mit einem linearen Modell im Sinne des Koopman-Operators darzustellen. Durch die sehr spezielle Formgebung der Schwingung des zweiten Zustandes, ist eine solche bijektive Transformation nur mit dem Wissen des Zustandes selbst ohne eine weitere Information, wie beispielsweise dem Gradienten oder einer Zeitinformation, nicht möglich, wie sich leicht an Abbildung 5.18 erkennen lässt. Durch das Einführen der Time-Delay-Koordinaten wäre es theoretisch möglich, dass das Netz sich einen Gradienten aus den Daten erschließt, jedoch ist das Optimierungsgebirge einer komplexen Verlustfunktion, insbesondere im Rahmen von Deep-Learning, in der Regel nicht konvex, wie in Abschnitt 2.3.2 erläutert. Daher ist man an dieser Stelle auf Glück mit den Anfangswerten der Schichtgewichte angewiesen, damit sich eine solche Funktion ergeben kann.

Mit den bei der Untersuchung des Van der Pol Oszillators gewonnenen Erfahrungen soll nun der dynamisch komplexere Duffing Oszillator analysiert und geregelt werden.

### 5.1.2 Duffing Oszillator

Da die grundsätzliche Methodik nun aus dem vorangegangenen Beispiel des Van der Pol Systems bekannt ist, werden in den kommenden beiden Beispielen weniger grundsätzliche Erklärung gegeben und sich auf das eigentliche Beispiel konzentriert.

Beim Duffing Oszillator handelt es sich, wie bereits beim Van der Pol System, um ein nichtlineares (und je nach gewählten Parametern) schwingungsfähiges System. Das hier verwendete zugehörige nichtlineare Zustandsraum-Modell wird um einen Eingang erweitert und lautet

$$\dot{\underline{x}} = \begin{bmatrix} x_2 \\ \kappa_1 x_1 + \kappa_2 x_1^3 + \delta x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \beta \end{bmatrix} u. \quad (5.2)$$

Dieser kann unter Anderem als ein Feder-Masse-Dämpfer System mit nichtlinearer, wegabhängiger Feder nach Abbildung 5.19 aufgefasst werden<sup>26</sup>. Über die Parameter können verschiedenste Modellcharakteristika eingestellt werden. Die hier verwendete Konfiguration mit  $\kappa_1 = 1$ ,  $\kappa_2 = -1$ ,  $\delta = -0,4$  und  $\beta = 1$  ergibt ein System mit drei Ruhelagen bei  $\underline{x}_{R,1/2} = \left[ \pm \sqrt{-\kappa_1/\kappa_2}, 0 \right] = [\pm 1, 0]$  und  $\underline{x}_{R,3} = \underline{0}$  (für  $u = 0$ ) von denen  $\underline{x}_{R,1/2}$  asymptotisch stabil und  $\underline{x}_{R,3}$  instabil ist. Durch

<sup>26</sup>Um einen physikalischen Feder-Masse-Dämpfer System sinnvoll zu entsprechen, muss  $\delta < 0$  sowie die Parameter  $\kappa_1$  und  $\kappa_2$  so gewählt werden, dass für den nichtlinearen Federkoeffizient  $\kappa_1 + \kappa_2 x_1^2 < 0 \forall x_1 \in \mathcal{X}$  gilt. Andernfalls kann es zu unterschiedlichem Verhalten kommen [82]. Dazu gehört auch die in dieser Arbeit untersuchte Konfiguration, mit der sich drei Ruhelagen ergeben.

die drei unterschiedlichen Arbeitspunkte in der gegebenen Konfiguration ist bei der Analyse des Modells und der Regelung die Frage interessant, inwieweit es mit nur einem linearen identifizierten Modell möglich ist, mittels einer Regelung die drei verschiedenen Arbeitspunkte anzufahren und (stabil) zu halten. Die Einzugsgebiete der zwei asymptotisch stabilen Ruhelagen ist für das analysierte Gebiet in Abbildung 5.20 aufgetragen.

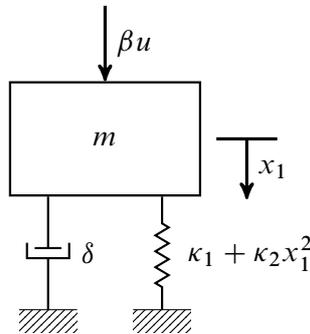


Abbildung 5.19: Feder-Masse-Dämpfer System mit nichtlinearer, wegabhängiger Feder nach Gleichung (5.2) mit  $m = 1$ .

Das System wird mit einem normalverteilten Signal mit  $\mathcal{N}(0, 0,1)$  und einem Anfangszustand vom Rand des Raumes  $\mathcal{X}_0 = \{\underline{x} \mid x_1 = [-2, \dots 2], x_2 = [-1, \dots 1]\}$  angeregt und mit einer Abtastzeit von  $T_s = 0,2$  s für  $T = 30$  s simuliert. Von besonderem Interesse für die Identifikation ist das transiente Verhalten, da sich der Duffing Oszillator in der Nähe der beiden stabilen Ruhelagen ohnehin nahezu linear verhält. Dazu werden die Daten in zwei Gruppen eingeteilt. Die Selektion erfolgt dabei durch den euklidischen Abstand der letzten letzten  $N = 20$  Messpunkte zu den Arbeitspunkten. Die Daten werden rauschfrei generiert.

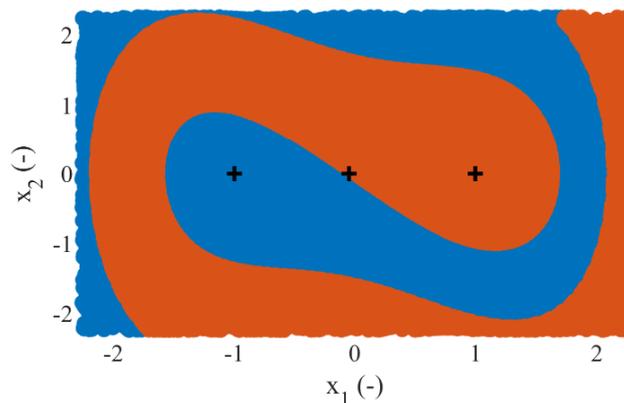


Abbildung 5.20: Visualisierung der Einzugsgebiete beider stabiler Ruhelagen. Die beiden Einzugsgebiete sind symmetrisch. Die leichten Größenunterschiede zwischen den Gebieten ergeben sich durch die Plotmarker. Die drei Ruhelagen sind durch schwarze Kreuze markiert.

### Identifikation

Da es in erster Näherung auch mit der Koopman-Methodik durch ihre Darstellungsweise als lineares Zustandsraumssystem nicht möglich ist, nichtlineare Systeme mit mehreren Arbeitspunkten

ganzheitlich darzustellen, muss auch hier ein Arbeitspunkt festgelegt werden. Betrachtet man die Linearisierung

$$\Delta \dot{\underline{x}} = \begin{bmatrix} 0 & 1 \\ \kappa_1 + 3\kappa_2 x_{1,R}^2 & \delta \end{bmatrix} \Delta \underline{x} + \begin{bmatrix} 0 \\ \beta \end{bmatrix} u$$

von Gleichung (5.2) jedoch genauer, so fällt auf, dass diese für beide stabilen Arbeitspunkte bei  $\underline{x}_{R,1/2} = \left[ \pm \sqrt{-\kappa_1/\kappa_2}, 0 \right]$  bei  $u_R = 0$  identisch ist. Dieser Fakt wird direkt in die Transformation des Zustandes mit einbezogen. Dazu werden die Daten zwar wie üblich normiert, jedoch wird keiner der Arbeitspunkte abgezogen. Es verbleibt also ein Gleichanteil in den Trajektorien zu den einzelnen Attraktoren. Aus einer anderen Sichtweise gesprochen, werden die Daten auf den instabilen Arbeitspunkt bei  $\underline{x}_{R,3} = \underline{0}$  normiert. Die so normierten Daten werden durch die spezielle Beschaffenheit einer gleichen linearisierten Dynamik an den beiden stabilen Arbeitspunkten direkt für eine Schätzung mittels DKLC verwendet.

Eine erste Testidentifikation mit den Daten ohne weitere Vorkehrungen liefert in den meisten Fällen kein valides und in einigen sogar ein instabiles Identifikationsergebnis, was der Systemdynamik widerspricht. Da die Dynamik um beide stabilen Arbeitspunkte jedoch dieselbe ist, ist die Idee, dass man den Encodern mit Hilfe eines weiteren Hilfszustandes mitteilt, dass die Daten für eine bestimmte Ruhelage gelten. Genauer bedeutet dies, dass ein weiterer, künstlicher Schaltzustand  $x_3$  dem Zustandsvektor hinzugefügt wird, der  $-1$  bei der linken Ruhelage in  $\underline{x}_{R,1} = [-1, 0]$  und  $+1$  bei der rechten Ruhelage  $\underline{x}_{R,2} = [1, 0]$  annimmt. Obwohl dieser Gleichanteil bereits in ähnlicher Weise in den gemessenen Daten vorhanden ist, scheint dies nicht auszureichen, ein valides Modell zu schätzen. Womöglich liegt dies daran, dass der Gleichanteil zwar vorhanden, jedoch von „Oberschwingungen“ überlagert ist und der Algorithmus beziehungsweise der Encoder keine designierten Gleichanteil herausarbeiten kann.

Durch den speziellen Aufbau des Modells ist eine erste Abschätzung der Modellordnung durch die Subspace-Methodik nicht unmittelbar mit den Daten möglich. Daher wird eine Ordnungsschätzung mit einzelnen Trajektorien für jeden der beiden stabilen Arbeitspunkte durchgeführt und bewertet. Das Ergebnis der Subspace-Identifikation ist in Abbildung 5.21 aufgetragen.

Die Aussagen beider betrachteter Methoden, PBSID<sub>opt</sub> und OKID, sind schlecht miteinander vergleichbar. Es ist jedoch deutlich zu sehen, dass die Ordnung für beide Verfahren und für beide Ruhelagen vergleichbar geschätzt wird. Je nach betrachteter Methode kommen hier Ordnungen von zwei bis fünf in Frage. Höhere Ordnungen sind allerdings ebenso denkbar, da sich besonders beim OKID-Algorithmus, kein klares „Plateau“ ausbildet. Daher wird erneut auf denselben Bereich an Ordnungen wie bereits beim Van der Pol Oszillator zurückgegriffen. Mit diesen Ordnungen wird wieder eine AIC<sub>c</sub> Analyse durchgeführt, welche in Abbildung 5.22 dargestellt ist. Die zum Training der Netze verwendeten Parameter sind in Abschnitt A.3.4 im Anhang aufgeführt.

Abbildung 5.23 zeigt die Fehlerhistogramme zu den nach AIC<sub>c</sub> besten Modellen. Es zeigt sich eine klare Tendenz über alle Modellansätze hinweg. Über den trainierten Prädiktionshorizont ist der Fehler vergleichsweise gering. Über den trainierten Horizont von  $p_{DKLC} = 50$  kommt es bei

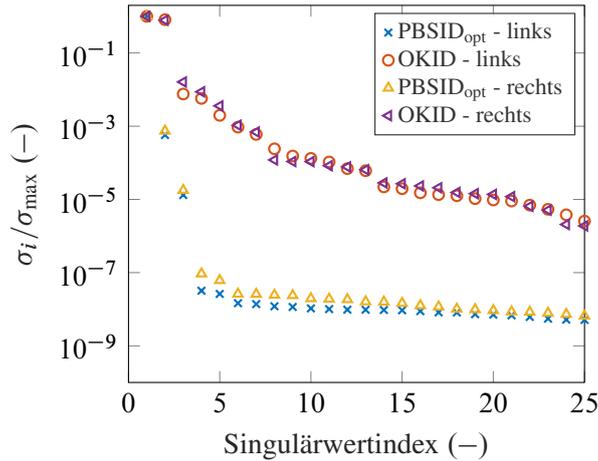


Abbildung 5.21: Ordnungsschätzung des Duffing Oszillators mittels Subspace-Methoden um beide Arbeitspunkte mit  $p_{\text{PBSID}} = 10$ ,  $f_{\text{PBSID}} = 40$  und  $p_{\text{OKID}} = 50$

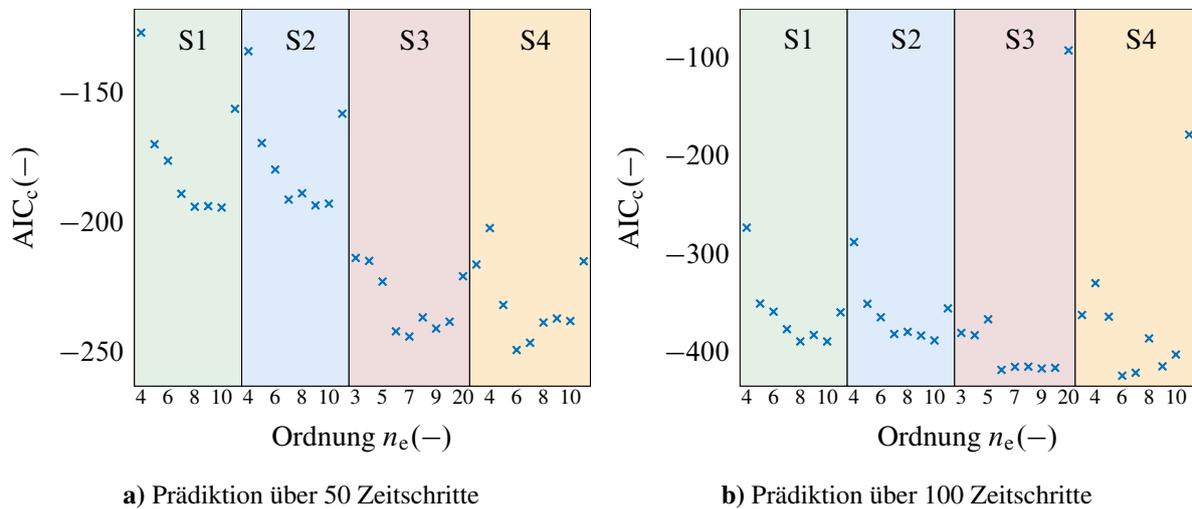


Abbildung 5.22:  $AIC_c$  bei verschiedenen Systemordnungen für die DKLC-Schätzung des Duffing Systems. Durch die eng beieinander liegenden Datenpunkte ist nur jede zweite Ordnung auf der Abszisse aufgetragen.

einigen Trajektorien zu großen Fehlern, was zu einer erhöhten Varianz führt. Die Masse der Testtrajektorien verläuft jedoch mit einem kleinen Fehler. Besonders auffällig ist allerdings der große Fehler in den ersten zwei Prädiktionsschritten bei den Modellansätzen S1 und S2 mit Zustandsdurchgriff. Dieser Fehler schwingt jedoch sehr schnell ein. Die Streuung des Fehlers nimmt nach dem Ende des Trainingshorizontes von  $p_{\text{DKLC}} = 50$  bei diesem System bei fast allen Modellen ab. Dies ist auf die Stabilität der geschätzten Modelle zurück zu führen, welche sich ebenso im nicht-linearen Modell selbst zeigt. Die Stabilität lässt sich außerdem in den Eigenwerten der geschätzten Systeme in Abbildung 5.24 feststellen.

Durch die Ähnlichkeit der Modelle mit und ohne Eingangscoder bleibt eine weitergehende Analyse an der Stelle aus. Eine genauere Betrachtung der Dynamik in Form der Eigenwerte ist jedoch

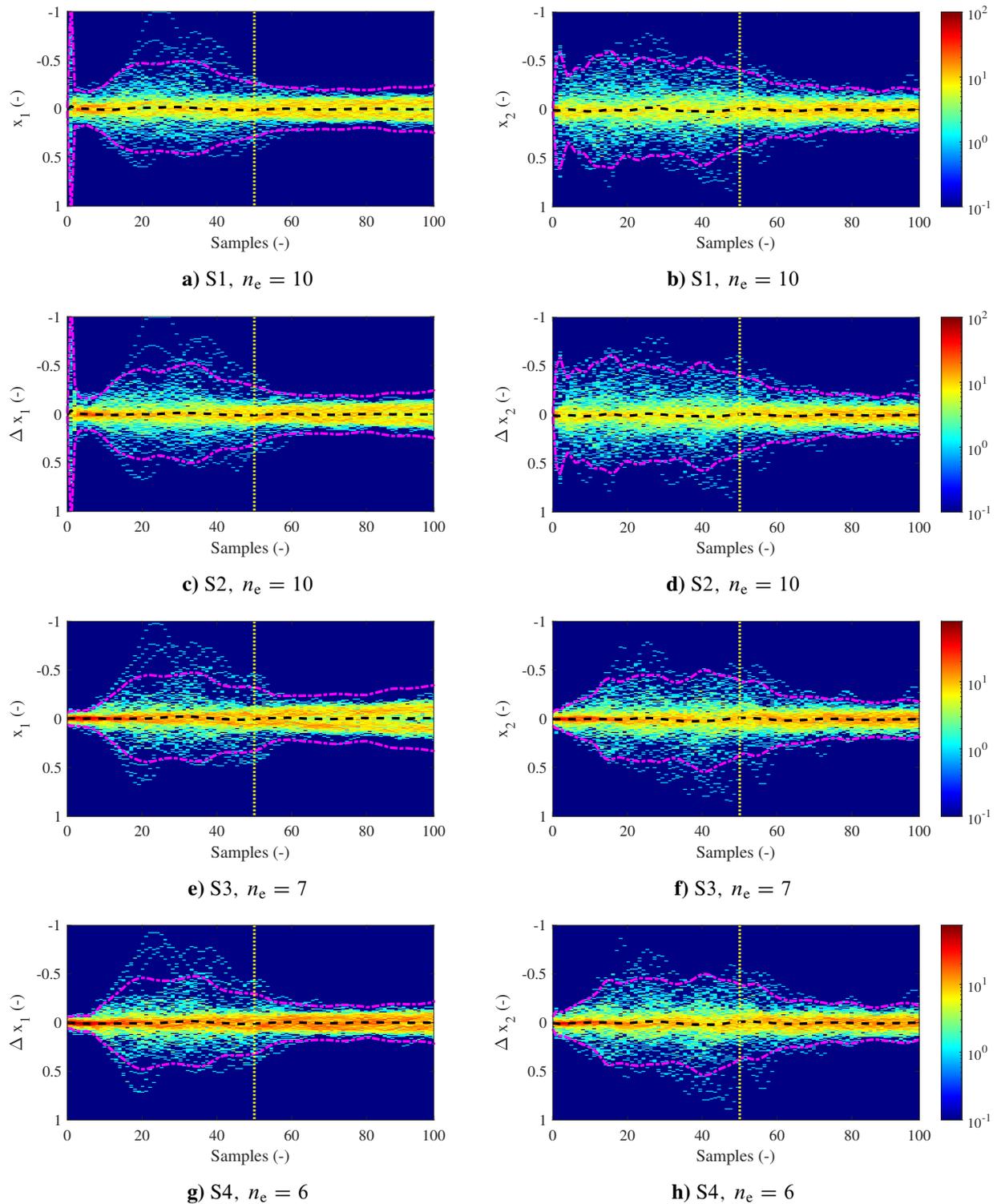


Abbildung 5.23: Fehler Histogramme der nach  $AIC_c$  mit  $p = 100$  besten Modelle des Duffing Oszillators für verschiedene Modellkonfigurationen

für die Plausibilisierung und die spätere Regelung der geschätzten Modelle wichtig. Auf die Stabilität der Modelle wurde bereits bei der Analyse der Fehlerhistogramme hingewiesen. Als Referenz für die Eigenwerte aller identifizierten Modelle in Abbildung 5.24 sind wieder die Eigenwer-

te des linearisierten Modells aufgetragen. Diese korrespondieren außerdem mit der dominanten Frequenz, die das System beim Einlaufen in einen der beiden stabilen Arbeitspunkte einnimmt, welche bei den gegebenen Parametern bei etwa  $f_{\text{dom}} \approx 0,2 \text{ Hz}$  liegt.

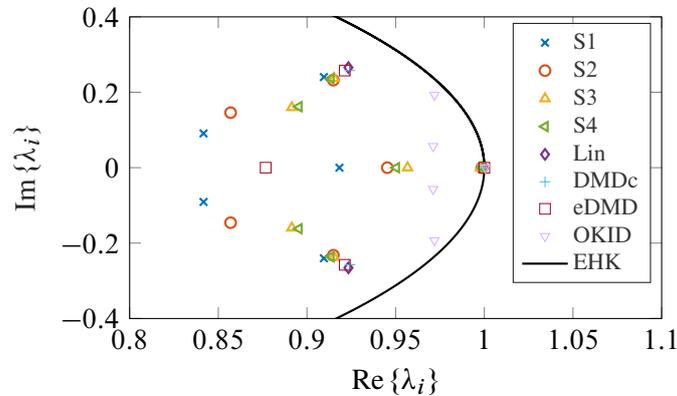


Abbildung 5.24: Zeitdiskrete Eigenwerte der identifizierten Modelle des Duffing Oszillators. Für DKLc werden jeweils die nach  $AIC_c$  besten DKLc-Modelle, OKID ( $n_e = 5$ ), DMDc ( $n_e = n = 3$ ) und randomized-eDMDc ( $n_e = 4$ ) verwendet.

Die Eigenwerte der konkurrierenden DMD basierten Ansätze besitzen teilweise eine sehr ähnliche, scheinbar passende, Dynamik, liefern jedoch ein durchwachsenes Bild hinsichtlich der Fehlerhistogramme, wie Abbildung A.4 im Anhang zeigt. Eine OKID-Schätzung liefert im eingeschwungenen Zustand ein vielversprechendes Prädiktionsverhalten, jedoch eine eher abwegige Eigendynamik, wie im Plot der Eigenwerte zu sehen ist. Eine Folge ist, dass das transiente Verhalten nur sehr präzisiert werden kann. Der ARX-Modell basierte Ansatz sowie die PBSID<sub>opt</sub>-Modelle liefern dabei keine Modelle, welche das Systemverhalten näherungsweise darstellen können. Ein Plot bleibt daher an dieser Stelle aus. Des Weiteren ist es bei den Subspace-Verfahren notwendig, nur einzelne Trajektorien für die Identifikation auszuwählen. Somit ist eine generalisierte Darstellung über diese Trajektorien, und somit den gesamten Zustandsraum, ohnehin schwierig, was den teilweise starken Unterschied zwischen abgeschätzter (Abbildung 5.21) und tatsächlich ausgewählter Ordnung (Abbildung 5.22) erklären kann.

Durch den schwachen dynamischen und prädiktionsseitigen Rückhalt der zu DKLc konkurrierenden Modelle werden diese nicht zur Regelung verwendet. Dies zeigt jedoch, dass der DKLc-Identifikationsansatz bei diesem Modell große Vorteile gegenüber seinen linearen, beziehungsweise randomisierten Konkurrenten besitzt. Zudem stellt sich durch eine Erhöhung der Systemordnung bei diesen Ansätzen keine Verbesserung ein. Es werden lediglich Oberwellen geschätzt, die teilweise einen negativen Effekt auf die Prädiktionsfähigkeit der Modelle haben.

Besonders interessant ist jedoch die konsistente Abweichung der DKLc-Modelle hinsichtlich der dominanten Eigenwerte in Abbildung 5.24. Diese Modelle schätzen ein etwas schnelleres und stärker gedämpftes System, als es die Linearisierung vorgibt, während die DMD basierten Ansätze einen sehr viel geringeren Fehler gegenüber der Linearisierung hinsichtlich der geschätzten Eigenwerte aufweisen. Betrachtet man die Fehlerhistogramme, so scheint diese Abweichung zur

Linearisierung Vorteile in der Prädiktionsfähigkeit, vor allem im transienten Verhalten, mit sich zu bringen.

Wie schon beim Van der Pol Oszillator gibt es neben einigen Ähnlichkeiten, wie dem Eigenwert in der Nähe der dominanten Schwingung, auch gravierende Unterschiede. Beide Ansätze, mit und ohne Zustandsdurchgriff schätzen eine weitere Oberschwingung im dargestellten dominanten Bereich. Die übrigen, nicht dargestellten Eigenwerte liegen alle um null herum und sind damit nicht dominant, aber dennoch mit dem Blick auf Abbildung 5.22 scheinbar wichtig für die Prädiktion. Die durch die Ansätze S3 und S4 geschätzten Eigenwerte entsprechen dabei langsameren, schwächer gedämpften Oberschwingungen im Vergleich zu S1 und S2. Insbesondere für den zweiten Zustand scheint die Wahl dieser Dynamik von großem Vorteil zu sein, wenn der oben besprochene Verlauf der Trajektorienfehler einbezogen wird. Gemein haben die Modelle wieder einen Eigenwert in nächster Nähe zum Punkt 1. Dieser ist, wie in Abschnitt 3.3.2 besprochen, auf die gewählten Daten zurückzuführen, da der Eingang ein mittelwertfreies Signal darstellt und die Zustandsdaten über die Trajektorien einen Mittelwert beinhalten. Für den gewählten Modellansatz ist dieser jedoch auch nötig, um den verbleibenden Mittelwert darstellen zu können.

Dieser Eigenwert ist, wie Abbildung 5.25 zeigt, auch wie zu erwarten, nicht, oder nur sehr schlecht, steuerbar. In den Modellen ist dies zudem kein exakter Integrator, sondern eine sehr langsame, aber stabile Dynamik. Dies ist für die Regelung von Vorteil, da der Regler im Koopman-Raum keine instabile Dynamik stabilisieren muss, die im realen System gar nicht vorhanden ist.

## Regelung

Durch die drei unterschiedlichen Ruhelagen ist der Duffing Oszillator in der vorgestellten Konfiguration ein interessantes System für die Erprobung einer Regelung auf Basis der vorgestellten Methode. Bevor sie jedoch entworfen werden kann, muss eine Steuerbarkeitsanalyse durchgeführt werden. Zu diesem Zweck werden für die Systemansätze S3 und S4 die Steuerbarkeitsmaße in Abbildung 5.25 für alle Eigenwerte aufgetragen, wobei die Modelle mit den besten  $AIC_c$ -Maßen ausgewählt werden. Wie erwartet, stellen sich die Eigenwerte, welche mit der dominanten Frequenz beziehungsweise den Eigenwerten des linearisierten Systems korrespondieren, als deutlich besser steuerbar heraus, als die übrigen. Anders als beim Van der Pol Oszillator liefert auch das Gilbert-Kriterium das gleiche Ergebnis in ähnlicher Deutlichkeit wie der modale Ansatz.

Mit den gewonnenen Ergebnissen bezüglich der Steuerbarkeit werden im Folgenden die Regler berechnet und miteinander verglichen. Wie bereits beim Van der Pol System führt eine erhöhte Gewichtung des zweiten Zustandes  $x_2$  zu einem Kriechen beziehungsweise Nichterreichen des beabsichtigten Arbeitspunktes für den ersten Zustand  $x_1$ , was bezüglich des Systemmodells (5.2) nachvollziehbar ist. Durch die unterschiedlich einstellbaren Arbeitspunkte für  $x_1$  ist es daher nicht sinnvoll,  $x_2$  höher zu gewichten als  $x_1$ , was im Folgenden als Bedingung für die Wahl der Gewichtungen verwendet wird.

Wie eingangs erwähnt, ist das Anfahren aller drei Arbeitspunkte mit nur einem Modell ein Ziel der im Folgenden vorgestellten Regelung. Insbesondere der instabile Arbeitspunkt ist hierbei be-

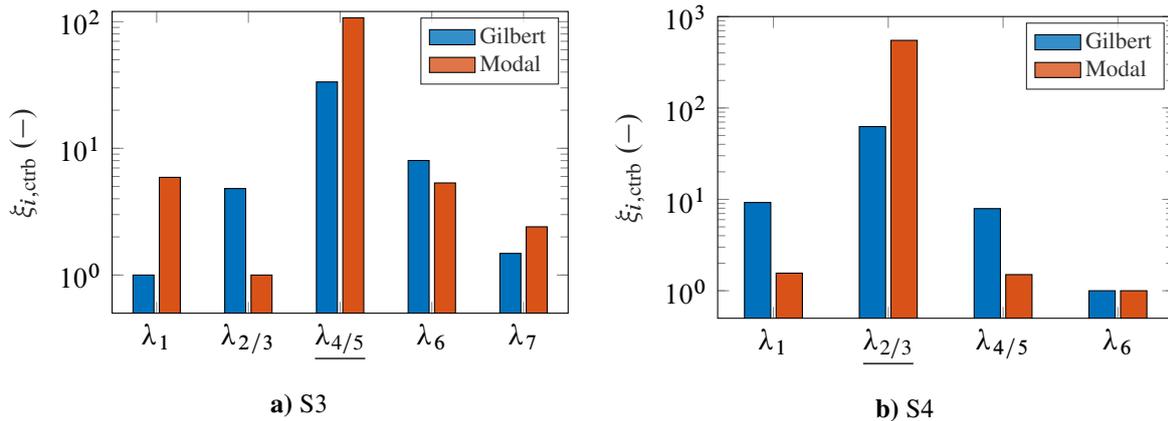


Abbildung 5.25: Steuerbarkeitsmaße  $\xi_{i,ctrl}$  geschätzter Duffing Systeme der einzelnen Eigenwerte. Die Maße sind für jedes Verfahren auf das kleinste Maß über alle Eigenwerte normiert. Die Eigenwerte sind aufsteigend nach Realteil sortiert. Die zur dominanten Eigenfrequenz passenden Eigenwerte ist unterstrichen dargestellt.

sonders interessant, da dessen Dynamik nicht direkt mitgelernt wird. Abbildung 5.26 zeigt das Anfahren der beiden stabilen Arbeitspunkte  $\underline{x}_{R,1/2} = [\pm 1, 0]$ . Dabei können beide Ruhelagen ohne eine Oszillation um sie herum angefahren und gehalten werden. Tabelle 5.5 zeigt die dafür gewählten Gewichtungsmatrizen. Diese werden so gewählt, dass sich eine ähnliche Einschwingzeit  $T_{5\%} \approx 3,2$  s für  $x_1$  ergibt. Da die Optimierung nach dem Optimierungsmaß (5.1) zu keinem sinnvollen Ergebnis konvergiert, werden die Matrizen für dieses Beispiel händisch eingestellt. Insbesondere fällt hier im Vergleich zum Van der Pol System auf, dass die Gewichte teilweise sehr unterschiedlich zueinander gewählt werden müssen, um dasselbe Ergebnis zu erhalten. Dies könnte auf die notwendige Linearisierung bei der Transformation zurück zu führen sein, da diese insbesondere im transienten Verhalten weniger mit der Dynamik am eigentlichen Arbeitspunkt übereinstimmt. Der Hauptteil der Daten kommt aus dem Bereich um die Ruhelagen herum, sodass auch das Modell selbst zum überwiegenden Teil in diesem Bereich trainiert wird. Der Zusammenhang zwischen der Aufteilung der Gewichte auf die einzelnen Zustände bleibt jedoch erhalten. Das heißt, dass beispielsweise eine Vergrößerung des Gewichtes zu  $x_2$  in  $\underline{Q}'_x$  zu einem schnelleren Einschwingen dieses Zustandes führt.

Bezüglich der Stellenergie kann für den Großteil der DKLC-Regler kein Vorteil gegenüber dem Zustandsregler auf Basis der Linearisierung ermittelt werden, wie Abbildung 5.27 zeigt. Lediglich für die Modellstruktur S3 wird für eine Anfahrt der Ruhelage  $\underline{x}_{R,2} = [1, 0]$  deutlich weniger Energie benötigt. Dies ist im Hinblick auf Abbildung 5.26b erklärbar, da ein Regler auf Basis dieses Modells das Duffing-System mit deutlich geringerem Überschwingen und damit weniger Stelleingriff in die Ruhelage befördert.

Zusammengefasst kann für das Anfahren der stabilen Arbeitspunkte kein konkreter Vorteil gegenüber dem Zustandsregler auf Basis der Linearisierung herausgearbeitet werden. Ein Nutzen der Nichtlinearität für einen energetisch besseren Regler ist für den betrachteten Anwendungsfall beim Duffing Oszillator nicht erkennbar. Dies mag jedoch auch daran liegen, dass sich der Duffing

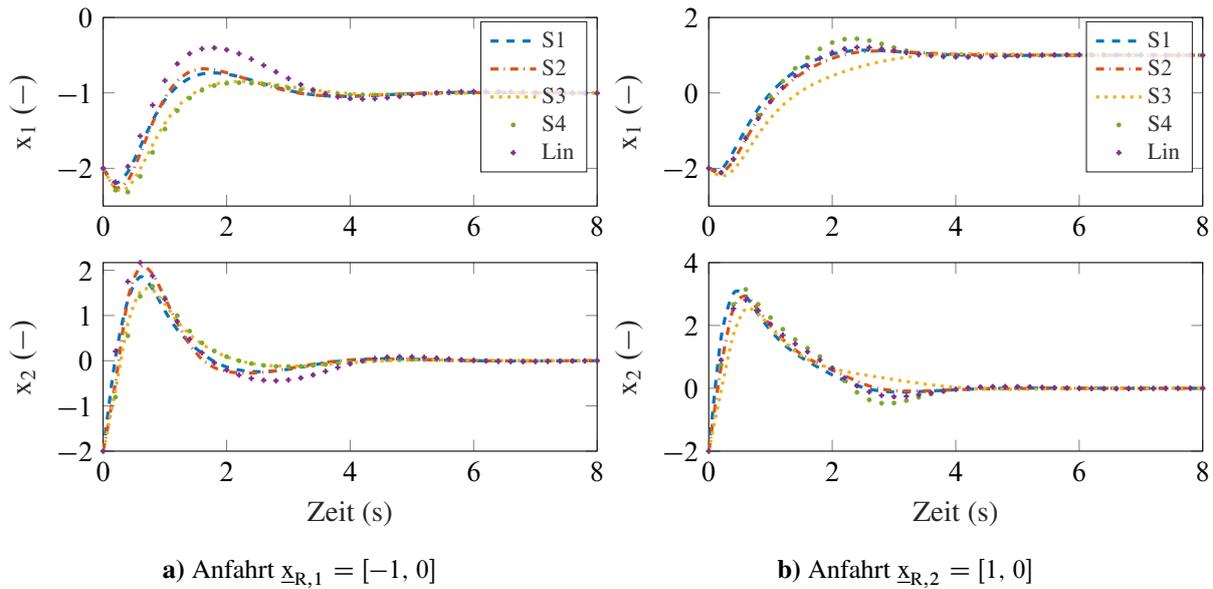


Abbildung 5.26: Anfahrt beider stabiler Arbeitspunkte des Duffing Systems mittels DKLc-Regler, sowie einem Zustandsregler auf Basis der Linearisierung von Gleichung (5.2)

Tabelle 5.5: Gewichte der LQ-Regelung für die ausgewählten Modelle vor der Konvertierung nach Gleichung (3.52). Wenn Gewichte nicht angegeben sind, sind diese null

| Arbeitspunkte   | Gewicht                          | S1            | S2            | S3             | S4             | Linear                        |
|-----------------|----------------------------------|---------------|---------------|----------------|----------------|-------------------------------|
| [1, 0]          | $\underline{Q}'_{\underline{x}}$ | $q_{11} = 11$ | $q_{11} = 15$ | $q_{11} = 14$  | $q_{11} = 600$ | $q_{11} = 11$<br>$q_{22} = 1$ |
|                 | $\underline{R}'_{\underline{u}}$ | 1             | 1             | 2              | 1              | 4                             |
| [-1, 0]         | $\underline{Q}'_{\underline{x}}$ | $q_{11} = 20$ | $q_{11} = 20$ | $q_{11} = 50$  | $q_{11} = 15$  | $q_{11} = 10$                 |
|                 | $\underline{R}'_{\underline{u}}$ | 1             | 1             | 1              | 1              | 2                             |
| $\underline{0}$ | $\underline{Q}'_{\underline{x}}$ | $q_{11} = 5$  | $q_{11} = 10$ | $q_{11} = 100$ | $q_{11} = 10$  | $q_{11} = 15$                 |
|                 | $\underline{R}'_{\underline{u}}$ | 1             | 1             | 2              | 1              | 2                             |

Oszillator nahe der stabilen Ruhelagen wie ein gedämpftes lineares schwingungsfähiges System verhält. Für andere Anfangszustände oder eine andere Wahl der Parameter in Gleichung (5.2) ist eine performantere Regelung mittels des Koopman-Operators jedoch nicht ausgeschlossen.

Die interessantere Betrachtung ist jedoch das Stabilisieren der instabilen Ruhelage bei  $\underline{x} = \underline{0}$ . Abbildung 5.28a zeigt zu diesem Zweck den zeitlichen Verlauf der Zustände. Wie schon bei den beiden stabilen Arbeitspunkten ist es auch hier möglich, das System mit den DKLc-Reglern mit einer ähnlichen Dynamik wie mit einem linearen Zustandsregler in die instabile Ruhelage zu führen und das System dort stabil zu halten.

Hier zeigt sich wieder das Potenzial eines DKLc-Reglers gegenüber einem regulären Zustandsregler. Bei ähnlicher, teilweise leicht schnellerer Dynamik der geschlossenen Regelkreise mittels

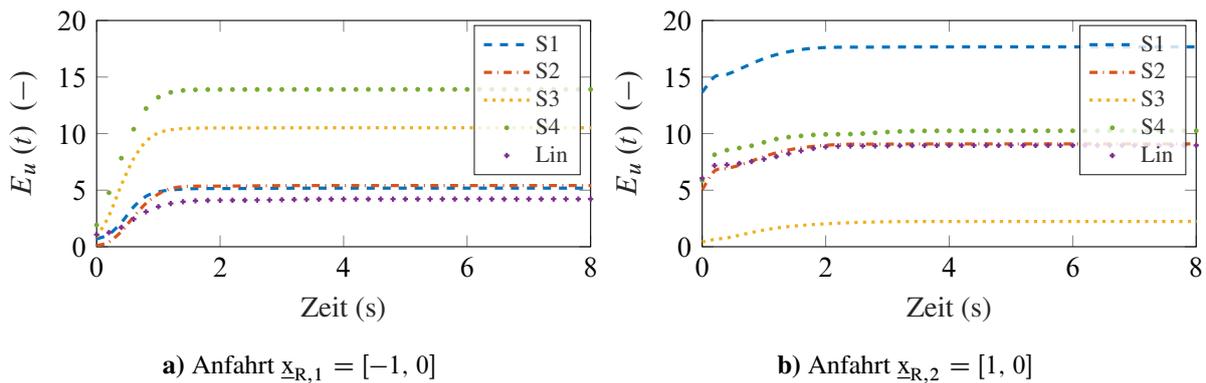


Abbildung 5.27: Stellenergie für die Anfahrt beider stabiler Arbeitspunkte des Duffing Systems mittels DKLc-Regler, sowie einem Zustandsregler auf Basis der Linearisierung von Gleichung (5.2)

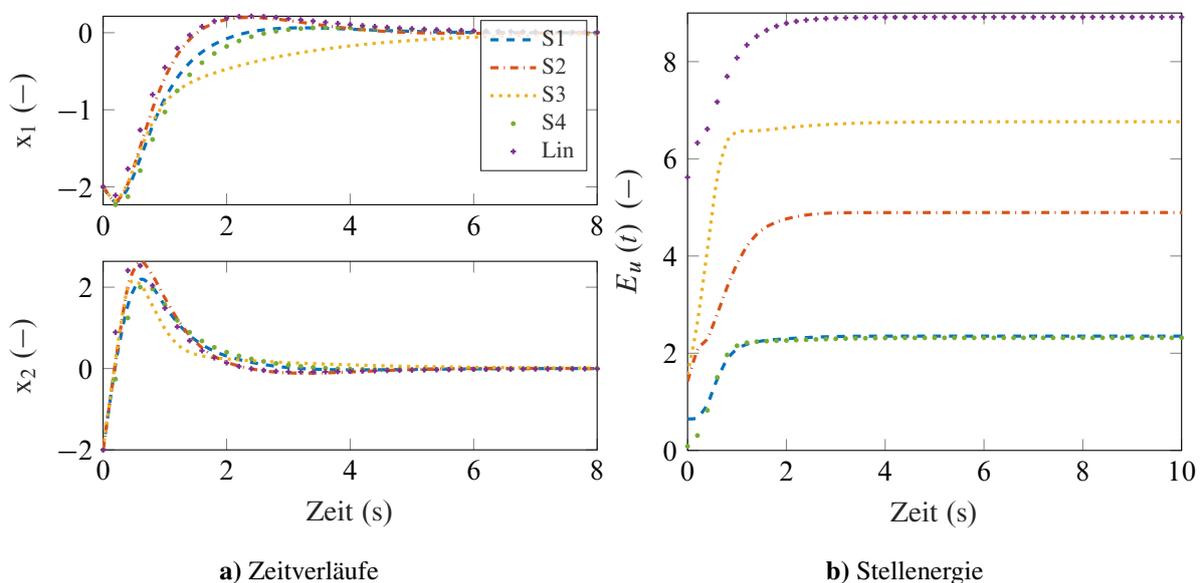


Abbildung 5.28: Regelergebnis für die Anfahrt des instabilen Arbeitspunktes  $\underline{x}_{R,3} = 0$  des Duffing Systems mittels DKLc-Regler, sowie einem Zustandsregler auf Basis der Linearisierung von Gleichung (5.2)

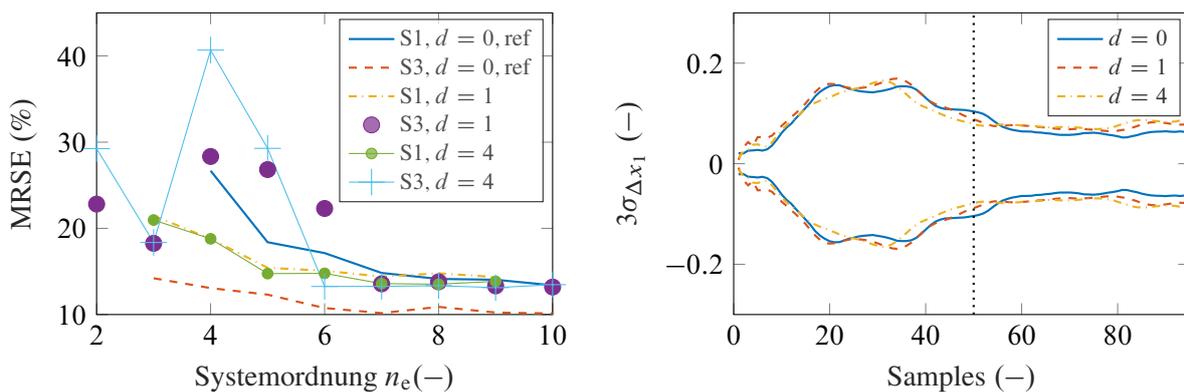
Zustandsrückführung über die verschiedenen Regler erreichen die DKLc-Regler ein sehr viel besseres Resultat hinsichtlich des Energieverbrauchs. Hierbei sei insbesondere der Regler auf Basis des Modells mit der Struktur S2 hervorgehoben, da dieser, verglichen mit dem linearen Zustandsregler, bei nahezu identischem Verhalten etwa 50 % weniger Energie benötigt. Der größere Stellaufwand beim Regler auf Basis der Linearisierung ist maßgeblich auf die vergleichsweise große Stellgröße beim Zeitschritt  $k = 0$  zurückzuführen. Abgesehen von der Energie würde ein DKLc-Regler somit gegebenenfalls das Stellglied schonen, beziehungsweise ein offensichtlich unnötiges Einfahren in eine Sättigung vermeiden.

Durch die, verglichen mit dem Van der Pol System, weniger exakte Prädiktionsfähigkeit kann auch für den Duffing Oszillator mit den LQ-Matrizen aus Tabelle 5.5 sowie  $\underline{P}_{\text{lyap}} = \underline{I}$  kein Bereich garantierter Stabilität für keine der Modellstrukturen im geregelten Fall angegeben werden. Des Weiteren sind die Stabilitätsparameter  $\alpha$  und  $\gamma$  durch verschiedene Eigenwerte der geregelten Koopman-Matrix sehr nahe des Einheitskreises im Bereich von  $10^{-4}$ , wobei stets  $\alpha < \gamma$  gilt.

### Time-Delay Koordinaten

Auch bei diesem, dynamisch komplexeren Beispiel kann durch Messung des gesamten Zustandsvektors ein valides Modell zur Schätzung und zum Aufbau von Zustandsreglern aufgebaut werden. Des Weiteren soll an dieser Stelle auf eine Schätzung und Regelung eingegangen werden, die auf Basis einer zeitlichen Verschiebung der Zustände ermittelt werden. Durch dieselbe Ordnung ergibt sich wie schon beim Van der Pol System die Anzahl an Delay-Schritten zu  $d = 1$  sowie  $d = 4$ .

Wie bereits beim Van der Pol Oszillator, ist durch das Einführen von Delay-Koordinaten bei der Messung des gesamten Zustandsvektors keine Verbesserung der Schätzqualität zu erreichen und wird daher für dieses Beispiel nicht weiter betrachtet. Aufgrund derselben Argumentation wie beim Van der Pol System wird auch beim Duffing System eine reine Messung und Regelung von  $x_2$ , hinsichtlich des Systemmodells aus Gleichung (5.2), in der Regel nicht sinnvoll sein. Des Weiteren könnten durch die unterschiedlichen Arbeitspunkte in  $x_1$  bei  $u_{R,1/2/3} = 0$  kaum vorhergesehen werden, welcher der drei verschiedene Arbeitspunkte durch eine solche Regelung angefahren wird. Daher wird ausschließlich die Messung und Regelung des Zustandes  $x_1$  betrachtet. Zusätzlich sind durch das Einführen einer Eingangskodierung keine Verbesserungen zu erzielen, wie Abbildung 5.22 zeigt. Daher werden lediglich die Modellstrukturen S1 und S3 für die Schätzung mit Time-Delay-Koordinaten verwendet.



a) MRSE über verschiedenen Systemordnungen

b) Varianz in Form der  $3\sigma$ -Grenzen für S3,  $n_c = 7$

Abbildung 5.29: Fit und Varianz der DKLc-Schätzung nach Einführung von Time-Delay-Koordinaten bei Messung des Zustandes  $x_1$

Abbildung 5.29a zeigt den MRSE über verschiedene Modellordnungen gegenüber den Referenzmodellen aus dem vorherigen Abschnitt ohne Delay-Koordinaten. Für Modellstruktur S1 kann für kleinere Modellordnungen durch das Einführen der Delay-Schritte sogar eine bessere Schätzung

hinsichtlich des Prädiktionsfehlers erreicht werden als bei der Nutzung des gesamten Zustandsvektors. Für größer werdende Modellordnungen ist die Modellqualität für diesen Zustand nahezu identisch. Des Weiteren wird die Schätzung durch zusätzliche Delay-Schritte nur leicht verbessert. Bei einer realen Anwendung wäre daher die erhöhte Komplexität durch mehr Delay-Schritte nur in Ausnahmefällen zu rechtfertigen. Für Struktur S3 die die Schätzung mittels  $d = 4$  bis zur erweiterten Systemordnung  $n_e = 6$  gleich bis schlechter im Vergleich mit einer Schätzung mit einem Delay-Schritt. Ab einer Modellordnung  $n_e = 7$  ist praktisch kein Unterschied in der Schätzqualität mit Delay-Koordinaten festzustellen. Auf die Varianz der Schätzung des ersten Zustandes hat das Einführen von Delay-Koordinaten für das vorliegende Beispiel keinen erwähnenswerten Effekt, wie Abbildung 5.29b zeigt.

Wie bereits beim Van der Pol System soll auch der Duffing Oszillator mittels eines Zustandsreglers auf Basis der Time-Delay-Koordinaten geregelt werden. Die beiden stabilen Arbeitspunkte werden durch alle Modellansätze bei geeigneter Wahl der Gewichtungen in endlicher Zeit ohne konkreten Vorteil gegenüber anderen Methoden eingestellt. Eine nähere Betrachtung ergibt damit keinen Mehrwert. Aufgrund derselben Gegebenheiten bleibt eine Beurteilung der Steuerbarkeitsmaße aus. Um allerdings die Grenzen und Potenziale der beschriebenen Methode bei der Nutzung von Time-Delay-Koordinaten aufzuzeigen, wird die Regelung der instabilen Ruhelage bei  $\underline{x}_{R,3} = \underline{0}$  im Folgenden näher betrachtet.

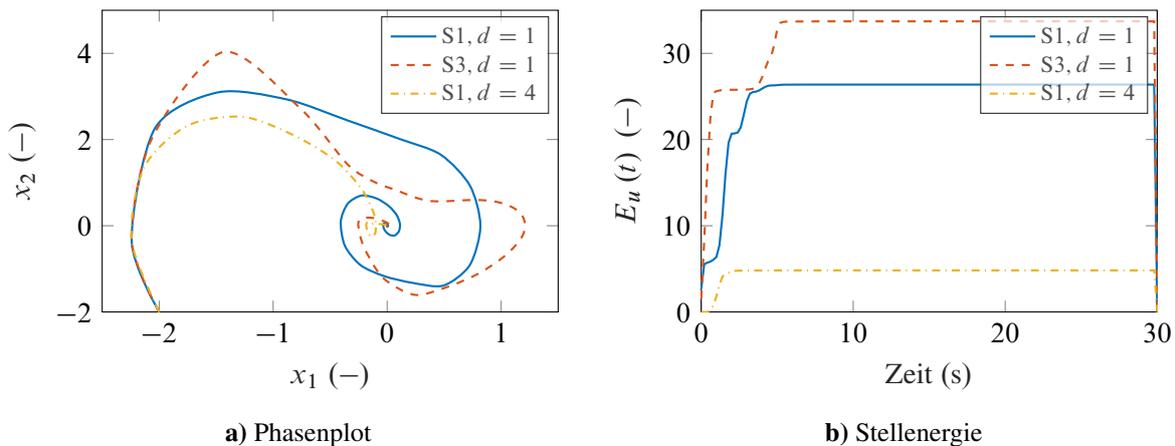


Abbildung 5.30: Phasenplot und Stellenergie des mittels Delay-DKLC-Reglern in den instabilen Arbeitspunkt  $\underline{x}_{R,3} = \underline{0}$  geregelten Duffing Oszillators,  $n_e = 8$

Abbildung 5.30a zeigt den Phasenplot des in den instabilen Arbeitspunkt geregelten Duffing Oszillators unter Verwendung verschiedener Modellstrukturen sowie Delay-Schritte. Die gewählte Systemordnung von  $n_e = 8$  ergibt sich anhand des besten  $AIC_c$  in Verbindung mit dem besten Regelergebnis. Das Regelergebnis mit in die Auswahl der Modellordnung einzubeziehen zeigt erneut die Nichtlinearität des Ansatzes, wie sie bereits bei der Identifikation mit dem gesamten Zustand auftritt. Dies äußert sich maßgeblich in einer Abhängigkeit des Regelergebnisses von der Systemordnung. Je nach Ordnung kommt es dabei dazu, dass kein DKLC-Regler gefunden werden kann, der den instabilen Arbeitspunkt stabilisiert und stattdessen in einen der stabilen Arbeitspunkte „rutscht“. Im Phasenplot ist je nach Modell und verwendeter Anzahl an Delay-Schritten ein völ-

lig unterschiedliches Verhalten erkennbar, was sich ebenso in der aufzuwendenden Stellenergie in Abbildung 5.30b zeigt.

Das mit Abstand beste Reglergebnis wird mit einem Modell auf Basis von Struktur S1 mit  $d = 4$  erreicht. Die Verwendung der anderen Regler führt zu einem erheblich größeren Überschwingen. Somit scheint es für dieses Modell durch die erhöhte Anzahl an Delay-Schritten zu einem Dämpfungseffekt zu kommen, der sich positiv auf das Einregelverhalten sowie die aufzuwendende Stellenergie auswirkt. Mit diesem Effekt ist es möglich, auch den Zustand  $x_2$  im Vergleich zu den anderen Reglern schneller in Richtung seiner Ruhelage zu bringen. Mit Modellstruktur S3 bei  $d = 4$  ist es mit keiner Modellordnung möglich, einen stabilisierenden Regler zu berechnen.

Es ist jedoch festzuhalten, dass all diese Regler aufgrund ihrer völlig verschiedenen Verhaltensweisen schlecht miteinander zu vergleichen sind. Des Weiteren zeigt dieses Beispiel anschaulich, dass man auf das Testen der hier vorgestellten Entwurfsmethodik angewiesen ist, auch wenn durch Satz 3.11 eine gewisse Stabilitätsgarantie erreichbar ist. Eine allgemein gültige Aussage lässt sich aus diesem Beispiel selbstverständlich nicht ziehen, zumal das Van der Pol System performantere Ergebnisse liefert.

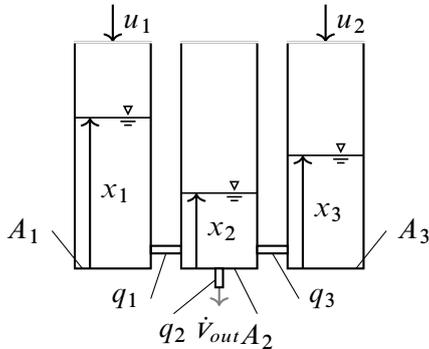
Mit den gewonnenen Ergebnissen und Erfahrungen anhand der rein mathematischen Systeme soll die Methodik nun im nachfolgenden Abschnitt an einem realen Prüfstand getestet werden, um so die Anwendbarkeit der Methode auf reale Systeme zu validieren.

## 5.2 Drei-Tank Prüfstand

Durch die vielversprechenden Ergebnisse der bisher betrachteten theoretischen Beispielsysteme soll die in dieser Arbeit vorgestellte Methode auf einen realen Prüfstand angewendet werden. Hierzu steht ein sogenannter Drei-Tank Prüfstand zur Verfügung, mit dem mittels zweier Pumpen Wasser von einem Reservoir von oben in die beiden äußeren Tanks gefördert werden kann. Der mittlere Tank wird über jeweils einen Zulauf von den äußeren, direkt aktuierten Tanks gespeist. Die Flüssigkeit kann dann über einen Ablauf am mittleren Tank zurück in das Reservoir fließen. Eine Konzeptskizze der verwendeten Konfiguration sowie ein Foto des Prüfstandes sind in Abbildung 5.31 dargestellt. Aus dieser Konzeptskizze lässt sich mit Hilfe des Abflussgesetzes von Torricelli  $v = \sqrt{2gh}$ , welches sich aus der Bernoulli-Gleichung ableiten lässt und die Abflussgeschwindigkeit  $v$  abhängig vom Füllstand  $h$  eines gefüllten Zylinders bei vorhandener Erdbeschleunigung  $g$  beschreibt, die in Gleichung (5.3) dargestellte Dynamik herleiten. In dieser Gleichung gibt der Zustand  $\underline{x}$  die Füllstände der drei Tanks an. Bei der angegebenen Gleichung gilt im Unterschied zu Abbildung 5.31a  $A_1 = A_2 = A_3 = A$ . Tabelle A.1 im Anhang zeigt die Parameter, welche über Abflussmessungen für das angegebene Modell ermittelt werden können.

$$\dot{\underline{x}} = \frac{1}{A} \begin{bmatrix} u_1 - q_1 \sqrt{2g(x_1 - x_2)} \\ q_1 \sqrt{2g(x_1 - x_2)} + q_3 \sqrt{2g(x_3 - x_2)} - q_2 \sqrt{2gx_2} \\ u_2 - q_3 \sqrt{2g(x_3 - x_2)} \end{bmatrix} \quad \text{für } \begin{matrix} x_2 < x_1 \\ x_2 < x_3 \end{matrix}. \quad (5.3)$$

Der mittlere Tank des Aufbaus ist lediglich indirekt über die beiden äußeren, direkt aktuierten, Tanks einstellbar. Durch die beiden Eingänge handelt es sich um ein Mehrgrößensystem, was die Komplexität gegenüber den vorangegangenen Beispielen erhöht. Der Prüfstand mit verwendeter Hard- sowie Software ist in Anhang A.2 mit weiteren Details beschrieben.



a) Prinzipskizze des Drei-Tanks



b) Foto des Drei-Tank Prüfstandes

Abbildung 5.31: Visualisierungen des Drei-Tank Prüfstandes

### Identifikation

Für die Identifikation wird das System mit einer Sequenz von Sprüngen des geförderten Wassers angeregt. Um den unterschiedlichen Einfluss der beiden Eingänge auf das Gesamtsystem besser identifizierbar zu machen, wird der linke Tank alle 60 s mit einer Anregung  $u_1$  im Bereich  $[-0,8, -0,1]$  V variiert. Der zweite Eingang  $u_2$  wird dagegen nicht nur 30 s zeitversetzt, sondern auch alle 30 s zwischen  $[-1, 0]$  V etwas stärker variiert. Die Amplituden der Sprünge beider Eingänge sind normalverteilt mit  $\mathcal{N}(-0,55 \text{ V}, 0,25 \text{ V})$  und werden an den oben angegebenen Grenzen gesättigt. Dadurch kommt es zu einer leichten Erhöhung in der Häufigkeit in Richtung der maximalen Anregungen.

Die Parameter der Anregung wurden experimentell ermittelt und so gewählt, dass das Reservoir und die einzelnen Tanks nicht leerlaufen können. Die Messung wurde über einen Zeitraum von  $T = 95,71 \text{ h}$  durchgeführt. Mit den gegebenen mittleren Anregungen stellt sich somit ein Arbeitspunkt bei etwa  $\underline{x}_{\text{AP}}^T = [195,0, 82,9, 197,2]$  mm ein.

Nach der Analyse des Einflusses von Rauschen auf die DKLC-Schätzung, methodisch auf Seite 69 und beispielhaft anhand des Van der Pol Oszillators, ist ein besonderes Augenmerk auf die Vorverarbeitung der Daten zu legen, da diese beim Drei-Tank-Prüfstand nur erheblich verrauscht gemessen werden. Daher werden alle gemessenen Eingangs- und Ausgangsdaten mittels eines Tiefpass-Chebyshev-Filters vom Typ II mit einer Grenzfrequenz bei 10 Hz und Filterordnung  $n_f = 6$  sowohl vor der Identifikation als auch vor deren Verwendung bei den später analysierten Reglern gefiltert<sup>27</sup>. Nach dieser Filterung erfolgt zur Identifikation eine Unterabtastung der Daten.

<sup>27</sup>Die Rohdaten der Messungen stehen unter <https://1drv.ms/u/s!AtaxmCt9qCAr1P1shsCj92pQP8wsyw?e=gL14eu> zum Download zur Verfügung.

Die Mikrocontroller Karte tastet die Messgrößen des Systems mit  $f_s = 1000$  Hz ab. Dies ist jedoch aus zwei Gründen für die Identifikation ungeeignet. Der erste Grund ist, dass bei einer hohen Abtastfrequenz im Vergleich zur Systemdynamik alle Eigenwerte des zeitdiskreten Modelläquivalents in Richtung eins „rutschen“. Dies führt zu einer schlechten Unterscheidbarkeit der Eigenwerte der Systemmatrix und erschwert daher die Identifizierbarkeit in praktischer Hinsicht. Des Weiteren muss eine Messung durch die sehr langsame Dynamik des Drei-Tank-Systems mehrere Tage durchgeführt werden. Eine zu hohe Abtastfrequenz generiert somit sehr viele Daten ohne praktischen Informationsgehalt. Die Unterabtastung auf  $T_s = 5$  s wird anhand zweier Aspekte festgelegt. Zum einen müssen für die Identifikation ausreichend Messpunkte vorhanden sein. Die identifizierten DKLC-Modelle besitzen bis zu 22.500 zu schätzende Parameter. Bei der gegebenen Messzeit von  $T = 95,71$  h ergibt sich damit eine untere Grenze für die Abtastzeit von  $T_{s,\min} = 7,6566$  s, um auf mindestens zwei Messpunkte pro Parameter zu kommen. Hinsichtlich der Dynamik empfiehlt Ljung in [65, S. 452] eine Abtastzeit, welche etwa dem zehnfachen der Bandbreite ( $-3$  dB Amplitudenabfall) eines linearen Systems entspricht. Dies würde bei einer Bandbreite von etwa  $0,00484$  Hz, welche anhand der Linearisierung des nichtlinearen Modells (5.3) um den oben gegebenen Arbeitspunkt ermittelt wird, zu einer Abtastzeit von  $20,661$  s führen. Damit ergeben sich zeitdiskrete Eigenwerte von etwa  $\lambda_1 = 0,3927$ ,  $\lambda_2 = 0,8219$  und  $\lambda_3 = 0,9121$ , was die Eigenwerte der Systemmatrix zwar gut unterscheidbar macht, aber bei der gegebenen Messzeit zu  $N \approx 16.500$  Messpunkten führt, was weniger als einem Messpunkt pro Parameter entspricht. Daher wird die Abtastzeit zu den genannten  $T_s = 5$  s gewählt. Dies führt zu einer zeitdiskreten Systemmatrix der Linearisierung von Gleichung (5.3), bei der die Eigenwerte bei  $\lambda_1 = 0,7976$ ,  $\lambda_2 = 0,9537$  und  $\lambda_3 = 0,9780$  sich für die beiden größten bereits in der zweiten Nachkommastelle verschieden und damit klar voneinander unterscheidbar sind. Bei dieser Wahl verbleiben zusätzlich eine ausreichende Anzahl von Messpunkten von etwa  $N \approx 69.000$ . Die Aufteilung der Daten in Trainings-, Validierungs- und Testdaten wird wie bei den anderen Beispielen zu  $85/10/5$  % gewählt, um die Menge an Trainingsdaten groß zu halten. Die zum Training der Netze verwendeten Parameter sind in Abschnitt A.3.5 im Anhang aufgeführt

Abbildung 5.32 zeigt die Abschätzung der Systemordnung mittels  $\text{PBSID}_{\text{opt}}$  und OKID anhand der normierten Singulärwerte der zur Systemordnung charakteristischen Matrix der einzelnen Algorithmen. Hierbei kann wieder keine eindeutige Aussage über die Systemordnung getroffen werden, wie es insbesondere bei der Analyse von verrauschten Daten und nichtlinearen Systemen oft der Fall ist [99]. Die Darstellung lässt einen Spielraum von  $n_e = 3$  bis  $n_e = 9$ , oder gar Ordnungen im Bereich von  $n_e \approx 20$ , zu. Wie bei den obigen Beispielen sind die Schätzungen der Ordnungen bei den beiden verwendeten Verfahren nicht gleich. Der OKID-Algorithmus weist die charakteristischen Abfälle in den Singulärwerten bei einer um eins reduzierten Ordnung auf.

Der ermittelte Bereich an Systemordnungen wird wieder mittels des  $\text{AIC}_c$  analysiert. Die Ergebnisse davon sind in Abbildung 5.33 aufgetragen. Anders als bei den vorherigen Beispielen ist das Bild zwischen trainiertem und erweiterten Prädiktionshorizont weniger verzerrt. Genauer heißt das, dass die nach  $\text{AIC}_c$  beste Ordnung bei den Modellen mit affinem Eingang auch bei einem Prädiktionshorizont gleich bleibt, der größer als der trainierte Horizont ist. Auch bei einer Eingangstransformation bleibt das prinzipielle Bild ähnlich. Durch das Einführen einer Eingangsko-

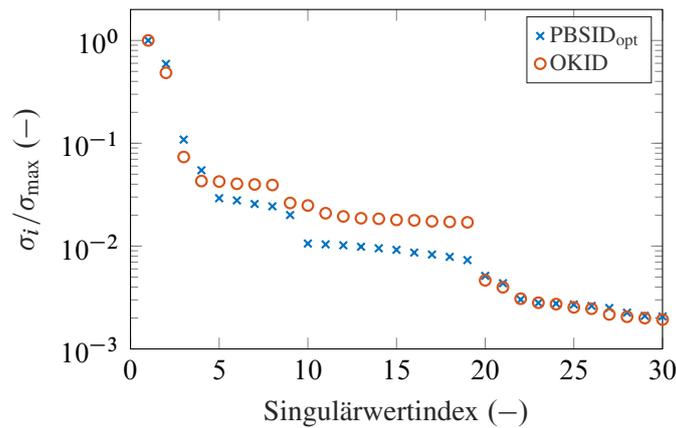


Abbildung 5.32: Ordnungsschätzung des Drei-Tank Systems mittels Subspace-Methoden mit  $p_{\text{PBSID}} = 10$ ,  $f_{\text{PBSID}} = 40$  und  $p_{\text{OKID}} = 50$

dierung kann zusätzlich eine verbesserte Prädiktion erreicht werden. Es ist jedoch dazuzusagen, dass die  $AIC_c$  Werte so ähnlich sind, dass sämtliche Modelle einen ähnlichen Support besitzen [29].

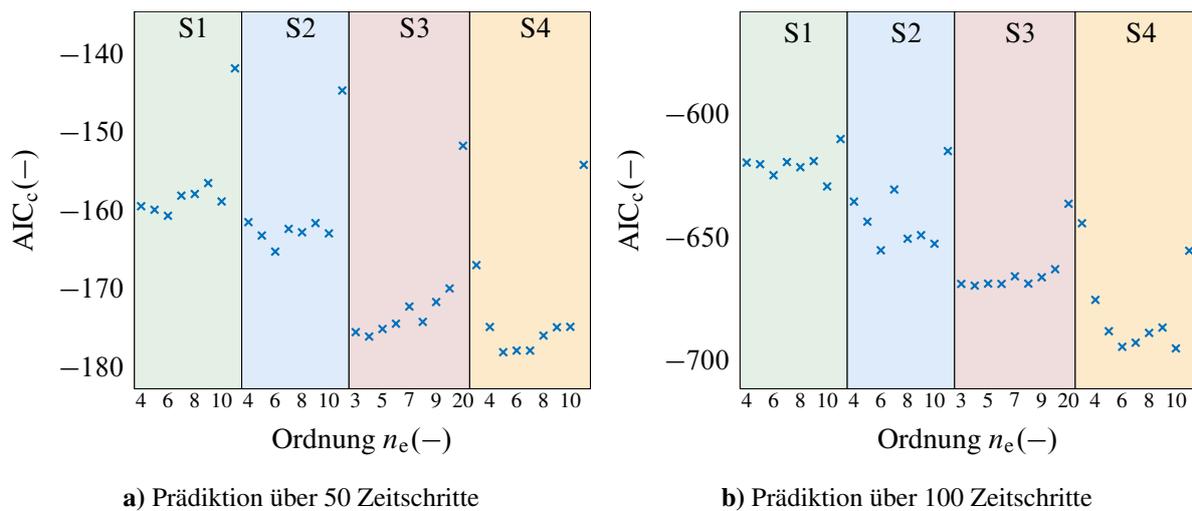


Abbildung 5.33:  $AIC_c$  bei verschiedenen Systemordnungen für die DKLC-Schätzung des Drei-Tank-Systems. Durch die eng beieinander liegenden Datenpunkte ist nur jede zweite Ordnung auf der Abszisse aufgetragen.

Abbildung 5.34 stellt die Fehlerhistogramme des mittleren Tanks dar. Eine Darstellung der Fehlerhistogramme der anderen beiden Tanks ist im Anhang in Abbildung A.6 dargestellt. Ein Vergleich mit einer DMD-Schätzung aus Abbildung A.7 zeigt, dass die DKLC-Schätzung Modelle hervorbringt, die einer wesentlich geringeren Streuung unterliegen, ohne dabei das Modell überanzupassen, was am ausbleibenden Anstieg des Fehlers nach dem trainierten Prädiktionshorizont bei  $p_{\text{DKLC}} = 50$  zu erkennen ist.

Für die DKLC-Modelle fällt zunächst auf, dass unabhängig vom Modellansatz das Systemverhalten auch über das Erreichen des Prädiktionshorizontes hinaus bei allen Modellstrukturen hervorragend prädiziert werden kann. Der Fehler steigt über die Zeit nicht erheblich an und variiert zwischen den einzelnen Modellansätzen kaum. Bei genauerer Betrachtung der einzelnen Fehlerhistogramme in Abbildung 5.34 zeigt sich jedoch, dass die beiden Ansätze S2 und S4, welche eine Nichtlinearität im Eingang zulassen, einen etwas geringeren Anstieg der Streuung nach Erreichen des Prädiktionshorizont bei  $p = 50$  besitzen. Möglicherweise hängt dies mit der, zwar wenig ausgeprägten aber vorhandenen, Nichtlinearität im Eingang zusammen, wie in Abbildung A.1 im Anhang dargestellt.

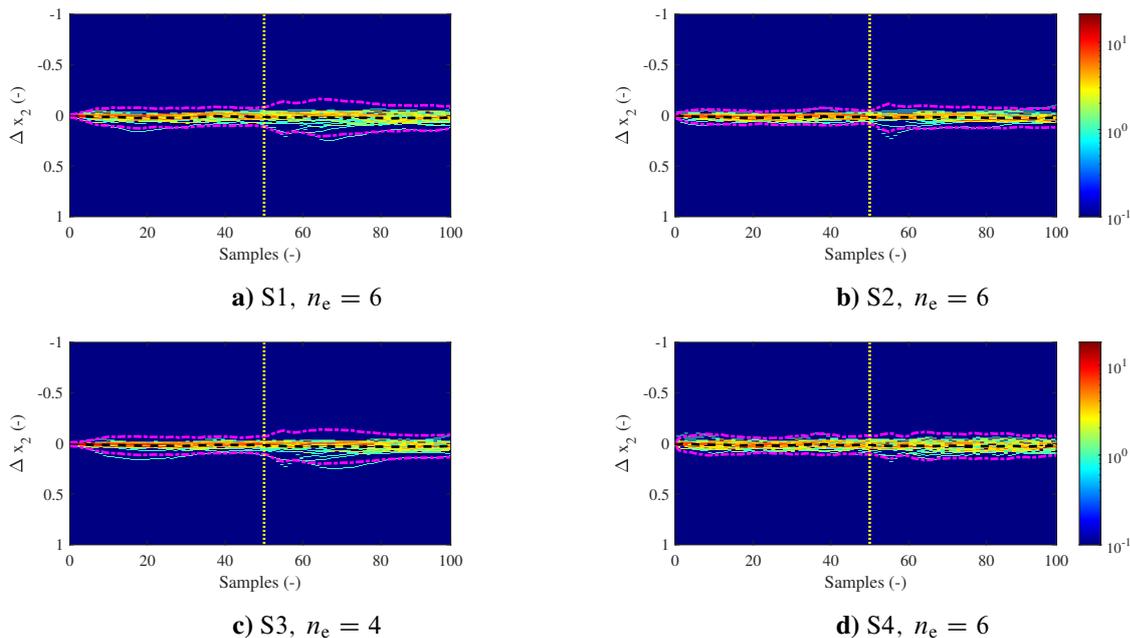


Abbildung 5.34: Fehler Histogramme der besten Modelle nach  $AIC_c$  mit  $p = 100$  des mittleren Tanks des Drei-Tank Systems für verschiedene Modellkonfigurationen

Abbildung 5.35 zeigt zum Zweck der weiteren Analyse dieser Vermutung die Eingangscoder als Funktion der Eingänge. Für die obere Zeile, welche Modellstruktur S2 beschreibt, liegt, genau wie es die Kennlinien aus Abbildung A.1 zeigen, nahezu lineares Verhalten vor. Dies lässt sich daran erkennen, dass es sich bei den Funktionen näherungsweise um Ebenen handelt. Der Fehler gemäß MRSE gegenüber einer Ebene beträgt dabei nur etwa 7 %. Für die Modellstruktur S4 liegt dieser Fehler bei noch niedrigeren 2 % gegenüber einer Ebene. Die Fehlerspitzen liegen für beide Beispiele jedoch vornehmlich am Rand, wie Abbildung A.8 im Anhang zeigt. Hier stößt die gesamte Approximation ohnehin an ihre Grenzen, da der normierte Bereich von etwa  $-1$  bis  $1$  reicht. Allerdings kommt es teilweise zu kleineren Unterschieden zwischen linearer Approximation und tatsächlicher Encoder-Funktion. Diese scheinen jedoch Grund genug zu sein, eine Erhöhung des Fehlers über die Testtrajektorien bei den Modellstrukturen S1 und S3 auszulösen, wie die zugehörigen Fehlerhistogramme in Abbildung 5.34 zeigen. Wie bereits beim Van der Pol Oszillator (zu sehen in Abbildung 5.8) kommt es bei der Transformation des Eingangs bei Modellstruktur S4

dazu, dass der Nullpunkt der Eingangsfunktion verschoben wird. Im Allgemeinen reduziert das Einführen einer Eingangskodierung Varianz der Schätzung.

Interessant ist außerdem, dass die beiden Eingangsfunktionen bei S2 um etwa  $82^\circ$  zueinander gedreht sind<sup>28</sup>, was in etwa dem eigentlichen Systemmodell entspricht. Dies ist dadurch gegeben, dass durch die Unabhängigkeit beider Eingänge voneinander die „Verdrehung“ im realen Modell  $90^\circ$  entspricht. Bei Modellstruktur S4 ist dies durch eine Verdrehung von etwa  $41^\circ$  nur bedingt gegeben, sodass sich diese Darstellung vom eigentlichen Systemmodell unterscheidet. Dies ist jedoch im Rahmen der Koopman-Operator-Theorie nicht ausgeschlossen und kann sogar zu einer Qualitätsverbesserung des Modells führen wie auch in [18] und [93] dargestellt.

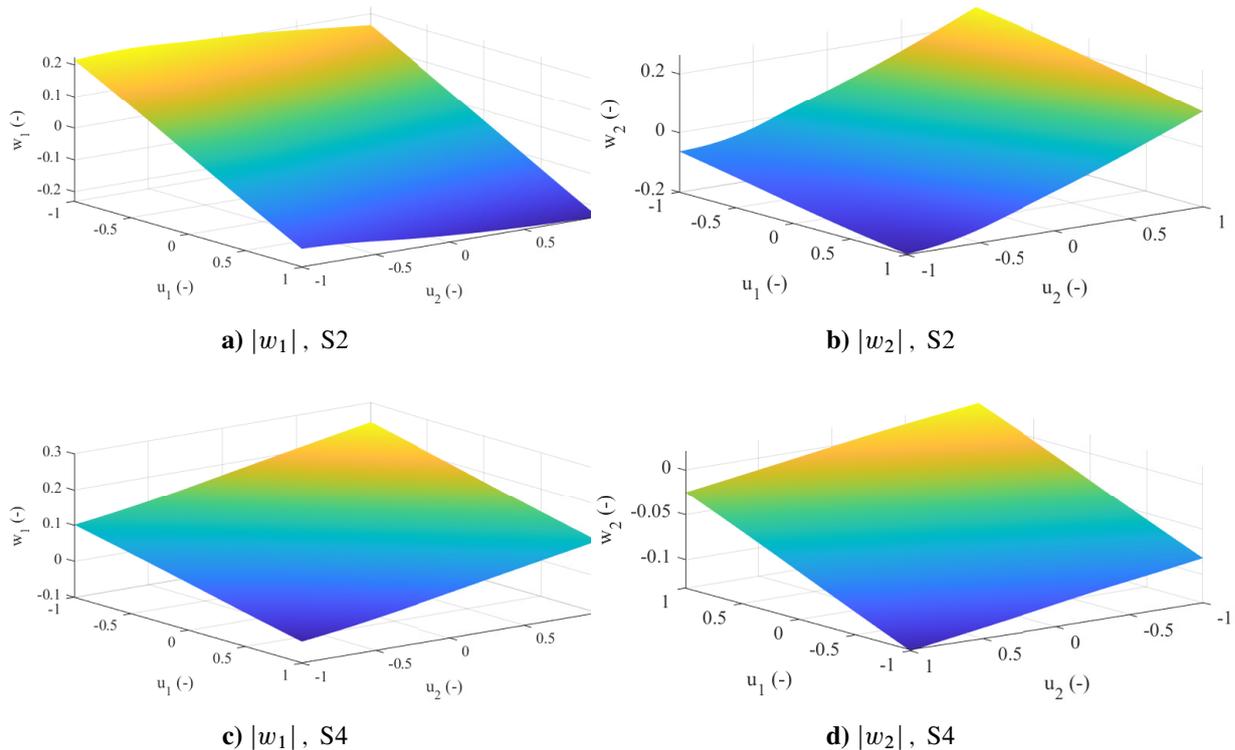


Abbildung 5.35: Betrag der Eingangscoder als Funktion der beiden Eingänge für  $n_e = 6$

Nach der Darstellung der Prädiktionsfähigkeit folgt eine Bewertung der dynamischen Eigenschaften der geschätzten Modelle. In Abbildung 5.36 sind die Eigenwerte verschiedener Modell unterschiedlicher Identifikationsverfahren und die Eigenwerte, welche sich aus der Linearisierung von Gleichung (5.3) am gegebenen Arbeitspunkt ergeben, aufgetragen. Das DKLC-, genau wie das eDMD-Verfahren, finden die ersten beiden dominanten Eigenbewegungen in ähnlicher Qualität. Die dritte reale, am wenigsten dominante, Dynamik unterliegt einer wesentlichen Streuung bei der Schätzung mittels verschiedener Ansätze. Die DKLC-Modelle S1 und S2 finden eine weitere Dynamik im Bereich um  $\lambda = 0,9$ , also zwischen den beiden weniger dominanten Eigenwerten, was nicht mit der Linearisierung vereinbar ist. Im Rückblick auf die Fehlerhistogramme aus Abbildung 5.34 scheint dies jedoch für die Prädiktion weder Nachteile noch Vorteile mit sich zu bringen. Mit

<sup>28</sup>Die Verdrehungen berechnen sich aus den in Abbildung 5.35 dargestellten Ebenen.

dem ARX-Verfahren ist es möglich, die dominanteste und die am wenigsten dominante Dynamik zu identifizieren. Der Eigenwert bei circa  $\lambda = 0,95$  wird hingegen nicht identifiziert, was sich in einer schlechten Schätzung des Systemverhaltens niederschlägt, wie Abbildung A.5 im Anhang zeigt.

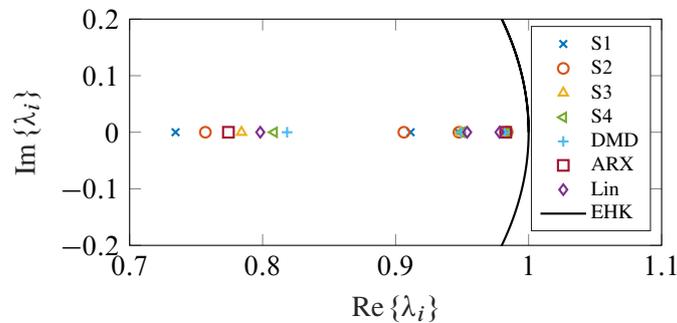


Abbildung 5.36: Zeitdiskrete Eigenwerte der identifizierten Modelle des Drei-Tanks. Für DKLc und MISO-ARX Modelle gilt  $n_e = 6$ , für DMDc gilt  $n = 3$ .

## Regelung

Für die Regelung wurde bisher stets die vollständige Steuerbarkeit vorausgesetzt. Das Drei-Tank System ist allerdings nicht vollständig steuerbar im erreichbarem  $\mathcal{X}$ , da aus der gegebenen Konfiguration heraus und durch die ausschließlich aktive Zuführung ( $u_1, u_2 > 0$ ) von Wasser kein Betriebspunkt mit  $x_2 > x_1$  oder  $x_2 > x_3$  eingestellt werden kann. Somit sind abnehmende Füllstände nur durch Abfließen realisierbar, sodass während der Zeit in der der Füllstand sinkt wenig bis kein Zufluss gestellt wird. Diese Punkte sind bei der Festlegung der anzufahrenden Arbeitspunkte zu berücksichtigen, da sonst gegebenenfalls unprädictibles Verhalten eintritt.

Bevor eine Regelung mit der in der vorliegenden Arbeit beschriebenen Methodik durchgeführt wird, ist wie zuvor eine Bewertung der Steuerbarkeit der einzelnen Eigenwerte notwendig. Die beiden verwendeten Steuerbarkeitsmaße sind in Abbildung 5.37 gegenübergestellt. Wie es bei den vorangegangenen Beispielen bereits häufig der Fall war, ist es mit dem Maß nach Gilbert nicht möglich, die signifikanten Eigenwerte (die auf Basis der Linearisierung) als die dominant steuerbaren zu identifizieren. Für die Modellstruktur S3 sind die Eigenwerte, welche nahe denen der Linearisierung liegen, klar als die dominant steuerbaren Eigenwerte zu erkennen. Bei Modellstruktur S4 ist dies zwar rein numerisch nach wie vor der Fall, jedoch könnte ohne genaueres Systemwissen auch  $\lambda_1$  als potenziell steuerbar erkannt werden.

Spätestens bei Testläufen stellt sich jedoch heraus, dass bei Verwendung von  $\lambda_1$  als Eigenwert beziehungsweise der zugehörigen Eigenfunktion die Regelung nicht die erwartete Güte erreicht, wie Abbildung 5.38 in Form einer Regelung am Prüfstand zeigt. Dargestellt ist die Arbeitspunktanfahrt für ein gewähltes  $\underline{Q}'_x = \text{diag}([10, 10, 10])$  und  $\underline{R}_u = \underline{I}$ . Die äußeren Tanks werden im Gegensatz zur später betrachteten Regelung höher gewichtet, um dem Effekt entgegen zu wirken, dass diese nicht den Zielzustand erreichen. Allerdings bringt auch diese Maßnahme keine Verbesserung mit

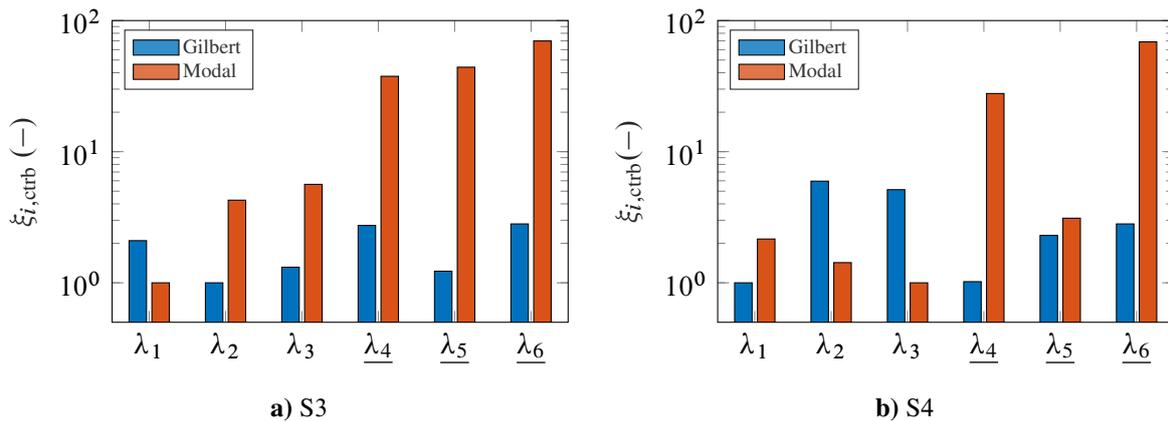


Abbildung 5.37: Steuerbarkeitsmaße  $\xi_{i,ctrb}$  geschätzter Drei-Tank Systeme bei  $n_e = 6$  der einzelnen Eigenwerte. Die Maße sind für jedes Verfahren auf das kleinste Maß über alle Eigenwerte normiert. Die Eigenwerte sind aufsteigend nach Realteil sortiert. Die zu den Eigenwerten der Linearisierung passenden Eigenwerte sind unterstrichen dargestellt.

sich und die Regelung erzielt nicht das beabsichtigte Regelergebnis der nahezu fehlerfreien Arbeitspunktanfahrt aller Tanks. Der Zielfüllstand des mittleren Tank wird hingegen auch mit der „falschen“ Wahl der Eigenwerte erreicht. Für beide Regler gilt, dass durch den fehlenden Integratoranteil im geschlossenen Regelkreis der Arbeitspunkt nicht exakt stationär genau eingeregelt werden kann. Dieses Beispiel zeigt jedoch, wie wichtig die Wahl geeigneter Eigenwerte und eine nachfolgende Validierung des Regler ist.

Da der mittlere Tank zwar gemessen, aber nur indirekt durch die beiden äußeren Tanks, beziehungsweise Pumpen, beeinflusst werden kann, wird der Füllstand dieses Tanks für die nachfolgenden Untersuchungen als maßgebliches Regelziel definiert. Abbildung 5.39 zeigt die Tiefpassgefilterten Ergebnisse einer Reihe von Reglern, welche anhand der Wahl  $\underline{Q}'_x = \text{diag}([1, 10, 1])$  and  $\underline{R}_u = \underline{I}$  berechnet werden. Der verwendete Tiefpassfilter entspricht dem auf Seite 107 beschriebenen. Wie bereits beim Beispiel zu Abbildung 5.38 gilt auch hier, dass der einzustellende Arbeitspunkt durch den fehlenden Integrator im Regler nicht exakt stationär genau erreicht werden kann. Dies ist jedoch auch nicht das Ziel der gegebenen Untersuchung. Es soll vielmehr die Leistungsfähigkeit der Koopman-Regelung gegenüber anderen Methoden aufgezeigt werden.

Wie Abbildung 5.39 zeigt, kann durch das Einführen der Koopman-Regelung am Prüfstand keine signifikante Verbesserung erreicht werden. Das Systemverhalten ist verglichen zum DMD-Ansatz praktisch identisch. Dies gilt für den Zustands- wie auch den Eingangsverlauf. Dieser fehlende Unterschied zwischen den Reglern ist womöglich auf das weitestgehend lineare Verhalten des Systems bei gegebener Anregung und gegebenem Betriebszustand zurückzuführen. Somit kann sich, wie bei den vorangegangenen Beispielen, das nichtlineare Verhalten nur sehr marginal durch die Observables ausprägen und der potenzielle dynamische oder energetische Vorteil der Regler geht verloren beziehungsweise kommt nicht zum Vorschein. Zur Einordnung der durch die Regler induzierten Dynamik ist eine gesteuerte Trajektorie gezeigt, für die das System ausschließlich mit der Stellgröße im Arbeitspunkt angeregt wird.

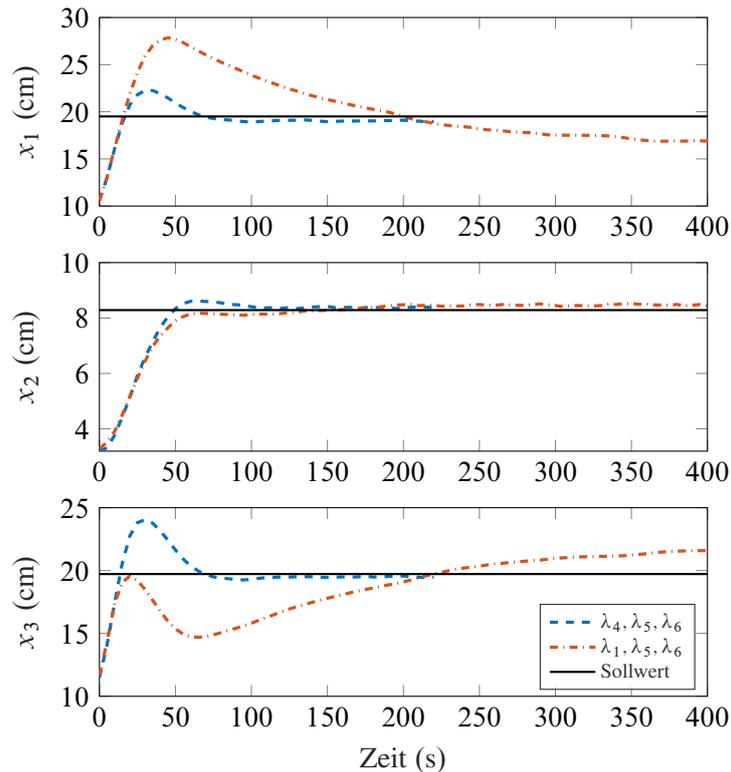


Abbildung 5.38: Einschwingvorgang des Reglers bei Arbeitspunktanfahrt für „falsch“ gewählte steuerbare Eigenwerte bei S4 und  $n_e = 6$

Dieses Beispiel ist daher als *proof-of-concept* für reale Systeme zu verstehen und zeigt die Anwendbarkeit für diese. Gegebenenfalls muss für eine Anwendung eines Koopman-Reglers die Komplexität der Netze reduziert werden, um so eine geringere Anzahl an Matrixmultiplikationen zu erhalten und somit auf Mikrocontrollern lauffähig zu sein. Je nach Systemverhalten kann auch mit weniger tiefen Netzen eine ähnliche Schätzqualität erreicht werden.

Abschließend soll die Stabilität des Systems unter der Nutzung von Koopman-Zustandsreglern analysiert werden. Leider kann auch beim Drei-Tank Prüfstand kein Bereich garantierter Stabilität angegeben werden. Der betragsmäßig größte Eigenwert der geregelten Koopman-Matrix liegt mit der Modellstruktur S3 bei  $n_e = 7$  und den LQ-Matrizen  $\underline{Q} = \text{diag}([1, 10, 1])$  und  $\underline{R} = \underline{I}$  bei etwa  $\lambda_{\max} \approx 0,899$ , was zu einem  $\gamma_{\max} = 0,0057$  führt. Diese obere Fehlerschranke ist verglichen mit den vorangegangenen Beispielen zwar etwas größer, übertrifft aber dennoch den relativen Fehler zwischen Einschrittprädiktion und Observables aus Satz 3.11. Auch in diesem Beispiel gilt für alle untersuchten Systemmodelle  $\alpha < \gamma$ .

### Time-Delay Koordinaten

Das System ist grundsätzlich über alle Zustände schwach beobachtbar<sup>29</sup>, solange die Füllstände nicht alle gleich sind. Diese Konfiguration ist aus dem einfachen Betrieb heraus nicht erreichbar,

<sup>29</sup>Die Erklärung der schwachen Beobachtbarkeit, sowie eine Einordnung ist in Anhang A.1.8 dargestellt.

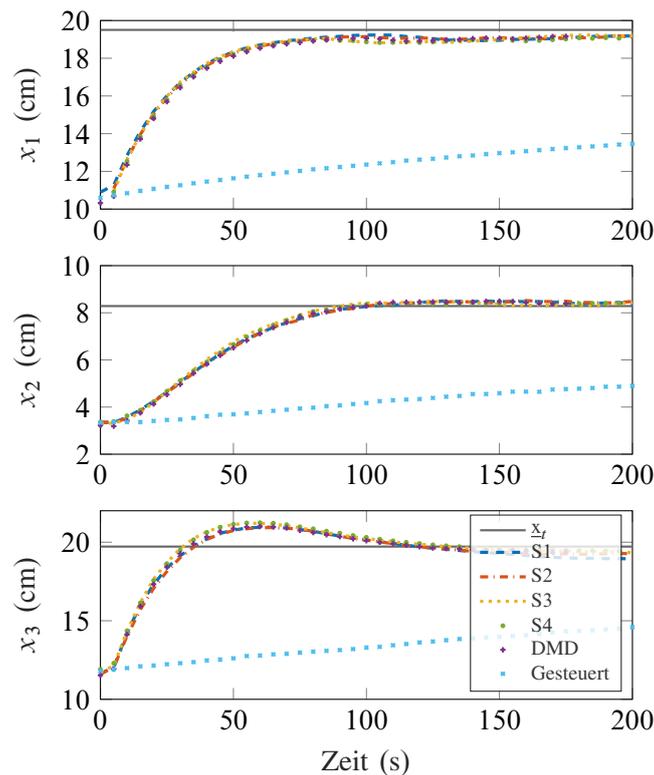


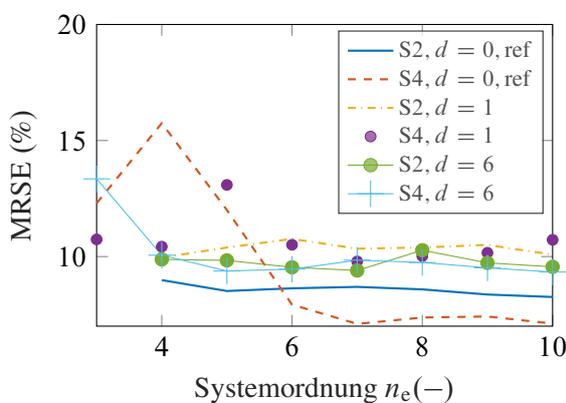
Abbildung 5.39: Tiefpassgefilterte Ergebnisse des Drei-Tank Systems bei verschiedenen Regelelungskonzepten mit Arbeitspunktfahrt nach  $\underline{x}_{AP} = [19,50, 8,29, 19,72]$  cm

da die äußeren Tanks keinen Ablauf besitzen. Sie sind damit immer höher als der mittlere Tank gefüllt. Daher sollte es möglich sein, sämtliche Eigenwerte aus einer Identifikation zu ermitteln, welche auf einer Messung des Füllstandes von lediglich einem Tank beruht. Die Anzahl an Delay-Schritten ergibt sich nach der Argumentation, welche beim Van der Pol System besprochen wurde, zu  $d = 1$  sowie  $d = 6$ .

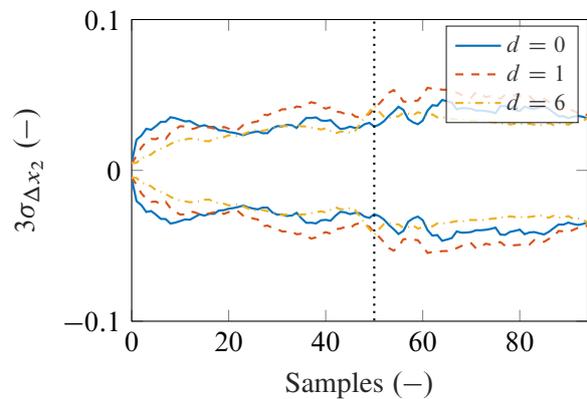
Bei der isolierten Messung einzelner Füllstände geht die Schätzung der beiden äußeren Tanks stets mit einer enormen Vergrößerung der Varianz einher, wie beispielhaft für  $x_1$  in Abbildung A.10 im Anhang dargestellt. Für den mittleren Tank ist dies allerdings nicht der Fall, wie Abbildung 5.40b beispielhaft für die gleichen Modellkonfigurationen nach Abbildung 5.34 aus dem Beispiel ohne Delay-Schritte zeigt. Es ist deutlich zu sehen, dass die anfängliche Varianz der Schätzung mit der Erhöhung der Delay-Schritte reduziert wird. Für  $d = 1$  erhöht sich jedoch nach etwa 20 Prädiktionsschritten wieder. Für den größeren Delay-Horizont ist ein geringer positiver Effekt auf die Streuung der Schätzung über den kompletten Verlauf zu beobachten.

Die Schätzqualität hinsichtlich des Fits, durch das Einführen der Delay-Schritte, zeigt ein schlüssiges Bild, wie Abbildung 5.40a zeigt. Bei der Schätzung mittels aller gemessenen Zustände kann das beste Schätzergebnis über beide betrachtete Modellstrukturen erreicht werden. Außerdem geht eine Erhöhung der Delay-Schritte bei fast jeder Ordnung mit einer Verbesserung der Schätzqualität einher. Jedoch ist diese nicht signifikant, sodass in der Anwendung daher zu überlegen ist,

ob die Erhöhung der Komplexität durch mehr Delay-Schritte gegenüber den vernachlässigbaren Verbesserungen zu rechtfertigen ist.



a) MRSE über verschiedenen Systemordnungen



b) Varianz des Schätzfehlers in Form der  $3\sigma$ -Grenzen für S4,  $n_e = 7$

Abbildung 5.40: Fit und Varianz der DKLc-Schätzung nach Einführung von Time-Delay-Koordinaten bei Messung des Zustandes  $x_2$

Hinsichtlich der Dynamik der identifizierten Modelle und ihrer Steuerbarkeit sind zwei weitere Punkte zu besprechen. Zum einen werden für sämtliche Konfigurationen unabhängig von der Messgröße und den Delay-Schritten lediglich zwei Eigenwerte geschätzt, die in der Nähe der Eigenwerte der Linearisierung liegen. Dies widerspricht zunächst den theoretischen Ergebnissen der Beobachtbarkeit des Systems. Außerdem werden bei der Schätzung mit den beiden äußeren Tanks lediglich die beiden langsamsten Eigenwerte bei etwa  $[0,9538, 0,9780]$  geschätzt. Bei der Identifikation mit dem mittleren Tank als einzige Messgröße wird hingegen der schnellste, sowie der langsamste Eigenwert geschätzt. Dies ist außerdem, anders als es zunächst erscheinen mag, kein Widerspruch, da der zweite Tank hinsichtlich seiner Eigendynamik durch den etwa doppelt so großen Ablaufquerschnitt schneller abfließen kann, wie die Parameter in Tabelle A.1 im Anhang aufzeigen.

Dieses Beispiel zeigt anschaulich, dass auch wenn das Modell aus der Theorie heraus über einen einzelnen Ausgang (schwach) beobachtbar ist, es dennoch zu einem Informationsverlust kommen kann. Für dieses Beispiel ist dieser „Informationsverlust“ vermutlich auf die ähnliche Dynamik der beiden äußeren Tanks, oder deren Einfluss auf die übrigen Zustände zurückzuführen. Die Folge ist, dass nicht alle dominanten Eigenwerte aus der Messung lediglich eines Zustandes rekonstruiert werden können, wie es bei der vollständigen Messung des gesamten Zustandes möglich ist.

Die Konfiguration der Messung, Schätzung und Regelung der beiden äußeren Tanks ist allerdings weniger interessant, da der Füllstand praktisch direkt über die einzelnen Zuläufe gesteuert werden kann und wird daher nicht weiter beleuchtet. Der Vollständigkeit halber sei gesagt, dass bei Testläufen der erwartete Füllstand eingestellt wird.

Für die Regelung ist die Besonderheit zu beachten, dass durch die Schätzung mittels einzelner Messgrößen nur zwei der drei dominanten Eigenwerte geschätzt werden können. Aus der Erfah-

rung der letzten Beispiele heraus sollten für die Regelung daher nur diese verwendet werden. Es zeigt sich zudem eine weitere Problematik bei der vollständigen Kodierung der Zustände mittels der Modellansätze S3 und S4. Wird die Steuerbarkeit der einzelnen Eigenwerte betrachtet, so fällt auf, dass für die meisten Modelle nicht die dominanten Eigenwerte, sondern völlig andere als steuerbar erkannt werden, wie Abbildung A.9 im Anhang zeigt. Des Weiteren werden eine Vielzahl weiterer Moden als ähnlich gut steuerbar erkannt und nicht, wie in den voran gegangenen Beispielen, meist  $n$ , wie es Heuristik 3.14 vorgibt. Dies führt bei der Regelung zu dem in Abbildung 5.38 aufgezeigten Effekt.

Für das folgende Beispiel wird die manuelle Auswahl von lediglich zwei Eigenwerten verwendet, da die dritte potenziell steuerbare Dynamik bei der Identifikation unterzugehen scheint. Es wird diese Methode gewählt, um effektiv die beiden ausgewählten Eigenfunktionen des Koopman-Operators zu beeinflussen, welche bereits in der vorangegangenen Untersuchungen zu guten Ergebnissen führt. Des Weiteren ist diese Wahl wichtig, um konsistent und vergleichbar mit den bisher betrachteten Analysen und Beispielen zu bleiben. Dieses Beispiel zeigt jedoch deutlich, dass die Übertragung der Steuerbarkeitseigenschaft aus Heuristik 3.14 nicht wie bei den vorangegangenen Beispielen bei Verwendung der Time-Delay-Koordinaten während der Identifikation erhalten bleibt.

Abbildung 5.41 zeigt die Regelung des Drei-Tanks in Form einer Arbeitspunktanfahrt für die nach  $AIC_c$  besten Modelle. Es ist gegenüber Abbildung 5.39 zu beobachten, dass die beiden äußeren Tanks deutlich weiter überschwingen. Dies hat anschaulich den Grund, da diese Tanks nicht mehr effektiv mitgeregelt werden, da sie lediglich als Zulauf für den mittleren Tank dienen. Der Zielarbeitspunkt wird verglichen zur Systemdynamik schnell eingeregelt und verharrt dort. Auch die anderen Tanks werden mit der Zeit den Zielarbeitspunkt erreichen, da mit Erreichen des Füllstandes in mittleren Tank keine Differenzstellgröße mehr abseits des konstanten Volumenstroms  $\underline{u}_{AP}$  zufließt, der im Arbeitspunkt gilt. Allerdings kann dies, aufgrund dessen dass die Füllstände nicht geregelt werden, eine lange Zeit in Anspruch nehmen.

### 5.3 Zusammenfassende Bewertung der Beispiele

Die beiden theoretischen Beispiele, sowie das reale Prüfstand Beispiel haben gezeigt, dass die in dieser Arbeit entwickelte Methode zu validen Ergebnissen und leistungsfähigen Reglern führen kann. Die Übertragung des interpretierbaren Zustands- in den Koopman-Raum gelingt, so wie es die Herleitungen angeben.

Der Teil der Ordnungsabschätzung der vorgestellten Kette an Methoden aus Abbildung 4.5, welcher die Subspace-Verfahren integriert, hat für alle Beispiele grundsätzlich sinnvolle Bereiche für die anzunehmende Ordnung ergeben. Diese liegen über alle Beispiele jedoch in einer ähnlichen Größenordnung. So kann fälschlicher Weise der Eindruck entstehen, dass die Subspace-Verfahren immer in diesem Bereich eine entsprechende lineare Ordnung finden. Für die besprochenen Beispiele liegt dies jedoch maßgeblich daran, da die hier analysierten Systemen alle ähnliche Ordnung

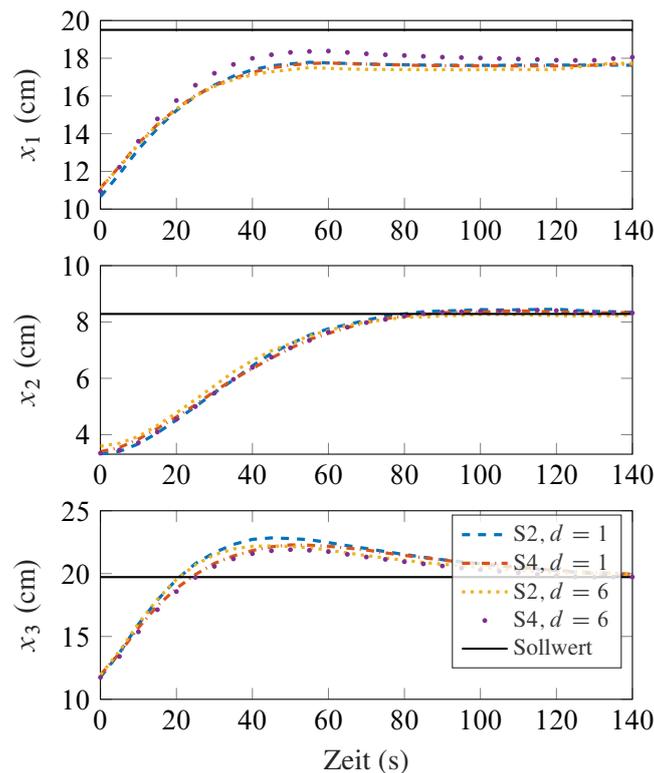


Abbildung 5.41: Tiefpassgefilterte Ergebnisse des Drei-Tank Systems bei verschiedenen Regelungskonzepten mit Arbeitspunktanfahrt nach  $\underline{x}_{AP} = [19,50, 8,29, 19,72]$  cm mittels isolierter Messung von  $x_2$

aufweisen. Als ein Ausblick gebendes Beispiel sei Abbildung 5.42 gezeigt, welches die Singulärwerte aus der Subspace-Identifikation anhand eines Hefe Glykolyse Systems nach [36] zeigt, welches ein nichtlineares Zustandsraumsystem siebter Ordnung darstellt. Der Plot zeigt, dass für dieses System, je nach Anzahl der Ausgänge, verschiedene Ordnungen abgeschätzt werden können. Dies ist unter dem Gesichtspunkt, dass mehrere unterschiedlich geartete Schwingungen dargestellt werden müssen nachvollziehbar. Alle Schätzungen haben jedoch gemein, dass es um den Singulärwertindex bei circa 13 einen gemeinsamen Abfall in den Singulärwerten mit anschließendem Plateau gibt. Dieses Plateau ist jedoch weit weniger deutlich, als bei den übrigen besprochenen Beispielen. Dies liegt jedoch maßgeblich an der ausgeprägten Nichtlinearität des Systems [99].

Bei den mathematischen Modellen ist ein weiterer erwähnenswerter Punkt hinsichtlich der Ordnungsschätzung mittels  $AIC_c$  erkennbar. Durch das Einführen des Eingangscoders kann bei allen Modellstrukturen über alle betrachteten Horizonte eine Systemordnung eingespart werden. Hierbei ist in der Anwendung ein gewisser Kompromiss zwischen erhöhter Komplexität durch die Modellordnung oder durch das Einführen des Eingangscoders abzuschätzen. In der Regel wird sich jedoch durch die Gestalt des Encoders dieser Mehraufwand nicht lohnen, sodass es vor allem in der Anwendung an realen Systemen und auf Mikrocontrollern sinnvoller sein kann die Systemordnung um eins zu erhöhen und auf den Eingangscoder zu verzichten. Dieses Verhalten kann beim Drei-Tank in dieser Form nur wenig ausgeprägt beobachtet werden. Hierzu sei erwähnt, dass

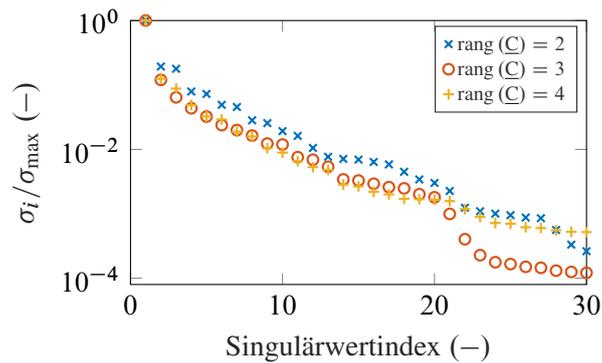


Abbildung 5.42: Ordnungsschätzung des Glycolyse-Systems mittels  $\text{PBSID}_{\text{opt}}$  um beide Arbeitspunkte

dies durch das weitestgehend lineare Verhalten ausgelöst sein kann, da eine Komplexitätserhöhung nur eine nahezu vernachlässigbare Verbesserungen mit sich bringt.

Basierend auf den analysierten Systemen lässt sich bezüglich der Steuerbarkeit der geschätzten Eigenwerte Folgendes festhalten. In so fern eine Systemgleichung mit Linearisierung, beziehungsweise eine grobe dynamische Vorkenntnis vorhanden ist, wo die maßgebliche Dynamik liegt, können die zugehörigen Eigenwerte für eine erste Analyse einer Koopman-Regelung als steuerbar angesehen werden. Sollte diese nicht vorhanden sein, so ist eine genauere Betrachtung durch eine Steuerbarkeitsanalyse der einzelnen Eigenwerte sinnvoll, wie sie in den vorangegangenen Beispielen vorgestellt wird. Hierbei ist, basierend auf den untersuchten Beispielen, das Verfahren nach Lückel und Müller gegenüber dem Kriterium nach Gilbert zu bevorzugen. Die einfache Annahme, dass der dominanteste Eigenwert beziehungsweise das dominanteste Eigenwertpaar steuerbar ist, kann nur in Ausnahmefällen angewendet werden, wie das Beispiel des Van der Pol Oszillators zeigt.

Für die Wahl der Reglergewichte  $\underline{Q}'_x$  und  $\underline{R}'_u$  fällt zudem auf, dass diese für einen ähnlichen Effekt bei der vollständigen Kodierung der Zustände wesentlich höher gewählt werden müssen als für die Modellstrukturen S1 und S2. Ein Grund dafür kann die Skalierung der Observables sein, da die Einträge je nach geschätztem System vergleichsweise klein sind und somit die Auswirkung der Einträge in  $\underline{Q}_z$  reduziert wird. Des Weiteren ist insbesondere bei Verwendung von Time-Delay-Koordinaten bei dieser Methodik Vorsicht geboten. Die Ergebnisse für den Van der Pol Oszillator und den Drei-Tank sind zwar vielversprechend, das Beispiel des Duffing Oszillators zeigt jedoch, dass ein Testen der Regler an einem realen Prüfstand nur unter strengen Aufsicht durchzuführen ist, da es gegebenenfalls zu starken Ausschlägen in den nicht gemessenen Zustandskoordinaten kommen kann.

Als letzten Punkt soll die Stabilität des Systems anhand von Satz 3.11 bewertet werden. In der Erläuterung zu den Lemmas in Abschnitt 3.1.1, auf denen Satz 3.11 aufbaut, wurde aufgrund der Vereinfachungen in ihren Beweisen die Vermutung aufgestellt, dass der Parameter  $\gamma$  weniger Restriktionen unterliegt als  $\alpha$ . Dies sollte dazu führen, dass sich eine höhere Fehlerschranke einstellt. Diese Vermutung hat sich in allen untersuchten Beispielen bestätigt. Dennoch sind die

Fehlerschranken so niedrig, dass eine nahezu perfekte Schätzung im Koopman-Raum notwendig ist, um den Bedingungen zu entsprechen. Die praktische Anwendung von Satz 3.11 ist somit womöglich nur in Ausnahmefällen sinnvoll. Die reale Fehlerschranke liegt allerdings vermutlich deutlich höher. Diese Vermutung basiert auf der Beobachtung, dass, solange das Modell plausibel ist und die steuerbaren Eigenwerte in der Berechnung von  $\underline{Q}_z$  ausgewählt werden, sich kein geschätztes Modell ergeben hat, auf dessen Basis ein Koopman-Zustandsregler das nichtlineare System destabilisiert.

## 6 Zusammenfassung und Ausblick

### Zusammenfassung

Die vorliegende Arbeit beschäftigt sich mit der Einordnung des Koopman-Operators in den Kontext angeregter nichtlinearer dynamischer Systeme. Hierzu wurden theoretische Grundlagen aufbereitet und erweitert. Die beiden wichtigsten theoretischen Erkenntnisse sind dabei die eingeschränkte Steuerbarkeit Koopman-linearer Äquivalente sowie die Sicherstellung der Stabilität des geschlossenen Regelkreises unter der Nutzung von linearen Zustandsreglern auf Basis der Koopman-Operator-Theorie und den somit geschätzten Koopman-linearen Modellen.

Zur Schätzung solcher Modelle wurde eine Deep-Learning-basierte Methode erarbeitet, welche auf dem Grundprinzip eines Autoencoders, einem bekannten Modellansatz aus dem Bereich des *Machine learning*, basiert. Hierbei werden die Messgrößen zunächst transformiert, dann mit einem linearen Zustandsraumssystem, welches das eigentliche Koopman-lineare System darstellt, in der Zeit entwickelt und anschließend wieder rücktransformiert, um die Schätzqualität besser beurteilen zu können. Anders als bisherige Verfahren fasst die in dieser Arbeit entwickelte Methode außerdem Eingangsgrößen in die Schätzung mit ein und ermöglicht so eine Regelung des identifizierten, nichtlinearen Systems.

Der Nachteil der Deep-Learning-basierten Methode ist allerdings ihre Rechenzeit. Für umfassende Analysen mit verschiedenen Ordnungen, Modellstrukturen und Hyperparameteroptimierung können mehrere Wochen reine Rechenzeit notwendig sein. Gegebenenfalls kann diese Rechenzeit auf leistungsfähigerer Hardware beschleunigt werden. Die andauernde Weiterentwicklung in der Grundlagenforschung künstlicher neuronaler Netze kann dabei ebenso potenziell in Zukunft durch effizientere Algorithmen Abhilfe schaffen.

Für die Regelung mittels linearer Zustandsregler wurde der linear quadratische Ansatz, unter anderem aufgrund seiner anschaulichen Funktionsweise, ausgewählt. Für diesen müssen bestimmte Gewichtungsmatrizen gewählt werden, deren Wahl sich für gewisse Konfigurationen des erweiterten Zustandsvektors im Koopman-Raum als schwierig bis unmöglich erweist. Um diesem Problem zu begegnen, wurde eine Möglichkeit vorgestellt, die im anschaulichen Zustandsraum gewählten Gewichte in die Basis des Koopman-linearen Systems zu transformieren. Hiermit ist es außerdem möglich, gezielt Einfluss auf bestimmte Eigenwerte zu nehmen. Sämtliche Identifikations- und Regelungsansätze wurden außerdem für den Fall untersucht, dass nicht alle Zustände messbar sind.

Die erarbeiteten Grundlagen sowie Identifikationsansätze wurden dann anhand von zwei theoretischen Modellen sowie einem realen Prozess evaluiert. Die Ergebnisse des Van der Pol Systems sind schlüssig und die Regelung des instabilen Arbeitspunktes erfolgt mit einer besseren Energie-

ausnutzung, als es mit einem Regler auf Basis der Linearisierung der Systemgleichung der Fall ist. Für das zweite theoretische Beispiel, den Duffing Oszillator, sind die Ergebnisse gemischt. Eine konkrete Verbesserung der Regelung durch einen DKLc-Regler gegenüber einem Regler auf Basis der Linearisierung kann nicht erreicht werden. Die Regelung des Drei-Tanks, welcher einen realen Prüfstand darstellt, erzielt ebenso die praktisch gleichen Ergebnisse hinsichtlich Qualität und Energienutzung wie ein Zustandsregler auf Basis der Linearisierung des Systemmodells beziehungsweise einer Identifikation eines linearen Modells. Dies liegt jedoch maßgeblich daran, dass das Systemverhalten bereits äußerst gut linear beschrieben werden kann, wie die angegebene DMD-Schätzung zeigt.

Die reine Prädiktion sämtlicher DKLc-Modelle liefert jedoch in vielen Fällen ein deutlich besseres Ergebnis als andere Identifikationsansätze. Insbesondere bei den theoretischen Beispielen versagen klassische Identifikationsansätze, wohingegen der DKLc-Ansatz valide Modelle hervorbringt. Dies liegt im Wesentlichen daran, dass mit linearen Ansätzen keine nichtlinearen Dynamiken dargestellt werden können.

Des Weiteren zeigt sich bei den in der Literatur häufig verwendeten DMD basierten Methoden zur Schätzung von Koopman-linearen Approximationen mit zunehmender Systemordnung keine Konvergenz, aufgrund der schlecht wählbaren Observables. Für die in dieser Arbeit als Referenz herangezogenen Monome sowie trigonometrischen Ansatzfunktionen als Observables für eine DMD-Schätzung verschlechtert sich das Schätzergebnis im Falle des Van der Pol Oszillators mit der Vergrößerung der linearen Systemordnung sogar. Im Gegensatz dazu stellt sich durch die simultane Optimierung der Observables und der Koopman-Matrix durch den DKLc-Ansatz zumindest für die trainierten Prädiktionshorizonte bei allen beschriebenen Beispielen eine Konvergenz ab einer systemabhängigen Mindestordnung ein.

Für die Regelung ist zusammenfassend zu sagen, dass der Koopman-Ansatz rein auf der Identifikation anhand von Messdaten beruht. Die Linearisierung, welche bessere Ergebnisse in zwei der drei Beispielen erreichen konnte, beruht auf einem ganzheitlichen Systemwissen, was in vielen Fällen in der praktischen Anwendung nicht in diesem Maße vorliegt. Daher stellt der vorgestellte Ansatz und die gesamte Kette an Verfahren, welche sich zu einem gesamten Framework verbinden, eine geeignete Möglichkeit dar, Modelle zu nichtlinearen Systemen zu schätzen und diese mit den ermittelten Modellen mittels linearer Zustandsregler zu regeln. Dieses Framework ist in Abbildung 4.5 übersichtlich als Strukturbild zusammengefasst.

## Ausblick

Durch die Komplexität des Koopman-Frameworks sowie der vorgestellten Methode lässt sich eine Vielzahl von weiteren Forschungsschwerpunkten ableiten.

Die Erweiterung des Eingangsraumes kann für eine verbesserte Prädiktion des Zustands verwendet werden, wie in [18] gezeigt wird. Für die Regelung ist diese Erweiterung jedoch zunächst nicht direkt übertragbar, da sich so ein überbestimmtes Gleichungssystem für das Stellsignal er-

gibt. Hier kann weiter untersucht werden, ob eine Erweiterung unter Berücksichtigung dieser Überbestimmtheit zu einem besseren Regelergebnis führen kann.

Des Weiteren kann die DKLc-Methode auf reine Eigenfunktionen erweitert werden. So kann beispielsweise auch die reduzierte Steuerbarkeit der  $n$  von  $n_e$  Eigenwerten direkt berücksichtigt werden und in die Schätzung mit einfließen.

Durch die starke Empfindlichkeit der DKLc-Methode gegenüber Messrauschen kann dies, je nach Möglichkeit der Filterung der Daten, neue Wege der Forschung eröffnen. Hierbei ist zu untersuchen, in wie fern die DKLc-Methode direkt beeinflusst werden kann, um besser mit Rauschen umzugehen. Die Empfindlichkeit von Deep-Learning Methoden gegenüber Rauschen ist auch in andern Disziplinen, wie beispielsweise der Bilderkennung, bekannt und Gegenstand aktueller Forschung [45, 127]. Gegebenenfalls können hier Analogien gefunden und zur Verbesserung der DKLc-Methode eingesetzt werden.

Durch die Dynamisierung der Koopman-Regler in Form von Time-Delay-Koordinaten ergeben sich Möglichkeiten weiterer Forschungsschwerpunkte. Hier lassen sich gegebenenfalls ebenfalls Vorteile gegenüber klassischen dynamischen Reglern, wie beispielsweise einem PID-Regler, aufzeigen.

# Anhang

## A.1 Herleitungen

### A.1.1 Herleitung der eigenwertabhängigen Gleichung zur Zeitlösung eines linearen Zustandsraumsystems

Ein lineares autonomes Zustandsraumsystem nach

$$\dot{\underline{x}} = \underline{A}\underline{x}$$

besitzt die Zeitlösung

$$\underline{x}(t) = \exp(\underline{A}t)\underline{x}_0.$$

Durch die Erweiterung  $\underline{V}\underline{W}^T = \underline{I}_n$  mit  $\underline{V}$  als Rechts- und  $\underline{W}$  als Linkseigenvektormatrix der Matrix  $\underline{A}$  kann obige Gleichung zu

$$\begin{aligned}\underline{x}(t) &= \exp(\underline{V}\underline{W}\underline{A}\underline{V}\underline{W}t)\underline{x}_0 \\ &= \underline{V}\exp(\underline{\Lambda}t)\underline{W}\underline{x}_0\end{aligned}$$

umgeformt werden. Die Matrix  $\exp(\underline{\Lambda}t)$  hat dabei lediglich Diagonaleinträge der Form  $e^{\lambda_i t}$  mit den Eigenwerten  $\lambda_i$  der Matrix  $\underline{A}$ , da  $\underline{\Lambda}$  eine Diagonalmatrix ist. Werden nun lediglich die Spalten von  $\underline{V}$  beziehungsweise Zeilen von  $\underline{W}$  betrachtet, so erhält man

$$\begin{aligned}\underline{x}(t) &= [\underline{v}_1 \quad \underline{v}_2 \quad \cdots \quad \underline{v}_n] \begin{bmatrix} e^{\lambda_1 t} & 0 & 0 & \cdots & 0 \\ 0 & e^{\lambda_2 t} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & e^{\lambda_n t} \end{bmatrix} \begin{bmatrix} \underline{w}_1^T \\ \underline{w}_2^T \\ \vdots \\ \underline{w}_n^T \end{bmatrix} \underline{x}_0 \\ &= \sum_{i=1}^n \underline{v}_i e^{\lambda_i t} \underline{w}_i^T \underline{x}_0 \\ &= \sum_{i=1}^n \tilde{x}_{0,i} \underline{v}_i e^{\lambda_i t}.\end{aligned}\tag{A.1}$$

Die Größe  $\tilde{x}_{0,i}$  besitzt einen Akzent als Tilde um klarzumachen, dass es sich hierbei nicht um den *direkten* Anfangszustand, sondern um den modalen Anfangszustand zum  $i$ -ten Eigenwert, handelt.

Für ein zeitdiskretes Zustandsraumsystem existiert die analoge Lösung

$$\underline{x}_k = \sum_{i=1}^n \tilde{x}_{0,i} \underline{v}_i \lambda_i^k. \quad (\text{A.2})$$

Diese ergibt sich aus der Identität, dass die Eigenwerte eines zeitdiskreten Systems über die Eigenwerte des äquivalenten zeitkontinuierlichen Systems nach

$$\lambda_d = e^{\lambda T_s}$$

berechnet werden können. Dabei ist  $T_s$  die Abtastzeit zum zeitdiskreten Zustandsraumsystem. Da des Weiteren  $t = T_s k$  gilt, ergibt sich Gleichung (A.2) mit dem Eigenwert in  $k$ -ter Potenz

### A.1.2 Stabilität von nichtlinearen Systemen und ihren exakten Koopman-linearen Äquivalenten

Das folgende Lemma zeigt, dass die Stabilität der Koopman-Matrix einer exakten Koopman-linearen Darstellung mit exakter Umkehrfunktion  $\underline{g}_z(\underline{z}) = \underline{x}$  einer Ruhelage eines nichtlinearen Systems hinreichend die Stabilität des nichtlinearen Systems sicherstellt.

**Lemma A.1.** Gegeben ist eine  $C^1$  differenzierbare, bijektive, kontinuierliche Transformation  $\underline{z} = \underline{g}_z : \mathcal{X} \rightarrow \mathcal{Z}$  mit  $\mathcal{X} \subset \mathbb{R}^n$  und  $\mathcal{Z} \subset \mathbb{R}^{n_e}$  mit  $n_e \geq n$ . Weiter sei  $\underline{0}$  ein innerer Punkt von  $\mathcal{Z}$  und es gilt  $\underline{g}_z(\underline{0}) = \underline{0}$ . Existiert eine Lyapunov-Funktion für eine Ruhelage mit Umgebung  $\mathcal{U}$  eines nichtlinearen Zustandsraumsystems  $\dot{\underline{x}} = \underline{f}(\underline{x})$ , für die die Ruhelage stabil im Sinne von Lyapunov ist, so besitzt das dazugehörige Koopman-lineare System  $\dot{\underline{z}} = \underline{K}_z \underline{z}$  nur Eigenwerte  $\lambda_i$  mit  $Re\{\lambda_i\} \leq 0$ .

*Beweis.* Die dritte Bedingung der Lyapunov Stabilitätstheorie ([2, S. 97]) zu einem autonomen nichtlinearen System lautet

$$\dot{V}_{\underline{x}} = \nabla_{\underline{x}} V_{\underline{x}} \underline{f} \leq 0 \quad \forall \underline{x} \in \mathcal{X}_r$$

mit einer Lyapunov-Funktion  $V_{\underline{x}}(\underline{x})$ . Diese kann zu

$$\dot{V}_{\underline{x}}(\underline{x}) = \nabla_{\underline{z}} V_{\underline{x}} \left( \underline{x} = \underline{g}_z(\underline{z}) \right) \underline{K}_z \underline{z} \leq 0$$

umformuliert werden. Für das Koopman-lineare System gilt die Lyapunov Gleichung

$$\dot{V}_z(z) = \nabla_z V_z(z) \underline{K}_z z \leq 0.$$

Mit der Wahl  $V_z(z) := V_x(\bar{g}_z(z))$  erhält man also eine Lyapunov-Funktion, die beiden Gleichungen entspricht. Somit gilt, dass wenn das nichtlineare System stabil im Sinne von Lyapunov mit einer Lyapunov-Funktion  $V_x(x)$  ist, auch das lineare System stabil in Sinne von Lyapunov mit der Lyapunov-Funktion  $V_z(z) = V_x(\bar{g}_z(z))$  ist. Da jede Systemmatrix eines linearen Systems, welches stabil im Sinne von Lyapunov ist, nur Eigenwerte  $\lambda_i$  mit  $Re\{\lambda_i\} \leq 0$  besitzt ([2]), hat die Matrix  $\underline{K}_z$  auch nur Eigenwerte, für die das gleiche gilt.  $\square$

Mit dergleichen Definition  $V_z(z_k) := V_x(\bar{g}_z(z_k))$  gilt dasselbe für zeitdiskrete Systeme. Denn wenn die betroffene Ruhelage des nichtlinearen Systems stabil im Sinne von Lyapunov mit der Lyapunov  $V_x(x_k)$  ist, dann gilt  $V_x(x_{k+1}) - V_x(x_k) < 0$ . Mit obiger Definition gilt dies somit auch für das Koopman-lineare System und die Koopman-Matrix besitzt nur Eigenwerte  $\lambda$  mit  $|\lambda| \leq 1$ .

### A.1.3 Herleitung der zeitdiskreten Gleichung des Koopman-Linearen Systems auf Basis von nichtlinearen Systemen mit Eingangsnichtlinearitäten

Für Gleichung (3.32) wird zunächst das zeitdiskrete Pendant

$$\underline{x}_{k+1} = \bar{f}_{\underline{x}, \underline{u}}(\underline{x}_k, \underline{u}_k) + \sum_{i=1}^{m_x} \bar{f}_{\underline{u}, i}(\underline{u}_k) \bar{f}_{\underline{x}, i}(\underline{x}_k) = \bar{f}_d(\underline{x}_k, \underline{u}_k) \quad (\text{A.3})$$

geschrieben. Die darin auftretenden Funktionen werden mit einem Strich notiert, um sie von denen aus Gleichung (3.32) zu unterscheiden. Des Weiteren gilt für Gleichung (3.32) die Substitution  $\dot{\underline{x}} = \underline{f}_c$ . Mit dem Vorwärtsdifferenzenquotient mit Schrittweite 1 gilt mit den Systemgleichungen (3.32) und (A.3)

$$\dot{\underline{x}} \approx \frac{\underline{x}_{k+1} - \underline{x}_k}{h} = \frac{\bar{f}_d(\underline{x}_k, \underline{u}_k) - \underline{x}_k}{1} \approx \underline{f}_c.$$

Die zeitliche Ableitung der Observables wird ebenso mit dem Vorwärtsdifferenzenquotient angenähert, sodass sich

$$\begin{aligned} \dot{\underline{g}}_z &\approx \frac{\underline{g}_z(\underline{x}_{k+1}, \underline{u}_{k+1}) - \underline{g}_z(\underline{x}_k, \underline{u}_k)}{h} \approx \nabla_{\underline{x}} \underline{g}_z \underline{f}_c \\ \rightarrow \underline{g}_z(\underline{x}_{k+1}, \underline{u}_{k+1}) &\approx \underline{g}_z(\underline{x}_k, \underline{u}_k) + h \nabla_{\underline{x}} \underline{g}_z \underline{f}_c \\ &\approx \underline{g}_z(\underline{x}_k, \underline{u}_k) + \nabla_{\underline{x}} \underline{g}_z (\bar{f}_d(\underline{x}_k, \underline{u}_k) - \underline{x}_k) \end{aligned}$$

ergibt. Dabei wurde der Term mit  $\dot{\underline{u}}$ , wie bei Gleichung (3.32) durch die Annahme 3.4, vernachlässigt. Darin wird Gleichung A.3 eingesetzt und es ergibt sich

$$\begin{aligned}\underline{g}_z(\underline{x}_{k+1}, \underline{u}_{k+1}) &= \underline{g}_z(\underline{x}_k, \underline{u}_k) + \nabla_{\underline{x}} \underline{g}_z \left( \tilde{f}_{\underline{x}, \underline{u}}(\underline{x}_k, \underline{u}_k) + \sum_{i=1}^{m_x} \tilde{f}_{\underline{u}, i}(\underline{u}_k) \tilde{f}_{\underline{x}, i}(\underline{x}_k) - \underline{x}_k \right) \\ &= \underline{g}_z(\underline{x}_k, \underline{u}_k) + \nabla_{\underline{x}} \underline{g}_z \left( \tilde{f}_{\underline{x}, \underline{u}}(\underline{x}_k, \underline{u}_k) + \sum_{i=1}^{m_x} \tilde{f}_{\underline{u}, i}(\underline{u}_k) \tilde{f}_{\underline{x}, i}(\underline{x}_k) \right) \\ &= \underline{g}_z(\underline{x}_k, \underline{u}_k) + L_{\tilde{f}_{\underline{x}, \underline{u}}, \underline{x}} \underline{g}_z + \sum_{i=1}^{m_x} \tilde{f}_{\underline{u}, i} L_{\tilde{f}_{\underline{x}, i}, \underline{x}} \underline{g}_z.\end{aligned}$$

Wie beim zeitkontinuierlichen Fall werden die Lie-Ableitungen über Matrizen approximiert, so dass sich

$$\begin{aligned}\underline{g}_z(\underline{x}_{k+1}, \underline{u}_{k+1}) &= \underline{g}_z(\underline{x}_k, \underline{u}_k) + \tilde{\underline{K}}_z \underline{g}_z + \sum_{i=1}^{m_x} \tilde{f}_{\underline{u}, i} \tilde{\underline{K}}_{w, i} \underline{g}_z \\ &= \underbrace{(\underline{I} + \tilde{\underline{K}}_z)}_{\tilde{\underline{K}}_z} \underline{g}_z + \sum_{i=1}^{m_x} \tilde{f}_{\underline{u}, i} \tilde{\underline{K}}_{w, i} \underline{g}_z.\end{aligned}$$

ergibt. Diese Gleichung besitzt dieselbe Struktur, wie im zeitkontinuierlichen Fall. Mit der Wahl der Basis aus (3.34) ergibt sich somit das zeitdiskrete Koopman-System aus Gleichung (3.36). Im Verlauf der Arbeit wird der Übersicht halber der Strich-Akzent über den zeitdiskreten Größen weggelassen. Dieser Akzent dient nur der Unterscheidung zwischen zeitkontinuierlich und zeitdiskret für die obige Herleitung.

#### A.1.4 Herleitung des Gradienten von Observables für DKLc

Durch die spezielle Struktur der Gleichungen der Observables bei einer DKLc-Schätzung können diese symbolisch in Abhängigkeit der Gewichte des Netzes angegeben werden. Die Observables werden über

$$\underline{g}(\underline{x}) = \underline{W}_{d-1} \underline{f}(\underline{W}_{d-2} \underline{f}(\cdots \underline{f}(\underline{W}_0 \underline{x} + \underline{b}_0))) + \underline{b}_{d-2}) + \underline{b}_{d-1}$$

angegeben. Für die folgende Herleitung wird die Abkürzung  $\underline{y}_i = \underline{W}_i \underline{f}(\underline{y}_{i-1}) + \underline{b}_i$  verwendet. Für die Ableitung gilt

$$\begin{aligned} \nabla_{\underline{x}} \underline{g}(\underline{x}) &= \frac{\partial \underline{g}(\underline{x})}{\partial \underline{x}} = \frac{\partial \underline{g}(\underline{x})}{\partial \underline{y}_{d-2}} \frac{\partial \underline{y}_{d-2}}{\partial \underline{x}} = \frac{\partial \underline{g}(\underline{x})}{\partial \underline{f}(\underline{y}_{d-2})} \frac{\partial \underline{f}(\underline{y}_{d-2})}{\partial \underline{y}_{d-2}} \frac{\partial \underline{y}_{d-2}}{\partial \underline{x}} \\ &= \underline{W}_d \frac{\partial \underline{f}(\underline{y}_{d-2})}{\partial \underline{y}_{d-2}} \frac{\partial \underline{y}_{d-2}}{\partial \underline{x}}. \end{aligned} \quad (\text{A.4})$$

Für die Aktivierungsfunktion in der  $i$ -ten Schicht gilt

$$\underline{f}(\underline{y}_{i-1}) = \begin{bmatrix} f(y_{i-1,1}) \\ \vdots \\ f(y_{i-1,w_{i-1}}) \end{bmatrix}$$

mit der Breite  $w_{i-1}$  dieser Schicht. Die Ableitung dieser Funktion ist demnach eine Diagonalmatrix der Form

$$\underline{T}_{i-1} = \frac{\partial \underline{f}(\underline{y}_{i-1})}{\partial \underline{y}_{i-1}} = \text{diag}([f'(y_{i-1,1}) \quad \cdots \quad f'(y_{i-1,w_{i-1}})]).$$

Die Ableitungskette aus Gleichung (A.4) kann weitergeführt werden, sodass sich für den Gradient der Observables die Bestimmungsgleichung

$$\nabla_{\underline{x}} \underline{g}(\underline{x}) = \underline{W}_{d-1} \prod_{i=2}^d \underline{T}_{d-i} \underline{W}_{d-i}$$

ergibt.

### A.1.5 Exaktes Minimierungsmaß der Netzprädiktion

Zur Formulierung des Optimierungsfunktionals aus Gleichung (4.1) werden auch die Prädiktionen durch das Netz und durch das erweiterte Zustandsraumsystem aus der linearen Schicht des Netzes verwendet. Die Prädiktion ist durch die Eingänge  $\underline{u}_i$  über den gesamten Prädiktionshorizont und den Anfangszustand zu Beginn des Prädiktionshorizontes  $\underline{x}_1$  gegeben.

$$J_{\text{pred},\underline{x}} = \frac{1}{p} \sum_{i=1}^p \|\underline{x}_{i+1} - \hat{\underline{x}}_{i+1}(\underline{x}_1, \underline{u}_1, \dots, \underline{u}_i)\|_2$$

Die Funktion  $\hat{\underline{x}}_{i+1}(\underline{x}_1, \underline{u}_1, \dots, \underline{u}_i)$  ist eine iterative Funktionskomposition, welche durch

$$\begin{aligned}\underline{x}_i &= \underline{g}_z \circ \underline{z}_i \\ \underline{z}_i &= \underline{K}_z \underline{g}_z \circ \underline{x}_{i-1} + \underline{K}_w \underline{g}_w \circ \underline{u}_{i-1}\end{aligned}$$

beschrieben wird. Um die Ausdrücke in Gleichung (4.1) zu erhalten, wird angenommen, dass  $\underline{g}_z(\underline{g}_z(\underline{x})) = \underline{x}$  gilt. Diese Eigenschaft ist ebenfalls Teil des Optimierungsfunktionals (4.1).

### A.1.6 LMI Formulierung der zeitdiskreten Lyapunov-Gleichung

Der Nachweis erfolgt über eine zeitdiskrete Version des Kalman-Yakubovich-Popov Lemmas [38, 85]. Die folgenden beiden Aussagen sind für rein reelle Matrizen eines linearen Zustandsraum-systems ( $\underline{A}$ ,  $\underline{B}$ ,  $\underline{C}$ ,  $\underline{D}$ ) äquivalent:

- (i)  $\underline{A}$  ist Schur (für alle Eigenwerte gilt  $|\lambda| < 1$ )
- (ii) Es existiert eine Diagonalmatrix  $\underline{Q} \succ 0$  sodass

$$\begin{bmatrix} \underline{Q} & \underline{0} \\ \underline{0} & \underline{I} \end{bmatrix} - \begin{bmatrix} \underline{A} & \underline{B} \\ \underline{C} & \underline{D} \end{bmatrix}^T \begin{bmatrix} \underline{Q} & \underline{0} \\ \underline{0} & \underline{I} \end{bmatrix} \begin{bmatrix} \underline{A} & \underline{B} \\ \underline{C} & \underline{D} \end{bmatrix} \succ 0.$$

Mit  $\underline{A} = \underline{K}_z$ ,  $\underline{B} = \underline{I}$ ,  $\underline{C} = \gamma \underline{I}$  und  $\underline{D} = \underline{0}$  ergibt sich nach ausmultiplizieren obiger Matrixungleichung

$$\begin{bmatrix} \underline{Q} - \underline{K}_z^T \underline{Q} \underline{K}_z - \gamma^2 \underline{I} & -\underline{K}_z^T \underline{Q} \\ -\underline{Q} \underline{K}_z & \underline{I} - \underline{Q} \end{bmatrix} \succ 0.$$

Der Ausgang über  $\underline{C}$  ist damit ein Hilfsausgang, um den Fehler mit einem Parameter zu verknüpfen. Mit dem Schur-Komplement ist diese Matrixungleichung äquivalent zu

$$\underline{Q} - \underline{K}_z^T \underline{Q} \underline{K}_z - \gamma^2 \underline{I} - \underline{K}_z^T \underline{Q} (\underline{I} - \underline{Q})^{-1} \underline{Q} \underline{K}_z \succ 0.$$

Damit gilt die Einschränkung, dass  $\underline{Q}$  keine Eigenwerte in 1 haben darf. Wenn  $\underline{Q}$  die Lösung einer Lyapunov-Gleichung ist, ist die letzte Ungleichung die Nullmatrix ([78]) und es gilt die Ersetzung im Beweis von Satz 3.3.

### A.1.7 Existenz eines Integrators bei mittelwertfreier Anregung

Es wird ein lineares Zustandsraumsystem

$$(\underline{x}_{k+1} + \bar{\underline{x}}) = \underline{A}(\underline{x}_k + \bar{\underline{x}}) + \underline{B}u_k$$

mit den mittelwertfreien Signalen  $\underline{x}_k \in \mathbb{R}^n$  und  $\underline{u}_k \in \mathbb{R}^m$ , sowie einem Gleichanteil  $\bar{\underline{x}} \neq \underline{0}$  zugrunde gelegt. Die Anregung erfolgt also über einen mittelwertfreien Zufallsprozess. Mit dem Erwartungswert der obigen Gleichung ergibt sich

$$\begin{aligned} E \{ \underline{x}_{k+1} + \bar{\underline{x}} \} &= E \{ \underline{A} (\underline{x}_k + \bar{\underline{x}}) + \underline{B} \underline{u}_k \} \\ E \{ \underline{x}_{k+1} \} + E \{ \bar{\underline{x}} \} &= \underline{A} E \{ \underline{x}_k \} + \underline{A} E \{ \bar{\underline{x}} \} + \underline{B} E \{ \underline{u}_k \} \\ E \{ \bar{\underline{x}} \} &= \underline{A} E \{ \bar{\underline{x}} \} \\ \bar{\underline{x}} &= \underline{A} \bar{\underline{x}}. \end{aligned}$$

Aus der letzten Gleichung lässt sich

$$(\underline{I} - \underline{A}) \bar{\underline{x}} = \underline{0}$$

ableiten. Da  $\bar{\underline{x}} \neq \underline{0}$  gilt, muss  $\underline{A}$  daher mindestens einen Eigenwert in 1 besitzen, welcher für den zeitdiskreten Ansatz einem Integrator entspricht.

### A.1.8 Nachweis der schwachen Beobachtbarkeit des Drei-Tank Prüfstandes

Für den Nachweis der schwachen Beobachtbarkeit wird diese zunächst definiert.

**Definition A.2** (Schwache Beobachtbarkeit angeregter, nichtlinearer Systeme [2, S. 534]). Das nichtlineare System nach (3.26) mit  $\mathcal{X} \subseteq \mathbb{R}^n$  und  $\mathcal{C}_{\underline{u}} \subseteq \mathcal{C}^{n-1}$ , wobei  $\mathcal{C}^{n-1}$  der Raum der  $n - 1$  fach ableitbaren (Eingangs-)Funktionen ist, heißt schwach beobachtbar, wenn für alle  $\underline{x} \in \mathcal{X}$  und  $\underline{u} \in \mathcal{C}_{\underline{u}}$  die Bedingung

$$\text{rang} (\underline{Q}_w) = \text{rang} \begin{bmatrix} \nabla_{\underline{x}} h_0 (\underline{x}, \underline{u}) \\ \nabla_{\underline{x}} h_1 (\underline{x}, \underline{u}, \dot{\underline{u}}) \\ \vdots \\ \nabla_{\underline{x}} h_{n-1} (\underline{x}, \underline{u}, \dot{\underline{u}}, \dots, \underline{u}^{(n-1)}) \end{bmatrix} = n \quad (\text{A.5})$$

erfüllt ist. Dabei gilt für die Hilfsgrößen  $h_i$

$$h_i = \nabla_{\underline{x}} h_{i-1} \underline{f} (\underline{x}, \underline{u}) + \sum_{j=1}^i \nabla_{\underline{u}^{(j-1)}} h_{i-1} \underline{u}^{(j)}.$$

Des Weiteren ist  $h_0 = y$  die untersuchte Ausgangsgröße.

Für die Messung der einzelnen Zustände des Drei-Tank-Systems aus Gleichung (5.3) ergeben sich die folgenden Determinanten der Matrix  $\underline{Q}_w$  aus Gleichung (A.5).

$$\det(\underline{O}_{W,x_1}) = \frac{\sqrt{2g^3q_1^2q_3}}{4A^3(x_1-x_2)\sqrt{x_3-x_2}}$$

$$\det(\underline{O}_{W,x_2}) = \frac{\sqrt{2g^2q_1q_3}}{8A^3\sqrt{gx_2}\sqrt{g(x_1-x_2)^3}\sqrt{g(x_3-x_2)^3}} \left( 2gq_2x_1x_2 - 2gq_2x_2x_3 \right. \\ \left. + \sqrt{2}u_1x_2\sqrt{gx_2} + \sqrt{2}u_2x_1\sqrt{gx_2} - \sqrt{2}u_1x_3\sqrt{gx_2} - \sqrt{2}u_2x_2\sqrt{gx_2} \right. \\ \left. - 2q_1x_1\sqrt{gx_2}\sqrt{g(x_1-x_2)} + 2q_1x_3\sqrt{gx_2}\sqrt{g(x_1-x_2)} \right. \\ \left. - 2q_3x_1\sqrt{gx_2}\sqrt{g(x_3-x_2)} + 2q_3x_3\sqrt{gx_2}\sqrt{g(x_3-x_2)} \right)$$

$$\det(\underline{O}_{W,x_3}) = \frac{\sqrt{2g^3q_1q_3^2}}{4A^3(x_3-x_2)\sqrt{x_1-x_2}}$$

Es ist deutlich erkennbar, dass die Determinante für die Ausgänge  $x_1$  und  $x_3$  für gleiche Füllstände von  $x_1$  und  $x_2$  beziehungsweise  $x_3$  und  $x_2$  nicht definiert ist. Dieser Betriebspunkt ist allerdings ohnehin in der gegebenen Konfiguration nicht erreichbar. Die Interpretation für den Ausgang  $x_2$  ist praktisch nicht möglich. Daher wird der Verlauf der Identifikationsdaten in diese Gleichung eingesetzt. Es zeigt sich, dass für diese Daten die Determinante niemals null wird. Für die Analyse sind alle Parameter, Zustände und Eingänge auf Millimeter umgerechnet worden. Somit befinden sich sämtliche Größen in einer Größenordnung von etwa 1 bis 1.000. Die Determinante zum Ausgang  $x_2$  besitzt ihr Minimum im Bereich von  $10^{-3}$ . Die Minima der Determinanten zu den Ausgängen  $x_1$  und  $x_3$  besitzen Größenordnungen von etwa  $10^2$ . Die schwache Beobachtbarkeit ist somit über den Ausgang  $x_2$  verhältnismäßig schlecht, jedoch nicht null. Aus der Anschauung heraus ist das zunächst kaum intuitiv, da insbesondere der Einfluss der äußeren Tanks auf den jeweils anderen äußeren Tank kaum erkennbar ist.

Für die Identifikation und Analyse in Abschnitt 5.2 ist diese Gegebenheit, sowie der Fakt zu beachten, dass sich der reale Prüfstand vom theoretischen Modell unterscheidet. Daher könnte es dazu kommen, dass bei den gegebenen Verläufen eine Beobachtbarkeit in Teilen nicht gegeben ist.

## A.2 Hardware Aufbau des Drei-Tank Prüfstandes

Der Drei-Tank wird über zwei 4-Kammer-Druckwassermembranpumpen der Firma Schurfloh mit dem notwendigen Volumenstrom durch einen Zufluss an der Oberseite der äußeren Tanks versorgt. Diese werden über eine Steuerspannung  $U_C$  je Pumpe zwischen  $[-1, 1]$  V angesteuert. Die Kennlinien der Pumpen ist in Abbildung A.1 dargestellt. Die Kennlinien sind als lineare Funktionen in Gleichung (A.6) aufgetragen.

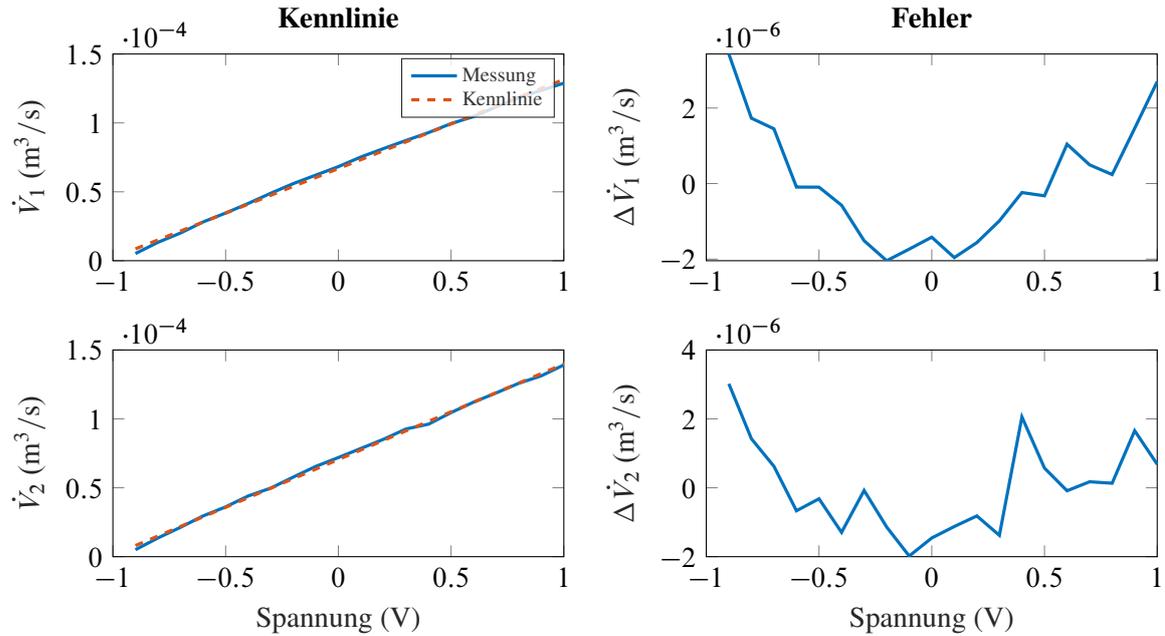


Abbildung A.1: Kennlinie der beiden Schurfloh 4-Kammer-Druckwassermembranpumpen

$$\begin{aligned} \dot{V}_1(U_{C,1}) &= \left( 0,6479 \frac{1}{V} \cdot U_{C,1} + 0,6685 \right) \cdot 10^{-4} \frac{\text{m}^3}{\text{s}} \\ \dot{V}_2(U_{C,2}) &= \left( 0,6934 \frac{1}{V} \cdot U_{C,2} + 0,7049 \right) \cdot 10^{-4} \frac{\text{m}^3}{\text{s}} \end{aligned} \quad (\text{A.6})$$

Der Füllstand der Tanks wird über Drucksensoren ermittelt. Bei diesen handelt es sich um VEGABAR 14 Prozessdruckumformer der Firma VEGA. Die Kennlinien der Sensoren für den betriebsrelevanten Bereich in Abhängigkeit einer Sensorspannung  $U_S$  je Sensor sind in Abbildung A.2 gezeigt. Die Kennlinien sind als lineare Funktionen in Gleichung (A.7) dargestellt.

$$\begin{aligned} h_1(U_{S,1}) &= \left( -0,3357 \frac{1}{V} \cdot U_{S,1} + 0,3032 \right) \text{ m} \\ h_2(U_{S,2}) &= \left( -0,3335 \frac{1}{V} \cdot U_{S,2} + 0,2943 \right) \text{ m} \\ h_3(U_{S,3}) &= \left( -0,3322 \frac{1}{V} \cdot U_{S,3} + 0,2891 \right) \text{ m} \end{aligned} \quad (\text{A.7})$$

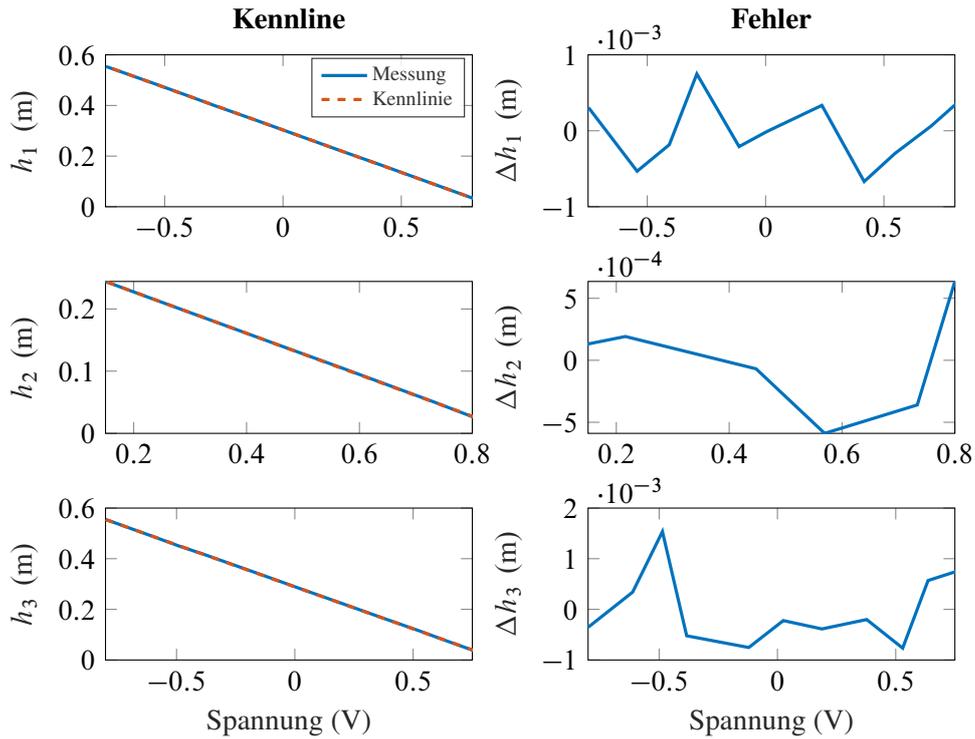


Abbildung A.2: Kennlinie der drei verwendeten VEGABAR 14 Drucksensoren

Die Ansteuerung der Pumpen, sowie das Auslesen der Sensoren geschieht über eine dSpace<sup>®</sup> ds1104 PCIe Karte mit der dazugehörigen Software dSpace<sup>®</sup> ControlDesk 7.2.

Tabelle A.1 zeigt des Weiteren die Parameter des Drei-Tank Modells, welche mit Abflussmessungen hinsichtlich der Systemmodells (5.3) ermittelt werden können.

Tabelle A.1: Parameter des Drei-Tank Prüfstandes hinsichtlich des Systemmodells aus (5.3)

| Parameter | Wert                   |
|-----------|------------------------|
| $A$       | 153,94 cm <sup>2</sup> |
| $q_1$     | 22,23 mm <sup>2</sup>  |
| $q_2$     | 42,45 mm <sup>2</sup>  |
| $q_3$     | 22,18 mm <sup>2</sup>  |

## A.3 Weiterführende Plots und Modellkonfigurationen

### A.3.1 Amplituden Verlauf der dominanten Frequenzen des mathematischen Pendels

Ergänzend zum Beispiel des mathematischen Pendels zur Erläuterung eines kontinuierlichen Spektrums in Abschnitt 2.2.3 zeigt Abbildung A.3 die Amplituden der ersten beiden dominanten Frequenzanteile. Insbesondere im Bereich um  $x_{1,0} = 0$  wird die Schwingung von der ersten Frequenz dominiert, sodass eine nahezu reine Sinusschwingung bei einer Frequenz vorliegt.

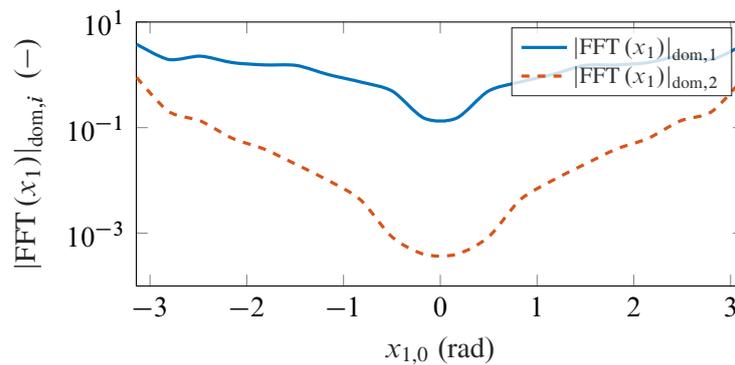


Abbildung A.3: Visualisierung der Amplituden der ersten beiden dominanten Frequenzen der FFT des Winkels des mathematischen Pendels in Abhängigkeit des Anfangszustandes bei  $x_{2,0} = 0$

### A.3.2 Batch-Size und Bias-Varianz-Analyse von DKLc

Die Konfiguration der Netze und des Trainings, dargestellt in Tabelle A.2, ergibt sich aus einer Hyperparameteroptimierung, welche in [5] durchgeführt wurde. Die Schichtbreiten sind im Encoder [41, 62] zu und im Decoder zu [51, 37] gewählt.

Tabelle A.2: Parameter des Netzes unter Verwendung von Modellstruktur S3 (beschrieben in Tabelle 5.1), sowie der Trainingsparameter nach (4.1)

| Parameter  | Wert | Parameter  | Wert | Parameter  | Wert              |
|------------|------|------------|------|------------|-------------------|
| $\alpha_1$ | 1    | $\delta_1$ | 0,99 | $\gamma_1$ | $1 \cdot 10^{-6}$ |
| $\alpha_2$ | 0,2  | $\delta_2$ | 0,99 | $\gamma_2$ | $1 \cdot 10^{-6}$ |
| $\alpha_3$ | 1    | $\beta_1$  | 0,01 | h          | $1 \cdot 10^{-4}$ |
| $\alpha_4$ | 1    | $\beta_2$  | 0    |            |                   |
| $\alpha_5$ | 0,1  | $\beta_3$  | 0    |            |                   |

### A.3.3 Van der Pol Oszillator

Tabelle A.3 gibt die Netzkonfigurationen der DKLC-Netze an, mit denen der in Abschnitt 5.1.1 besprochene Van der Pol Oszillator gelernt wird. Die Schichtbreiten sind im Encoder zu [20, 40, 60] und im Decoder zu [60, 40, 20] gewählt. Ab einer Ordnung von  $n_e = 9$  wird die Schicht mit 60 Neuronen auf 100 erhöht. Für die Ordnung  $n_e = 20$  wird diese erneut auf 120 angehoben. Die Anhebung wird vorgenommen, um der erhöhten Menge an Observables mehr Variationsmöglichkeiten zu geben. Wird eine Transformation des Eingangs vorgenommen, so werden diese mit im Encoder mit [10, 20] und im Decoder mit [20, 10] Neuronen kodiert.

Tabelle A.3: Hyperparameter für die Optimierung des Van der Pol Systems

| Parameter  | Wert | Parameter  | Wert | Parameter  | Wert              |
|------------|------|------------|------|------------|-------------------|
| $\alpha_1$ | 1    | $\delta_1$ | 0,99 | $\gamma_1$ | $1 \cdot 10^{-6}$ |
| $\alpha_2$ | 1    | $\delta_2$ | 0,99 | $\gamma_2$ | $1 \cdot 10^{-6}$ |
| $\alpha_3$ | 1    | $\beta_1$  | 0,01 | h          | $1 \cdot 10^{-4}$ |
| $\alpha_4$ | 1    | $\beta_2$  | 0    |            |                   |
| $\alpha_5$ | 0,1  | $\beta_3$  | 0    |            |                   |

### A.3.4 Duffing Oszillator

Tabelle A.4 gibt die Netzkonfigurationen der DKLC-Netze an, mit denen der in Abschnitt 5.1.2 besprochene Duffing Oszillator gelernt wird. Die Schichtbreiten sind im Encoder zu [20, 40, 60] und im Decoder zu [60, 40, 20] gewählt. Ab einer Ordnung von  $n_e = 9$  wird die Schicht mit 60 Neuronen auf 100 erhöht. Für die Ordnung  $n_e = 20$  wird diese erneut auf 120 angehoben. Die Anhebung wird vorgenommen, um der erhöhten Menge an Observables mehr Variationsmöglichkeiten zu geben. Wird eine Transformation des Eingangs vorgenommen, so werden diese mit im Encoder mit [10, 20] und im Decoder mit [20, 10] Neuronen kodiert.

Tabelle A.4: Hyperparameter für die Optimierung des Duffing Oszillators

| Parameter  | Wert | Parameter  | Wert | Parameter  | Wert              |
|------------|------|------------|------|------------|-------------------|
| $\alpha_1$ | 1    | $\delta_1$ | 0,99 | $\gamma_1$ | $1 \cdot 10^{-7}$ |
| $\alpha_2$ | 3    | $\delta_2$ | 0,99 | $\gamma_2$ | $1 \cdot 10^{-7}$ |
| $\alpha_3$ | 1    | $\beta_1$  | 0,01 | h          | $1 \cdot 10^{-4}$ |
| $\alpha_4$ | 0,7  | $\beta_2$  | 0    |            |                   |
| $\alpha_5$ | 0,5  | $\beta_3$  | 0    |            |                   |

Die folgenden Darstellungen zeigen Fehlerhistogramme mittels einer DMD- ( $n_e = n + 1 = 3$ ) sowie randomisierten eDMD-Schätzung ( $n_e = 4$ ).

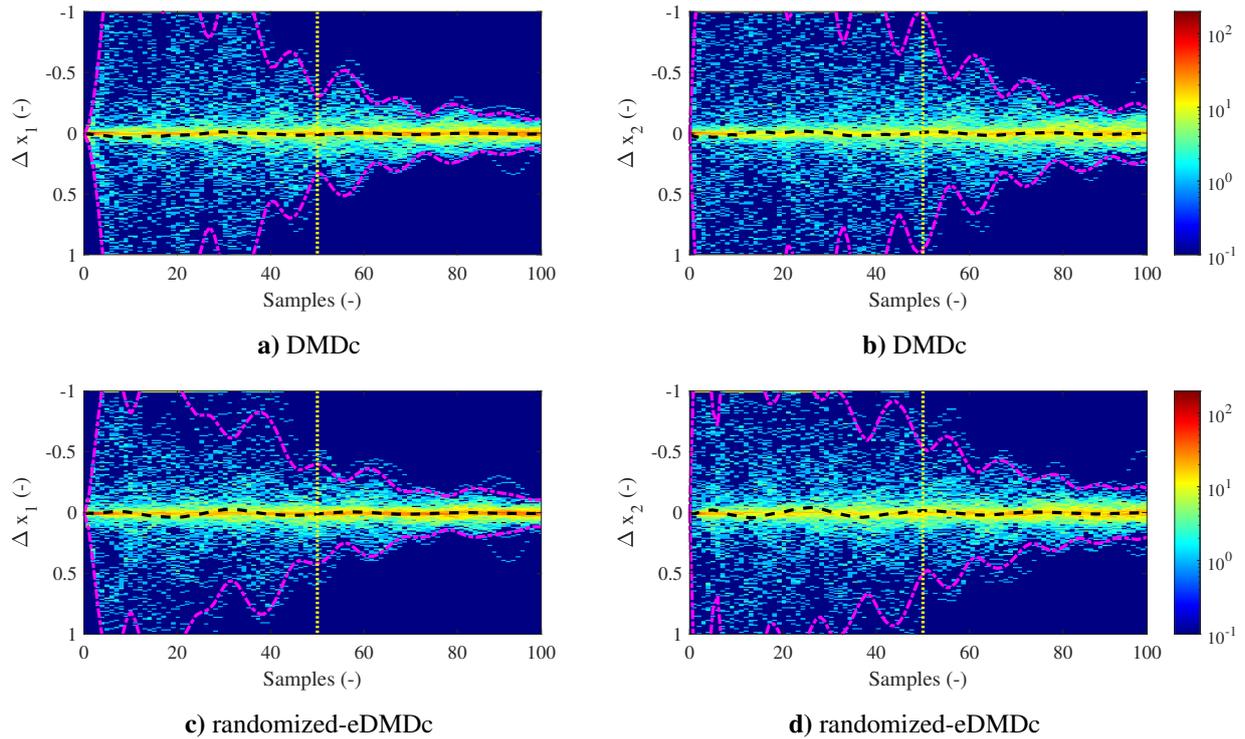


Abbildung A.4: Fehler Histogramme der DMD ( $n_e = n + 1 = 3$ ) und eDMD ( $n_e = 4$ ) des Duffing Oszillators

### A.3.5 Drei-Tank

Tabelle A.5 gibt die Netzkonfigurationen der DKLc-Netze an, mit denen der in Abschnitt 5.2 besprochene Drei-Tank gelernt wird. Die Schichtbreiten sind im Encoder zu [40, 80] und im Decoder zu [80, 40] gewählt. Die Verkleinerung der Netzarchitektur gegenüber den anderen Beispielen ist dem Umstand geschuldet, dass die Menge an Rechenoperationen eine gewisse Grenze nicht überschreiten darf, um auf dem verwendeten dSpace<sup>®</sup> Mikrocontroller zur Applikation am Prüfstand in Echtzeit lauffähig zu sein. Wird eine Transformation des Eingangs vorgenommen, so werden diese mit im Encoder mit [10, 20] und im Decoder mit [20, 10] Neuronen kodiert.

Tabelle A.5: Hyperparameter für die Optimierung des Drei-Tank Systems

| Parameter  | Wert | Parameter  | Wert | Parameter  | Wert              |
|------------|------|------------|------|------------|-------------------|
| $\alpha_1$ | 0,8  | $\delta_1$ | 0,99 | $\gamma_1$ | $1 \cdot 10^{-7}$ |
| $\alpha_2$ | 1    | $\delta_2$ | 0,99 | $\gamma_2$ | $1 \cdot 10^{-6}$ |
| $\alpha_3$ | 1    | $\beta_1$  | 0,01 | h          | $1 \cdot 10^{-4}$ |
| $\alpha_4$ | 0,7  | $\beta_2$  | 0    |            |                   |
| $\alpha_5$ | 0,2  | $\beta_3$  | 0    |            |                   |

Zur besseren Veranschaulichung der Schätzqualität des ARX-Modells sind im Folgenden die Fehlerhistogramme zu den einzelnen Zuständen angegeben. Die angegebene Ordnung von  $n_e = 6$  be-

zieht sich hier auf die Ordnung des Nennerpolynoms der Übertragungsfunktionen und steht nicht um Zusammenhang mit einem erweiterten Zustandsraum im Sinne der Observables.

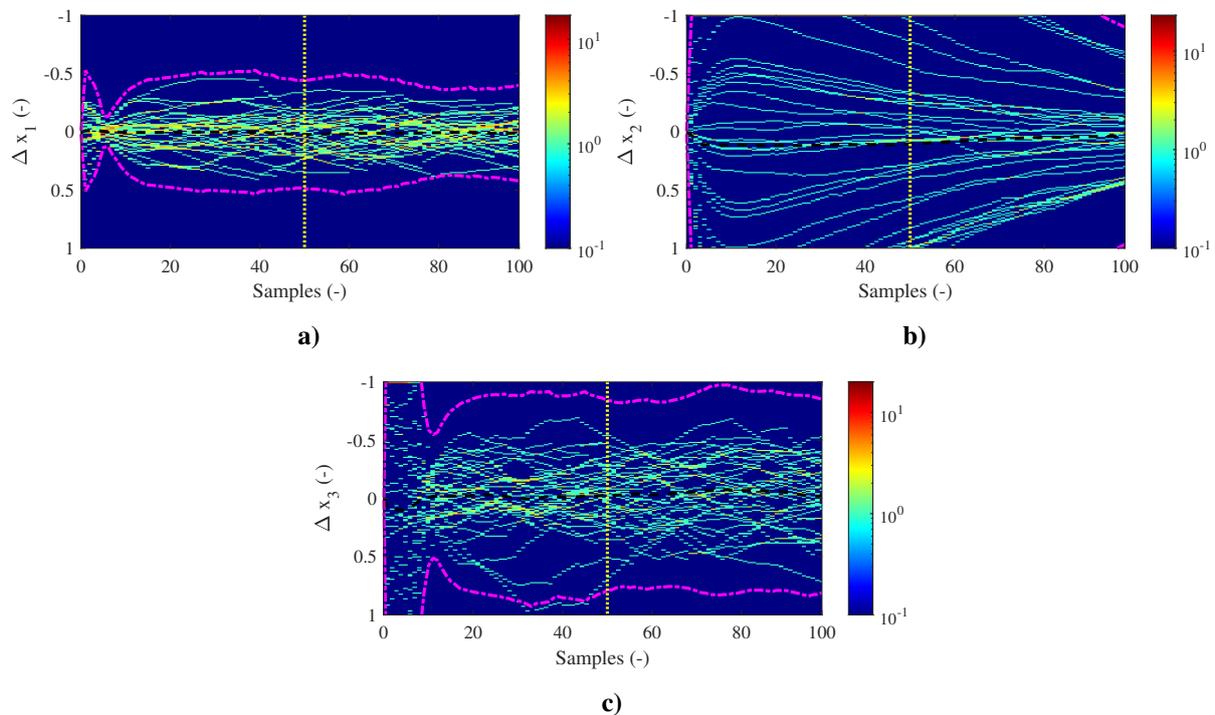


Abbildung A.5: Fehlerhistogramme eines MIMO-ARX-Modells bei  $n_e = 6$

Die Darstellung der übrigen Fehlerhistogramme zu Abbildung 5.34 sind in Abbildung A.6 dargestellt.

Die nachfolgende Darstellung zeigt den Verlauf Fehlers zwischen gemessener und geschätzter Zeitreihe über verschiedene Beispieltrajektorien dar.

Die folgende Darstellung zeigt den Fehler zwischen einer approximierten Ebene und der tatsächlichen Eingangscoderfunktion

Die nachfolgende Darstellung zeigt die Steuerbarkeitsmaße für eine beispielhafte Ordnung von  $n_e = 7$  bei S4 unter Verwendung von Time-Delay-Koordinaten.

Die im Folgenden dargestellten Verläufe der Varianz in Abbildung A.10 zeigen die Erhöhung selbiger bei der Schätzung mittels Time-Delay-Koordinaten auf. Ein Grund hierfür kann schlicht die fehlende Messgröße  $x_2$  sein, sodass weniger Information zur Ermittlung des Füllstandes vorhanden ist. Gegenüber den anderen Fehlerhistogrammen ist hierbei jedoch die veränderte Skalierung der Ordinate zu beachten, sodass der Fehler zwar größer wird, die gesamte Schätzqualität jedoch ähnlich gut bleibt. Selbst mit dieser größeren Varianz weist die Schätzung nach wie vor eine geringere Streuung als die der DMD-Schätzung aus Abbildung A.7 auf.

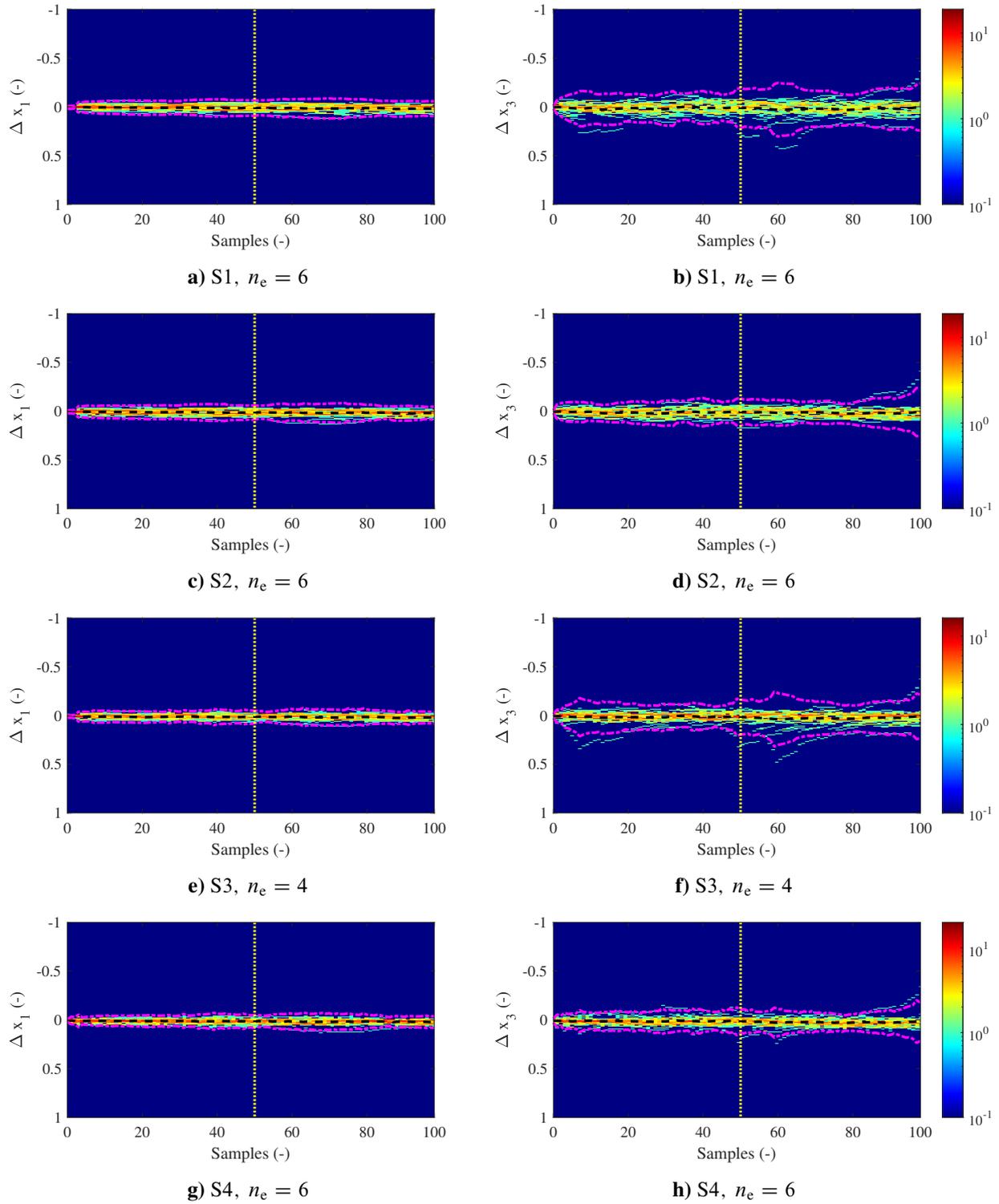


Abbildung A.6: Fehler Histogramme der besten Modelle des Drei-Tank Systems für verschiedene Modellkonfigurationen

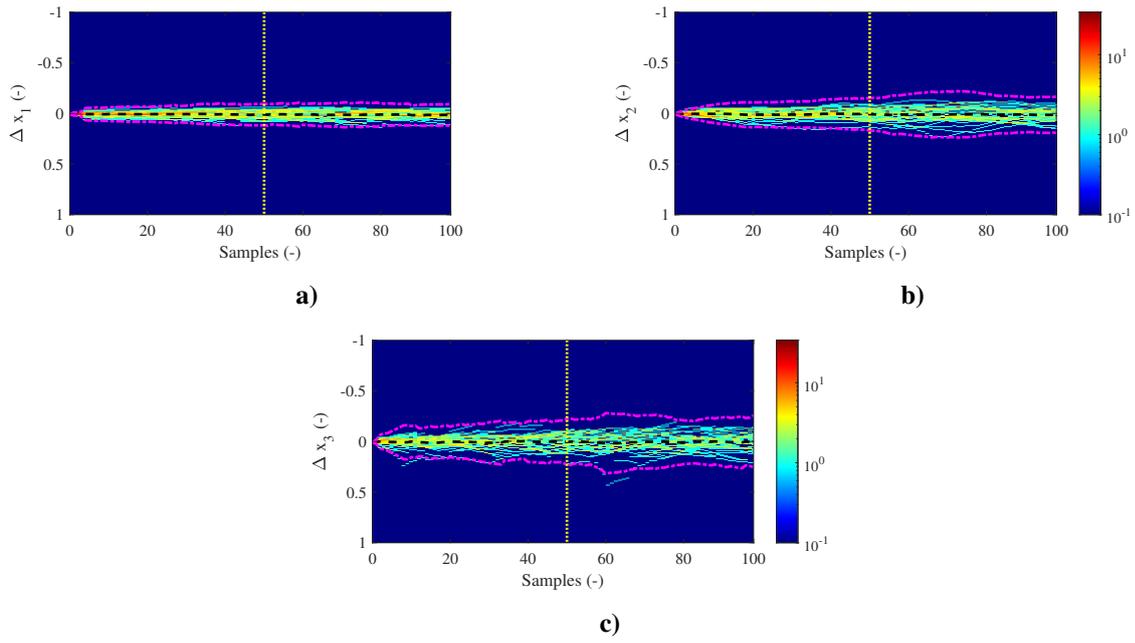


Abbildung A.7: DMD-Schätzung des Dreitanks ohne Observables ( $n_e = n = 3$ )

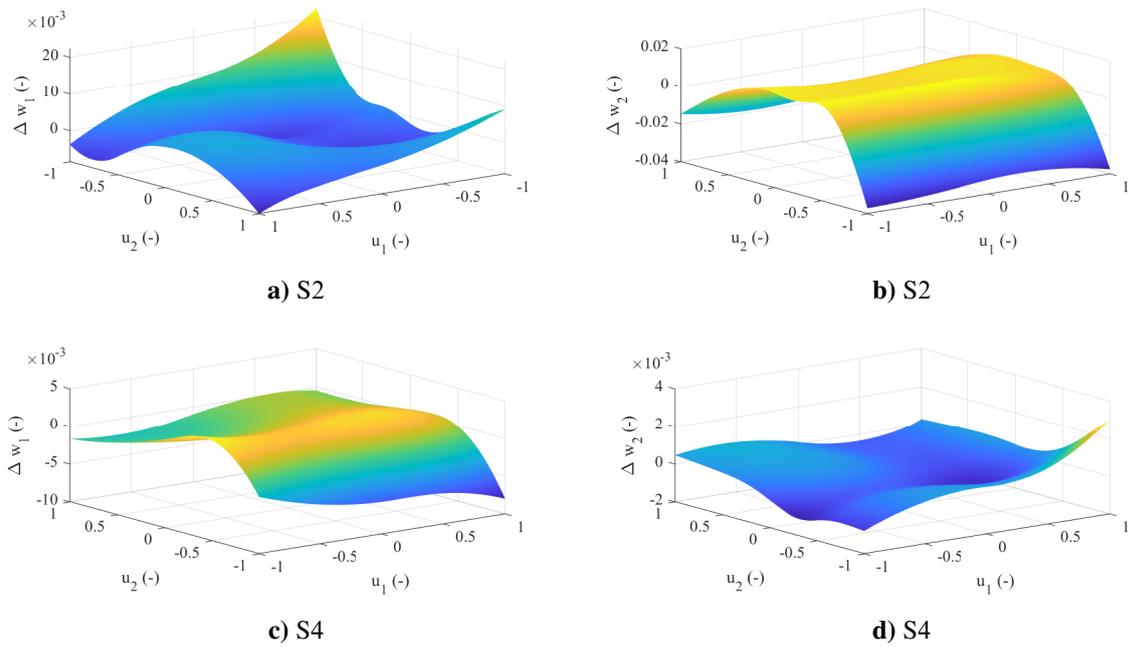


Abbildung A.8: Absoluter Fehler  $\Delta w_i = w_i - \tilde{w}_i$  zwischen der tatsächlichen Eingangscodierfunktion  $w_i$  und einer angenäherten Ebene  $\tilde{w}_i$  für  $n_e = 6$

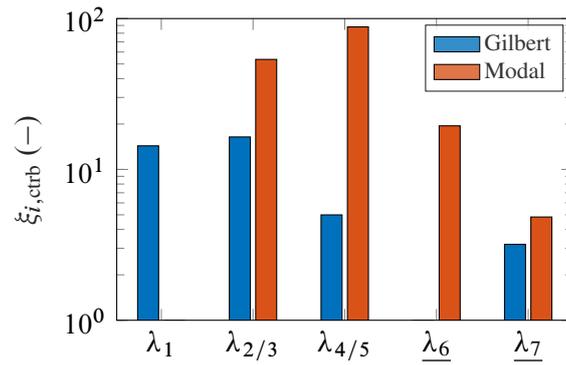


Abbildung A.9: Steuerbarkeitsmaße  $\xi$  eines mittels Time-Delay-Koordinaten geschätzten Dreitanke Systems bei  $n_e = 7$  der einzelnen Eigenwerte, normiert auf das kleinste Maß bei Modellstruktur S4. Eigenwerte aufsteigend nach Realteil sortiert. Die zur Linearisierung passenden Eigenwerte sind unterstrichen dargestellt

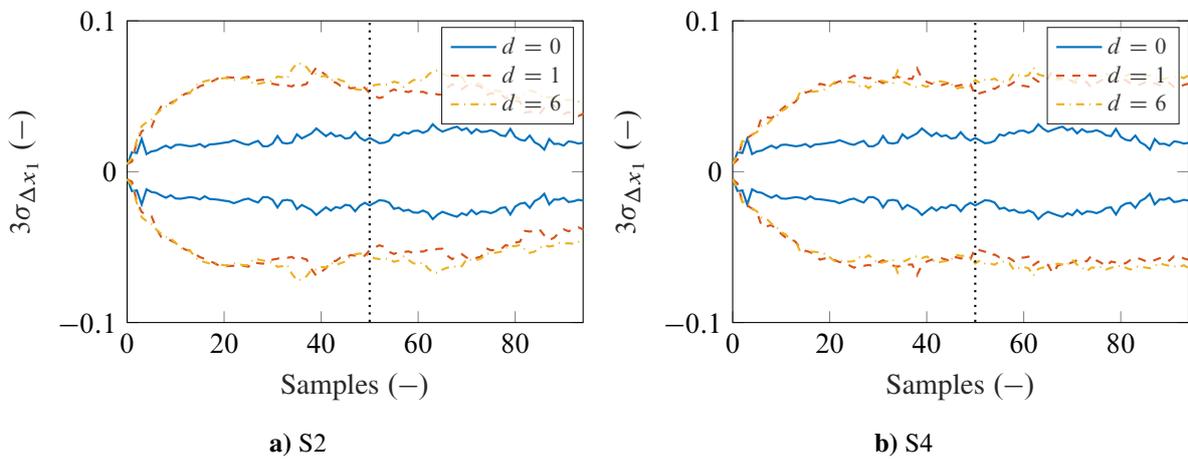


Abbildung A.10: Varianz der DKLc-Schätzung nach Einführung von Time-Delay-Koordinaten bei Messung des Zustandes  $x_1$

# Literaturverzeichnis

- [1] Ian Abraham, Gerardo de La Torre und Todd Murphey. „Model-Based Control Using Koopman Operators“. In: *Robotics: Science and Systems XIII*. Robotics: Science and Systems Foundation, July 12-16, 2017.
- [2] Jürgen Adamy. *Nichtlineare Systeme und Regelungen*. 3., aktualisierte Auflage. Berlin: Springer Vieweg, 2018.
- [3] Charu C. Aggarwal. *Neural Networks and Deep Learning*. Cham: Springer International Publishing, 2018.
- [4] Luis Antonio Aguirre und Christophe Letellier. „Observability of multivariate differential embeddings“. In: *Journal of Physics A: Mathematical and General* 38.28 (2005), S. 6311–6326.
- [5] Alexander Stephan Gräfe. „Analyse und Entwicklung von Deep Learning Methoden zur Schätzung von invarianten Unterräumen des Koopman Operators nichtlinearer angeregter dynamischer Systeme“. Master-Thesis. Darmstadt: Technische Universität Darmstadt, 2020.
- [6] Herbert Amann und Joachim Escher. *Analysis II*. Zweite, korrigierte Auflage. Grundstudium Mathematik. Basel: Birkhäuser Verlag, 2006.
- [7] Martin Anthony und Michele Harvey. *Linear Algebra: Concepts and Methods*. Cambridge: Cambridge University Press, 2012.
- [8] Tilo Arens u. a. *Mathematik*. Springer Berlin Heidelberg, 2018.
- [9] Karl-Johan Åström und Torsten Bohlin. „Numerical Identification of Linear Dynamic Systems from Normal Operating Records“. In: *IFAC Proceedings Volumes* 2.2 (1965), S. 96–111.
- [10] Karl-Johan. Åström und Pieter Eykhoff. „System identification—A survey“. In: *Automatica* 7.2 (1971), S. 123–162.
- [11] Karl-Johan. Åström, Per Hagander und Jan Sternby. „Zeros of sampled systems“. In: *Automatica* 20.1 (1984), S. 31–38.
- [12] Pierre Baldi und Kurt Hornik. „Neural networks and principal component analysis: Learning from examples without local minima“. In: *Neural Networks* 2.1 (1989), S. 53–58.
- [13] Giuseppe Basile und Giovanni Marro. *Controlled and conditioned invariants in linear system theory*. Englewood Cliffs, NJ: Prentice Hall, 1992.
- [14] Dietmar Bauer. „Order estimation for subspace methods“. In: *Automatica* 37.10 (2001), S. 1561–1573.

- [15] Tal Ben-Nun und Torsten Hoeffler. „Demystifying Parallel and Distributed Deep Learning“. In: *ACM Computing Surveys* 52.4 (2019), S. 1–43.
- [16] Mouhacine Benosman, Hassan Mansour und Vahan Huroyan. „Koopman-operator Observer-based Estimation of Pedestrian Crowd Flows“. In: *IFAC-PapersOnLine* 50.1 (2017), S. 14028–14033.
- [17] Marcel Bonnert. „Subspace-Identifikation dynamischer Systeme: Explizite Berücksichtigung von Vorwissen und Bewertung von Messdaten“. Master-Thesis. Darmstadt: Technische Universität Darmstadt, 2016.
- [18] Marcel Bonnert und Ulrich Konigorski. „Estimating Koopman Invariant Subspaces of Excited Systems Using Artificial Neural Networks“. In: *IFAC-PapersOnLine* 53.2 (2020), S. 1156–1162.
- [19] Santos D. M. Borjas und Claudio Garcia. „Subspace Identification for Industrial Processes“. In: *TEMA - Tendências em Matemática Aplicada e Computacional* 12.3 (2011), S. 183–194.
- [20] Hervé. Bourlard und Yves Kamp. „Auto-association by multilayer perceptrons and singular value decomposition“. In: *Biological cybernetics* 59.4-5 (1988), S. 291–294.
- [21] Luitzen E. J. Brouwer. „Beweis der Invarianz der Dimensionenzahl“. In: *Mathematische Annalen* 70.2 (1911), S. 161–165.
- [22] Adam L. Bruce, Vera M. Zeidan und Dennis S. Bernstein. „What Is the Koopman Operator? A Simplified Treatment for Discrete-Time Systems“. In: *2019 American Control Conference (ACC)*. IEEE, 10.07.2019 - 12.07.2019, S. 1912–1917.
- [23] Steven L. Brunton und J. Nathan Kutz. *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge, United Kingdom und New York, NY: Cambridge University Press, 2019.
- [24] Steven L. Brunton, Joshua L. Proctor und J. Nathan Kutz. „Discovering governing equations from data by sparse identification of nonlinear dynamical systems“. In: *Proceedings of the National Academy of Sciences of the United States of America* 113.15 (2016), S. 3932–3937.
- [25] Steven L. Brunton, Joshua L. Proctor und J. Nathan Kutz. „Sparse Identification of Nonlinear Dynamics with Control (SINDYc)“. In: *IFAC-PapersOnLine* 49.18 (2016), S. 710–715.
- [26] Steven L. Brunton u. a. „Chaos as an intermittently forced linear system“. In: *Nature communications* 8 (2017).
- [27] Steven L. Brunton u. a. „Koopman Invariant Subspaces and Finite Linear Representations of Nonlinear Dynamical Systems for Control“. In: *PloS one* 11 (2016), e0150171.
- [28] Marko Budisić, Ryan Mohr und Igor Mezić. „Applied Koopmanism“. In: *Chaos (Woodbury, N.Y.)* 22.4 (2012), S. 047510.
- [29] Kenneth P. Burnham und David R. Anderson. *Model Selection and Multimodel Inference*. New York, NY: Springer New York, 2004.

- [30] Torsten Carleman. „Application de la théorie des équations intégrales linéaires aux systèmes d'équations différentielles non linéaires“. In: *Acta Mathematica* 59.0 (1932), S. 63–87.
- [31] Xiangyi Chen u. a. „On the Convergence of A Class of Adam-Type Algorithms for Non-Convex Optimization“. In: *International Conference on Learning Representations*. 2019.
- [32] Alessandro Chiuso. „Asymptotic Variance of Closed-Loop Subspace Identification Methods“. In: *IEEE Transactions on Automatic Control* 51.8 (2006), S. 1299–1314.
- [33] Alessandro Chiuso. „The Role of Vector AutoRegressive Modeling in Subspace Identification“. In: *Proceedings of the 45th IEEE Conference on Decision and Control*, S. 3646–3651.
- [34] François Chollet. *Deep learning with Python*. Safari Tech Books Online. Shelter Island, NY: Manning, 2018.
- [35] Anna Choromanska u. a. „The Loss Surfaces of Multilayer Networks“. In: *AISTATS*. 2015.
- [36] Bryan C. Daniels und Ilya Nemenman. „Efficient inference of parsimonious phenomenological models of cellular dynamics using S-systems and alternating regression“. In: *PloS one* 10.3 (2015), e0119821.
- [37] John Duchi, Elad Hazan und Yoram Singer. „Adaptive Subgradient Methods for Online Learning and Stochastic Optimization“. In: *Journal of Machine Learning Research* 12.61 (2011), S. 2121–2159.
- [38] Geir E. Dullerud u. a. *A Course in Robust Control Theory*. Bd. 36. New York, NY: Springer New York, 2000.
- [39] Bryan Eisenhower u. a. „Decomposing Building Systems Data for Model Validation and Analysis Using the Koopman Operator“. In: *Proceedings of SimBuild*. Bd. 4, S. 434–441.
- [40] Pieter Eykhoff. „Identification in measurement and instrumentation“. In: *Concise Encyclopedia of Measurement and Instrumentation*. Hrsg. von L. Finkelstein und K.T.V. Grattam. Pergamon, 1994, S. 137–142.
- [41] Pieter Eykhoff. *System identification: Parameter and state estimation*. Wiley-Interscience publication. London: Wiley, 1974.
- [42] M. Sami Fadali und Antonio Visioli. „Properties of State–Space Models“. In: *Digital Control Engineering*. Elsevier, 2009, S. 281–333.
- [43] Jean Gallier. *Logarithms and Square Roots of Real Matrices*. 2008. arXiv: 0805.0245v2.
- [44] Elmer G. Gilbert. „Controllability and Observability in Multivariable Control Systems“. In: *Journal of the Society for Industrial and Applied Mathematics Series A Control* 1.2 (1963), S. 128–151.
- [45] Ian J. Goodfellow, Jonathon Shlens und Christian Szegedy. *Explaining and Harnessing Adversarial Examples*. 2014. arXiv: 1412.6572v3.

- [46] Priya Goyal u. a. *Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour*. 2017. arXiv: 1706.02677v2.
- [47] Yiqiang Han, Wenjian Hao und Umesh Vaidya. *Deep Learning of Koopman Representation for Control*. 2020. arXiv: 2010.07546v1.
- [48] Kurt Hornik. „Approximation capabilities of multilayer feedforward networks“. In: *Neural Networks* 4.2 (1991), S. 251–257.
- [49] Rolf Isermann. *Identifikation dynamischer Systeme 1*. Springer Berlin Heidelberg, 1992.
- [50] Rolf Isermann. *Identifikation dynamischer Systeme 2*. Springer Berlin Heidelberg, 1992.
- [51] Rolf Isermann. *Mechatronische Systeme*. Springer Berlin Heidelberg, 2008.
- [52] Rolf Isermann und Marco Münchhof. *Identification of dynamic systems: An introduction with applications*. Berlin [u.a.]: Springer, 2011.
- [53] Eurika Kaiser, J. Nathan Kutz und Steven L. Brunton. *Data-driven discovery of Koopman eigenfunctions for control*. 2018. arXiv: 1707.01146v2.
- [54] Eurika Kaiser, J. Nathan Kutz und Steven L. Brunton. „Sparse identification of nonlinear dynamics for model predictive control in the low-data limit“. In: *Proceedings. Mathematical, physical, and engineering sciences* 474.2219 (2018), S. 20180335.
- [55] Mason Kamb u. a. „Time-Delay Observables for Koopman: Theory and Applications“. In: *SIAM Journal on Applied Dynamical Systems* 19.2 (2020), S. 886–917.
- [56] Nitish Shirish Keskar u. a. *On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima*. 2016. arXiv: 1609.04836v2.
- [57] Bernard O. Koopman und John von Neumann. „Dynamical Systems of Continuous Spectra“. In: *Proceedings of the National Academy of Sciences* 18.3 (1932), S. 255–263.
- [58] Bernard. O. Koopman. „Hamiltonian Systems and Transformation in Hilbert Space“. In: *Proceedings of the National Academy of Sciences* 5 (1931), S. 315–318.
- [59] Milan Korda und Igor Mezić. „Optimal construction of Koopman eigenfunctions for prediction and control“. In: *IEEE Transactions on Automatic Control* 65 (2020), S. 5114–5129.
- [60] Milan Korda und Igor Mezić. „Koopman Model Predictive Control of Nonlinear Dynamical Systems“. In: *The Koopman Operator in Systems and Control*. Hrsg. von Alexandre Mauroy, Igor Mezić und Yoshihiko Susuki. Bd. 484. Lecture Notes in Control and Information Sciences. Cham: Springer International Publishing, 2020, S. 235–255.
- [61] Milan Korda und Igor Mezić. „Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control“. In: *Automatica* 93 (2018), S. 149–160.
- [62] Carlos S. Kubrusly. *Spectral Theory of Operators on Hilbert Spaces*. Birkhäuser Boston, 2012.

- [63] Qianxiao Li u. a. „Extended dynamic mode decomposition with dictionary learning: A data-driven adaptive spectral decomposition of the Koopman operator“. In: *Chaos (Woodbury, N.Y.)* 27.1 (2017), S. 103111.
- [64] Yunzhu Li u. a. *Learning Compositional Koopman Operators for Model-Based Control*. 2019.
- [65] Lennart Ljung. *System identification: Theory for the user*. Upper Saddle River, NJ: Prentice-Hall, 1999.
- [66] Boris Lohmann. „Order reduction and determination of dominant state variables of nonlinear systems“. In: *Mathematical Modelling of Systems* 1 (1995), S. 77–90.
- [67] Jan Lunze. *Regelungstechnik 1*. 12. Auflage. Berlin, Heidelberg: Springer Berlin Heidelberg, 2020.
- [68] Jan Lunze. *Regelungstechnik 2*. 10. Auflage. Berlin, Heidelberg: Springer Berlin Heidelberg, 2020.
- [69] Bethany Lusch, J. Nathan Kutz und Steven L. Brunton. „Deep learning for universal linear embeddings of nonlinear dynamics“. In: *Nature communications* 9 (2018).
- [70] Xu Ma, Bowen Huang und Umesh Vaidya. „Optimal Quadratic Regulation of Nonlinear System Using Koopman Operator“. In: *2019 American Control Conference (ACC)*. IEEE, 10.07.2019 - 12.07.2019, S. 4911–4916.
- [71] Giorgos Mamakoukas u. a. *Learning Data-Driven Stable Koopman Operators*. 2020. arXiv: 2005.04291v1.
- [72] Niall M. Mangan u. a. „Model selection for hybrid dynamical systems via sparse regression“. In: *Proceedings. Mathematical, physical, and engineering sciences* 475.2223 (2019), S. 20180534.
- [73] Andreas Mardt u. a. „VAMPnets for deep learning of molecular kinetics“. In: *Nature communications* 9 (2018).
- [74] J. E. Marsden u. a. *Manifolds, Tensor Analysis, and Applications*. Bd. 75. New York, NY: Springer New York, 1988.
- [75] Martin Abadi u. a. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. 2015.
- [76] Daniele Masti und Alberto Bemporad. „Learning Nonlinear State-Space Models Using Deep Autoencoders“. In: *2018 IEEE Conference on Decision and Control (CDC)*. IEEE, 17.12.2018 - 19.12.2018, S. 3862–3867.
- [77] Alexandre Mauroy und Igor Mezic. „Global Stability Analysis Using the Eigenfunctions of the Koopman Operator“. In: *IEEE Transactions on Automatic Control* 61.11 (2016), S. 3356–3369.
- [78] Alexandre Mauroy, Igor Mezic und Yoshihiko Susuki. *The Koopman operator in systems and control: Concepts, methodologies and applications*. Lecture Notes in Control and Information Sciences. 2020.

- [79] Igor Mezić. „Spectral Properties of Dynamical Systems, Model Reduction and Decompositions“. In: *Nonlinear Dynamics* 41.1-3 (2005), S. 309–325.
- [80] Igor Mezić. „Spectrum of the Koopman Operator, Spectral Expansions in Functional Spaces, and State-Space Geometry“. In: *Journal of Nonlinear Science* (2019).
- [81] Igor Mezić und Andrzej Banaszuk. „Comparison of systems with complex behavior“. In: *Physica D: Nonlinear Phenomena* 197.1-2 (2004), S. 101–133.
- [82] Francis C. Moon. *Chaotic Vibrations*. Wiley, 2004.
- [83] Jeremy Morton u. a. *Deep Dynamical Modeling and Control of Unsteady Fluid Flows*. 2018. arXiv: 1805.07472v2.
- [84] Peter C. Müller und Joachim Lückel. „Modale Maße für Steuerbarkeit, Beobachtbarkeit und Störbarkeit dynamischer Systeme“. In: *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik* 54.12 (1974), S. 57–58.
- [85] Federico Najson. „On the Kalman–Yakubovich–Popov lemma for discrete-time positive linear systems: a novel simple proof and some related results“. In: *International Journal of Control* 86 (2013), S. 1813–1823.
- [86] Oliver Nelles. „LOLIMOT - Lokale, lineare Modelle zur Identifikation nichtlinearer, dynamischer Systeme“. In: *at - Automatisierungstechnik* 45.4 (1997).
- [87] Oliver Nelles. *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*. Berlin und Heidelberg: Springer, 2001.
- [88] Oliver Nelles, Alexander Fink und Rolf Isermann. „Local Linear Model Trees (LOLIMOT) Toolbox for Nonlinear System Identification“. In: *IFAC Proceedings Volumes* 33.15 (2000), S. 845–850.
- [89] Oliver Nelles und Rolf Isermann. „Basis function networks for interpolation of local linear models“. In: *Proceedings of 35th IEEE Conference on Decision and Control*. IEEE, 11-13 Dec. 1996, S. 470–475.
- [90] Samuel E. Otto und Clarence W. Rowley. „Linearly Recurrent Autoencoder Networks for Learning Dynamics“. In: *SIAM Journal on Applied Dynamical Systems* 18.1 (2019), S. 558–593.
- [91] Elad Plaut. *From Principal Subspaces to Principal Components with Linear Autoencoders*. 2018. arXiv: 1804.10253v3.
- [92] Joshua L. Proctor, Steven L. Brunton und J. Nathan Kutz. „Dynamic Mode Decomposition with Control“. In: *SIAM Journal on Applied Dynamical Systems* 15 (2016), S. 142–161.
- [93] Joshua L. Proctor, Steven L. Brunton und J. Nathan Kutz. „Generalizing Koopman Theory to Allow for Inputs and Control“. In: *SIAM Journal on Applied Dynamical Systems* 1 (2018), S. 909–930.
- [94] S. Joe Qin. „An overview of subspace identification“. In: *Computers & Chemical Engineering* 30 (2006), S. 1502–1513.

- [95] Markus Quade u. a. „Sparse identification of nonlinear dynamics for rapid model recovery“. In: *Chaos (Woodbury, N.Y.)* 28.6 (2018), S. 063116.
- [96] Clarence W. Rowley u. a. „Reduced-order models for flow control: balanced models and Koopman modes“. In: *Seventh IUTAM Symposium on Laminar-Turbulent Transition*. Hrsg. von PHILIPP SCHLATTER und Dan S. Henningson. Bd. 18. IUTAM Bookseries. Dordrecht: Springer Netherlands, 2010, S. 43–50.
- [97] Clarence W. Rowley u. a. „Spectral analysis of nonlinear flows“. In: *Journal of Fluid Mechanics* 641 (2009), S. 115–127.
- [98] Sebastian Ruder. *An overview of gradient descent optimization algorithms*. 2017. arXiv: 1609.04747v2.
- [99] Amirali Sadeqi, Shapour Moradi und Kourosch Heidari Shirazi. „Nonlinear subspace system identification based on output-only measurements“. In: *Journal of the Franklin Institute* 357 (2020), S. 12904–12937.
- [100] Schmid, Peter J. Sesterhenn, Jörn. „Dynamic mode decomposition of numerical and experimental data“. In: *Journal of Fluid Mechanics* 656 (2010), S. 5–28.
- [101] Christof Schütte, Péter Koltai und Stefan Klus. „On the numerical approximation of the Perron-Frobenius and Koopman operator“. In: *Journal of Computational Dynamics* 3 (2016), S. 1–12.
- [102] Jaroslav Stark. „Delay Embeddings for Forced Systems. I. Deterministic Forcing“. In: *Journal of Nonlinear Science* 9.3 (1999), S. 255–332.
- [103] Jaroslav Stark u. a. „Takens embedding theorems for forced and stochastic systems“. In: *Nonlinear Analysis: Theory, Methods & Applications* 30 (1997), S. 5303–5314.
- [104] Steven Arigunawa Tjong. „Quantitive Comparison of Different Identification Methods using Measured Data“. Bachelor-Thesis. Darmstadt: Technische Universität Darmstadt, 2017.
- [105] Amit Surana. „Koopman operator based observer synthesis for control-affine nonlinear systems“. In: *2016 IEEE 55th Conference on Decision and Control (CDC)*. IEEE, 12.12.2016 - 14.12.2016, S. 6492–6499.
- [106] Amit Surana und Andrzej Banaszuk. „Linear observer synthesis for nonlinear systems using Koopman Operator framework“. In: *IFAC-PapersOnLine* 49 (2016), S. 716–723.
- [107] Yoshihiko Susuki und Igor Mezic. „Nonlinear Koopman Modes and a Precursor to Power System Swing Instabilities“. In: *IEEE Transactions on Power Systems* 27.3 (2012), S. 1182–1191.
- [108] Yoshihiko Susuki und Igor Mezic. „Nonlinear Koopman Modes and Coherency Identification of Coupled Swing Dynamics“. In: *IEEE Transactions on Power Systems* 26.4 (2011), S. 1894–1904.
- [109] Yoshihiko Susuki und Igor Mezic. „Nonlinear Koopman modes of coupled swing dynamics and coherency identification“. In: *IEEE PES General Meeting*. IEEE, 25.07.2010 - 29.07.2010, S. 1–8.

- [110] Yoshihiko Susuki, Igor Mezić und Takashi Hikiyara. „Coherent Swing Instability of power systems and cascading failures“. In: *IEEE PES General Meeting*. IEEE, 25.07.2010 - 29.07.2010, S. 1–8.
- [111] Naoya Takeishi, Yoshinobu Kawahara und Takehisa Yairi. „Learning Koopman Invariant Subspaces for Dynamic Mode Decomposition“. In: *Advances in Neural Information Processing Systems (Proc. of NIPS)* 30 (2017).
- [112] Floris Takens. „Detecting strange attractors in turbulence“. In: *Dynamical Systems and Turbulence, Warwick 1980*. Hrsg. von David Rand und Lai-Sang Young. Bd. 898. Lecture Notes in Mathematics. Berlin und Heidelberg: Springer, 1981, S. 366–381.
- [113] Harry L. Trentelman, Anton A. Stoorvogel und Malo Hautus. *Control Theory for Linear Systems*. London: Springer London, 2001.
- [114] Jochen Trumf. „On the geometry and parametrization of almost invariant subspaces and observer theory“. Würzburg, Univ., Diss., 2002.
- [115] Jan van Mill. „Domain Invariance in Infinite-Dimensional Linear Spaces“. In: *Proceedings of the American Mathematical Society* 101 (1987), S. 173.
- [116] Peter van Overschee und Bart de Moor. „A unifying theorem for three subspace system identification algorithms“. In: *Automatica* 31 (1995), S. 1853–1864.
- [117] Peter van Overschee und Bart de Moor. *Subspace Identification for Linear Systems*. Boston, MA: Springer US, 1996.
- [118] Vega Del Valle, Iliusi Donaji. *Reconstruction and analysis of the state space for the identification of dynamical states in real-world time series*. 2017.
- [119] Francesco Vicario. „OKID as a general approach to linear and bilinear system identification“. Diss. 2014.
- [120] Alejandro F. Villaverde. „Observability and Structural Identifiability of Nonlinear Biological Systems“. In: *Complexity* 2019 (2019), S. 1–12.
- [121] Jan C. Willems und Collectif. „Almost  $A(\text{mod } \beta)$ -invariant subspaces“. In: *Analyse des systèmes*. Astérisque. Société mathématique de France, 1980, S. 239–248.
- [122] Matthew O. Williams, Ioannis G. Kevrekidis und Clarence W. Rowley. „A Data-Driven Approximation of the Koopman Operator: Extending Dynamic Mode Decomposition“. In: *Journal of Nonlinear Science* 25 (2015), S. 1307–1346.
- [123] Zhewei Yao u. a. *Large batch size training of neural networks with adversarial training and second-order information*. 2018. arXiv: 1810.01021v3.
- [124] Enoch Yeung, Soumya Kundu und Nathan Hodas. „Learning Deep Neural Network Representation for Koopman Operators of Nonlinear Dynamical Systems“. In: *2019 American Control Conference (ACC)* (2019), S. 4832–4839.
- [125] L. Zadeh. „From Circuit Theory to System Theory“. In: *Proceedings of the IRE* 50.5 (1962), S. 856–865.

- [126] Matthew D. Zeiler. *ADADELTA: An Adaptive Learning Rate Method*. 2012. arXiv: 1212.5701v1.
- [127] Long Zhang u. a. *A Noise-Sensitivity-Analysis-Based Test Prioritization Technique for Deep Neural Networks*. 2019. arXiv: 1901.00054v3.