Fortschritt-Berichte VDI

VDI

Reihe 8

Mess-,
Steuerungs- und
Regelungstechnik

Nr. 1252

Dipl. Wirtsch.-Ing. Andreea Violeta Röthig,
Frankfurt am Main

# Nicht-konservative weiche strukturvariable Regelungen und Methoden zur Performance-Analyse in nichtlinearen Regelkreisen

Berichte aus dem

Institut für
Automatisierungstechnik
und Mechatronik
der TU Darmstadt

# Nicht-konservative weiche strukturvariable Regelungen und Methoden zur Performance-Analyse in nichtlinearen Regelkreisen

Dem Fachbereich
Elektrotechnik und Informationstechnik
der Technischen Universität Darmstadt
zur Erlangung des akademischen Grades
einer Doktor-Ingenieurin (Dr.-Ing.)
vorgelegte Dissertation

von

**Dipl. Wirtsch.-Ing. Andreea Violeta Röthig**

geboren am 21. Dezember 1980 in Bukarest, Rumänien

| | |
|---|---|
| Referent: | Prof. Dr.-Ing. Jürgen Adamy |
| Korreferent: | Prof. Dr.-Ing. Ulrich Konigorski |
| Tag der Einreichung: | 9. März 2016 |
| Tag der mündlichen Prüfung: | 21. Juli 2016 |

D17

Darmstadt 2016

Fortschritt-Berichte VDI

Reihe 8

Mess-, Steuerungs- und Regelungstechnik

Dipl. Wirtsch.-Ing.
Andreea Violeta Röthig,
Frankfurt am Main

Nr. 1252

# Nicht-konservative weiche strukturvariable Regelungen und Methoden zur Performance-Analyse in nichtlinearen Regelkreisen

Berichte aus dem

Institut für
Automatisierungstechnik
und Mechatronik
der TU Darmstadt

Röthig, Andreea Violeta
**Nicht-konservative weiche strukturvariable Regelungen und Methoden zur Performance-Analyse in nichtlinearen Regelkreisen**
Fortschr.-Ber. VDI Reihe 8 Nr. 1252. Düsseldorf: VDI Verlag 2016.
220 Seiten, 21 Bilder, 10 Tabellen.
ISBN 978-3-18-525208-2, ISSN 0178-9546,
€ 76,00/VDI-Mitgliederpreis € 68,40.

**Für die Dokumentation:** Reglerentwurf – Stellgrößenbeschränkungen – Nichtlineare Systeme – Lineare Matrixungleichung – Konvexe Optimierung – Ljapunov – Polynomiale Implizite Regelung – Computerexperimente

Die vorliegende Dissertation wendet sich an Ingenieure und Wissenschaftler im Bereich der Regelungstechnik. Sie befasst sich mit der Synthese nichtlinearer Regelungen für lineare Systeme mit Stellgrößenbeschränkungen. Im ersten Teil der Arbeit wird die Regelung solcher Systeme durch nicht-konservative weiche strukturvariable Regelungen vorgestellt. Die Nichtkonservativität bezieht sich auf die Eigenschaft der Stabilitätsbedingungen, sowohl notwendig als auch hinreichend zu sein. Im zweiten Teil der Arbeit wird die Performance-Analyse solcher Regelkreise mithilfe des Konzepts der Computerexperimente durchgeführt. Mittels Bayes'scher Interpolationsmethoden wird die Performance-Prädiktion eines Regelstreckenensembles ermöglicht. Damit ist es möglich, die erwartete Performance einer Regelmethode für eine bestimmte Strecke anzugeben. In diesem Zusammenhang wird auch eine Sensitivitätsanalyse vorgestellt, die Aussagen darüber zulässt, welchen Einfluss einzelne Streckenparameter auf die Performance einer Regelmethode erwartungsgemäß haben.

**Bibliographische Information der Deutschen Bibliothek**
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet unter http://dnb.ddb.de abrufbar.

**Bibliographic information published by the Deutsche Bibliothek**
(German National Library)
The Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliographie (German National Bibliography); detailed bibliographic data is available via Internet at http://dnb.ddb.de.

D 17



ISSN 0178-9546
ISBN 978-3-18-525208-2

# Vorwort

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftliche Mitarbeiterin am Fachgebiet Regelungsmethoden und Robotik des Instituts für Automatisierungstechnik und Mechatronik der Technischen Universität Darmstadt.

Mein besonderer Dank geht an Herrn Professor Dr.-Ing. Jürgen Adamy für seine Motivation, Betreuung und stets positive Einstellung bezüglich aller meiner wissenschaftlichen Bemühungen.

Die vorangegangenen Dissertationen meiner Kollegen Dr.-Ing. Boris Fischer (geb. Jasniewicz), Dr.-Ing. Hendrik Lens und Dr.-Ing. Dilyana Domont-Yankulova haben mir durch deren klare Struktur einen sehr guten Einstieg in das Thema ermöglicht. Zudem möchte ich mich bei meinen Kollegen Dr.-Ing. Dieter Lens, Kalina Olhofer-Karova, Dr.-Ing. Thomas Gußner, Dr.-Ing. Andreas Ortseifen, Dr.-Ing. Klaus Kefferpütz, Dr.-Ing. Arne Wahrburg, Dr.-Ing. Stefan Gering, Kerstin Groß, Dr.-Ing. Dilyana Domont-Yankulova, Dr.-Ing. Volker Willert, Saman Khodaverdian, Tatiana Tatarenko und Dimitri Chayka für viele spannende Diskussionen sehr bedanken. Des Weiteren möchte ich mich bei allen Kollegen, vor allem bei Andrea Schnall, Benjamin Reichhard, Martin Buczko, Nicolai Schweizer, Ivan Popov, Moritz Schneider, Florian Damerow, Diego Madeira, Dr.-Ing. Matthias Schreier, Birgit Heid, Susanne Muntermann und Sylvia Gelman, für die angenehme und freundschaftliche Atmosphäre am Fachgebiet bedanken.

Ein besonderer Dank geht an Herrn Professor Dr.-Ing. Ulrich Konigorski für die Übernahme des Korreferats und an Herrn Professor Dr. Carl Chiarella für seine freundschaftliche und sehr lehrreiche Betreuung während meines Forschungsaufenthaltes an der Technischen Universität Sydney.

Ferner möchte ich mich bei meinem Ehemann Andreas für seine unermüdliche Unterstützung und Motivation sowie für das Korrekturlesen der Arbeit bedanken.

Frankfurt am Main, September 2016 Andreea Röthig

# Inhaltsverzeichnis

# Symbole und Funktionen

| | |
|---|---|
| Symbole | |
| $\exists$ | es existiert |
| $\exists!$ | es existiert und ist eindeutig |
| $\forall$ | für alle |
| $:=$ | definiert als |
| $\overline{1,n}$ | $1,\ldots,n$ |
| $\lceil m \rceil$ | die kleinste ganze Zahl, die größer oder gleich $m \in \mathbb{R}$ ist |
| $\partial_t g(\mathbf{x}(t),v)$ | $\Leftrightarrow \dfrac{\partial g(\mathbf{x},v)}{\partial \mathbf{x}} \cdot \dot{\mathbf{x}}$ |
| $\partial_v g(\mathbf{x},v)$ | $\Leftrightarrow \dfrac{\partial g(\mathbf{x},v)}{\partial v}$ |
| Mengen | |
| $\mathcal{B}_\epsilon(\mathbf{x})$ | offene Kugel um $\mathbf{x} \in \mathbb{R}^n$ mit Radius $\epsilon$, vgl. Def. 5 (Anhang) |
| $\mathbb{R}^n$ | $n \times 1$ Spaltenvektoren |
| $\mathbb{R}^n_*$ | von null verschiedene Spaltenvektoren, d.h. $\mathbb{R}^n \backslash \{\mathbf{0}\}$ |
| $\mathbb{R}^{n\times m}$ | $n \times m$ reelle Matrizen |
| $\mathbb{F}^{n\times m}$ | $n \times m$ reelle oder komplexe Matrizen |
| $\mathbb{H}^n$ | $n \times n$ hermitesche Matrizen |
| $\mathrm{Sym}^n$ , $\mathrm{Skew}^n$ | $n \times n$ reelle symmetrische, schiefsymmetrische Matrizen |
| $\mathbb{P}^n(\mathbb{N}^n)$ | $n \times n$ symmetrische und positiv definite (semidefinite) Matrizen |
| $\varnothing$ | die leere Menge |
| $\mathcal{N}(\mathbf{A})$ | Kern (Nullraum) einer Matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathcal{N}(\mathbf{A}) := \{\mathbf{x} \in \mathbb{R}^n \vert \mathbf{A}\mathbf{x} = \mathbf{0}\}$ |
| $\mathcal{G}_\star(v)$ | Gebiet $\{\mathbf{x} \in \mathbb{R}^n \vert g_\star(\mathbf{x},v) < 0\}$ |
| $\mathcal{E}_\star(v)$ | Ellipsoid $\{\mathbf{x} \in \mathbb{R}^n \vert \mathbf{x}^\top \mathbf{P}(v)\mathbf{x} < 0\}$ |
| $\partial\mathcal{G}$ | Rand einer Menge $\mathcal{G}$, vgl. Def. 8 (Anhang) |

| | |
|---|---|
| $\operatorname{co}\mathcal{M}$ | bezeichnet die konvexe Hülle einer Menge $\mathcal{M}$, vgl. Def. 12 (Anhang) |
| $\dim\mathcal{M}$ | Dimension einer Menge |
| $\mathcal{L}(u,\beta)$ | das Gebiet $\mathcal{L}(u,\beta) := \{\mathbf{x}\in\mathbb{R}^n \,\|\, \|u(\mathbf{x})\| \leq \beta\}$ wo die Stellgrößenbegrenzung $\beta$ eingehalten wird |

Spezielle Matrizen

| | |
|---|---|
| $\mathbf{0}_{m,n}$ | $m\times n$ Nullmatrix |
| $\mathbf{I}_n$ | $n\times n$ Einheitsmatrix |
| $\mathbf{E}_n$ | $n\times n$ Elementarmatrix |
| $\mathbf{A}^+$ | Pseudoinverse einer Matrix $\mathbf{A}\in\mathbb{C}^{m\times n}$, auch Moore-Penrose-Inverse genannt, vgl. [8, S. 397] |
| $\tilde{\mathbf{A}}_{(i,j)}$ | der Kofaktor zum Element $a_{(i,j)}$ der Matrix $\mathbf{A}\in\mathbb{R}^{n\times n}$, d.h. die Matrix, die entsteht, wenn bei der Matrix $\mathbf{A}$ die $i$-te Zeile und $j$-te Spalte gestrichen werden |
| $\operatorname{adj}(\mathbf{A}), \mathbf{A}^A$ | die adjungierte Matrix zu $\mathbf{A}\in\mathbb{R}^{n\times n}$, d.h. $\mathbf{A}^A=\mathbf{C}^\top$, mit $c_{(i,j)}=(-1)^{i+j}\det(\tilde{\mathbf{A}}_{(i,j)})$ |
| $x_{(i)}$ | Element $(i)$ eines Vektors $\mathbf{x}\in\mathbb{R}^n$ |
| $a_{(i,j)}$ | Element $(i,j)$ einer Matrix $\mathbf{A}\in\mathbb{R}^{m\times n}$ |
| $\mathbf{x}_\lambda\in\mathbb{R}^n$ | zum Eigenwert $\lambda$ gehörender Rechtseigenvektor $\mathbf{x}_\lambda$ |
| $\mathbf{z}^{[k]}\in\mathbb{R}^k$ | $\begin{bmatrix} 1 & z & z^2 & \cdots & z^{k-1} \end{bmatrix}^\top$ |
| $\mathbf{A}\|\mathcal{A}$ | Schur-Komplement der Matrix $\mathbf{A}\in\mathbb{R}^{m\times n}$ bzgl. der Blockmatrix $\mathcal{A}=\begin{bmatrix}\mathbf{A}&\mathbf{B}\\\mathbf{C}&\mathbf{D}\end{bmatrix}$, definiert als $\mathbf{A}\|\mathcal{A}=\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}$, $\mathbf{A}$ nichtsingulär |

Matrixfunktionen

| | |
|---|---|
| $\operatorname{Rang}(\mathbf{A})$ | Rang einer Matrix $\mathbf{A}\in\mathbb{R}^{m\times n}$ |
| $\operatorname{sr}(\mathbf{A})$ | Spaltenrang einer Matrix $\mathbf{A}\in\mathbb{R}^{m\times n}$ |
| $\operatorname{zr}(\mathbf{A})$ | Zeilenrang einer Matrix $\mathbf{A}\in\mathbb{R}^{m\times n}$ |
| $\operatorname{Spec}(\mathbf{A})$ | Spektrum einer Matrix $\mathbf{A}\in\mathbb{R}^{n\times n}$, d.h. die Menge aller Eigenwerte bei Nichtbeachtung der Vielfachheit |
| $\lambda(\mathbf{A})\in\mathbb{C}$ | Eigenwert einer Matrix $\mathbf{A}\in\mathbb{R}^{n\times n}$ |

| | |
|---|---|
| $\kappa(\mathbf{A})$ | Konditionszahl einer Matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$, definiert als $\kappa(\mathbf{A}) := \lambda_{\max}(\mathbf{A})/\lambda_{\min}(\mathbf{A})$ |
| Re $\lambda(\mathbf{A})$ | Realteil des Eigenwertes $\lambda \in \mathrm{Spec}(\mathbf{A})$ |
| Im $\lambda(\mathbf{A})$ | Imaginärteil des Eigenwertes $\lambda \in \mathrm{Spec}(\mathbf{A})$ |
| $\lambda_i(\mathbf{A}) \in \mathrm{Spec}(\mathbf{A}) \subset \mathbb{R}$ | $i$.-größter Eigenwert einer Matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ mit reellen Eigenwerten |
| $\lambda_{\max}(\mathbf{A}) \in \mathrm{Spec}(\mathbf{A}) \subset \mathbb{R}$ | größter Eigenwert einer Matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ mit reellen Eigenwerten |
| $\lambda_{\min}(\mathbf{A}) \in \mathrm{Spec}(\mathbf{A}) \subset \mathbb{R}$ | kleinster Eigenwert einer Matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ mit reellen Eigenwerten |
| $\mathrm{vec}(\mathbf{A})$ | Spalten-Vektorisierung einer Matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, vgl. Def. 22 (Anhang) |
| $\|\mathbf{x}\| = \sqrt{\sum_{i=1}^{n} x_{(i)}^2}$ | Euklidische Norm eines Vektors $\mathbf{x} \in \mathbb{R}^n$, vgl. Def. 23 (Anhang) |
| $\|\mathbf{A}\|_\infty := \max_{\substack{i\in\{1,\dots,n\}\\ j\in\{1,\dots,m\}}} \|\mathbf{A}_{(i,j)}\|$ | Maximum Norm einer Matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$, vgl. Def. 24 (Anhang) |
| $\|\mathbf{A}\|_{\sigma\infty} := \sigma_{\max}(\mathbf{A})$ | Spektralnorm einer Matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$, vgl. Def. 24 (Anhang) |
| $\mathbf{A} \otimes \mathbf{B}$ | Kronecker Produkt, d.h. Multiplikation jedes Elements der Matrix $\mathbf{A}$ mit der Matrix $\mathbf{B}$, vgl. [74] für verschiedene Eigenschaften |
| $\mathbf{A} \oplus \mathbf{B}$ | Kronecker Summe, vgl. Def. 21 (Anhang) |
| $\mathbf{A}^\top$ | Transponierte einer Matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ |
| $\mathbf{C}^H$ | konjugiert komplexe und transponierte Matrix $\mathbf{C} \in \mathbb{C}^{m\times n}$ |
| $\mathbf{X}_v$ | von $v \in \mathbb{R}$ abhängige Matrix $\mathbf{X}(v) \in \mathbb{R}^{m\times n}$ |
| <u>Matrixrelationen</u> | |
| $\mathbf{A} \succ \mathbf{B}\,(\mathbf{A} \succeq \mathbf{B})$ | $\mathbf{A} - \mathbf{B} \in \mathbb{P}^n\,(\mathbf{A} - \mathbf{B} \in \mathbb{N}^n)$ |
| $\mathbf{A} \sim \mathbf{B}$ | die Matrizen $\mathbf{A}$ und $\mathbf{B}$ sind *ähnlich*, vgl. Def. 18 (Anhang) |
| <u>Stochastik</u> | |
| $\mathcal{H}(\,\cdot\,)$, $\mathcal{Z}(\,\cdot\,)$ | skalare Zufallsfelder |
| $h(\,\cdot\,)$, $z(\,\cdot\,)$ | Realisierungen (Pfade) der skalaren Zufallsfelder $\mathcal{H}(\,\cdot\,)$ bzw. $\mathcal{Z}(\,\cdot\,)$ |
| $\boldsymbol{\mathcal{C}}(\,\cdot\,)$ | mehrdimensionales Zufallsfeld |

| | |
|---|---|
| $\mathbf{c}(\cdot)$ | Realisierung (Pfad) eines mehrdimensionalen Zufallsfeldes $\boldsymbol{\mathcal{C}}(\cdot)$ |
| $H$, $\mathbf{H}$ | Zufallsvariable, Zufallsvektor |
| $\eta$, $\boldsymbol{\eta}$ | Realisierung einer Zufallsvariable bzw. eines Zufallsvektors |
| $[\cdot]$ | Wahrscheinlichkeitsverteilung (kurz: Verteilung), die durch ihre Wahrscheinlichkeitsdichte (oder kurz Dichte) gegeben ist |
| $[X] \cdot [Y]$ | Multiplikation von zwei Dichten, d.h. $f_X(x) \cdot f_Y(y)$ |
| $\propto$ | proportionale Verteilungen |
| $\mathcal{D}_{\boldsymbol{\zeta}_x}$ | Definitionsbereich der Variable $\boldsymbol{\zeta}_x$ |
| $\mathcal{N}_m(\boldsymbol{\mu},\boldsymbol{\Sigma})$ | $m$-dimensionale multivariate Normalverteilung mit Erwartungswertvektor $\boldsymbol{\mu} \in \mathbb{R}^m$ und positiv-definiter Kovarianzmatrix $\boldsymbol{\Sigma} \in \mathrm{Sym}^m$, vgl. Abschnitt B.1 (Anhang) |
| $\chi^2_n$ | eindimensionale Chi-Quadrat Verteilung mit $n$ Freiheitsgraden, vgl. Abschnitt B.2 (Anhang) |
| $\chi^{-2}_n$ | eindimensionale Inverse-Chi-Quadrat Verteilung mit $n$ Freiheitsgraden, vgl. Abschnitt B.2 (Anhang) |
| $\mathcal{T}_1(\nu,\mu,\sigma^2)$ | eindimensionale nicht-zentrale $t$-Verteilung mit $\nu$ Freiheitsgraden, Nichtzentralitätsparameter $\mu$ und Skalierungsparameter $\sigma^2$, vgl. Abschnitt B.3 (Anhang) |

Abkürzungen

| | |
|---|---|
| o.B.d.A. | ohne Beschränkung der Allgemeinheit |
| WSVR | weiche strukturvariable Regelung |
| iLF | implizite Ljapunov-Funktion |
| PPDQ Funktion | polynomiell parameterabhängige quadratische Funktion |
| LTI System | lineares zeitinvariantes System |
| LHS | *Latin-Hypercube-Sampling* |
| ERMSPE | ***E**mpirical **R**oot **M**ean **S**quared **P**rediction **E**rror* |
| AC | ***A**chieved **C**overage* |

# Kurzfassung

Die vorliegende Arbeit beschäftigt sich mit der Synthese weicher strukturvariabler Regelungen (WSVR) zur Stabilisierung linearer zeitinvarianten Systemen (LTI-Systeme) mit Stellgrößenbeschränkung und einer neuen Methode zur Performance-Analyse in nichtlinearen Regelkreisen. Im Rahmen der Regelsynthese werden zum ersten Mal notwendige und hinreichende Stabilisierbarkeitsbedingungen solcher Strecken durch WSVRs mittels impliziter Ljapunov-Funktionen (iLF) vorgestellt. Die erzielte Regelung ist also nicht-konservativ. Aus der Notwendigkeit der Bedingungen folgt, dass im Fall deren Nichterfüllung überhaupt kein Regler aus der untersuchten Klasse existiert. Die Notwendigkeit stellt eine wesentliche Erweiterung gegenüber bereits existierenden Stabilitätsbedingungen dar. Eine zweite Erweiterung der WSVR auf solche Regelungen mittels *invers-polynomialer* Selektionsstrategien wird ebenfalls vorgestellt. Darüber hinaus werden die nicht-konservativen Regler bezüglich der Konvergenzrate optimiert. Es wird gezeigt, dass der konvergenzoptimale Regler ein Zweipunktregler mit einer parameterabhängigen Umschaltstrategie ist, der ein sehr schnelles Ausregelverhalten aufweist. Die Formulierung der Bedingungen mittels äquivalenter linearer Matrixungleichungen (LMIs) wird ebenfalls vorgestellt. Dies erlaubt einen numerisch effizienten Entwurf der Regler für Strecken beliebiger Ordnung. Die Reglersynthese endet mit den Stabilisierbarkeitsbedingungen solcher Regler für Regelstreckenensembles, die durch parametrische LTI-Systeme beschrieben sind.

Der zweite Teil der Arbeit beschäftigt sich mit der Performance-Analyse in nichtlinearen Regelkreisen. Während für lineare Regelkreise exakte (auch frequenzbasierte) Methoden zur Performance-Analyse existieren, bezieht sich die kleine Menge an Analysemethoden für nichtlineare Regelkreise meistens auf experimentelle Aussagen über das dynamische Verhalten eines einzelnen Regelkreises. In dieser Arbeit wird zum ersten Mal das in der Praxis weitverbreitete Konzept der *Computerexperimente* auf die Performance-Analyse in nichtlinearen Regelkreisen angewandt. Mittels Bayes'scher Interpolationsmethoden wird die Performance-Prädiktion eines gesamten Streckenensembles ermöglicht. Sehr wichtig sind dabei

die angegebenen Konfidenzintervalle der Prädiktion. Damit ist es möglich, die erwartete Performance einer Regelmethode für eine bestimmte Strecke anzugeben, ohne dabei einen Regler entwerfen zu müssen. In diesem Zusammenhang wird auch eine Sensitivitätsanalyse vorgestellt, die Aussagen darüber zuläßt, welchen Einfluß einzelne Streckenparameter auf die Performance einer Regelmethode erwartungsgemäß haben. Die Arbeit endet mit einem empirischen Vergleich verschiedener Prädiktoren anhand mehrerer Streckenensembles. Es wird gezeigt, dass die Prädiktionsgenauigkeit abhängig von der Wahl der prädiktiven *A-posteriori*-Verteilung, von der Wahl der Korrelationsfunktion zwischen verschiedenen Strecken, sowie von der Wahl der empirischen Schätzmethode für die Parameter der Korrelationsfunktion ist.

# Abstract

The thesis deals with the non-conservative synthesis of soft variable structure controls (SVSC) for stabilizing linear time invariant systems (LTI-systems) with input saturation, and with a new method for the performance analysis in nonlinear control systems. The non-conservative control synthesis yields some necessary and sufficient stability conditions for these plants, employing implicit Lyapunov-functions (iLF). From the necessity of the conditions follows that if they are not fulfilled, then there exists no control from this class which stabilizes the given plant. This is an essential benefit of the proposed controls relative to the already existing SVSC employing iLFs. Furthermore, an extension of the SVSC to ones that employ inverse-polynomial selection strategies is presented. In addition, both (non-conservative) controls are being optimized relative to the convergence rate. The maximal convergence control is a bang-bang type control with a parameter-dependent switching scheme, that achieves very short settling times. The formulation of the stability conditions in form of equivalent linear matrix inequalities (LMI) is also a benefit of the proposed control methods. This allows for an efficient numerical control design for plants of any given order. The control synthesis part of the thesis ends with the (non-conservative) design of SVSC for a plant ensemble, that is described by a parameter dependent LTI-System.

The second part of the thesis deals with the performance analysis in nonlinear control systems. While for linear systems there exists a large number of exact (also frequency-based) methods for the performance-analysis, the number of methods for the performance analysis of nonlinear systems is very small, and deals mainly with the analysis of a single plant for a given control. In this thesis the concept of the design and analysis of computer experiments is applied to the performance analysis of nonlinear control systems. By employing Bayesian interpolation methods, one can make a prediction of the performance of the nonlinear control method for an ensemble of nonlinear closed-loop systems. An important benefit of employing this statistical framework is that the prediction is given together with some confidence bounds on the expected performance. Consequently,

it is possible to make a prediction of the performance of a control method without designing the control. In this context we present also a sensitivity analysis, which gives some insight on how the expected performance of the control method changes, if one changes the parameters of the plant. The thesis ends with an empirical comparison of some predictors, which shows that the prediction depends on the *a-posteriori* distribution of the model, on the correlation function employed and on its parameters, respectively on the empirical estimation method for these unknown parameters.

# 1 Einleitung

Jedes reale Stellglied ist aus physikalischen Gründen mit einer Amplitudenbeschränkung behaftet. In der Regelungstechnik sind Systeme mit Stellgrößenbeschränkungen seit längerer Zeit Gegenstand vieler Untersuchungen, vgl. [10] und die darin enthaltenen Verweise. Die lineare Zustandsrückführung ist eines der meist verwendeten Regelgesetze, deren Anwendung nichtsättigende bis hin zu *High-Gain*-Reglern erzielt, wobei auftretende Sättigungseffekte mittels Anti-Windup Strukturen reduziert [70] oder direkt in der Stabilitätsanalyse [35] berücksichtigt werden. Der generelle Nachteil der linearen Zustandsrückführung ist, dass der Stellgrößenbereich im Bereich kleiner Auslenkungen von der Ruhelage aufgrund konstanter Reglerverstärkung nicht optimal ausgenutzt werden kann, und der jeweilige Regler somit oft zu langsamen Zeitverläufen führt.

Unter den nichtlinearen Regelmethoden erzielt die zeitoptimale Regelung, in diesem Fall ein schaltendes Regelgesetz, auch Zweipunktregler genannt, den schnellsten Zeitverlauf. Da seine Umschaltstrategie aber generell nicht berechenbar ist, wird diese Methode oft nur bei Systemen niedrigerer Ordnung verwendet. Außerdem kann in der Praxis die Diskontinuität des Regelgesetzes technische Probleme verursachen. Beispielsweise kann die ununterbrochene Aktivität des Reglers aufgrund unvermeidbaren Rauschens die Aktoren beschädigen. Eine Alternative zu den beiden Regelungsmethoden bieten die strukturvariablen Regelungen an, welche durch variable Verstärkungen den Nachteil der linearen Zustandsrückführung überwinden und durch einfache Entwurfsverfahren bei Systemen beliebiger Ordnung verwendet werden können.

Die Geschwindigkeit des Zeitverlaufs wird bei exponentiell stabilen Systemen oft mit der Konvergenzrate in Verbindung gebracht, welche als kleinster Abklingfaktor der Norm einer Trajektorie definiert ist. Für lineare Systeme ist die Konvergenzrate konstant im gesamten Zustandsraum und entspricht dem Betrag des Realteils des Eigenwertes, der am nächsten zur Imaginärachse liegt. Für nichtlineare Systeme hängt sie von dem Abstand zur Ruhelage ab. In der Literatur wird die Konvergenzrate daher mit Hilfe von invarianten Gebieten analysiert, vgl. z.B. [13]. Dabei sind die meist

verwendeten invarianten Gebiete ellipsoidaler Form, da diese mit effizienten Hilfsmitteln, wie der Ljapunov Gleichung und, generell, der linearen Matrixungleichungen (LMI) analysiert werden können. Die Vorteile der ellipsoidalen Gebiete werden auch im Fall weicher strukturvariabler Regelungen genutzt, indem die Variation der Reglerverstärkungen mit solchen Gebieten im Zustandsraum verbunden wird.

Das Regelgesetz, das in der Form einer sättigenden linearen Zustandsrückführung die Abklingrate einer quadratischen Ljapunov-Funktion $V(\mathbf{x}) = \mathbf{x}^\top \mathbf{P}\mathbf{x}$ entlang der Trajektorien des Gesamtsystems maximiert, ist ein Zweipunktregler mit einer einfachen Umschaltstrategie, vgl. z.B. [35]. Das Regelgesetz hat die Form $u = -\text{sgn}(\mathbf{b}^\top \mathbf{P}\mathbf{x})$, wobei $\mathbf{b}$ der Steuervektor der Strecke ist. Die maximale Konvergenzrate hängt dabei von der Matrix $\mathbf{P}$ ab, welche auch das invariante Ellipsoid determiniert. In [35] wird gezeigt, dass die Größe des erzielten Ellipsoids im Konflikt mit der Höhe der Konvergenzrate steht, d.h. eine hohe Konvergenzrate erzielt ein kleines invariantes Ellipsoid und umgekehrt. Durch eine parameterabhängige Matrix $\mathbf{P}_v$, wie im Fall strukturvariabler Regelungen, besteht die Möglichkeit, die Größe des Ellipsoids an den Abstand zur Ruhelage anzupassen und somit die Konvergenzrate insgesamt zu verbessern. Vgl. auch [36] für eine solche Verbesserung durch eine parameterabhängige Matrix. Es wird gezeigt, dass auch in diesem Fall die Optimierung der Konvergenzrate einen Zweipunktregler (auch *Bang-Bang*-Regler genannt) mit einer parameterabhängigen Umschaltstrategie erzielt. Um die Nachteile der Diskontinuität des Regelgesetzes zu umgehen, wird eine stetige Approximation des konvergenzoptimalen Regelgesetzes vorgestellt, die auf Kosten einer leichten Verschlechterung der Konvergenzrate einen stetigen Stellgrößenverlauf erzielt. Dabei wird auch ein nichtkonservativer Entwurf des konvergenzoptimalen Regelgesetzes vorgestellt. Dieser beinhaltet Stabilitätsbedingungen, welche sowohl notwendig als auch hinreichend für die Stabilisierbarkeit einer linearen Strecke mit Stellgrößenbeschränkung durch eine weiche strukturvariable Regelung dieser Klasse sind.

Ein weiterer Aspekt dieser Arbeit ist die Performance-Analyse nichtlinearer Regelkreise. Während für lineare Regelkreise exakte (frequenzbasierte) Methoden zur Performance-Analyse existieren, beziehen sich die Performance-Methoden für nichtlineare Regelkreise meistens auf experimentelle Aussagen über das dynamische Verhalten eines einzelnen Regelkreises. Die Schwierigkeit der Performance-Analyse entsteht aufgrund der komplexeren Eigenschaften nichtlinearer Systeme relativ zu denen linearer Systeme, wie zum Beispiel fehlende Gültigkeit des Superpositionsprinzips,

oder bezüglich der Stabilität der Ruhelage, die abhängig von den Anfangsbedingungen und von den einwirkenden Eingangsgrößen ist. Ein weiterer Grund ist, dass für nichtlineare Regelkreise, die mit Hilfe von nichtlinearen Differentialgleichungen formuliert werden, im Allgemeinen keine exakte Zeitlösung bekannt ist.

Die in der Literatur vorgestellten Methoden zur Performance-Analyse werden meistens in exakte und approximative Methoden klassifiziert. Diese basieren auf der exakten bzw. approximativen Zeitlösung des Systems und sind jeweils auf einen einzelnen Regelkreis anwendbar. Um die Ergebnisse zu verallgemeinern, konzentriert sich die vorliegende Arbeit auf die Entwicklung einer Methode zur Performance-Analyse in Regelkreis-Ensembles. Ein Regelkreis-Ensemble besteht aus linearen Strecken, wobei die Systemmatrix und der Steuervektor jeder Strecke von einem Parameter aus einer kompakten Menge polynomiell abhängen. Diese Methode knüpft dabei an die Analyse durch *Computerexperimente* an, welche in der Industrie eine breite Anwendung findet. Demnach wird die Performance einer Regelungsmethode für ein Regelkreis-Ensemble an einzelnen Strecken (*Design-Strecken*) exakt oder approximativ überprüft und im übrigen analysierten Bereich statistisch interpoliert. Wesentlich dabei ist die Angabe von Konfidenzintervallen für die erwartete Performance im gesamten Ensemble.

## 1.1 Beiträge der Arbeit

Wie bereits erwähnt, beschäftigt sich die vorliegende Arbeit in einem ersten Teil mit der nicht-konservativen Synthese weicher strukturvariabler Regelungen (WSVR) mittels impliziter Ljapunov-Funktionen (iLF) zur Stabilisierung linearer Systeme mit Stellgrößenbeschränkung. Obwohl sich eine große Anzahl an Regelungsmethoden mit der Stabilisierung solcher Systeme beschäftigt, stellen die hier entwickelten Regelungsmethoden einen deutlichen Fortschritt bezüglich des Entwurfsaufwands und der Performance relativ zu der zeitoptimalen Regelung dar. Für die *klassische* WSVR mittels iLF, welche auf [2] zurückgeht, werden erstmals notwendige und hinreichende Bedingungen vorgestellt. Aus der Notwendigkeit der Bedingungen folgt, dass im Fall deren Nichterfüllung überhaupt kein Regler dieser Klasse die jeweilige Strecke stabilisieren kann. Dies stellt einen wesentlichen Vorteil gegenüber bereits existierenden Entwurfsbedingungen für diese Klasse dar.

Daran anschließend wird die WSVR mittels iLF auf WSVR mit *invers-polynomialen* Selektionsstrategien erweitert. Somit wird der (implizite) Parameter der weichen strukturvariablen Regelung anhand einer Selektionsstrategie berechnet, welche durch die Inverse einer (in diesem Parameter) polynomialen Matrix bestimmt ist. Eine ähnliche *invers-polynomiale* Selektionsstrategie wurde in [36] vorgestellt. Die wesentlichen Vorteile der in dieser Arbeit entwickelten Regelungsmethode stellen einerseits der frei wählbare Grad der polynomialen Matrix und andererseits die nicht-konservativen Entwurfsbedingungen dar.

Eine darauf aufbauende Optimierung der Konvergenzrate erzeugt jeweils Zweipunktregler mit einer parameterabhängigen Selektionsstrategie. Diese Regler werden auch *Bang-Bang*-Regler genannt und unterscheiden sich von der zeitoptimalen Regelung durch deren Umschaltstrategie. Da in der Praxis die Diskontinuität des Regelgesetzes technische Probleme verursachen kann, wird in dieser Arbeit eine stetige Approximation des konvergenzoptimalen Regelgesetzes vorgestellt, die auf Kosten einer leichten Verschlechterung der Konvergenzrate einen stetigen Stellgrößenverlauf erzielt.

Im zweiten Teil der Arbeit werden verschiedene Methoden zur Performance-Analyse in nichtlinearen Regelkreisen vorgestellt. Der Hauptbeitrag dieses Teils stellt die Anwendung der Theorie über das Design von *Computerexperimenten* auf die Performance-Analyse in nichtlinearen Regelkreisen dar. Darüber hinaus wird eine Sensitivitätsanalyse für eine Regelmethode bezüglich eines Streckenensembles mit unendlich vielen Strecken eingeführt. Diese soll Aufschluß darüber geben, wie die Streckenparameter die Performance einer Regelmethode beeinflussen. In diesem Zusammenhang erfährt die in dieser Arbeit neu-entwickelte konvergenzoptimale Regelung durch weiche strukturvariable Regelungen mit *invers-polynomialen* Selektionsstrategien eine besondere Berücksichtigung.

## 1.2 Gliederung

Die Arbeit besteht aus zwei Teilen. Im ersten Teil werden die nicht-konservativen weichen strukturvariablen Regelungen vorgestellt. Kapitel 2 enthält eine Einleitung über die Stabilisierung linearer Systeme mit Stellgrößenbeschränkung und die Stabilitätsanalyse mittels impliziter Ljapunov-Funktionen. Im Kapitel 3 werden die notwendigen und hinreichenden Stabilitätsbedingungen der *klassischen* WSVR mittels iLFs und im Kapitel 4

die *invers-polynomiale* WSVR vorgestellt. Im Kapitel 5 wird die Optimierung der Konvergenzrate analysiert. Das letzte Kapitel aus diesem Teil, Kapitel 6, stellt die notwendigen und hinreichenden Stabilitätsbedingungen der *klassischen* und *invers-polynomialen* WSVR für Regelstreckenensembles vor.

Der zweite Teil der Arbeit beschäftigt sich mit Methoden zur Performance-Analyse in nichtlinearen Regelkreisen. In dem einleitenden Kapitel 7 werden die wesentlichen Unterschiede zwischen linearen und nichtlinearen Regelkreisen dargestellt, sowie verschiedene Performance-Maße für nichtlineare Regelkreise klassifiziert. Eine besondere Berücksichtigung erfährt dabei die Konvergenzrate, welche für die im ersten Teil vorgestellten Regelungsmethoden analysiert wird. Kapitel 8 stellt die Anwendung des Designs von *Computerexperimenten* auf die Performance-Analyse von Regelungsmethoden vor. Das letzte Kapitel dieser Arbeit, Kapitel 9, fasst die wesentlichen Ergebnisse dieser Arbeit zusammen und gibt einen Ausblick über mögliche Weiterentwicklungen. Schließlich enthalten die Anhänge verschiedene Definitionen und Hilfssätze, sowie die Parameter der vorgestellten Beispiele. Anhang A enthält einige allgemeine Definitionen über Mengen, Funktionen und Matrizen, sowie mehrere Hilfssätze für die Reglersynthese. Anhang B enthält mehrere stochastische Grundlagen und Anhang C enthält die Parameter der Beispiele.

# Teil I

# Nicht-konservative WSVR-Synthese für LTI-Systeme mit Stellgrößenbeschränkung

# 2 Einleitende Hilfssätze

Für weiche strukturvariable Regelungen mit kontinuierlich parameterabhängiger Zustandsrückführung wurden in [1, 2] hinreichende Stabilisierbarkeitsbedingungen linearer Systeme mit Stellgrößenbeschränkung mittels impliziter Ljapunov-Funktionen (iLF) vorgestellt. Der Überblicksartikel [4] bietet eine Beschreibung solcher Regelungen. Bei diesen Regelungen wird zwischen verschiedenen linearen Zustandsrückführungen kontinuierlich während des Ausregelvorgangs umgeschaltet. Die Umschaltung ist so ausgelegt, dass während des Ausregelvorgangs mit kleiner werdendem Abstand zur Ruhelage immer mehr Einfluß auf die Strecke ausgeübt wird. Diese Regelungen arbeiten nicht im Sättigungsbereich, d.h. die Stellgrößenverläufe tangieren höchstens die Begrenzungen, bleiben aber nicht für längere Zeit in der Sättigung. Sättigende WSVR wurden beispielsweise in [19, 36], solche mittels impliziter Ljapunov-Funktionen wurden in [25, 40, 44] vorgestellt. Die Arbeit von [25] beinhaltet dabei eine besondere Berücksichtigung linearer Matrixungleichungen (LMI) zur Formulierung der Stabilisierbarkeitsbedingungen. Die Arbeit von [40] stellt darüber hinaus eine Generalisierung der WSVR mittels iLF dar, welche polynomiale Selektionsstrategien beliebigen Grades sowie den Spezialfall der expliziten LF beinhaltet. Da diese Stabilisierbarkeitsbedingungen jedoch nicht notwendig sondern nur hinreichend sind, kann der Entwurf konservative Regler erzeugen. Die Größe des erzielten invarianten Gebietes kann z.B. klein sein, oder die Einschwingzeit des Ausregelvorgangs kann lang sein. Daher werden in diesem Teil der Arbeit erstmals Bedingungen vorgestellt, die für die Stabilisierbarkeit der Strecke auch notwendig sind. Die Bedingungen bauen auf den notwendigen und hinreichenden Bedingungen für die Existenz eines linearen Zustandsreglers für eine lineare Strecke mit Stellgrößenbeschränkung auf, welche in [39] vorgestellt wurden. Durch eine konstruktive Methode werden notwendige und hinreichende Existenzbedingungen einer WSVR aufgebaut, welche darüber hinaus in äquivalente LMIs transformiert werden können. Sind diese Bedingungen erfüllt, so wird ein beschränkter Regler angegeben, der das System stabilisiert.

In Folgendem werden mehrere Hilfssätze vorgestellt, welche bei der nicht-konservativen WSVR-Synthese verwendet werden. Dieses Kapitel ist wie folgt gegliedert: im Abschnitt 2.1 wird die Stabilisierbarkeit linearer Strecken mit Stellgrößenbeschränkung diskutiert und im Abschnitt 2.2 werden zwei Hilfssätze über die Stabilitätsuntersuchung eines nichtlinearen Systems mittels impliziter Ljapunov-Funktionen vorgestellt.

## 2.1 Stabilisierbarkeit linearer Systeme mit Stellgrößenbeschränkung

Falls die Stellgröße unbeschränkt ist, ist jedes vollständig steuerbare LTI-System stabilisierbar, d.h. die Ruhelage des Systems kann durch die Aufschaltung eines geeigneten Regelgesetzes asymptotisch stabil werden. Im Fall eines nicht vollständig steuerbaren Systems muss das nichtsteuerbare Systemteil bereits asymptotisch stabil sein. Dies wird im folgenden Lemma verdeutlicht.

**Lemma 2.1 [Stabilisierbarkeit eines LTI-Systems].** *Gegeben sei das LTI-System*

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b}u, \quad \mathbf{x} \in \mathbb{R}^n, u \in \mathbb{R}, \mathbf{A} \in \mathbb{R}^{n\times n}, \mathbf{b} \in \mathbb{R}^n,$$

*welches vollständig beschrieben durch das Paar* $(\mathbf{A},\mathbf{b})$ *ist. Folgende Aussagen sind äquivalent:*

*i) Das Paar* $(\mathbf{A},\mathbf{b})$ *ist stabilisierbar.*

*ii)* $\exists\, \mathbf{T} \in \mathbb{R}^{n\times n}$, $\mathbf{T}$ *nichtsingulär, sodass* $\mathbf{A} = \mathbf{T}\begin{bmatrix}\mathbf{A}_1 & \mathbf{A}_{12}\\ \mathbf{0} & \mathbf{A}_2\end{bmatrix}\mathbf{T}^{-1}$ *und* $\mathbf{b} = \mathbf{T}\begin{bmatrix}\mathbf{b}_1\\ \mathbf{0}\end{bmatrix}$, *wobei* $\mathbf{A}_1 \in \mathbb{R}^{q\times q}$, $\mathbf{b}_1 \in \mathbb{R}^q$ *und* $(\mathbf{A}_1,\mathbf{b}_1)$ *vollständig steuerbar und* $\mathbf{A}_2 \in \mathbb{R}^{(n-q)\times(n-q)}$ *asymptotisch stabil ist.*

*Beweis.* Der Beweis kann in [8, Proposition 12.8.3] gefunden werden. Die folgende Bemerkung veranschaulicht die Bedingung *ii*). □

*Bemerkung* 2.1. Wie in [65, Abschnitt 3.4] gezeigt, existiert für jedes LTI-System eine eineindeutige Transformationsvorschrift

$$\tilde{\mathbf{x}} = \mathbf{T}_{\mathrm{g}}\mathbf{x},$$

sodass der transformierte Zustandsvektor in der Form

$$\tilde{\mathbf{x}} = \begin{bmatrix} \mathbf{x}_{\mathrm{sb}} \\ \mathbf{x}_{\mathrm{s}\overline{\mathrm{b}}} \\ \mathbf{x}_{\overline{\mathrm{s}}\mathrm{b}} \\ \mathbf{x}_{\overline{\mathrm{s}}\overline{\mathrm{b}}} \end{bmatrix} \begin{array}{l} \text{- steuerbares und beobachtbares Systemteil} \\ \text{- steuerbares aber nicht beobachtbares Systemteil} \\ \text{- nicht steuerbares aber beobachtbares Systemteil} \\ \text{- weder steuerbares noch beobachtbares Systemteil} \end{array}$$

vorliegt. Diese Systemform wird auch *Kalman-kanonische Form* genannt. Fasst man die beiden steuerbaren bzw. nichtsteuerbaren Systemteile zusammen, ergibt sich das transformierte System

$$\begin{bmatrix} \dot{\tilde{\mathbf{x}}}_{\mathrm{s}} \\ \dot{\tilde{\mathbf{x}}}_{\overline{\mathrm{s}}} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_{12} \\ \mathbf{0} & \mathbf{A}_2 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}_{\mathrm{s}} \\ \tilde{\mathbf{x}}_{\overline{\mathrm{s}}} \end{bmatrix} + \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{0} \end{bmatrix} u, \quad \tilde{\mathbf{x}}_{\mathrm{s}} = \begin{bmatrix} \mathbf{x}_{\mathrm{sb}} \\ \mathbf{x}_{\mathrm{s}\overline{\mathrm{b}}} \end{bmatrix}, \quad \tilde{\mathbf{x}}_{\overline{\mathrm{s}}} = \begin{bmatrix} \mathbf{x}_{\overline{\mathrm{s}}\mathrm{b}} \\ \mathbf{x}_{\overline{\mathrm{s}}\overline{\mathrm{b}}} \end{bmatrix},$$

aus Punkt *ii*). Die Matrix $\mathbf{A}_1$ und die Vektoren $\tilde{\mathbf{x}}_{\mathrm{s}}$ und $\mathbf{b}_1$ haben $n_s$ Zeilen, wobei $n_s$ der Rang der Steuerbarkeitsmatrix des Gesamtsystems ist. Es ist ersichtlich, dass das zweite Systemteil $\dot{\tilde{\mathbf{x}}}_{\overline{\mathrm{s}}} = \mathbf{A}_2\tilde{\mathbf{x}}_{\overline{\mathrm{s}}}$ nicht steuerbar ist. Da die Matrix $\mathbf{A}_2$ wesentlich für die Stabilität des Gesamtsystems ist,[1)] ist es sowohl notwendig als auch hinreichend, dass dieses Systemteil bereits asymptotisch stabil ist. △

Die Stabilisierbarkeit linearer Systeme mittels linearer Regelgesetze kann auch durch die Existenz quadratischer Ljapunov-Funktionen der Form $V(\mathbf{x}) = \mathbf{x}^\top\mathbf{P}\mathbf{x}$ überprüft werden. Dabei ist die Existenz einer quadratischen Ljapunov-Funktion hinreichend und notwendig für die Stabilität eines linearen Systems. Dies wird im folgenden Lemma verdeutlicht. Das Resultat wird in den nächsten Kapiteln für die nicht-konservative WSVR erweitert.

**Lemma 2.2 [Stabilisierbarkeit eines LTI-Systems mittels Ljapunov-Funktionen].** *Gegeben sei das LTI-System*

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b}u, \quad \mathbf{x} \in \mathbb{R}^n, u \in \mathbb{R}, \mathbf{A} \in \mathbb{R}^{n\times n}, \mathbf{b} \in \mathbb{R}^n,$$

*welches vollständig beschrieben durch das Paar* $(\mathbf{A},\mathbf{b})$ *ist. Folgende Aussagen sind äquivalent:*

[1)] Die Systemmatrix des Gesamtsystems hat eine obere Dreiecksform. Somit sind ihre Eigenwerte durch die Eigenwerte der Matrizen $\mathbf{A}_1$ und $\mathbf{A}_2$ bestimmt.

***i)*** *Das Paar* $(\mathbf{A},\mathbf{b})$ *ist durch ein lineares Zustandsregelgesetz* $u(\mathbf{x}) = -\mathbf{k}^\top\mathbf{x}$ *stabilisierbar.*

***ii)*** $\exists\, \mathbf{P} \succ \mathbf{0}$*, sodass* $\mathbf{AP} + \mathbf{PA}^\top \prec \mathbf{bb}^\top$*.*

*Beweis.* Das Lemma kann als Spezialfall des Satzes 3.1. aus [39] gesehen werden. Der Beweis wird daher an dieser Stelle weggelassen. □

Im Fall von LTI-Systemen mit Stellgrößenbeschränkung ist das Problem der Stabilisierbarkeit durch die Beschränkung der Stellgröße erschwert. Dies gilt natürlich nicht für Systeme, die bereits stabil sind. Die Schwierigkeit entsteht bei instabilen Systemen, da es für eine große Auslenkung von der Ruhelage möglich ist, dass die begrenzte Stellgröße nicht ausreicht um das System zu stabilisieren. Daher kann nicht jedes vollständig steuerbare System durch eine beschränkte Stellgröße im gesamten Zustandsraum, d.h. global, stabilisiert werden.

Der Bereich im Zustandsraum, der damit stabilisierbar ist, wird *null-steuerbare* Region, vgl. [35], genannt. Diese bildet zwar für stabile und semi-stabile[2)] LTI-Systeme den gesamten Zustandsraum, für instabile Systeme ist die Region jedoch beschränkt, konvex und offen, vgl. z.B. [35, Proposition 2.2.1]. Die Region beinhaltet dabei die Ruhelage des Systems. Beispiele für die Bestimmung dieser Region für lineare Systeme zweiter und dritter Ordnung kann in [35, Abschnitt 2.3] gefunden werden.

Im Fall nicht vollständig steuerbarer Systeme, wobei der nichtsteuerbare Systemteil asymptotisch stabil ist, wird die Region *asymptotisch-null-steuerbar* genannt, vgl. [35, Abschnitt 2.7]. Diese beinhaltet die null-steuerbare Region des steuerbaren Systemteils und den gesamten Unterraum des nichtsteuerbaren aber asymptotisch stabilen Systemteils.

Im Folgenden beschränken wir uns auf vollständig steuerbare LTI-Systeme mit Stellgrößenbeschränkung. Da die hier untersuchten Regelmethoden nichtlinear sind, ist Lemma 2.2 nicht mehr anwendbar und muss erweitert werden. Dabei wird der im nächsten Unterabschnitt vorgestellte Satz aus [1, 2] verwendet.

[2)] Bei linearen Systemen heißt ein System *semi-stabil*, falls dessen Systemmatrix Eigenwerte in der geschlossenen linken Halbebene besitzt. Dies umfasst auch die Imaginärachse und, im Gegensatz zum Fall grenzstabiler Systeme, die rein imaginären Eigenwerte können eine beliebige Vielfachheit besitzen. Die grenzstabilen Systeme bilden also ein Spezialfall semi-stabiler Systeme.

## 2.2 Stabilität mittels impliziter Ljapunov-Funktionen (iLF)

Auch implizite Ljapunov-Funktionen können für die Überprüfung der Stabilität dynamischer Systeme verwendet werden. Diese entstehen beispielsweise durch eine Einteilung des Zustandsraumes in Regionen, z.B. Ellipsoide, wobei diese in analytischer Form, d.h. $\mathbf{x}^\top \mathbf{P}(v)\mathbf{x} = 1$, mit $v \in [\underline{v},\overline{v}]$, vorliegen. In dieser Arbeit ist $\mathbf{P}(v)$ beispielsweise eine polynomiale Matrix in $v$. Die zustandsabhängige Variation des Parameters $v$, d.h. $v(\mathbf{x})$, kann zwar als Ljapunov-Funktion fungieren, diese kann jedoch in den meisten Fällen nicht explizit angegeben werden. Folgender Satz aus [2] stellt hinreichende Bedingungen für die asymptotische Stabilität der Ruhelage eines nichtlinearen Systems mittels impliziter Ljapunov-Funktionen dar. Der Satz verknüpft die direkte Methode von Ljapunov mit dem Satz über implizite Funktionen, vgl. dazu [18].

**Satz 2.3** [Vgl. [2]] *Gegeben sei die stetige Funktion* $\mathbf{h}(\mathbf{x})$ *und das System* $\dot{\mathbf{x}} = \mathbf{h}(\mathbf{x})$*, mit* $\mathbf{x} \in \mathbb{R}^n$ *und der Ruhelage* $\mathbf{x}_\mathrm{R} = \mathbf{0}$ *und eindeutiger Lösung für jeden Anfangswert, sowie eine stetige und differenzierbare Funktion*

$$g(\mathbf{x},v) : \mathcal{V}(0) \to \mathbb{R}, \quad \mathcal{V}(0) := \{(\mathbf{x},v) | \mathbf{x} \in \mathcal{U}_0 \backslash \{\mathbf{0}\} \subseteq \mathcal{B}_\delta(\mathbf{0}), 0 < v < 1\},$$

*welche folgende Bedingungen erfüllt:*

$$\begin{aligned}
&(i) && \text{Aus } g(\mathbf{x},v) = 0 \text{ folgt: } \mathbf{x} = \mathbf{0} \Leftrightarrow v \to 0^+,\\
&(ii) && \lim_{v\to 0^+} g(\mathbf{x},v) > 0 \text{ und } \lim_{v\to 1^-} g(\mathbf{x},v) < 0, \forall\, \mathbf{x} \in \mathcal{U}_0 \backslash \{\mathbf{0}\},\\
&(iii) && -\infty < \partial_v g(\mathbf{x},v) < 0, \quad \forall (\mathbf{x},v) \in \mathcal{V}(0),\\
&(iv) && \partial_t g(\mathbf{x}(t),v) < 0, \quad \forall (\mathbf{x},v) \in \mathcal{V}(0).
\end{aligned}$$

*Dann ist die Ruhelage* $\mathbf{x}_\mathrm{R} = \mathbf{0}$ *asymptotisch stabil, und die Gebiete*

$$\mathcal{G}(v) := \{\mathbf{x} \in \mathbb{R}^n \,|\, g(\mathbf{x},v) < 0\} \subseteq \mathcal{U}_0$$

*sind verschachtelt und kontraktiv invariant für alle* $v \in (0,1)$*.*[a)]

[a)] Vgl. Def. 15 (Anhang) bzw. Def. 14 (Anhang).

*Beweis.* Der Beweis des Satzes bezüglich der asymptotischen Stabilität der Ruhelage kann in [2, Satz 4] gefunden werden. Die kontraktive In-

varianz der Gebiete $\mathcal{G}(v)$ wird in [2, Satz 5] und deren Verschachtelung in [2, Abschnitt III] nachgewiesen. Im Folgenden wird daher lediglich die Beweisidee für die asymptotische Stabilität der Ruhelage skizziert.

Die Bedingungen $(ii)$ und $(iii)$ stellen sicher,[3)] dass die Selektionsstrategie $g(\mathbf{x},v) = 0$ eine eindeutige Lösung für jedes $\mathbf{x} \in \mathcal{U}_0 \backslash \{\mathbf{0}\}$ hat, welche eine stetige Funktion $v = v(\mathbf{x})$ ist. Darüber hinaus stellt die Funktion $0 < v(\mathbf{x}) < 1$, welche durch $g(\mathbf{x},v) = 0$ implizit definiert ist, und aufgrund der Bedingung $(i)$ für $\mathbf{x} = \mathbf{0}$ stetig erweiterbar mit $v(\mathbf{0}) = 0$ ist, eine implizite Ljapunov-Funktion des Systems dar. Dies resultiert aus der Tatsache, dass $v(\mathbf{x}) > 0$ und $\dot{v}(\mathbf{x}) < 0$, da auf Grund der Bedingungen $(iii)$ und $(iv)$ gilt, dass

$$\dot{v}(\mathbf{x}(t)) = -\frac{\partial_t g(\mathbf{x}(t),v)}{\partial_v g(\mathbf{x},v)} < 0, \quad \forall\, (\mathbf{x},v) \in \mathcal{V}(0). \tag{2.1}$$

□

**Korollar 2.4 [Vgl. [2]].** *Alle Trajektorien, die auf dem Rand eines Gebietes $\mathcal{G}(v)$ - d.h. für $\mathbf{x}(0) \in \partial\mathcal{G}(v)$, mit $v \in (0,1)$ - starten, laufen für $t > 0$ in das Gebiet hinein.*

*Beweis.* Dies folgt unmittelbar aus Bedingung $(iv)$ des Satzes 2.3, da für ein beliebiges $\Delta t > 0$ gilt $g(\mathbf{x}(t+\Delta t),v) < g(\mathbf{x}(t),v) = 0$, $\forall\, \mathbf{x} \in \partial\mathcal{G}(v)$, $v \in (0,1)$. Der Zustand $\mathbf{x}(t+\Delta t)$ wird in das Ellipsoid $\mathcal{G}(v)$ hineinlaufen. □

Unter Verwendung des Satzes 2.3 werden in den nächsten Kapiteln dieses Teils der Arbeit nicht-konservative Stabilitätsbedingungen für weiche strukturvariable Regelungen (WSVR) vorgestellt. Kapitel 3 beschäftigt sich mit dem Entwurf nicht-konservativer *klassischer* WSVR mittels iLF und Kapitel 4 führt die *invers-polynomiale* WSVR ein. Anschließend werden die vorgestellten Regelgesetze bezüglich der Konvergenzrate der damit verbundenen Ljapunov-Funktion optimiert. Die somit entstandere *konvergenzoptimale* Regelung wird in Kapitel 5 vorgestellt. Schließlich zeigt Kapitel 6 eine einfache Erweiterung der Regelgesetze für Regelstreckenensembles.

[3)] Vgl. [2, Satz 3].

# 3 Die *klassische* WSVR mittels iLF

Die erste weiche strukturvariable Regelung mittels impliziter Ljapunov-Funktionen wurde in [2] vorgestellt und geht auf [42] zurück. Diese Regelungsmethode wird im Weiteren als *klassische* WSVR bezeichnet. Es handelt sich um eine kontinuierlich parameterabhängige Zustandsrückführung, wobei der Parameter gleichzeitig eine implizite Ljapunov-Funktion des Systems darstellt. Der Parameter teilt darüber hinaus den stabilisierbaren Zustandsraum in infinitesimal dicht ineinander verschachtelte Ellipsoide, vgl. Def. 15 (Anhang). Dabei entspricht jedem Ellipsoidenrand ein eindeutiger Parameterwert, und je kleiner der Ellipsoid ist, desto höher ist die mit dessem Rand verbundene Reglerverstärkung.

Im Abschnitt 3.1 wird die Definition einer *klassischen* WSVR mittels iLF aus [2] erweitert und im Abschnitt 3.2 wird das erste Hauptergebnis dieser Arbeit, die nicht-konservativen Stabilitätsbedingungen der *klassischen* WSVR mittels iLFs vorgestellt. Das Kapitel endet mit dem Abschnitt 3.3 über mögliche Entwurfsschritte des vorgestellten Reglers.

## 3.1 Definition einer *klassischen* WSVR mittels iLF

Betrachtet werden lineare Strecken mit Stellgrößenbeschränkung in Steuerungsnormalform[1)], gegeben durch

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b}u, \quad \mathbf{x} \in \mathbb{R}^n, u \in \mathbb{R}, |u| \leq 1, \mathbf{A} \in \mathbb{R}^{n\times n}, \mathbf{b} \in \mathbb{R}^n. \tag{3.1}$$

Bei der strukturvariablen Regelung wird zwischen verschiedenen Regelgesetzen während des Ausregelvorgangs umgeschaltet. Dabei unterscheidet man zwischen parameter- und strukturumschaltenden Regelgesetzen. In dieser Arbeit wird eine Unterklasse der Ersteren betrachtet. Eine Übersicht solcher Regelungen bietet [4]. Die erste systematisch entwickelte Regelung dieser Art findet sich in [42]. Bei der untersuchten Regelung lautet

[1)] Dies ist keine Einschränkung, da jede steuerbare Regelstrecke in diese Form transformiert werden kann.

das Regelgesetz in allgemeiner Form

$$u(\mathbf{x}) = -f(\mathbf{x},v(\mathbf{x})), \tag{3.2}$$

wobei der Parameter $v \in \mathbb{R}$ aus einer vorgegebenen Selektionsgleichung $g(v,\mathbf{x}) = 0$ implizit bestimmt wird. Ein Regelgesetz aus dieser Klasse ist die nichtsättigende implizite WSVR, die auf [1] zurückgeht. Diese wird als *klassische* WSVR bezeichnet. Es handelt sich dabei um eine vom Parameter $v$ abhängige lineare Zustandsrückführung[2)]

$$u(\mathbf{x}) = -\mathbf{h}_v^\top \mathbf{x}. \tag{3.3}$$

Das Regelgesetz ist so ausgelegt, dass während des Ausregelvorgangs mit kleiner werdendem Abstand zur Ruhelage immer mehr Einfluß auf die Strecke ausgeübt wird, d.h., dass der Abstand zwischen den *momentanen Eigenwerten* des geschlossenen Regelkreises und denjenigen der Regelstrecke immer größer wird, aber die Stellgrößenbeschränkung nicht überschritten wird. Folglich wird der Stellgrößenbereich bei kleinen Auslenkungen von der Ruhelage im Vergleich zur linearen Regelung besser ausgenutzt. Betrachtet man den geschlossenen Regelkreis bei der Variation der Reglerverstärkung, so werden die *momentanen Eigenwerte* auf (vorgebbare) Bahnen *wandern*. Strahlenförmige Bahnen, wie in Bild 3.1 (links) gezeigt, können durch das folgende Regelgesetz [1] festgelegt werden:

$$\begin{aligned} &\mathbf{h}_v = \mathbf{D}_v^{-1}\hat{\mathbf{a}} - \mathbf{a},\ \mathbf{a}^\top = [0\cdots 0\ 1]\mathbf{A}, \\ &\hat{\mathbf{a}}^\top = [0\cdots 0\ 1]\hat{\mathbf{A}}_1,\ \hat{\mathbf{A}}_1 = \mathbf{A} - \mathbf{b}\mathbf{h}_1^\top, \\ &\mathbf{D}_v = \mathrm{diag}(v^n,v^{n-1},\ldots,v). \end{aligned} \tag{3.4}$$

In der allgemeinen Form ist die Selektionsgleichung auf einer Menge

$$\mathcal{V}_0 := \{(\mathbf{x},v)|\mathbf{x} \in U_0\backslash\{\mathbf{0}\}, 0 < v < \overline{v}\}$$

definiert, wobei $U_0 \subseteq \mathcal{B}_\delta(\mathbf{0})$ eine Umgebung der Ruhelage ist, und sie lautet

$$g(v,\mathbf{x}) = 0. \tag{3.5}$$

Eine spezielle Form der Gleichung (3.5) ist

$$g(\mathbf{x},v) = \mathbf{x}^\top \mathbf{P}_v \mathbf{x} - 1 = 0. \tag{3.6}$$

[2)] Der Übersichtlichkeit halber wird die Parameterabhängigkeit von $v$ als Index dargestellt, d.h. z.B. $\mathbf{h}_v = \mathbf{h}(v(\mathbf{x}))$.

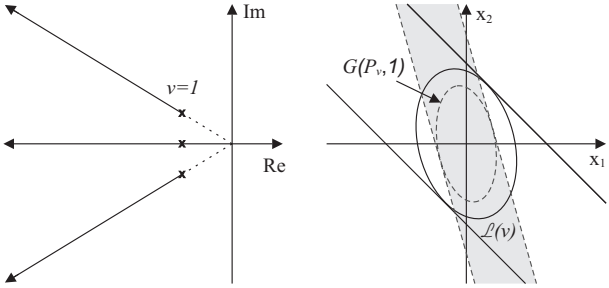


**Bild 3.1:** *Momentane Eigenwerte* (links) und kontraktiv invariante Gebiete (rechts) für den Fall $\mathbf{x} \in \mathbb{R}^2$, mit $v \in (0,1]$.

Folglich erzeugt die Selektionsgleichung ellipsoidale Gebiete im Zustandsraum, die durch die vom Parameter $v$ abhängige Matrix $\mathbf{P}_v$ skaliert und gedreht werden.

Die Veränderung des Selektionsparameters $v(\mathbf{x}) : U_0 \backslash \{\mathbf{0}\} \rightarrow (0,\overline{v}]$ ist an die Ränder der ellipsoidalen Gebiete $\partial\mathcal{E}(\mathbf{P}_v,1) \subset \mathbb{R}^n$ gekoppelt. Jedem Gebiet entspricht ein eindeutiger Wert des Parameters $v \in (0,\overline{v}]$, dem äußeren Gebiet der Wert $v(\mathbf{x}) = \overline{v}$, dem innersten Gebiet der Wert $v(\mathbf{0}) \rightarrow 0^+$. Bild 3.1 (rechts) veranschaulicht zwei solche Gebiete für den Fall $n = 2$. Entsprechend wird während des Ausregelvorgangs für jedes $\mathbf{x} \in U_0 \backslash \{\mathbf{0}\}$ der Wert des Parameters $v(\mathbf{x})$ aus der Selektionsgleichung numerisch bestimmt.

Die Matrix $\mathbf{P}_v$ kann in der Form $\mathbf{P}_v = e_v \mathbf{D}_v^{-1} \mathbf{P}_1 \mathbf{D}_v^{-1}$, mit $\mathbf{D}_v = \mathrm{diag}(v^n,\ldots,v^2,v)$, $e_v = \mathbf{h}_v^\top \mathbf{P}_v^{-1} \mathbf{h}_v$ und $\mathbf{P}_1 \succ \mathbf{0}$ gewählt werden, sodass sich die Überprüfung der kontraktiven Invarianz der ellipsoidalen Gebiete $\mathcal{E}(\mathbf{P}_v,1)$ auf die des äußeren Gebietes $\mathcal{E}(\mathbf{P}_1,1)$ reduziert. Die Skalierungsfunktion $e_v$ ist dabei so bestimmt, dass das jeweilige Gebiet $\mathcal{E}(\mathbf{P}_v,1)$ unter der Bedingung

$$\mathcal{E}(\mathbf{P}_v,1) \subset \left\{\mathbf{x} \in \mathbb{R}^n \middle| \, |\mathbf{h}^\top(v)\mathbf{x}| \leq 1\right\} =: \mathcal{L}(v) \tag{3.7}$$

maximiert wird. Die Gebiete $\mathcal{L}(v)$ werden ebenfalls im Bild 3.1 (rechts) gezeigt.

Die Selektionsgleichung kann man vereinfachen [41], indem man die Skalierungsfunktion $e(v)$ weglässt, d.h.

$$\mathbf{P}_v = \mathbf{D}_v^{-1} \mathbf{P}_1 \mathbf{D}_v^{-1}. \tag{3.8}$$

Durch diese Vereinfachung kann man zwar auch Regler für instabile Strecken entwerfen [41], es ist jedoch nicht mehr sichergestellt, dass die Stellgröße die Beschränkung nicht überschreitet. Dies muss durch eine zusätzliche Bedingung beim Regelungsentwurf sichergestellt werden.

Darüber hinaus ist es möglich, die Matrixpotenz von $-1$ auf eine Zahl zwischen $-1$ und 0 zu reduzieren. Dies ergibt eine Matrix[3)]

$$\mathbf{P}_v = \mathbf{D}_v^{-r}\mathbf{P}_1\mathbf{D}_v^{-r}, \quad r \in (0,1], \tag{3.9}$$

wobei der Skalar $r \in (0,1]$ konstant ist. Diese Matrix wird im Folgenden verwendet.

In dieser Arbeit ist die *Existenz* eines solchen Reglers für das LTI-System aus Gl. (3.1) definitionsgemäß äquivalent zu der Existenz einer Matrix $\mathbf{P}_1 \succ \mathbf{0}$, sodass die Gebiete

$$\mathcal{E}(v) := \{\mathbf{x} \in \mathbb{R}^n | \mathbf{x}^\top \mathbf{D}_v^{-r}\mathbf{P}_1\mathbf{D}_v^{-r}\mathbf{x} < 1\}$$

für alle $v \in (0,1]$ (infinitesimal dicht ineinander) verschachtelt und kontraktiv invariant sind, und $u(\mathbf{x}) \leq 1$, $\forall\, \mathbf{x} \in \mathcal{E}(v)$ mit $v \in (0,1]$ gilt. Die kontraktive Invarianz eines Gebietes zeichnet sich dadurch aus, dass Trajektorien, die in ein solches Gebiet hineinlaufen, das Gebiet nicht mehr verlassen und asymptotisch in die Ruhelage konvergieren.[4)] Zwei Gebiete sind verschachtelt, falls ein Gebiet vollständig innerhalb des anderen Gebietes ist und deren Ränder keine gemeinsame Punkte haben.[5)]

Basierend auf dieser Definition der *klassischen* WSVR werden im Folgenden notwendige und hinreichende Stabilitätsbedingungen vorgestellt.

## 3.2 Nicht-konservative Stabilitätsbedingungen

Der folgende Satz mitsamt einer konstruktiven Beweismethode stellt notwendige und hinreichende Bedingungen für die Existenz einer stabilisierenden *klassischen* WSVR mittels iLFs dar. Für eine beliebige lineare Strecke ergibt sich daraus (im Existenzfall) ein stabilisierendes Regelgesetz. Der Satz wurde in [61] zum ersten Mal vorgestellt.

[3)] Diese Form wurde beispielsweise in [55] verwendet.
[4)] Vgl. Def. 14 (Anhang).
[5)] Vgl. Def. 15 (Anhang).

**Satz 3.1** *Gegeben sei das folgende LTI-System in Steuerungsnormalform mit einer Eingangsgröße und Stellgrößenbeschränkung*

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b}u, \quad \mathbf{x} \in \mathbb{R}^n,\, u \in \mathbb{R},\, |u| \leq 1. \tag{3.10}$$

*Folgende Aussagen sind äquivalent:*

*i) Für jedes* $\varepsilon \in (0,1)$ *existieren ein Vektor* $\hat{\mathbf{a}} \in \mathbb{R}^n$*, ein Skalar* $r \in (0,1]$ *und eine Matrix* $\mathbf{P}_1 \in \mathbb{P}^n$*, sodass für alle* $v \in (0,1]$ *die Gebiete*

$$\mathcal{E}_\Delta(v) := \left\{ \mathbf{x} \in \mathbb{R}^n \,\middle|\, g_\Delta(\mathbf{x},v) := \mathbf{x}^\top \mathbf{D}_v^{-r} \mathbf{P}_1 \mathbf{D}_v^{-r} \mathbf{x} - 1 < 0 \right\} \tag{3.11}$$

*verschachtelt und kontraktiv invariant für das System in Gl.* (3.10) *mit dem Regelgesetz*

$$u = -\mathbf{k}_v^\top \mathbf{x}, \quad \mathbf{k}_v^\top := \mathbf{D}_v^{-r} \hat{\mathbf{a}} - \mathbf{a}, \tag{3.12}$$

*mit* $\mathbf{D}_v := \operatorname{diag}(v^n, \ldots, v)$ *und* $\mathbf{a} := -[0 \;\; 0 \cdots 1]\mathbf{A}$ *sind, wobei der Parameter* $v$ *implizit definiert durch die Gleichung*

$$g_\Delta(\mathbf{x},v) = 0$$

*ist. Darüber hinaus gilt*

$$|u| < 1, \quad \forall \mathbf{x} \in \mathcal{E}_\Delta(v),\; v \in [\varepsilon,1]. \tag{3.13}$$

*ii)* $\exists\, \mathbf{P} \in \mathrm{Sym}^n$*, sodass*

$$\mathbf{P} \succ \mathbf{0}, \tag{3.14}$$

$$\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^\top \prec \mathbf{b}\mathbf{b}^\top, \tag{3.15}$$

$$\mathbf{N}\mathbf{P} + \mathbf{P}\mathbf{N} \;\prec \mathbf{0}, \quad \mathbf{N} := \operatorname{diag}(-n, \ldots, -1). \tag{3.16}$$

*Falls ii) gilt, dann ist ein stabilisierendes Regelgesetz in i) gegeben durch*

$$\hat{\mathbf{a}} = \mathbf{a} + c\mathbf{P}^{-1}\mathbf{b}, \quad \mathbf{P}_1 = d\mathbf{P}^{-1}, \tag{3.17}$$

*wobei*

$$d > \mathbf{b}^\top \mathbf{P}^{-1} \mathbf{b}/4, \tag{3.18}$$

$$c = \nu \sqrt{d(\mathbf{b}^\top \mathbf{P}^{-1} \mathbf{b})^{-1}}, \nu \in \left[ \sqrt{\frac{\mathbf{b}^\top \mathbf{P}^{-1} \mathbf{b}}{4d}}, 1 \right), \tag{3.19}$$

*gilt, und* $r \in (0,1]$ *die Lösung des quasi-konvexen Optimierungsproblems*

$$\max_{r\in(0,1]} \; r, \text{ sodass}$$
$$(\hat{\mathbf{a}} - \mathbf{D}_w\mathbf{a})^\top \mathbf{P} (\hat{\mathbf{a}} - \mathbf{D}_w\mathbf{a}) < d, \forall w \in [\varepsilon^r,1], \tag{3.20}$$

*ist.*

*Bemerkung* 3.1. Der Parameter $d$ aus Gl. (3.18) kann zur Skalierung der Matrix $\mathbf{P}^{-1}$, und somit des kontraktiv invarianten Gebietes $\mathcal{E}_\Delta(1)$ verwendet werden. Ein kleinerer Wert von $d$ ergibt ein größeres Gebiet. Der Parameter $c$ aus Gl. (3.19) skaliert den linearen Bereich der Sättigung $\mathcal{L}(-\mathbf{k}_1^\top\mathbf{x},1)$, sodass dieser das kontraktiv invariante Gebiet $\mathcal{E}_\Delta(1)$ beinhaltet. Somit ergibt sich für $w = 1$ ein nichtsättigendes Regelgesetz. Ein größerer Wert von $\nu$ ergibt einen kleineren linearen Bereich der Sättigung. Bedingungen (3.18) und (3.19) können in vereinfachter Form als

$$\frac{d}{c^2} \geq (\mathbf{b}^\top\mathbf{P}^{-1}\mathbf{b}), \quad c > \frac{1}{2}$$

geschrieben werden. △

*Bemerkung* 3.2. Sind die Parameter $d$, $\nu$ und $c$ gegeben, dann kann Gl. (3.20) in eine äquivalente parameterunabhängige LMI-Bedingung transformiert werden. Darüber hinaus existiert immer ein Skalar $r \in (0,1]$, sodass Gl. (3.20) erfüllt ist. Dies resultiert aus der Tatsache, dass Gl. (3.20) für $w = 1$ erfüllt ist und die linke Seite für jedes $w \in [\varepsilon^r,1]$ endlich ist. △

*Bemerkung* 3.3. Falls Gl. (3.20) für einen Skalar $r^* \in (0,1]$ erfüllt ist, dann ist sie auch für $r \leq r^*$ erfüllt. Dies resultiert aus der Tatsache, dass $\varepsilon \in (0,1)$ und folglich, dass $\varepsilon^r \geq \varepsilon^{r^*}$, $\forall\, r \leq r^*$. Der Skalar $r \in (0,1]$ kann daher mit Hilfe des Bisektionsverfahrens berechnet werden.[6)] △

*Bemerkung* 3.4. Der Übersichtlichkeit halber wird der Beweis, der weiter unten erfolgt, vorerst kurz skizziert. In dem ersten Teil des Beweises wird $ii) \Rightarrow i)$ gezeigt, d.h. aus $ii)$ folgt $i)$. Mit Hilfe des Satzes 2.3 wird dabei bewiesen, dass die Bedingungen (3.14)-(3.16) aus Punkt $ii)$, unter Verwendung des Regelgesetzes aus Gl. (3.12), (3.17)-(3.20), hinreichend dafür sind, dass die ellipsoidalen Gebiete $\mathcal{E}_\Delta(v)$ aus Gl. (3.11) verschachtelt und kontraktiv invariant sind. Im zweiten Teil des Satzes wird die Notwendigkeit der Bedingungen (3.14)-(3.16) gezeigt, d.h. $i) \Rightarrow ii)$. Dies wird mit Hilfe von Finsler's Lemma, vgl. Satz A.5 (Anhang), bewiesen. △

[6)] Vgl. [17, Algorithm 4.1].

*Beweis.* $ii) \Rightarrow i)$ Aus Gl. (3.15) folgt

$$\begin{aligned}(\mathbf{A} - c\mathbf{b}\mathbf{b}^\top\mathbf{P}^{-1})\mathbf{P} + \mathbf{P}(\mathbf{A} - c\mathbf{b}\mathbf{b}^\top\mathbf{P}^{-1})^\top \\ = \mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^\top - 2c\mathbf{b}\mathbf{b}^\top \preceq \mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^\top - \mathbf{b}\mathbf{b}^\top \prec \mathbf{0},\end{aligned} \tag{3.21}$$

da wegen Gl. (3.18)-(3.19) $c \geq 1/2$ ist. Aus Lemma A.4 (Anhang), Seite 178, und Gl. (3.14) folgt, dass das LTI System

$$\dot{\mathbf{x}} = (\mathbf{A} - c\mathbf{b}\mathbf{b}^\top\mathbf{P}^{-1})\mathbf{x} \tag{3.22}$$

asymptotisch stabil ist. Nach links und rechts Multiplizieren der Gl. (3.21) mit der nichtsingulären Matrix $\mathbf{P}^{-1} \succ \mathbf{0}$ folgt darüber hinaus

$$\mathbf{P}^{-1}(\mathbf{A} - c\mathbf{b}\mathbf{b}^\top\mathbf{P}^{-1}) + (\mathbf{A} - c\mathbf{b}\mathbf{b}^\top\mathbf{P}^{-1})^\top\mathbf{P}^{-1} \prec \mathbf{0}.$$

Daher ist die quadratische Funktion $V(\mathbf{x}) = \mathbf{x}^\top\mathbf{P}^{-1}\mathbf{x}$ eine gültige Ljapunov-Funktion des Systems $\dot{\mathbf{x}} = (\mathbf{A} - c\mathbf{b}\mathbf{b}^\top\mathbf{P}^{-1})\mathbf{x}$.

Das LTI-System aus Gl. (3.22) beschreibt den geschlossenen Regelkreis aus $i)$ für $v = 1$. Dies resultiert aus Gl. (3.10) und (3.12), da

$$\begin{aligned}\dot{\mathbf{x}} &= (\mathbf{A} - \mathbf{b}\mathbf{k}_v^\top)\mathbf{x} \\ &= [\mathbf{A} - \mathbf{b}(\mathbf{D}_v^{-r}\hat{\mathbf{a}} - \mathbf{a})^\top]\mathbf{x} \\ &= \begin{bmatrix} \mathbf{0}_{n-1,1} & \mathbf{I}_{n-1} \\ -\hat{\mathbf{a}}^\top\mathbf{D}_v^{-r} & \end{bmatrix}\mathbf{x} \\ &= \frac{1}{v^r}\mathbf{D}_v^r\begin{bmatrix} \mathbf{0}_{n-1,1} & \mathbf{I}_{n-1} \\ -\hat{\mathbf{a}}^\top & \end{bmatrix}\mathbf{D}_v^{-r}\mathbf{x} \\ &= \frac{1}{v^r}\mathbf{D}_v^r(\mathbf{A} - c\mathbf{b}\mathbf{b}^\top\mathbf{P}^{-1})\mathbf{D}_v^{-r}\mathbf{x}.\end{aligned} \tag{3.23}$$

Darüber hinaus ist die zeitliche Änderung der Funktion $g_\Delta(\mathbf{x},v)$ aus Gl. (3.11), mit $\mathbf{P}_1 = d\mathbf{P}^{-1}$, entlang einer Trajektorie des geschlossenen Regelkreises aus Gl. (3.23) gegeben durch

$$\begin{aligned}\frac{\partial g_\Delta(v)(\mathbf{x}(t),v)}{\partial t} &= \dot{\mathbf{x}}^\top\mathbf{D}_v^{-r}d\mathbf{P}^{-1}\mathbf{D}_v^{-r}\mathbf{x} + \mathbf{x}^\top\mathbf{D}_v^{-r}d\mathbf{P}^{-1}\mathbf{D}_v^{-r}\dot{\mathbf{x}} \\ &= \frac{d}{v^r}\mathbf{x}^\top\mathbf{D}_v^{-r}\left[(\mathbf{A} - \mathbf{b}\mathbf{b}^\top\mathbf{P}^{-1})^\top\mathbf{P}^{-1}\right. \\ &\quad \left. + \mathbf{P}^{-1}(\mathbf{A} - \mathbf{b}\mathbf{b}^\top\mathbf{P}^{-1})\right]\mathbf{D}_v^{-r}\mathbf{x} < 0, \forall v \in (0,1].\end{aligned} \tag{3.24}$$

Ferner ist die Matrix $\mathbf{NP}+\mathbf{PN}$ negativ definit. Somit folgt, dass die Matrix $\mathbf{NP}^{-1}+\mathbf{P}^{-1}\mathbf{N}$ auch negativ definit ist und, dass

$$\begin{aligned}\frac{\partial g_{\vartriangle}(\mathbf{x},v)}{\partial v} &= \mathbf{x}^{\top}\frac{\partial(\mathbf{D}_v^{-r}\mathbf{P}^{-1}\mathbf{D}_v^{-r})}{\partial v}\mathbf{x}\\ &= \frac{d}{v^r}\mathbf{x}^{\top}\mathbf{D}_v^{-r}(\mathbf{NP}^{-1}+\mathbf{P}^{-1}\mathbf{N})\mathbf{D}_v^{-r}\mathbf{x} < 0, \forall v \in (0,1], \qquad (3.25)\end{aligned}$$

gilt. Aus Gl. (3.24) und Gl. (3.25) folgt schließlich, dass die Gebiete $\mathcal{E}_{\vartriangle}(v)$ verschachtelt und kontraktiv invariant sind.[7)] Somit sind die Bedingungen aus Gl. (3.14)-(3.16) hinreichend für die Existenz einer weichen strukturvariablen Regelung mittels impliziter Ljapunov-Funktionen. Ferner ist die implizite Funktion $0 < v(\mathbf{x}) \leq 1$ für $\mathbf{x} = \mathbf{0}$ stetig erweiterbar mit $v(\mathbf{0}) = 0$ und folglich eine zulässige Ljapunov-Funktion des Systems.

Mit der Notation $\mathbf{P}_v := \mathbf{D}_v^{-r}\mathbf{P}_1\mathbf{D}_v^{-r} = \mathbf{D}_v^{-r}d\mathbf{P}^{-1}\mathbf{D}_v^{-r}$ ist die Bedingung aus Gl. (3.13), $|u| < 1$, äquivalent zu

$$\max_{\mathbf{x}^{\top}\mathbf{P}_v\mathbf{x}\leq 1} |\mathbf{k}_v^{\top}\mathbf{x}| = \sqrt{\mathbf{k}_v^{\top}\mathbf{D}_v^r d^{-1}\mathbf{P}\mathbf{D}_v^r\mathbf{k}_v} < 1, \quad \forall v \in [\varepsilon,1],$$

und folglich zu

$$\mathbf{k}_v^{\top}\mathbf{D}_v^r\mathbf{P}\mathbf{D}_v^r\mathbf{k}_v < d, \quad \forall v \in [\varepsilon,1]. \qquad (3.26)$$

Der Term auf der linken Seite der Gl. (3.26) ist weiterhin äquivalent zu

$$\begin{aligned}\mathbf{k}_v^{\top}\mathbf{D}_v^r\mathbf{P}\,\mathbf{D}_v^r\mathbf{k}_v &= (\mathbf{D}_v^{-r}\hat{\mathbf{a}}-\mathbf{a})^{\top}\mathbf{D}_v^r\mathbf{P}\mathbf{D}_v^r(\mathbf{D}_v^{-r}\hat{\mathbf{a}}-\mathbf{a})\\ &= (\hat{\mathbf{a}}-\mathbf{D}_v^r\mathbf{a})^{\top}\mathbf{P}(\hat{\mathbf{a}}-\mathbf{D}_v^r\mathbf{a})\\ &= (\hat{\mathbf{a}}-\mathbf{D}_w\mathbf{a})^{\top}\mathbf{P}(\hat{\mathbf{a}}-\mathbf{D}_w\mathbf{a}), \quad w := v^r.\end{aligned}$$

Folglich garantiert Gl. (3.20), dass die Stellgrößenbeschränkung eingehalten ist, d.h. $|u| < 1$, $\forall \mathbf{x} \in \mathcal{E}_{\vartriangle}(w)$, $w \in [\varepsilon^r,1]$.

*i)* $\Rightarrow$ *ii)* Falls *i)* gilt, dann ist das Gebiet $\mathcal{E}_{\vartriangle}(1)$ kontraktiv invariant für das System $\dot{\mathbf{x}} = (\mathbf{A}-\mathbf{b}\mathbf{k}_1^{\top})\mathbf{x}$ und die Funktion $V(\mathbf{x}) := \mathbf{x}^{\top}\mathbf{P}_1\mathbf{x}$ ist eine gültige Ljapunov-Funktion des Systems. Folglich ist dieses System (global) asymptotisch stabil. Für $\mathbf{P} := \mathbf{P}_1^{-1} \succ \mathbf{0}$ folgt

$$(\mathbf{A}-\mathbf{b}\mathbf{k}_1^{\top})\mathbf{P}+\mathbf{P}(\mathbf{A}-\mathbf{b}\mathbf{k}_1^{\top})^{\top} \prec \mathbf{0}. \qquad (3.27)$$

[7)] Dies folgt unmittelbar aus dem Satz 2.3.

Sei $\mathbf{B}^\perp \in \mathbb{R}^{(n-1)\times n}$ eine Basis für den Nullraum von $\mathbf{b}$, sodass $\mathbf{B}^\perp\mathbf{b} = \mathbf{0}$ und $\mathbf{B}^\perp(\mathbf{B}^\perp)^\top \succ \mathbf{0}$ gilt. Nach Multiplizieren der Gl. (3.27) mit der Matrix $\mathbf{B}^\perp$ (von links) und der Matrix $(\mathbf{B}^\perp)^\top$ (von rechts) folgt, dass

$$\begin{aligned}&\mathbf{B}^\perp(\mathbf{A} - \mathbf{b}\mathbf{k}_1^\top)\mathbf{P}(\mathbf{B}^\perp)^\top + \mathbf{B}^\perp\mathbf{P}(\mathbf{A} - \mathbf{b}\mathbf{k}_1^\top)^\top(\mathbf{B}^\perp)^\top \\ &\quad = \mathbf{B}^\perp(\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^\top)(\mathbf{B}^\perp)^\top \prec 0. \end{aligned} \tag{3.28}$$

Da dies nicht offensichtlich ist, wird im Folgenden eine kurze Erklärung hinzugefügt. Weil $\mathbf{B}^\perp(\mathbf{B}^\perp)^\top \succ \mathbf{0}$ gilt, und dies äquivalent zu dem Fakt ist, dass $\mathrm{Rang}(\mathbf{B}^\perp) = \mathrm{Rang}(\mathbf{B}^\perp(\mathbf{B}^\perp)^\top) = n-1$ ist, folgt, dass ein Vektor $\mathbf{d} \in \mathbb{R}^n_*$ existiert, sodass die Matrix $\mathbf{S} := \begin{bmatrix}\mathbf{d} & (\mathbf{B}^\perp)^\top\end{bmatrix}^\top \in \mathbb{R}^{n\times n}$ nichtsingulär ist. Unter Verwendung der Notation

$$\mathbf{M} := (\mathbf{A} - \mathbf{b}\mathbf{k}_1^\top)\mathbf{P} + \mathbf{P}(\mathbf{A} - \mathbf{b}\mathbf{k}_1^\top)^\top$$

folgt dann aus Gl. (3.27), dass

$$\mathbf{S}\mathbf{M}\mathbf{S}^\top = \begin{bmatrix}\mathbf{d}^\top \\ \mathbf{B}^\perp\end{bmatrix}\mathbf{M}\begin{bmatrix}\mathbf{d} & (\mathbf{B}^\perp)^\top\end{bmatrix} = \begin{bmatrix}\mathbf{d}^\top\mathbf{M}\mathbf{d} & \mathbf{d}^\top\mathbf{M}(\mathbf{B}^\perp)^\top \\ \mathbf{B}^\perp\mathbf{M}\mathbf{d} & \mathbf{B}^\perp\mathbf{M}(\mathbf{B}^\perp)^\top\end{bmatrix} \prec \mathbf{0}$$

gilt. Daraus folgt unmittelbar Gl. (3.28), welche der unteren rechten Ecke dieser Blockmatrix entspricht.[8)]

Aufgrund von Finslers' Theorem[9)] ist Gl. (3.28) weiterhin äquivalent zu dem Sachverhalt, dass ein Skalar $\mu \in \mathbb{R}$ existiert, sodass

$$\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^\top \prec \mu\mathbf{b}\mathbf{b}^\top,$$

mit

$$\begin{aligned}&\mu > \mathbf{d}^\top\left[\mathbf{Q} - \mathbf{Q}(\mathbf{B}^\perp)^\top(\mathbf{B}^\perp\mathbf{Q}(\mathbf{B}^\perp)^\top)^{-1}\mathbf{B}^\perp\mathbf{Q}\right]\mathbf{d}, \\ &\mathbf{Q} := \mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^\top, \\ &\mathbf{d}^\top := |b_r|^{-1}\mathbf{b}_l^+, \quad b_r \in \mathbb{R}\backslash\{\mathbf{0}\}, \mathbf{b}_l^+ \in \mathbb{R}^{1\times n},\end{aligned}$$

wobei das Tupel $(\mathbf{b}_l, b_r)$ eine Voll-Rang-Faktorisierung des Vektors $\mathbf{b}$ ist.[10)] Schließlich folgt Gl. (3.15) nach der Skalierung der Matrix $\mathbf{P}$ mit einem $\tilde{\mu} > \max\{0,\mu\}$.

[8)] Vgl. [8, Prop. 8.2.4].
[9)] Vgl. [68, Theorem 2.3.10].
[10)] Vgl. Def. 20 (Anhang).

Darüber hinaus sind die Gebiete $\mathcal{E}_\Delta(v)$ verschachtelt, d.h. $\mathcal{E}_\Delta(v_2) \subset \mathcal{E}_\Delta(v_1)$, $\forall\, 0 < v_2 < v_1 \leq 1$ und $\partial\mathcal{E}_\Delta(v_1) \cap \partial\mathcal{E}_\Delta(v_2) = \emptyset$.[11] Dies bedeutet, dass für $\mathbf{x} \in \partial\mathcal{E}_\Delta(v_1)$ folgt, dass $g_\Delta(\mathbf{x},v_2) > 0$ und $g_\Delta(\mathbf{x},v_1) = 0$, und daher dass

$$g_\Delta(\mathbf{x},v_2) > g_\Delta(\mathbf{x},v_1), \quad \forall\, 0 < v_2 < v_1 \leq 1, \mathbf{x} \in \partial\mathcal{E}_\Delta(v_1).$$

Folglich ist

$$\begin{aligned}\frac{\partial g_\Delta(\mathbf{x},v)}{\partial v} &= \frac{r}{v}\mathbf{x}^\top \mathbf{D}_v^{-r}(\mathbf{N}\mathbf{P}^{-1} + \mathbf{P}^{-1}\mathbf{N})\mathbf{D}_v^{-r}\mathbf{x} \\ &< 0,\ \forall \mathbf{x} \in \mathcal{E}_\Delta(v) \text{ und } v \in (0,1].\end{aligned}$$

Es folgt auch, dass $\mathbf{x}^\top(\mathbf{N}\mathbf{P}^{-1} + \mathbf{P}^{-1}\mathbf{N})\mathbf{x} < 0$, $\forall \mathbf{x} \in \mathcal{E}_\Delta(1)$. Zudem ist $(k\mathbf{x})^\top(\mathbf{N}\mathbf{P}^{-1} + \mathbf{P}^{-1}\mathbf{N})(k\mathbf{x}) < 0$, $\forall k \in \mathbb{R}$, $\mathbf{x} \in \mathcal{E}_\Delta(1)$, d.h. die Matrix $\mathbf{N}\mathbf{P}^{-1} + \mathbf{P}^{-1}\mathbf{N}$ ist negativ definit. Nach Links- und Rechtsmultiplizieren der Matrix mit der nichtsingulären Matrix $\mathbf{P}$ folgt schließlich auch Gl. (3.16). Daher sind die Bedingungen aus Gl. (3.14)-(3.16) auch notwendig für die Existenz einer weichen strukturvariablen Regelung mittels impliziter Ljapunov-Funktionen. □

**Transformation der Bedingung aus Gl.** (3.20) **in eine LMI**

Gl. (3.20) kann in eine äquivalente parameterunabhängige LMI transformiert werden. Die Transformation basiert auf Lemma 4.4 aus [78], das eine Generalisierung der in [38] vorgestellten $S$-Prozedur verwendet. Dabei wird gezeigt, dass Gl. (3.20) ein Matrixpolynom in der Variablen $w \in [\varepsilon^r,1]$ ist, welches in ein weiteres Matrixpolynom in der Variablen $z \in [-1,1]$ transformiert werden kann. Darauf basierend wird eine parameterunabhängige LMI-Bedingung eingeführt, welche notwendig und hinreichend dafür ist, dass das letzte Matrixpolynom positiv für alle $z \in [-1,1]$ ist. Dies wird im Folgenden gezeigt.

Da $\mathbf{P} \in \mathbb{P}^n$, ist Gl. (3.20) äquivalent zu[12]

$$\mathbf{M}_w := \begin{bmatrix} \mathbf{P}^{-1} & \hat{\mathbf{a}} - \mathbf{D}_w\mathbf{a} \\ (\hat{\mathbf{a}} - \mathbf{D}_w\mathbf{a})^\top & d \end{bmatrix} \succ \mathbf{0}, \quad \forall w \in [\varepsilon^r,1]. \tag{3.29}$$

[11] Vgl. Def. 15 (Anhang).
[12] Vgl. [8, Fact 8.15.5].

Die Matrix $\mathbf{M}_w$ kann in Form eines Matrixpolynoms geschrieben werden, d.h.

$$\mathbf{M}_w = \sum_{i=0}^{n} w^i \mathbf{M}_i, \;\; \mathbf{M}_i \in \mathrm{Sym}^{n+1}, \quad w \in [\varepsilon^r,1], \tag{3.30}$$

mit

$$\mathbf{M}_0 = \begin{bmatrix} \mathbf{P}^{-1} & \hat{\mathbf{a}} \\ \hat{\mathbf{a}}^\top & d \end{bmatrix},$$
$$\mathbf{M}_i = -\begin{bmatrix} \mathbf{0}_{n,n} & \mathbf{I}_n \\ \mathbf{a}^\top & \mathbf{0}_{1,n} \end{bmatrix} \begin{bmatrix} \mathbf{e}_i\mathbf{e}_i^\top & \mathbf{0}_{n,n} \\ \mathbf{0}_{n,n} & \mathbf{e}_i\mathbf{e}_i^\top \end{bmatrix} \begin{bmatrix} \mathbf{I}_n & \mathbf{0}_{n,1} \\ \mathbf{0}_{n,n} & \mathbf{a} \end{bmatrix}, \quad i = 1,\ldots,n,$$

wobei $\mathbf{e}_i^\top = [0,\ldots,1,\ldots,0]$, mit $\mathbf{e}_{i_{(i)}} = 1$, der i-te kanonische Einheitsvektor ist. Eine zweite Variablensubstitution mit $z := \frac{1}{\alpha}w - \frac{\beta}{\alpha}$, wobei $\alpha := (1-\varepsilon^r)/2$ und $\beta = (1+\varepsilon^r)/2$, wird angewandt, um das Intervall $w \in [\varepsilon^r,1]$ auf das Intervall $z \in [-1,1]$ abzubilden. Folglich ist das Matrixpolynom aus Gl. (3.30) äquivalent zu

$$\begin{aligned} \mathbf{M}_z &= \sum_{i=0}^{n} (\alpha z + \beta)^i \mathbf{M}_i = \sum_{i=0}^{n} \left( \sum_{j=0}^{i} \binom{i}{j} z^{i-j} \alpha^{i-j} \beta^j \right) \mathbf{M}_i \\ &= \sum_{k=0}^{n} z^k \tilde{\mathbf{M}}_k, \quad \tilde{\mathbf{M}}_k \in \mathrm{Sym}^{n+1}, \; z \in [-1,1], \end{aligned} \tag{3.31}$$

mit

$$\tilde{\mathbf{M}}_k = \sum_{i=k}^{n} \binom{i}{k} \alpha^k \beta^{i-k} \mathbf{M}_i. \tag{3.32}$$

Das Matrixpolynom in Gl. (3.31) kann darüber hinaus geschrieben werden als

$$\mathbf{M}_z = \left( \mathbf{z}^{[k+1]} \otimes \mathbf{I}_n \right)^\top \mathbf{M}_\Sigma \left( \mathbf{z}^{[k+1]} \otimes \mathbf{I}_n \right),$$

mit $\mathbf{M}_\Sigma \in \mathrm{Sym}^{(k+1)(n+1)}$, $z \in [-1,1]$, $k = \lceil n/2 \rceil$,[13] sowie $\mathbf{z}^{[k+1]} = \begin{bmatrix} 1 & z & \cdots & z^k \end{bmatrix}^\top$, wobei $\mathbf{M}_\Sigma$ eine symmetrische Tridiagonalmatrix mit den

[13] $\lceil m \rceil$ ist die kleinste ganze Zahl, die größer oder gleich $m$ ist.

Elementen[14)]

$$\mathbf{M}_{\Sigma_{(i,j)}}=\begin{cases}\tilde{\mathbf{M}}_{2(i-1)}, & i=j\leq k,\\ \tilde{\mathbf{M}}_{2k}, & i=j=k+1,\ k=\text{gerade},\\ \mathbf{0}_{n+1,n+1}, & i=j=k+1,\ k=\text{ungerade},\\ \frac{1}{2}\tilde{\mathbf{M}}_{2(l-1)+1}, & |i-j|=1,\ l=\min\{i,j\},\\ \mathbf{0}_{n+1,n+1}, & |i-j|>1.\end{cases}$$

ist. Gl. (3.29) ist daher äquivalent zu

$$\left(\mathbf{z}^{[k+1]}\otimes\mathbf{I}_n\right)^{\top}\mathbf{M}_{\Sigma}\left(\mathbf{z}^{[k+1]}\otimes\mathbf{I}_n\right)\succ\mathbf{0}, \tag{3.33}$$
$$\mathbf{z}^{[k+1]}=\begin{bmatrix}1 & z & \cdots & z^k\end{bmatrix}^{\top},\forall\, z\in[-1,1].$$

Gl. (3.33) ist dann und nur dann erfüllt, wenn[15)] zwei Matrizen $\mathbf{D}\in\mathbb{P}^{nk}$ und $\mathbf{G}\in\text{Skew}^{nk}$ existieren, sodass

$$\mathbf{M}_{\Sigma}\succ\begin{bmatrix}\mathbf{C}\\ \mathbf{J}\end{bmatrix}^{\top}\begin{bmatrix}-\mathbf{D} & \mathbf{G}\\ \mathbf{G} & \mathbf{D}\end{bmatrix}\begin{bmatrix}\mathbf{C}\\ \mathbf{J}\end{bmatrix}, \tag{3.34}$$

mit $\mathbf{C}=[\mathbf{I}_k\quad\mathbf{0}_{k,1}]\otimes\mathbf{I}_n$ und $\mathbf{J}=[\mathbf{0}_{k,1}\quad\mathbf{I}_k]\otimes\mathbf{I}_n$ gilt. Gl. (3.34) stellt eine parameterunabhängige LMI-Bedingung dar, welche notwendig und hinreichend für die Bedingung aus Gl. (3.20) ist.

## 3.3 Regelungsentwurf

Die in Abschnitt 3 vorgestellten Regelgesetze können durch folgende Schritte entworfen werden:

**Schritt 1a** Löse das Validierungsproblem (3.14)-(3.16).

**Schritt 2a** Für die resultierende Matrix $\mathbf{P}$ und einen frei gewählten Parameter $\varepsilon\in(0,1)$, wähle einen Skalar $d>\mathbf{b}^{\top}\mathbf{P}^{-1}\mathbf{b}/4$ um das größte erzielte Einzugsgebiet $\mathcal{G}_{\Delta}(1)$ zu bestimmen und einen Skalar $\nu\in(0,1)$ um die lineare Region der Sättigung $\mathcal{L}(u_f,1)$ zu bestimmen.

[14)] $\mathbf{M}_{\Sigma_{(i,j)}}$ bezeichnet eine Matrix in der $i$-ten Zeile und $j$-ten Spalte der Blockmatrix $\mathbf{M}_{\Sigma}$.

[15)] Vgl. [78, Lemma 4.4].

**Schritt 3a** Löse das Optimierungsproblem (3.20) mit Hilfe des Bisektionsverfahrens[16)], um ein Skalar $r \in (0,1]$ zu finden.

**Schritt 4a** Verwende das resultierende nichtsättigende Regelgesetz $u_f$ in der Form gegeben in Gl. (3.12) und (3.17), mit den Parametern $\mathbf{P}$, $d$, $\nu$, and $r$ aus den vorherigen Schritten.

Ein Beispiel eines solchen Reglers wird im Abschnitt 5.5 vorgestellt. Zusammenfassend kann man feststellen, dass die im Satz 3.1 vorgestellten nicht-konservativen Stabilitätsbedingungen der *klassischen* WSVR mittels iLF, so wie sie im Abschnitt 3.1 definiert wurde, einen nicht-konservativen Regler erzielen. Wären die oben genannten Bedingungen nicht erfüllt, so würde keine *klassische* WSVR mittels iLF existieren, die die untersuchte Regelstrecke stabilisieren würde. Eine Verbesserung des Ausregelverhaltens dieser nicht-konservativen *klassischen* WSVR mittels iLF wird im Kapitel 5 vorgestellt. Im nächsten Kapitel wird aber erstmals eine Weiterentwicklung der *klassischen* WSVR bezüglich der Selektionsstrategie vorgestellt.

[16)] Vgl. [17, Algorithm 4.1].

# 4 Die *invers-polynomiale* WSVR

Bei der in diesem Abschnitt analysierten Klasse von WSVR ergibt sich der Parameterwert aus einer Selektionsstrategie, die durch Inverse von polynomialen Matrizen definiert ist. Ein Vorläufer dieser Klasse von WSVR wurde in [40] eingeführt. Darin wurden hinreichende Stabilisierbarkeitsbedingungen linearer Systeme mit Stellgrößenbeschränkung durch WSVR mittels impliziter Ljapunov-Funktionen (iLF) und **polynomialer**[1)] Selektionsstrategien vorgestellt. Bei der in diesem Abschnitt vorgestellten Regelmethode stellt der implizite Parameter $v$ darüber hinaus keine implizite Ljapunov-Funktion des Systems mehr dar.

Außerdem kann diese Methode im Gegensatz zur *klassischen* WSVR auf Systeme in beliebiger Form direkt angewendet werden, d.h. die Systeme müssen nicht vorerst in die Steuerungsnormalform transformiert werden. Für lineare Systeme mit nur einer Eingangsgröße stellt dies keinen Vorteil dar, da jedes lineare System in die Steuerungsnormalform transformiert werden kann. Bei Mehrgrößensystemen existieren zwar auch Normalformen, jedoch ist der Entwurf der *klassischen* WSVR deutlich schwieriger, vgl. [32, 40]. Da die Transformation in die Steuerungsnormalform nicht mehr notwendig ist, kann eine Ausdehnung der *invers-polynomialen* WSVR auf Mehrgrößensysteme wahrscheinlich leichter erfolgen, vgl. auch [36] für eine ähnliche WSVR für Mehrgrößensysteme.

Die Inversion der polynomialen Matrix, welche die Selektionsstrategie bestimmt, hat zur Folge, dass die Stabilitätsbedingungen, welche im Folgenden vorgestellt werden, nicht nur notwendig und hinreichend sind, sondern auch polynomiale Matrizen darstellen. Dies stellt einen wesentlichen Vorteil dar, da die Definitheit einer polynomialen Matrix mittels einer Äquivalenztransformation durch die Definitheit einer konstanten Matrix überprüft werden kann, d.h. in Form einer linearen Matrixungleichung (LMI) formuliert werden kann.

Dieses Kapitel ist wie folgt gegliedert: im Abschnitt 4.1 wird die *invers-polynomiale* WSVR definiert und im Abschnitt 4.2 werden nicht-konser-

[1)] Die Selektionsstrategie ist auch in diesem Fall eine quadratische Gleichung der Form $\mathbf{x}^\top \mathbf{P}(v)\mathbf{x} = 1$. Dabei ist jedoch $\mathbf{P}(v)$ eine polynomiale Matrix in $v$.

vative Stabilitätsbedingungen vorgestellt. Das Kapitel endet mit dem Abschnitt 4.3 über mögliche Entwurfsschritte.

## 4.1 Definition einer stabilisierenden *invers-polynomialen* WSVR

Wie in den vorherigen Abschnitten betrachten wir LTI-Systeme mit einer Eingangsgröße und Stellgrößenbeschränkung, gegeben durch

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b}u, \quad \mathbf{x} \in \mathbb{R}^n, u \in \mathbb{R}, |u| \leq 1. \tag{4.1}$$

Die allgemeine Form der im nächsten Abschnitt analysierten WSVR mit *invers-polynomialer* Selektionsstrategie ist[2)]

$$u = -\mathbf{k}_v^\top \mathbf{x}, \tag{4.2}$$

wobei der Parameter $v \in [\varepsilon,1]$, mit $\varepsilon \in (0,1)$, für ein gegebenes $\mathbf{x} \in \mathbb{R}_*^n$ durch die Gleichung

$$g_{\mathrm{P}}(\mathbf{x},v) := \mathbf{x}^\top \mathbf{R}_v^{-1}\mathbf{x} - 1 = 0, \quad \mathbf{R}_v := \sum_{i=\mathrm{M_l}}^{\mathrm{M_u}} v^i \mathbf{R}_{c_i}, \tag{4.3}$$

mit $\mathrm{M_l}, \mathrm{M_u} \in \mathbb{Z}$, $\mathrm{M_l} < 0$, $\mathrm{M_l} \leq \mathrm{M_u}$ und $\mathbf{R}_v \succ \mathbf{0}$, $\forall\, v \in [\varepsilon,1]$, bestimmt wird. Dabei ist $\mathbf{R}_v$ eine polynomiale Matrix.[3)]

Das Regelgesetz aus Gl. (4.2)-(4.3) für das System aus Gl. (4.1) heißt *stabilisierende WSVR mit invers-polynomialer Selektionsstrategie* (oder kurz *invers-polynomiale WSVR*), wenn die Gebiete

$$\mathcal{E}_{\mathrm{P}}(v) := \{\mathbf{x} \in \mathbb{R}^n | g_{\mathrm{P}}(\mathbf{x},v) < 0\} \tag{4.4}$$

für alle $v \in [\varepsilon,1]$ verschachtelt und kontraktiv invariant sind,[4)] und $|u(\mathbf{x})| \leq 1$, $\forall\, \mathbf{x} \in \mathcal{E}_{\mathrm{P}}(v)$, $v \in [\varepsilon,1]$ gilt.

[2)] Der Übersichtlichkeit halber stellt in dieser Arbeit der Index $v$ eine Notation dar. Diese bedeutet, dass der Vektor $\mathbf{k}_v$ eine Vektorfunktion in $v$ darstellt, d.h. $\mathbf{k}_v := \mathbf{k}(v)$.

[3)] Strenggenommen handelt es sich um Laurent-Polynome, welche auch negative Exponenten zulassen. Diese Unterscheidung spielt jedoch im Weiteren keine Rolle.

[4)] Vgl. Def. 15 (Anhang) bzw. Def. 14 (Anhang).

## 4.2 Nicht-konservative Stabilitätsbedingungen

Der folgende Satz mitsamt einer konstruktiven Beweismethode stellt die notwendigen und hinreichenden Bedingungen für die Existenz einer stabilisierenden weichen strukturvariablen Regelung mit *invers-polynomialer* Selektionsstrategie dar. Für eine beliebige lineare Strecke mit einer Eingangsgröße und Stellgrößenbeschränkung ergibt sich daraus (im Existenzfall) ein stabilisierendes Regelgesetz. Dieses kann z.B. verwendet werden, wenn eine beliebige stabilisierende Regelung für die Initialisierung eines Optimierungsalgorithmus gesucht wird. Der Satz wurde zum ersten Mal in [60] vorgestellt.

**Satz 4.1** *Gegeben sei das folgende LTI-System mit einer Eingangsgröße und Stellgrößenbeschränkung*

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b}u, \quad \mathbf{x} \in \mathbb{R}^n, u \in \mathbb{R}, |u| \leq 1, \tag{4.5}$$

*sowie eine reelle Zahl* $\varepsilon \in (0,1)$. *Folgende Aussagen sind äquivalent:*

*i) Es existieren die ganzen Zahlen* $\mathrm{M_l}, \mathrm{M_u} \in \mathbb{Z}$, *mit* $\mathrm{M_l} < 0$, $\mathrm{M_l} \leq \mathrm{M_u}$, *die Matrizen* $\mathbf{R}_{c_i} \in \mathrm{Sym}^n$, *mit* $i = \mathrm{M_l}, \ldots, \mathrm{M_u}$, *sowie eine rationale Vektorfunktion* $\mathbf{k}_v : [\varepsilon,1] \to \mathbb{R}^n$, *sodass für alle* $v \in [\varepsilon,1]$ *die Gebiete*

$$\mathcal{E}_\mathrm{P}(v) := \{\mathbf{x} \in \mathbb{R}^n | g_\mathrm{P}(\mathbf{x},v) < 0\}, \tag{4.6}$$

*mit*

$$\begin{aligned} g_\mathrm{P}(\mathbf{x},v) &:= \mathbf{x}^\top \mathbf{Q}_v \mathbf{x} - 1, \\ \mathbf{Q}_v &:= \mathbf{R}_v^{-1}, \\ \mathbf{R}_v &:= \sum_{i=\mathrm{M_l}}^{\mathrm{M_u}} v^i \mathbf{R}_{c_i} \succ \mathbf{0}, \quad \forall\, v \in [\varepsilon,1], \end{aligned} \tag{4.7}$$

*für das System aus Gl.* (4.5) *mit dem Regelgesetz*

$$u = -\mathbf{k}_v^\top \mathbf{x}, \quad \mathbf{k}_v : [\varepsilon,1] \to \mathbb{R}^n, \tag{4.8}$$

*verschachtelt und kontraktiv invariant sind, und*

$$|u(\mathbf{x})| \leq 1, \quad \forall \mathbf{x} \in \mathcal{E}_{\mathrm{P}}(v),\ v \in [\varepsilon,1], \tag{4.9}$$

*gilt.*

*ii)* $\exists\, \mathrm{m_l}, \mathrm{m_u} \in \mathbb{Z}$*, mit* $\mathrm{m_l} < 0$*,* $\mathrm{m_l} \leq \mathrm{m_u}$*, sowie die Matrizen* $\mathbf{P}_{\mathrm{c}_i} \in \mathrm{Sym}^n$*,* $i = \mathrm{m_l}, \ldots, \mathrm{m_u}$*, sodass für jedes* $v \in [\varepsilon,1]$ *gilt*

$$\mathbf{P}_v = \sum_{i=\mathrm{m_l}}^{\mathrm{m_u}} v^i \mathbf{P}_{\mathrm{c}_i} \succ \mathbf{0}, \tag{4.10}$$

$$\partial_v \mathbf{P}_v = \sum_{i=\mathrm{m_l}}^{\mathrm{m_u}} i v^{i-1} \mathbf{P}_{\mathrm{c}_i} \succ \mathbf{0}, \tag{4.11}$$

$$\mathbf{A}\mathbf{P}_v + \mathbf{P}_v \mathbf{A}^\top \prec \mathbf{b}\mathbf{b}^\top. \tag{4.12}$$

*Falls ii) gilt, dann ist ein stabilisierendes Regelgesetz in i) gegeben durch*

$$\mathbf{k}_v^\top = c\mathbf{b}^\top \mathbf{P}_v^{-1}, \quad \mathbf{R}_v = d^{-1}\mathbf{P}_v, \tag{4.13}$$

*sowie*

$$\frac{d}{c^2} \geq \mathbf{b}^\top \mathbf{P}_\varepsilon^{-1} \mathbf{b} > 0, \quad c > \frac{1}{2}. \tag{4.14}$$

*Bemerkung* 4.1. Die Parameter $d$ und $c$ skalieren wie im vorigen Kapitel die kontraktiv invarianten Ellipsoide gegeben durch $\mathbf{P}_v$ bzw. den linearen Bereich der Sättigung, sodass dieser die kontraktiv invarianten Gebiete beinhaltet.[5)] △

*Bemerkung* 4.2. Die parameterabhängigen Bedingungen aus Gl. (4.10)-(4.12) stellen polynomiale Matrizen in $v \in [\varepsilon,1]$ dar und können in äquivalente parameterunabhängige LMIs transformiert werden. Die hier verwendete Transformation wurde in [77] vorgestellt. Sie beruht auf der verallgemeinerten $S$-Prozedur aus [38]. In [40, Anhang A.4] werden auch einfache Matlab-Funktionen zur Verfügung gestellt, mit denen man solche Bedingungen in äquivalente parameterunabhängige LMIs transformieren kann. △

*Bemerkung* 4.3. Der Übersichtlichkeit halber wird der Beweis, der weiter unten erfolgt, vorerst kurz skizziert. Mit Hilfe des Satzes 2.3 aus [2] wird erstens bewiesen, dass die Bedingungen (4.10)-(4.12), unter Verwendung

[5)] Vgl. Bemerkung 3.1.

des Regelgesetzes aus Gl. (4.8), (4.13)-(4.14), hinreichend dafür sind, dass die ellipsoidalen Gebiete $\mathcal{E}_{\mathrm{P}}(v)$ aus Gl. (4.6) verschachtelt und kontraktiv invariant sind. Die Notwendigkeit der Bedingungen (4.10)-(4.12) wird anschließend mit Hilfe von Finsler's Lemma, vgl. Satz A.5 (Anhang), gezeigt. △

*Beweis. ii)* $\Rightarrow$ *i)* Für den Beweis wird folgende Notation verwendet:

$$\mathbf{A}_v := \mathbf{A} - c\mathbf{b}\mathbf{b}^\top \mathbf{P}_v^{-1}.$$

Aus Gl. (4.12) folgt, dass

$$\begin{aligned}\mathbf{A}_v\mathbf{P}_v + \mathbf{P}_v\mathbf{A}_v^\top &= \mathbf{A}\mathbf{P}_v + \mathbf{P}_v\mathbf{A}^\top - 2c\mathbf{b}\mathbf{b}^\top \\ &\prec \mathbf{A}\mathbf{P}_v + \mathbf{P}_v\mathbf{A}^\top - \mathbf{b}\mathbf{b}^\top \prec \mathbf{0}, \quad \forall\, v \in [\varepsilon,1], \end{aligned} \tag{4.15}$$

da $c > 0.5$ ist. Durch Links- und Rechtsmultiplizieren der Gl. (4.15) mit der positiv definiten Matrix $\mathbf{P}_v^{-1}$ ergibt sich

$$\mathbf{P}_v^{-1}\mathbf{A}_v + \mathbf{A}_v^\top \mathbf{P}_v^{-1} \prec \mathbf{0}, \quad \forall\, v \in [\varepsilon,1]. \tag{4.16}$$

Aus $\mathbf{Q}_v = \mathbf{R}_v^{-1} = d\,\mathbf{P}_v^{-1}$ folgt, dass eine Umgebung $\mathcal{U}_0 \subseteq \mathcal{B}_\delta(\mathbf{0})$ der Ruhelage existiert, sodass für die zeitliche Änderung der Funktion $g_{\mathrm{P}}(\mathbf{x},v)$ aus Gl. (4.7) entlang einer Trajektorie des Systems $\dot{\mathbf{x}} = \mathbf{A}_v\mathbf{x}$[6)]

$$\partial_t g_{\mathrm{P}}(\mathbf{x}(t),v) = \mathbf{x}^\top(\mathbf{A}_v^\top\mathbf{Q}_v + \mathbf{Q}_v\mathbf{A}_v)\mathbf{x} < 0, \quad \forall\,(\mathbf{x},v) \in \mathcal{V}(\varepsilon), \tag{4.17}$$

mit

$$\mathcal{V}(\varepsilon) := \{(\mathbf{x},v) | \mathbf{x} \in \mathcal{U}_0 \backslash \{\mathbf{0}\} \subseteq \mathcal{B}_\delta(\mathbf{0}), \varepsilon < v < 1\}$$

gilt. Aus Gl. (4.17) folgt, dass Bedingung $(iv)$ des Satzes 2.3 für alle $(\mathbf{x},v) \in \mathcal{V}(\varepsilon)$ erfüllt ist. Darüber hinaus gilt

$$\partial_v\mathbf{Q}_v = \partial_v(d\,\mathbf{P}_v^{-1}) = -d\,\mathbf{P}_v^{-1}(\partial_v\mathbf{P}_v)\mathbf{P}_v^{-1} \prec \mathbf{0}, \quad \forall\, v \in [\varepsilon,1],$$

d.h.

$$-\infty < \partial_v g_{\mathrm{P}}(\mathbf{x},v) = \mathbf{x}^\top(\partial_v\mathbf{Q}_v)\mathbf{x} < 0, \quad \forall\,(\mathbf{x},v) \in \mathcal{V}(\varepsilon). \tag{4.18}$$

Aus Gl. (4.18) folgt, dass die Bedingung $(iii)$ des Satzes 2.3 für alle $(\mathbf{x},v) \in \mathcal{V}(\varepsilon)$ erfüllt ist.

[6)] Die Ableitung wird in abgekürzter Form durch $\partial_t g_{\mathrm{P}}(\mathbf{x}(t),v)$ bezeichnet, d.h. $\partial_t g_{\mathrm{P}}(\mathbf{x}(t),v) := \frac{\partial g(\mathbf{x},v)}{\partial \mathbf{x}}\dot{\mathbf{x}}$.

Die Bedingung $(ii)$ ist darüber hinaus für alle $\mathbf{x} \in \mathcal{E}_{\mathrm{P}}(1)\backslash\mathcal{E}_{\mathrm{P}}(\varepsilon)$ erfüllt, da

$$\lim_{v\to 1^-} g_{\mathrm{P}}(\mathbf{x},v) = g_{\mathrm{P}}(\mathbf{x},1) < 0, \quad \forall\, \mathbf{x} \in \mathcal{E}_{\mathrm{P}}(1)\backslash\mathcal{E}_{\mathrm{P}}(\varepsilon),$$
$$\lim_{v\to \varepsilon^+} g_{\mathrm{P}}(\mathbf{x},v) = g_{\mathrm{P}}(\mathbf{x},\varepsilon) > 0, \quad \forall\, \mathbf{x} \in \mathcal{E}_{\mathrm{P}}(1)\backslash\mathcal{E}_{\mathrm{P}}(\varepsilon).$$

Die Bedingungen $(ii)$ und $(iii)$ stellen sicher, dass die Gleichung $g_{\mathrm{P}}(\mathbf{x},v) = 0$ eine eindeutige Lösung für jedes $\mathbf{x} \in \mathcal{E}_{\mathrm{P}}(1)\backslash\mathcal{E}_{\mathrm{P}}(\varepsilon)$ hat, welche eine stetige Funktion $v = v(\mathbf{x})$ für $\varepsilon < v < 1$ ist. Darüber hinaus stellt die Funktion

$$V_{\mathrm{P}}(\mathbf{x}) := \begin{cases} v(\mathbf{x}), \text{ mit } g_{\mathrm{P}}(\mathbf{x},v) = 0, & \mathbf{x} \in \mathcal{E}_{\mathrm{P}}(1)\backslash\mathcal{E}_{\mathrm{P}}(\varepsilon), \\ \varepsilon\, \mathbf{x}^\top \mathbf{Q}_\varepsilon \mathbf{x}, & \mathbf{x} \in \mathcal{E}_{\mathrm{P}}(\varepsilon) \end{cases} \tag{4.19}$$

eine Ljapunov-Funktion des Systems dar. Dies folgt aus der Anwendung der direkten Methode von Ljapunov.[7] Dabei gilt $V_{\mathrm{P}}(\mathbf{0}) = 0$, sowie $V_{\mathrm{P}}(\mathbf{x}) > 0$ für alle $\mathbf{x} \in \mathcal{E}_{\mathrm{P}}(1)\backslash\{\mathbf{0}\}$. Darüber hinaus gilt auf Grund der Bedingungen $(iii)$ und $(iv)$ $\dot{V}_{\mathrm{P}}(\mathbf{x}) < 0$ für alle $\mathbf{x} \in \mathcal{E}_{\mathrm{P}}(1)\backslash\mathcal{E}_{\mathrm{P}}(\varepsilon)$, sowie auf Grund der Gl. (4.16) $\dot{V}_{\mathrm{P}}(\mathbf{x}) < 0$ für alle $\mathbf{x} \in \mathcal{E}_{\mathrm{P}}(\varepsilon)\backslash\{\mathbf{0}\}$. Folglich ist die Ruhelage $\mathbf{x}_{\mathrm{R}} = \mathbf{0}$ asymptotisch stabil und die Gebiete $\mathcal{E}_{\mathrm{P}}(v)$ für alle $v \in [\varepsilon,1]$ verschachtelt und kontraktiv invariant.

Die Bedingung aus Gl. (4.9) bezüglich der Stellgrößenbeschränkung ist äquivalent zu

$$|u(\mathbf{x})| \le \max_{\mathbf{x}^\top \mathbf{Q}_v \mathbf{x} < 1} |\mathbf{k}_v^\top \mathbf{x}| = \sqrt{\mathbf{k}_v^\top \mathbf{Q}_v^{-1} \mathbf{k}_v} = c\sqrt{\mathbf{b}^\top \mathbf{P}_v^{-1} d^{-1} \mathbf{P}_v \mathbf{P}_v^{-1} \mathbf{b}}$$
$$= c\sqrt{d^{-1}\mathbf{b}^\top \mathbf{P}_v^{-1}\mathbf{b}} \le 1, \quad \forall\, v \in [\varepsilon,1],$$

und, folglich, zu

$$\mathbf{b}^\top \mathbf{P}_v^{-1} \mathbf{b} \le \frac{d}{c^2}, \quad \forall\, v \in [\varepsilon,1]. \tag{4.20}$$

Da $\partial_v \mathbf{P}_v \succ \mathbf{0}$, $\forall\, v \in [\varepsilon,1]$, folgt, dass mit steigendem $v$ die Matrixfunktion $\mathbf{P}_v$ monoton steigend ist, d.h. für alle $\varepsilon \le v_1 < v_2 \le 1$ gilt $\mathbf{P}_{v_1} \prec \mathbf{P}_{v_2}$. Somit ist die Matrixfunktion $\mathbf{P}_v^{-1}$ monoton fallend,[8] d.h. für alle $\varepsilon \le v_1 < v_2 \le 1$ gilt $\mathbf{P}_{v_1}^{-1} \succ \mathbf{P}_{v_2}^{-1}$. Daraus folgt, dass die Matrixfunktion

[7] Vgl. z.B. [3].

[8] Dies gilt weil $\partial_v(\mathbf{P}_v^{-1}) = -\mathbf{P}_v^{-1}(\partial_v \mathbf{P}_v)\mathbf{P}_v^{-1} \prec \mathbf{0}$, $\forall\, v \in [\varepsilon,1]$.

$\mathbf{b}^\top \mathbf{P}_v^{-1}\mathbf{b}$ ebenfalls monoton fallend ist, d.h. für alle $\varepsilon \leq v_1 < v_2 \leq 1$ gilt $\mathbf{b}^\top \mathbf{P}_{v_1}^{-1}\mathbf{b} > \mathbf{b}^\top \mathbf{P}_{v_2}^{-1}\mathbf{b}$.[9] Folglich gilt

$$\mathbf{b}^\top \mathbf{P}_v^{-1}\mathbf{b} \leq \mathbf{b}^\top \mathbf{P}_\varepsilon^{-1}\mathbf{b}, \quad \forall\, v \in [\varepsilon,1].$$

und daher

$$\mathbf{b}^\top \mathbf{P}_v^{-1}\mathbf{b} \leq \mathbf{b}^\top \mathbf{P}_\varepsilon^{-1}\mathbf{b} \leq \frac{d}{c^2}, \quad \forall\, v \in [\varepsilon,1],$$

d.h. Gl. (4.9) ist erfüllt.

*i)* $\Rightarrow$ *ii)* Falls *i)* gilt, dann ist für jedes $v \in [\varepsilon,1]$ das jeweilige Gebiet $\mathcal{E}_{\mathrm{P}}(v)$ kontraktiv invariant für das entsprechende System $\dot{\mathbf{x}} = (\mathbf{A} - \mathbf{b}\mathbf{k}_v^\top)\mathbf{x}$. Die Funktion $W(\mathbf{x}) := \mathbf{x}^\top \mathbf{Q}_v \mathbf{x}$ ist dabei eine gültige Ljapunov-Funktion des Systems, d.h.

$$(\mathbf{A} - \mathbf{b}\mathbf{k}_v^\top)^\top \mathbf{Q}_v + \mathbf{Q}_v(\mathbf{A} - \mathbf{b}\mathbf{k}_v^\top) \prec \mathbf{0}, \quad \forall\, v \in [\varepsilon,1],$$

und das System ist global asymptotisch stabil. Für

$$\mathbf{P}_v = \mathbf{Q}_v^{-1} \tag{4.21}$$

folgt

$$(\mathbf{A} - \mathbf{b}\mathbf{k}_v^\top)\mathbf{P}_v + \mathbf{P}_v(\mathbf{A} - \mathbf{b}\mathbf{k}_v^\top)^\top \prec \mathbf{0}, \quad \forall\, v \in [\varepsilon,1]. \tag{4.22}$$

Sei $\mathbf{B}^\perp \in \mathbb{R}^{(n-1)\times n}$ eine Basis für den Nullraum von $\mathbf{b}$, sodass $\mathbf{B}^\perp\mathbf{b} = \mathbf{0}$ und $\mathbf{B}^\perp(\mathbf{B}^\perp)^\top \succ \mathbf{0}$ gilt. Nach Multiplizieren der Gl. (4.22) mit $\mathbf{B}^\perp$ (von links) und $(\mathbf{B}^\perp)^\top$ (von rechts) folgt, dass[10]

$$\begin{aligned}\mathbf{B}^\perp(\mathbf{A} - \mathbf{b}\mathbf{k}_v^\top)\mathbf{P}_v(\mathbf{B}^\perp)^\top + \mathbf{B}^\perp\mathbf{P}_v(\mathbf{A} - \mathbf{b}\mathbf{k}_v^\top)^\top(\mathbf{B}^\perp)^\top \\ = \mathbf{B}^\perp(\mathbf{A}\mathbf{P}_v + \mathbf{P}_v\mathbf{A}^\top)(\mathbf{B}^\perp)^\top \prec \mathbf{0}, \quad \forall\, v \in [\varepsilon,1].\end{aligned}$$

Dies ist äquivalent zu dem Sachverhalt, dass für jedes $v \in [\varepsilon,1]$ ein Skalar $\mu_v \in \mathbb{R}$ existiert,[11] sodass

$$\mathbf{A}\mathbf{P}_v + \mathbf{P}_v\mathbf{A}^\top \prec \mu_v \mathbf{b}\mathbf{b}^\top,$$

mit

$$\begin{aligned}&\mu_v > \mathbf{d}^\top \left\{\mathbf{L}_v - \mathbf{L}_v(\mathbf{B}^\perp)^\top[\mathbf{B}^\perp\mathbf{L}_v(\mathbf{B}^\perp)^\top]^{-1}\mathbf{B}^\perp\mathbf{L}_v\right\}\mathbf{d},\\ &\mathbf{L}_v = \mathbf{A}\mathbf{P}_v + \mathbf{P}_v\mathbf{A}^\top,\\ &\mathbf{d}^\top := |b_r|^{-1}\mathbf{b}_l^+, \quad b_r \in \mathbb{R}\backslash\{\mathbf{0}\},\, \mathbf{b}_l^+ \in \mathbb{R}^{1\times n},\end{aligned}$$

[9] Vgl. [8, Proposition 8.6.13, $xv$)].
[10] Vgl. [8] oder die Ausführung auf Seite 20.
[11] Vgl. Finsler's Lemma, [68, Theorem 2.3.10].

wobei das Tupel ($\mathbf{b}_l$,$b_r$) eine Voll-Rang-Faktorisierung des Vektors $\mathbf{b}$ ist. Gl. (4.12) folgt durch Skalierung der Matrix $\mathbf{P}_v$ durch einen Skalar

$$\mu > \max\left\{0, \max_{v\in[\varepsilon,1]} \mu_v\right\}.$$

Darüber hinaus sind die Gebiete $\mathcal{E}_\mathrm{P}(v)$ verschachtelt, d.h.

$$\begin{aligned}\mathcal{E}_\mathrm{P}(v_2) &\subset \mathcal{E}_\mathrm{P}(v_1), \quad \forall\, \varepsilon \le v_2 < v_1 \le 1,\\ \mathcal{E}_\mathrm{P}(v_2) \cap \mathcal{E}_\mathrm{P}(v_1) &= \emptyset, \quad \forall\, \varepsilon \le v_2 < v_1 \le 1.\end{aligned}$$

Dies bedeutet, dass für $\mathbf{x} \in \partial\mathcal{E}_\mathrm{P}(v_1)$, $g_\mathrm{P}(\mathbf{x}, v_1) = 0$ und $g_\mathrm{P}(\mathbf{x}, v_2) > 0$, $\forall\, \varepsilon \le v_2 < v_1 \le 1$, folgt und daher, dass

$$g_\mathrm{P}(\mathbf{x}, v_2) > g_\mathrm{P}(\mathbf{x}, v_1), \quad \forall\, \varepsilon \le v_2 < v_1 \le 1,\, \mathbf{x} \in \partial\mathcal{E}_\mathrm{P}(v_1)$$

gilt. Folglich gilt

$$\partial_v g_\mathrm{P}(\mathbf{x},v) = \mathbf{x}^\top(\partial_v \mathbf{Q}_v)\mathbf{x} < 0, \quad \forall\, \mathbf{x} \in \partial\mathcal{E}_\mathrm{P}(v),\, v \in [\varepsilon,1],$$

und somit $\partial_v \mathbf{Q}_v \prec \mathbf{0}$, $\forall\, v \in [\varepsilon,1]$. Daraus folgt, dass

$$\partial_v \mathbf{P}_v = \partial_v(\mathbf{Q}_v^{-1}) = -\mathbf{Q}_v^{-1}(\partial_v \mathbf{Q}_v)\mathbf{Q}_v^{-1} \succ \mathbf{0}, \forall\, v \in [\varepsilon,1],$$

d.h. es folgt Gl. (4.11).

Gl. (4.10), d.h. $\mathbf{P}_v \succ \mathbf{0}$, folgt aus Gl. (4.21), wobei die polynomiale Form der Matrix $\mathbf{P}_v$ durch

$$\mathbf{P}_v = \mathbf{Q}_v^{-1} = \mathbf{R}_v$$

sichergestellt ist.

Die Bedingungen (4.10)-(4.12) sind somit auch notwendig für die Existenz einer WSVR mit *invers-polynomialen* Selektionsstrategien. □

Die Sätze 3.1 und 4.1 stellen also die hinreichenden und notwendigen Bedingungen für die Existenz einer stabilisierenden WSVR mittels impliziter Ljapunov-Funktionen bzw. mittels *invers-polynomialer* Selektionsstrategien dar. Die Stabilisierbarkeitsbedingungen sind somit nicht-konservativ. Die damit entworfenen Regelgesetze können als Startwerte für eine Regleroptimierung verwendet werden. Dies wird im Kapitel 5 für den Fall der Maximierung der Konvergenzrate gezeigt.

## 4.3 Regelungsentwurf

Das im vorigen Abschnitt vorgestellte Regelgesetz kann durch folgende Schritte entworfen werden:

**Schritt 1b** Löse das Validierungsproblem (4.10)-(4.12).

**Schritt 2b** Verwende das resultierende nichtsättigende Regelgesetz $u$ in der Form gegeben in Gl. (4.8) und (4.13), mit den Parametern $d$ und $c$ aus Gl. (4.14).

*Bemerkung* 4.4. Das Validierungsproblem aus Schritt 1 beinhaltet keine Bedingungen bezüglich der Größe des Ellipsoids. Möchte man vermeiden, dass das Validierungsproblem ein zu kleines Ellipsoid erzeugt, so kann man weitere Bedingungen einführen. Beispielsweise kann man zusätzlich fordern, dass $\mathbf{P}_1^{-1} \prec \mathbf{C}$, wobei die Matrix $\mathbf{C} \in \mathrm{Sym}^n$ eine vorgegebene Matrix ist. Damit stellt man sicher, dass das Volumen des Ellipsoids $\partial\mathcal{E}(1)$ größer als das Volumen eines vorgegebenen Ellipsoids $\partial\mathcal{E}_0 := \{\mathbf{x} \in \mathbb{R}^n | \mathbf{x}^\top \mathbf{C}\mathbf{x} < 1\}$ ist. Erzielen kann man dies durch die zusätzliche LMI

$$\begin{bmatrix} \mathbf{P}_1 & \mathbf{I}_n \\ \mathbf{I}_n & \mathbf{C} \end{bmatrix} \succ \mathbf{0}.$$

△

Ein Beispiel eines solchen Reglers wird im Abschnitt 5.5 vorgestellt. Das Regelgesetz aus Schritt 2 wird nicht notwendigerweise zu einem schnellen Ausregelverhalten führen. Aufgrund der Nicht-Konservativität der Bedingungen des Validierungsproblems aus Punkt 1, stellt dieses Regelgesetz jedoch sicher, dass für die analysierte Strecke überhaupt ein Regelgesetz dieser Klasse existiert. Dieses kann auch als Startregler für einen Optimierungsalgorithmus verwendet werden. Dies wird im nächsten Kapitel vorgestellt.

# 5 Die Konvergenzoptimale (*Bang-Bang*) WSVR

Wie in [35] gezeigt, kann man eine untere Grenze der Konvergenzrate eines nichtlinearen Systems anhand der Abklingrate (entlang der Trajektorien des Systems) einer für das System gültigen Ljapunov-Funktion untersuchen. In diesem Kapitel wird das Regelgesetz bestimmt, das diese Abklingrate maximiert. Das resultierende Regelgesetz ist dabei ein *Bang-Bang*-Regelgesetz mit einer parameterabhängigen Selektionsstrategie.

Unter der Voraussetzung, dass eine Selektionsstrategie

$$g_{\mathrm{u}}(\mathbf{x},v) := \mathbf{x}^\top \mathbf{Q}_v \mathbf{x} - 1 = 0$$

eine eindeutige und stetige Lösung $v(\mathbf{x}) : \mathcal{U}_0 \backslash \{\mathbf{0}\} \to (\varepsilon,1)$ für jedes $\mathbf{x} \in \mathcal{U}_0 \backslash \{\mathbf{0}\}$ aus einer Umgebung der Ruhelage hat, kann diese Umgebung in die ellipsoidalen Gebiete

$$\mathcal{E}_{\mathrm{u}}(v) := \{\mathbf{x} \in \mathbb{R}^n | \mathbf{x}^\top \mathbf{Q}_v \mathbf{x} < 1\}$$

geteilt werden, wobei dem Rand jedes Gebiets ein eindeutiges $v \in (\varepsilon,1)$ zugewiesen wird. Somit kann für die Optimierung der Konvergenzrate des Gesamtsystems die Funktion

$$V_{\mathrm{u}}(\mathbf{x}) := \begin{cases} v(\mathbf{x}), \text{ mit } g_u(\mathbf{x},v) = 0, & \mathbf{x} \in \mathcal{E}_{\mathrm{u}}(1) \backslash \mathcal{E}_u(\varepsilon) \\ \varepsilon \mathbf{x}^\top \mathbf{Q}_\varepsilon \mathbf{x}, & \mathbf{x} \in \mathcal{E}_{\mathrm{u}}(\varepsilon) \end{cases} \tag{5.1}$$

verwendet werden. Diese Funktion ist für alle $\mathbf{x} \in \mathcal{E}_{\mathrm{u}}(1) \backslash \{\mathbf{0}\}$ positiv definit. Entlang einer Trajektorie des Systems aus Gl. (4.1), Seite 27, gilt dabei

$$\dot{V}_{\mathrm{u}}(\mathbf{x},u) = \begin{cases} V_{\mathrm{u},1}(\mathbf{x},u), & \mathbf{x} \in \mathcal{E}_{\mathrm{u}}(1) \backslash \mathcal{E}_{\mathrm{u}}(\varepsilon) \\ V_{\mathrm{u},2}(\mathbf{x},u), & \mathbf{x} \in \mathcal{E}_{\mathrm{u}}(\varepsilon) \backslash \{\mathbf{0}\} \end{cases}$$

mit

$$V_{\mathrm{u},1}(\mathbf{x},u) = -\frac{\mathbf{x}^\top (\mathbf{A}^\top \mathbf{Q}_v + \mathbf{Q}_v \mathbf{A})\mathbf{x} + 2\mathbf{x}^\top \mathbf{Q}_v \mathbf{b} u}{\mathbf{x}^\top (\partial_v \mathbf{Q}_v)\mathbf{x}},$$
$$V_{\mathrm{u},2}(\mathbf{x},u) = \varepsilon \mathbf{x}^\top (\mathbf{A}^\top \mathbf{Q}_\varepsilon + \mathbf{Q}_\varepsilon \mathbf{A})\mathbf{x} + 2\mathbf{x}^\top \mathbf{Q}_\varepsilon \mathbf{b} u.$$

Daraus folgt, dass für jedes $\mathbf{x} \in \mathcal{E}_{\mathrm{u}}(1)\backslash\{\mathbf{0}\}$

$$\arg\min_{|u|\leq 1} \dot{V}_{\mathrm{u}}(\mathbf{x},u) = -\operatorname{sgn}(\mathbf{b}^{\top}\mathbf{Q}_v\mathbf{x})$$

gilt, d.h., dass das Regelgesetz, das die Funktion $\dot{V}_{\mathrm{u}}(\mathbf{x},u)$ in jedem Punkt $\mathbf{x} \in \mathcal{E}_{\mathrm{u}}(1)\backslash\{\mathbf{0}\}$ minimiert, ein *Bang-Bang*-artiges Regelgesetz mit einer parameterabhängigen Umschaltstrategie ist. Die Stabilitätsbedingungen dieses Regelgesetzes werden in Abschnitt 5.1 analysiert. Dabei wird ersichtlich, dass die Existenz eines stabilisierenden beschränkten Reglers, wie z.B. in Abschnitt 4.2 bestimmt, notwendig und hinreichend für die Stabilität der konvergenzoptimalen Regelung ist. Der Entwurf der konvergenzoptimalen Regelung ist also auch nicht-konservativ. Jedoch hat die Unstetigkeit des Regelgesetzes Nachteile in einer praktischen Implementierung, beispielsweise durch die ununterbrochene Aktivität des Reglers aufgrund unvermeidbaren Rauschens. Daher wird eine stetige Approximation des konvergenzoptimalen Regelgesetzes in Abschnitt 5.2 untersucht, welche auf Kosten einer geringeren Konvergenzrate einen stetigen Verlauf erzielt.

Das Hauptergebnis dieses Kapitels ist im Abschnitt 5.1 enthalten. Darin werden (nicht-konservative) Stabilitätsbedingungen der *Bang-Bang*-WSVRs vorgestellt. Anschließend wird eine stetige Approximation des konvergenzoptimalen Regelgesetzes im Abschnitt 5.2 und die jeweiligen Entwurfsschritte im Abschnitt 5.4 vorgestellt. Das Kapitel endet mit dem Abschnitt 5.5, in dem zwei Beispiele die neu-entwickelte Regelungsmethode veranschaulichen.

## 5.1 Nicht-konservative Stabilitätsbedingungen

Der folgende Satz stellt die notwendigen und hinreichenden Stabilitätsbedingungen für das konvergenzoptimale Regelgesetz vor. Dieser Satz ist ähnlich zu dem Satz aus [37], der die Stabilitätsbedingungen eines konvergenzoptimalen Reglers mit einer parameterunabhängigen **zustandslinearen** Umschaltstrategie untersucht. Darüber hinaus stellt dieser Satz eine Generalisierung des Satzes aus [58] dar, welches die Stabilitätsbedingungen eines konvergenzoptimalen Reglers mit einer *klassischen* WSVR[1)] unter-

[1)] Vgl. [2].

sucht. Der hier vorgestellte Satz gilt für beliebige WSVR-Regelgesetze, welche durch quadratische Selektionsstrategien determiniert werden.

Im Folgenden verwenden wir die Notation

$$\mathcal{V}(\varepsilon) := \{(\mathbf{x},v)\,|\mathbf{x} \in \mathcal{U}_0 \backslash \{\mathbf{0}\}, \varepsilon < v < 1\},$$

wobei $\mathcal{U}_0 \subseteq \mathcal{B}_\delta(\mathbf{0})$ eine Umgebung der Ruhelage $\mathbf{x}_\mathrm{R} = \mathbf{0}$ darstellt.

**Satz 5.1** *Gegeben sei das folgende LTI-System mit einer Eingangsgröße und Stellgrößenbeschränkung*

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b}u, \quad \mathbf{x} \in \mathbb{R}^n, \quad u \in \mathbb{R}, |u| \leq 1, \tag{5.2}$$

*sowie eine reelle Zahl* $\varepsilon \in (0,1)$*. Folgende Aussagen sind äquivalent:*

*a) Es existiert ein beschränktes Regelgesetz*

$$u_f := -f(\mathbf{x},v), \quad |u_f| \leq 1, \tag{5.3}$$

*eine Umgebung* $\mathcal{U}_0 \subseteq \mathcal{B}_\delta(\mathbf{0})$ *der Ruhelage und eine Matrix* $\mathbf{Q}_v \succ \mathbf{0}, \forall\, v \in (\varepsilon,1)$*, sodass die Funktion*

$$g_f(\mathbf{x},v) : \mathcal{V}(\varepsilon) \to \mathbb{R}, \quad g_f(\mathbf{x},v) := \mathbf{x}^\top \mathbf{Q}_v \mathbf{x} - 1, \tag{5.4}$$

*für alle* $(\mathbf{x},v) \in \mathcal{V}(\varepsilon)$ *die Bedingungen* $(ii) - (iv)$ *des Satzes 2.3 für das System* (5.2) *mit dem Regelgesetz* (5.3) *erfüllt.*

*b) Es existiert eine Umgebung der Ruhelage* $\mathcal{U}_0 \subseteq \mathcal{B}_\delta(\mathbf{0})$ *und eine Matrix* $\mathbf{Q}_v \succ \mathbf{0}, \forall\, v \in (\varepsilon,1)$*, sodass die Funktion*

$$g_s(\mathbf{x},v) : \mathcal{V}(\varepsilon) \to \mathbb{R}, \quad g_s(\mathbf{x},v) := \mathbf{x}^\top \mathbf{Q}_v \mathbf{x} - 1, \tag{5.5}$$

*für alle* $(\mathbf{x},v) \in \mathcal{V}(\varepsilon)$ *die Bedingungen* $(ii) - (iv)$ *des Satzes 2.3 für das System* (5.2) *mit dem Regelgesetz*

$$u_s := -sgn(\mathbf{b}^\top \mathbf{Q}_v \mathbf{x}). \tag{5.6}$$

*erfüllt.*

*c) Es existiert eine Funktion* $\kappa_v : [\varepsilon,1] \to \mathbb{R}_+$*, eine Umgebung der Ruhelage* $\mathcal{U}_0 \subseteq \mathcal{B}_\delta(\mathbf{0})$ *und eine Matrix* $\mathbf{Q}_v \succ \mathbf{0}, \forall\, v \in (\varepsilon,1)$*, sodass die Funktion*

$$g_{sat}(\mathbf{x},v) : \mathcal{V}(\varepsilon) \to \mathbb{R}, \quad g_{sat}(\mathbf{x},v) := \mathbf{x}^\top \mathbf{Q}_v \mathbf{x} - 1, \tag{5.7}$$

*für alle* $(\mathbf{x},v) \in \mathcal{V}(\varepsilon)$ *die Bedingungen* $(ii) - (iv)$ *des Satzes 2.3 für das System* (5.2) *mit dem Regelgesetz*

$$u_{sat} := -sat(\kappa_v \mathbf{b}^\top \mathbf{Q}_v \mathbf{x}), \tag{5.8}$$

*erfüllt.*

*Beweis.* $a) \Rightarrow b)$ Falls $a)$ gilt, dann ist die zeitliche Änderung der Funktion $g_s(\mathbf{x},v)$ entlang einer Trajektorie des Systems (5.2) mit dem Regelgesetz (5.6)

$$\begin{aligned}
&\partial_t g_s(\mathbf{x}(t),v) \\
&\quad = \mathbf{x}^\top(\mathbf{A}^\top\mathbf{Q}_v + \mathbf{Q}_v\mathbf{A})\mathbf{x} - 2|\mathbf{x}^\top\mathbf{Q}_v\mathbf{b}| \\
&\quad = \partial_t g_f(\mathbf{x}(t),v) + 2\mathbf{x}^\top\mathbf{Q}_v\mathbf{b}f(\mathbf{x},v) - 2|\mathbf{x}^\top\mathbf{Q}_v\mathbf{b}| \\
&\quad = \partial_t g_f(\mathbf{x}(t),v) + 2|\mathbf{x}^\top\mathbf{Q}_v\mathbf{b}| \cdot \left[\operatorname{sgn}(\mathbf{x}^\top\mathbf{Q}_v\mathbf{b})f(\mathbf{x},v) - 1\right] \\
&\quad = \partial_t g_f(\mathbf{x}(t),v) + 2|\mathbf{b}^\top\mathbf{Q}_v\mathbf{x}| \cdot \begin{cases} -|f(\mathbf{x},v)| - 1, \operatorname{sgn}(f(\mathbf{x},v)) \neq \operatorname{sgn}(\mathbf{x}^\top\mathbf{Q}_v\mathbf{b}) \\ |f(\mathbf{x},v)| - 1, \text{sonst} \end{cases} \\
&\quad < 0,\ \forall(\mathbf{x},v) \in \mathcal{V}(\varepsilon),
\end{aligned}$$

wobei $\partial_t g_f(\mathbf{x}(t),v)$ die zeitliche Änderung der Funktion $g_f(\mathbf{x},v)$ entlang einer Trajektorie des Systems (5.2) mit dem Regelgesetz (5.3) darstellt und negativ für alle $(\mathbf{x},v) \in \mathcal{V}(\varepsilon)$ ist. Das bedeutet, dass die Bedingung $(iv)$ erfüllt ist. Darüber hinaus folgt unmittelbar aus $a)$, dass die Bedingungen $(ii) - (iii)$ für alle $(\mathbf{x},v) \in \mathcal{V}(\varepsilon)$ für das System (5.2) mit dem Regelgesetz (5.6) erfüllt sind.

$b) \Rightarrow c)$ Für diesen Teil des Beweises verwenden wir Finsler's Lemma[2)]. Sei

$$\mathbf{s}_v := \mathbf{Q}_v\mathbf{b}, \quad \mathbf{s}_v : (\varepsilon,1) \to \mathbb{R}^n,$$

mit $\operatorname{Rang}(\mathbf{s}_v) = 1$, und $\mathbf{S}_v^\perp \in \mathbb{R}^{(n-1)\times n}$ eine Basis des Nullraums von $\mathbf{s}_v^\top$, sodass $\mathbf{S}_v^\perp\mathbf{s}_v = \mathbf{0}$ und $\mathbf{S}_v^\perp(\mathbf{S}_v^\perp)^\top \succ \mathbf{0}$.[3)] Wenn $b)$ gilt, dann folgt aus der Bedingung $(iv)$ des Satzes 2.3, Seite 11, dass

$$\partial_t g_s(\mathbf{x}(t),v) = \mathbf{x}^\top(\mathbf{A}^\top\mathbf{Q}_v + \mathbf{Q}_v\mathbf{A})\mathbf{x} - 2|\mathbf{x}^\top\mathbf{Q}_v\mathbf{b}| < 0, \quad \forall(\mathbf{x},v) \in \mathcal{V}(\varepsilon).$$

[2)] Vgl. [68, Theorem 2.3.10].

[3)] Eine notwendige und hinreichende Bedingung, dass $\mathbf{S}_v^\perp(\mathbf{S}_v^\perp)^\top \succ \mathbf{0}$ gilt, ist, dass $\operatorname{Rang}(\mathbf{S}_v^\perp) = n - 1$ ist. Dies resultiert aus der Tatsache, dass $\mathbf{S}_v^\perp(\mathbf{S}_v^\perp)^\top \succ \mathbf{0} \Leftrightarrow \lambda(\mathbf{S}_v^\perp(\mathbf{S}_v^\perp)^\top) > 0 \Leftrightarrow \operatorname{Rang}(\mathbf{S}_v^\perp(\mathbf{S}_v^\perp)^\top) = \operatorname{Rang}(\mathbf{S}_v^\perp) = n - 1$.

Für jedes $\mathbf{x} \in \mathcal{N}(\mathbf{s}_v^\top) := \{\mathbf{x} \in \mathbb{R}_*^n | \mathbf{s}_v^\top \mathbf{x} = 0\}$ gilt

$$\mathbf{x}^\top(\mathbf{A}^\top\mathbf{Q}_v + \mathbf{Q}_v\mathbf{A})\mathbf{x} < 0, \quad \forall(\mathbf{x},v) \in \{(\mathbf{x},v) | \mathbf{x} \in \mathcal{N}(\mathbf{s}_v^\top), v \in (\varepsilon,1)\}.$$

Dabei existiert für jedes $\mathbf{x} \in \mathcal{N}(\mathbf{s}_v^\top)$ ein $\mathbf{y} \in \mathbb{R}_*^{n-1}$, sodass $\mathbf{x} = (\mathbf{S}_v^\perp)^\top\mathbf{y}$. Daraus folgt, dass

$$\mathbf{y}^\top\mathbf{S}_v^\perp\mathbf{L}_v(\mathbf{S}_v^\perp)^\top\mathbf{y} < 0, \quad \forall(\mathbf{y},v) \in \{(\mathbf{y},v) | \mathbf{y} \in \mathbb{R}_*^{n-1}, v \in (\varepsilon,1)\}, \tag{5.9}$$

wobei

$$\mathbf{L}_v := \mathbf{A}^\top\mathbf{Q}_v + \mathbf{Q}_v\mathbf{A}.$$

Daraus folgt wiederum, dass

$$\mathbf{S}_v^\perp\mathbf{L}_v(\mathbf{S}_v^\perp)^\top \prec 0, \quad \forall\, v \in (\varepsilon,1).$$

Wir wenden eine Kongruenztransformation der Matrix $2\kappa_v\mathbf{s}_v\mathbf{s}_v^\top - \mathbf{L}_v$ an, wobei $\mu_v := 2\kappa_v$. Sei $\mathbf{T}_v \in \mathbb{R}^{n\times n}$ eine nichtsinguläre Matrix, definiert als

$$\mathbf{T}_v := \begin{bmatrix} \mathbf{d}_v & (\mathbf{S}_v^\perp)^\top \end{bmatrix}^\top$$

mit dem Vektor $\mathbf{d}_v \in \mathbb{R}^n$ definiert als $\mathbf{d}_v^\top := s_{v_r}^{-1}\mathbf{s}_{v_l}^+$, wobei das Tupel $(\mathbf{s}_{v_l}, s_{v_r})$ eine Voll-Rang-Faktorisierung[4] des Vektors $\mathbf{s}_v$ ist.[5] Daraus folgt, dass

$$\begin{aligned}
&\mathbf{T}_v(\mu_v\mathbf{s}_v\mathbf{s}_v^\top - \mathbf{L}_v)\mathbf{T}_v^\top \\
&= \begin{bmatrix} \mathbf{d}_v^\top \\ \mathbf{S}_v^\perp \end{bmatrix}(\mu_v\mathbf{s}_v\mathbf{s}_v^\top - \mathbf{L}_v)\begin{bmatrix} \mathbf{d}_v & (\mathbf{S}_v^\perp)^\top \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{d}_v^\top(\mu_v\mathbf{s}_v\mathbf{s}_v^\top - \mathbf{L}_v) \\ \mathbf{S}_v^\perp(\mu_v\mathbf{s}_v\mathbf{s}_v^\top - \mathbf{L}_v) \end{bmatrix}\begin{bmatrix} \mathbf{d}_v & (\mathbf{S}_v^\perp)^\top \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{d}_v^\top(\mu_v\mathbf{s}_v\mathbf{s}_v^\top - \mathbf{L}_v)\mathbf{d}_v & \mathbf{d}_v^\top(\mu_v\mathbf{s}_v\mathbf{s}_v^\top - \mathbf{L}_v)(\mathbf{S}_v^\perp)^\top \\ \mathbf{S}_v^\perp(\mu_v\mathbf{s}_v\mathbf{s}_v^\top - \mathbf{L}_v)\mathbf{d}_v & \mathbf{S}_v^\perp(\mu_v\mathbf{s}_v\mathbf{s}_v^\top - \mathbf{L}_v)(\mathbf{S}_v^\perp)^\top \end{bmatrix} \\
&= \begin{bmatrix} \mu_v - \mathbf{d}_v^\top\mathbf{L}_v\mathbf{d}_v & -\mathbf{d}_v^\top\mathbf{L}_v(\mathbf{S}_v^\perp)^\top \\ -\mathbf{S}_v^\perp\mathbf{L}_v\mathbf{d}_v & -\mathbf{S}_v^\perp\mathbf{L}_v(\mathbf{S}_v^\perp)^\top \end{bmatrix}.
\end{aligned} \tag{5.10}$$

[4] Vgl. Def. 20 (Anhang).

[5] Die Matrix $\mathbf{T}_v$ ist nichtsingulär, da $\mathcal{N}(\mathbf{d}_v) \cap \mathcal{N}(\mathbf{S}_v^\perp) = \{0\}$ gilt, vgl. dazu [8, Fakt 2.11.3]. Dies ist ersichtlich aus der Tatsache, dass per Definition $\mathbf{S}_v^\perp\mathbf{s}_v = \mathbf{0}$ gilt, und daher $\mathbf{s}_v \in \mathcal{N}(\mathbf{S}_v^\perp)$. Da noch $\mathbf{d}_v^\top\mathbf{s}_v = \mathbf{s}_v^\top\mathbf{s}_v/\|\mathbf{s}_v\|^2 = 1 \neq 0, \forall v \in [\varepsilon,1]$, folgt, dass $\mathbf{s}_v \notin \mathcal{N}(\mathbf{s}_v^\top)$.

Da $-\mathbf{S}_v^{\perp}\mathbf{L}_v(\mathbf{S}_v^{\perp})^{\top} \succ \mathbf{0}$, ist die Blockmatrix aus Gl. (5.10) positiv definit dann und nur dann, wenn für jedes $v \in (\varepsilon,1)$ ein Skalar $\mu_v \in \mathbb{R}$ existiert, sodass[6)]

$$\mu_v > \mathbf{d}_v^{\top}\left[\mathbf{L}_v - \mathbf{L}_v(\mathbf{S}_v^{\perp})^{\top}(\mathbf{S}_v^{\perp}\mathbf{L}_v(\mathbf{S}_v^{\perp})^{\top})^{-1}\mathbf{S}_v^{\perp}\mathbf{L}_v\right]\mathbf{d}_v. \tag{5.11}$$

In diesem Fall ist die Matrix $\mu_v\mathbf{s}_v\mathbf{s}_v^{\top} - \mathbf{L}_v$ auch positiv definit, da beide Matrizen *kongruent*[7)] sind. Für $\kappa_v = 0.5\mu_v$ folgt, dass

$$\mathbf{A}^{\top}\mathbf{Q}_v + \mathbf{Q}_v\mathbf{A} - 2\kappa_v\mathbf{Q}_v\mathbf{b}\mathbf{b}^{\top}\mathbf{Q}_v \prec \mathbf{0}.$$

Daraus folgt, dass die Funktion $g_{\text{sat}}(\mathbf{x},v)$ die Bedingung $(iv)$ für das System (5.2) mit dem Regelgesetz $u_\kappa := -\kappa_v\mathbf{b}^{\top}\mathbf{Q}_v\mathbf{x}$ erfüllt. Da diese Bedingung auch mit dem beschränkten Regelgesetz $u_s$ erfüllt ist, folgt $c)$ aus der Anwendung des Satzes 2 aus [45]. Darüber hinaus folgt unmittelbar aus $b)$, dass die Bedingungen $(ii) - (iii)$ für alle $(\mathbf{x},v) \in \mathcal{V}(\varepsilon)$ für das System (5.2) mit dem Regelgesetz (5.8) erfüllt sind.

$c) \Rightarrow a)$ ist offensichtlich. □

**Korollar 5.2.** *Für das System aus Gl.* (5.2) *mit dem Regelgesetz aus Gl.* (5.3), (5.6) *oder* (5.8) *seien die Bedingungen* $(ii) - (iv)$ *des Satzes 2.3 für alle* $(\mathbf{x},v) \in \mathcal{V}(\varepsilon)$ *mit* $\mathcal{U}_0 = \mathcal{E}_\star(1)\backslash\mathcal{E}_\star(\varepsilon)$ *erfüllt, wobei*

$$\mathcal{E}_\star(v) := \{\mathbf{x} \in \mathbb{R}^n | \mathbf{x}^{\top}\mathbf{Q}_v\mathbf{x} - 1 < 0\}. \tag{5.12}$$

*Auf Grund des Satzes 5.1 ist dies z.B. der Fall wenn die Bedingungen* (4.10)-(4.12) *aus Satz 4.1 erfüllt sind. Dann ist das Gebiet* $\mathcal{E}_\star(1)$ *kontraktiv invariant. Darüber hinaus konvergieren die Trajektorien, die in dem Gebiet starten, asymptotisch gegen die Ruhelage.*

*Beweis.* Man betrachte die Funktion

$$V_\star(\mathbf{x}) := \begin{cases} v(\mathbf{x}), \text{ mit } g_\star(\mathbf{x},v) = 0, & \mathbf{x} \in \mathcal{E}_\star(1)\backslash\mathcal{E}_\star(\varepsilon), \\ \varepsilon\mathbf{x}^{\top}\mathbf{Q}_\varepsilon\mathbf{x}, & \mathbf{x} \in \mathcal{E}_\star(\varepsilon), \end{cases} \tag{5.13}$$

[6)] Dies resultiert aus der Bedingung, dass das Schur-Komplement von $-\mathbf{S}_v^{\perp}\mathbf{L}_v(\mathbf{S}_v^{\perp})^{\top}$ bezüglich der Blockmatrix aus Gl. (5.10) positiv definit sein muss, vgl. [8, Proposition 8.2.4].

[7)] Vgl. Def. 19 (Anhang).

wobei die Bezeichnung $(\cdot)_\star$ für eins der Symbole $(\cdot)_\mathrm{f}$, $(\cdot)_\mathrm{s}$ oder $(\cdot)_\mathrm{sat}$ steht. Dabei gilt

$$V_\star(\mathbf{0}) = 0, \tag{5.14}$$

$$V_\star(\mathbf{x}) > 0, \quad \forall\, \mathbf{x} \in \mathcal{E}_\star(1)\backslash\{\mathbf{0}\}, \tag{5.15}$$

$$\dot{V}_\star(\mathbf{x}) < 0, \quad \forall\, \mathbf{x} \in \mathcal{E}_\star(1)\backslash\{\mathbf{0}\}. \tag{5.16}$$

Gl. (5.14) folgt aus der Definition der Funktion $V_\star(\mathbf{x})$ aus Gl. (5.13). Gl. (5.15) folgt aus $v \in (\varepsilon,1)$ und $\mathbf{Q}_\varepsilon \succ \mathbf{0}$. Für alle $\mathbf{x} \in \mathcal{E}_\star(1)\backslash\{\mathbf{0}\}$ folgt Gl. (5.16) aus der kontraktiven Invarianz der Gebiete $\mathcal{E}_\star(v)$ für alle $v \in [\varepsilon,1]$. Letzteres folgt aus Satz 2.3 und Korollar 2.4. Somit ist das Ellipsoid $\mathcal{E}_\star(1)$ kontraktiv invariant[8)], und die Trajektorien, die in dem Gebiet starten, konvergieren asymptotisch gegen die Ruhelage. □

Schließlich wird die Berechnung der notwendigen Verstärkung $\kappa_v$, bzw. der unteren Grenze der Funktion $\mu_v$ aus Gl. (5.11), gezeigt.

## 5.2 Entwurf einer stetigen Approximation des Regelgesetzes

Die Existenzbedingungen einer stetigen Approximation der *Bang-Bang* WSVR wurden in Satz 5.1 vorgestellt. Dabei lautete das stetige Regelgesetz

$$u_\mathrm{sat} = -\operatorname{sat}(\kappa_v \mathbf{b}^\top \mathbf{Q}_v \mathbf{x}), \tag{5.17}$$

wobei zur Bestimmung der notwendigen Verstärkung $\kappa_v$ eine Basis des Nullraums der Umschaltstrategie notwendig war. Dieser letzte Schritt, d.h. die Berechung einer Basis des Nullraums des Vektors $\mathbf{s}_v := \mathbf{Q}_v\mathbf{b}$ wird im Folgenden für beide weichen strukturvariablen Regelungen gezeigt.

### 5.2.1 *Klassische* WSVR mittels iLF

Im Fall der *klassischen* WSVR mittels iLF ist die parameterabhängige Matrix $\mathbf{Q}_v$ definiert als

$$\mathbf{Q}_v := \mathbf{D}_v^{-r}\mathbf{Q}_1\mathbf{D}_v^{-r}, \quad \mathbf{D}_v := \operatorname{diag}(v^n,\ldots,v^2,v). \tag{5.18}$$

[8)] Vgl. Def. 14 (Anhang).

Eine Basis des Nullraums von $\mathbf{s}_v^\top = \mathbf{b}^\top \mathbf{Q}_v$, d.h. eine Matrix $\mathbf{S}_v^\perp \in \mathbb{R}^{(n-1)\times n}$, wofür $\mathbf{S}_v^\perp \mathbf{s}_v = \mathbf{0}$ und $\mathbf{S}_v^\perp (\mathbf{S}_v^\perp)^\top \succ \mathbf{0}$ gilt, kann aus

$$\mathbf{S}_v^\perp \mathbf{Q}_v \mathbf{b} = \mathbf{S}_v^\perp \mathbf{D}_v^{-r} \mathbf{Q}_1 \mathbf{D}_v^{-r} \mathbf{b} = v^{-r} \mathbf{S}_v^\perp \mathbf{D}_v^{-r} \mathbf{Q}_1 \mathbf{b} = v^{-r} \mathbf{S}_v^\perp \mathbf{D}_v^{-r} \mathbf{s}_1 = \mathbf{0}$$

berechnet werden. Es ergibt sich

$$\mathbf{S}_v^\perp = v^r \mathbf{S}_1^\perp \mathbf{D}_v^r, \tag{5.19}$$

wobei $\mathbf{S}_1^\perp \in \mathbb{R}^{(n-1)\times n}$ eine Basis des Nullraums von $\mathbf{s}_1$ ist. Die Matrix $\mathbf{S}_1^\perp$ kann aus der Gleichung

$$\mathbf{S}_1^\perp = \mathbf{H}\mathbf{U}_2^\top$$

berechnet werden, wobei die Matrix $\mathbf{H} \in \mathbb{R}^{(n-1)\times(n-1)}$ eine beliebige nichtsinguläre Matrix ist, und die Matrix $\mathbf{U}_2$ aus der Singulärwertzerlegung von $\mathbf{s}_1$, d.h. aus

$$\mathbf{s}_1 = [\mathbf{U}_1 \;\; \mathbf{U}_2] \begin{bmatrix} \sigma(\mathbf{s}_1) & \mathbf{0} \end{bmatrix}^\top. \tag{5.20}$$

stammt.[9] Dies kann man wie folgt erklären. Da die Blockmatrix $[\mathbf{U}_1 \;\; \mathbf{U}_2]$ aus Gl. (5.20) *unitär* ist, folgt, dass $\mathbf{U}_2^\top \mathbf{U}_1 = \mathbf{0}$ gilt und dass, $\mathbf{S}_1^\perp \mathbf{s}_1 = \mathbf{H}\mathbf{U}_2^\top \mathbf{U}_1 \sigma(\mathbf{s}_1) = \mathbf{0}$ ist. Da die Matrix $\mathbf{U}_2$ auch *unitär* ist, folgt, dass $\mathrm{Rang}(\mathbf{U}_2) = n-1$. Daraus folgt, dass[10] $\mathrm{Rang}(\mathbf{S}_1^\perp) \geq \mathrm{Rang}(\mathbf{H}) + \mathrm{Rang}(\mathbf{U}_2) - (n-1)$ und somit, dass $\mathrm{Rang}(\mathbf{S}_1^\perp) = n-1$. Da ebenfalls $\mathrm{Rang}(\mathbf{D}_v^r) = n$, folgt schließlich, dass $\mathrm{Rang}(\mathbf{S}_v^\perp) = n-1$, und somit, dass $\mathbf{S}_v^\perp (\mathbf{S}_v^\perp)^\top \succ \mathbf{0}$.

### 5.2.2 *Invers-polynomiale* WSVR

Um die untere Grenze der Funktion $\mu_v$ aus Gl. (5.11) zu berechnen, muss eine Basis des Nullraums der Umschaltfunktion $\mathbf{s}_v^\top = \mathbf{b}^\top \mathbf{Q}_v$ berechnet werden, wobei $\mathrm{Rang}(\mathbf{s}_v) = 1$, $\forall\, v \in [\varepsilon,1]$ gilt. Die in diesem Abschnitt analysierte Form der parameterabhängigen Matrix $\mathbf{Q}_v$ ist

$$\mathbf{Q}_v := d\,\mathbf{R}_v^{-1} = \frac{d}{\det(\mathbf{R}_v)} \mathbf{R}_v^A,$$

mit $d > 0$ und

$$\mathbf{R}_v := \sum_{i=\mathrm{M_l}}^{\mathrm{M_u}} v^i \mathbf{R}_{c_i} = v^{\mathrm{M_l}} \sum_{i=0}^{l} v^i \mathbf{R}_{c_{i+\mathrm{M_l}}} \succ \mathbf{0}, \forall\, v \in [\varepsilon,1], \tag{5.21}$$

[9] Vgl. [68, Theorem 2.1.1].
[10] Vgl. [8, Korollar 2.5.10].

wobei $\mathbf{R}_{c_i} \in \mathrm{Sym}^n$ und $l := \mathrm{M_u} - \mathrm{M_l}$. In [39, Lemma A.1] wurde eine obere Grenze des Grades der Adjunkten einer polynomialen Matrix angegeben, vgl. Lemma A.1 (Anhang). Für die Adjunkte der polynomialen Matrix aus Gl. (5.21) gilt demnach

$$\begin{aligned}\mathbf{R}_v^A &= v^{\mathrm{M_l}\cdot(n-1)}\left(\sum_{i=0}^{l} v^i \mathbf{R}_{c_{i+\mathrm{M_l}}}\right)^A \\ &= v^{\mathrm{M_l}\cdot(n-1)}\sum_{i=0}^{\mu} v^i \mathbf{N}_{c_i}, \quad \mathbf{N}_{c_i} \in \mathrm{Sym}^n,\end{aligned}$$

mit den konstanten Matrizen $\mathbf{N}_{c_i}$, $i = 0,\ldots,\mu$, und

$$\mu \le l\cdot\min\{n-1, n-q\}, \quad q := \dim\left[\bigcap_{i=1}^{l}\mathcal{N}(\mathbf{R}_{c_{i+\mathrm{M_l}}})\right].$$

Folglich gilt

$$\mathbf{s}_v = \frac{d\cdot v^{\mathrm{M_l}\cdot(n-1)}}{\det(\mathbf{R}_v)}\sum_{i=0}^{\mu} v^i \mathbf{s}_{c_i},$$

mit $\mathbf{s}_{c_i} := \mathbf{N}_{c_i}\mathbf{b}$. Gesucht wird im Weiteren eine polynomiale Matrix $\mathbf{S}_v^\perp \in \mathbb{R}^{(n-1)\times n}$, wofür

$$\mathbf{S}_v^\perp \mathbf{s}_v = \mathbf{0}, \quad \forall\, v \in [\varepsilon,1], \tag{5.22}$$

$$\mathbf{S}_v^\perp (\mathbf{S}_v^\perp)^\top \succ \mathbf{0}, \quad \forall\, v \in [\varepsilon,1], \tag{5.23}$$

gilt. Diese wird in der Form

$$\mathbf{S}_v^\perp = \sum_{i=0}^{\nu} v^i \mathbf{S}_{c_i}^\perp, \quad \mathbf{S}_{c_i}^\perp \in \mathbb{R}^{(n-1)\times n}, \nu > 0, \tag{5.24}$$

angenommen. Sind die Koeffizienten des Polynoms $\mathbf{R}_v^A$ bekannt, so können die Koeffizienten des gesuchten Polynoms aus Gl. (5.24), wie im Folgenden gezeigt, analytisch berechnet werden. Steht nur die Matrix $\mathbf{Q}_v^{-1} = d^{-1}\,\mathbf{R}_v$ zur Verfügung, so wie es im Satz 4.1 der Fall ist, so muss vorerst die Adjunkte der Matrix $\mathbf{R}_v$, d.h. die Matrizen $\mathbf{N}_{c_i}$, $i = 0,\ldots,\mu$, berechnet werden. Diese können z.B. aus [39, Lemmas A.1], vgl. A.1 (Anhang), berechnet werden. Eine Alternative zur Berechnung der Adjunkten $\mathbf{R}_v^A$ ist

die numerische Berechnung von $\mathbf{S}_v^{\perp}$. Diese Vorgehensweise wird am Ende dieses Abschnittes vorgestellt.

Im Weiteren nehmen wir an, dass die Matrizen $\mathbf{N}_{c_i}, i = 0, \ldots, \mu$, bekannt sind. Durch die Multiplikation der polynomialen Matrizen aus Gl. (5.22) ergibt sich

$$\mathbf{S}_v^{\perp}\mathbf{s}_v = \sum_{i=0}^{\nu} v^i \mathbf{S}_{c_i}^{\perp} \sum_{j=0}^{l} v^j \mathbf{s}_{c_j} = \mathbf{0}, \quad \forall\, v \in [\varepsilon,1],$$

da $\det(\mathbf{R}_v) \neq 0$, $\forall\, v \in (\varepsilon,1)$, sowie $d$ und $v$ strikt positive Zahlen sind. Diese Gleichung ist offensichtlich erfüllt, wenn alle Koeffizienten des resultierenden polynomialen Vektors null sind. Dies ist äquivalent zu der Bedingung

$$\begin{bmatrix}\mathbf{S}_{c_\nu}^{\perp} & \cdots & \mathbf{S}_{c_0}^{\perp}\end{bmatrix} \begin{bmatrix}\mathbf{s}_{c_l} & \cdots & \mathbf{s}_{c_0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{s}_{c_l} & \cdots & \mathbf{s}_{c_0}\end{bmatrix} = \mathbf{0},$$

wobei die konstante Matrix

$$\mathbf{S} := \begin{bmatrix}\mathbf{S}_{c_\nu}^{\perp} & \cdots & \mathbf{S}_{c_0}^{\perp}\end{bmatrix} \in \mathbb{R}^{(n-1)\times n(\nu+1)} \tag{5.25}$$

die Koeffizienten des gesuchten Polynoms gruppiert, und die konstante Matrix

$$\mathbf{S}_{\mathrm{G}} := \begin{bmatrix}\mathbf{s}_{c_l} & \cdots & \mathbf{s}_{c_0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{s}_{c_l} & \cdots & \mathbf{s}_{c_0}\end{bmatrix} \in \mathbb{R}^{n(\nu+1)\times(l+1+\nu)},$$

mit $r := \mathrm{Rang}(\mathbf{S}_{\mathrm{G}})$, die Koeffizienten der bekannten Umschaltfunktion gruppiert. Die Bestimmung der Matrix $\mathbf{S}$ kann aus der Singulärwertzerlegung der Matrix $\mathbf{S}_{\mathrm{G}}$ gewonnen werden, d.h. aus

$$\mathbf{S}_{\mathrm{G}}^{\perp} = \mathbf{H}_1 \mathbf{U}_2^{\top},$$

wobei die unitäre Matrix $\mathbf{U}_2 \in \mathbb{R}^{n(\nu+1)\times(\nu+1)n-r}$ aus der Zerlegung

$$\mathbf{S}_{\mathrm{G}} = \begin{bmatrix}\mathbf{U}_1 & \mathbf{U}_2\end{bmatrix} \begin{bmatrix}\mathbf{\Sigma}(\mathbf{S}_{\mathrm{G}}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix} \begin{bmatrix}\mathbf{V}_1^{\top} \\ \mathbf{V}_2^{\top}\end{bmatrix},$$

entnommen wird, wobei $\Sigma(\mathbf{S}_\mathrm{G}) := \mathrm{diag}(\sigma_i(\mathbf{S}_\mathrm{G}))$, $i = 1, \ldots, \min\{n(\nu+1), l+1+\nu\}$ und $\sigma_i(\mathbf{S}_\mathrm{G})$ die Singulärwerte der Matrix $\mathbf{S}_\mathrm{G}$ darstellen. Um die gewünschte Dimension der Matrix $\mathbf{S}_\mathrm{G}^\perp$ zu erzielen, kann eine beliebige nichtsinguläre Matrix $\mathbf{H}_1 \in \mathbb{R}^{(\nu+1)n-r\times(\nu+1)n-r}$ gewählt werden, sodass für die resultierende Matrix $\mathbf{S}_\mathrm{G}^\perp \in \mathbb{R}^{n(\nu+1)-r\times n(\nu+1)}$ gilt. Somit ergibt sich

$$\mathbf{S}_\mathrm{G}^\perp \mathbf{S}_\mathrm{G} = \mathbf{H}_1 \mathbf{U}_2^\top \mathbf{S}_\mathrm{G} = \mathbf{H}_1 \mathbf{U}_2^\top \mathbf{U}_1 \mathbf{\Sigma}(\mathbf{S}_\mathrm{G}) \mathbf{V}_1^\top = \mathbf{0},$$

da $\mathbf{U}_2^\top \mathbf{U}_1 = \mathbf{0}$. Dabei gilt noch

$$\begin{aligned} n(\nu+1) - r &= \min\{\mathrm{Rang}(\mathbf{H}_1), \mathrm{Rang}(\mathbf{U}_2^\top)\} \\ &\geq \mathrm{Rang}(\mathbf{S}_\mathrm{G}^\perp) \\ &\geq \mathrm{Rang}(\mathbf{H}_1) + \mathrm{Rang}(\mathbf{U}_2^\top) - n(\nu+1) + r \\ &= n(\nu+1) - r, \end{aligned}$$

d.h. $\mathrm{Rang}(\mathbf{S}_\mathrm{G}^\perp) = n(\nu+1)-r$. Die gesuchte Matrix $\mathbf{S}$ ist schließlich gegeben durch

$$\mathbf{S} = \mathbf{H}_2 \mathbf{S}_\mathrm{G}^\perp, \quad \mathbf{H}_2 \in \mathbb{R}^{(n-1)\times n(\nu+1)-r}, \tag{5.26}$$

wobei die Matrix $\mathbf{H}_2$ eine beliebige Matrix mit $\mathrm{Rang}(\mathbf{H}_2) = \min\{n-1, (\nu+1)n - r\}$ ist. Es gilt folglich

$$\begin{aligned} &\min\{n-1, (\nu+1)n - r\} \\ &\quad = \min\{\mathrm{Rang}(\mathbf{H}_2), \mathrm{Rang}(\mathbf{S}_\mathrm{G}^\perp)\} \\ &\quad\quad \geq \mathrm{Rang}(\mathbf{S}) \\ &\quad\quad\quad \geq \mathrm{Rang}(\mathbf{H}_2) + \mathrm{Rang}(\mathbf{S}_\mathrm{G}^\perp) - n(\nu+1) + r \\ &\quad\quad\quad\quad = \min\{n-1, (\nu+1)n - r\}, \end{aligned}$$

d.h. $\mathrm{Rang}(\mathbf{S}) = \min\{n-1, (\nu+1)n-r\}$. Für das gesuchte Polynom ergibt sich

$$\mathbf{S}_v^\perp = \begin{bmatrix} \mathbf{S}_{c_\nu}^\perp & \cdots & \mathbf{S}_{c_0}^\perp \end{bmatrix} \begin{bmatrix} v^\nu \mathbf{I}_n \\ \vdots \\ \mathbf{I}_n \end{bmatrix} = \mathbf{S} \begin{bmatrix} v^\nu \mathbf{I}_n \\ \vdots \\ \mathbf{I}_n \end{bmatrix},$$

mit $\mathbf{S} \in \mathbb{R}^{(n-1)\times n(\nu+1)}$, wobei

$$\mathrm{Rang}(\mathbf{S}_v^\perp) \leq \min\{n-1, (\nu+1)n - r\}.$$

Eine analytische Form des gesuchten Polynoms steht somit zur Verfügung. Diese Methode stellt jedoch nicht sicher, dass die Bedingung (5.23) erfüllt

ist. Dies wäre der Fall, wenn $\mathrm{Rang}(\mathbf{S}_v^{\perp}) = n-1$ wäre. Die Bedingung aus Gl. (5.23) kann aber in eine LMI transformiert werden. Dies wird im Folgenden noch gezeigt. Dazu bildet man erstens die Polynommultiplikation

$$\mathbf{M}_v := \mathbf{S}_v^{\perp}(\mathbf{S}_v^{\perp})^{\top} = \sum_{j=0}^{2\nu} \mathbf{M}_{c_j} v^j,$$

wobei

$$\mathbf{M}_{c_j} := \sum_{\substack{i=0 \\ j-\nu \le i \le \nu}}^{j} \mathbf{S}_{c_i}^{\perp}(\mathbf{S}_{c_{j-i}}^{\perp})^{\top}, \quad 0 \le j \le 2\nu.$$

Gl. (5.23) ist somit äquivalent zu

$$\mathbf{M}_v = \sum_{j=0}^{2\nu} \mathbf{M}_{c_j} v^j \succ \mathbf{0}, \quad \forall\, v \in [\varepsilon,1].$$

Dies ist für $\tilde{v} := (1/\alpha)v - \beta/\alpha$, mit $\alpha := (1-\varepsilon)/2$ und $\beta := (1+\varepsilon)/2$, weiterhin äquivalent zu

$$\tilde{\mathbf{M}}_{\tilde{v}} = \sum_{j=0}^{2\nu} \tilde{\mathbf{M}}_{c_j} \tilde{v}^j \succ \mathbf{0}, \quad \forall\, \tilde{v} \in [-1,1], \tag{5.27}$$

mit

$$\tilde{\mathbf{M}}_{c_j} = \sum_{i=j}^{2\nu} \binom{i}{j} \alpha^j \beta^{i-j} \mathbf{M}_{c_i}, \quad 0 \le j \le 2\nu.$$

Zweitens wird die Matrix $\tilde{\mathbf{M}}_{\tilde{v}}$ aus Gl. (5.27) in der Form

$$\tilde{\mathbf{M}}_{\tilde{v}} = (\tilde{\mathbf{v}}^{[\nu+1]} \otimes \mathbf{I}_n)^{\top} \tilde{\mathbf{M}}_{\Sigma} (\tilde{\mathbf{v}}^{[\nu+1]} \otimes \mathbf{I}_n)$$

mit

$$\tilde{\mathbf{M}}_{\Sigma} := \frac{1}{2} \begin{bmatrix} 2\tilde{\mathbf{M}}_{c_0} & \tilde{\mathbf{M}}_{c_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \tilde{\mathbf{M}}_{c_1} & 2\tilde{\mathbf{M}}_{c_2} & \tilde{\mathbf{M}}_{c_3} & \ddots & \vdots \\ \mathbf{0} & \tilde{\mathbf{M}}_{c_3} & 2\tilde{\mathbf{M}}_{c_4} & \ddots & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \tilde{\mathbf{M}}_{c_{2\nu-1}} \\ \mathbf{0} & \cdots & \mathbf{0} & \tilde{\mathbf{M}}_{c_{2\nu-1}} & 2\tilde{\mathbf{M}}_{c_{2\nu}} \end{bmatrix}$$

geschrieben. Die Matrix $\tilde{\mathbf{M}}_{\tilde{v}}$ ist positiv definit für jedes $\tilde{v} \in [-1{,}1]$ dann und nur dann, wenn zwei Matrizen, $\mathbf{D} \in \mathbb{P}^{n\nu}$ und $\mathbf{G} \in \mathrm{Skew}^{n\nu}$ existieren, sodass

$$\tilde{\mathbf{M}}_{\tilde{v}} \succ \begin{bmatrix} \hat{\mathbf{J}}_\nu \otimes \mathbf{I}_n \\ \check{\mathbf{J}}_\nu \otimes \mathbf{I}_n \end{bmatrix}^\top \begin{bmatrix} -\mathbf{D} & \mathbf{G} \\ \mathbf{G}^\top & \mathbf{D} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{J}}_\nu \otimes \mathbf{I}_n \\ \check{\mathbf{J}}_\nu \otimes \mathbf{I}_n \end{bmatrix}, \tag{5.28}$$
$$\hat{\mathbf{J}}_\nu := [\mathbf{I}_\nu \;\; \mathbf{0}_{\nu,1}], \quad \check{\mathbf{J}}_\nu := [\mathbf{0}_{\nu,1} \;\; \mathbf{I}_\nu],$$

gilt.[11] Die Bedingung aus Gl. (5.28) ist eine parameterunabhängige LMI Bedingung, welche numerisch effizient überprüft werden kann.

Zusammenfassend ergibt sich die notwendige Verstärkung $\kappa_v = 0.5\mu_v$ aus Gl. (5.11), mit dem Polynom $\mathbf{S}_v^\perp$ aus Gl. (5.24), dessen Koeffizienten in der Matrix $\mathbf{S}$ aus Gl. (5.25) gruppiert sind. Die Matrix $\mathbf{S}$ kann aus der Singulärwertzerlegung der Matrix $\mathbf{S}_G$ berechnet werden. Anschließend muss Bedingung (5.23) überprüft werden.

Eine Alternative zur obigen Berechnung ist eine Berechnung während des Ausregelvorgangs. Für jedes $v^* \in [\varepsilon{,}1]$ kann die Matrix $\mathbf{S}_{v^*}^\perp$ aus der Singulärwertzerlegung des Vektors $\mathbf{s}_{v^*}$ berechnet werden. Folglich hat die Matrix $\mathbf{S}_{v^*}^\perp$ die Form

$$\mathbf{S}_{v^*}^\perp = \mathbf{H}\mathbf{U}_{2_{v^*}}^\top,$$

mit einer beliebigen nichtsingulären Matrix $\mathbf{H} \in \mathbb{R}^{(n-1)\times(n-1)}$ und Matrix $\mathbf{U}_{2_{v^*}} \in \mathbb{R}^{n\times(n-1)}$ aus der Singulärwertzerlegung[12] von $\mathbf{s}_{v^*}$, d.h.[13]

$$\mathbf{s}_{v^*} = [\mathbf{U}_{1_{v^*}} \;\; \mathbf{U}_{2_{v^*}}] \begin{bmatrix} \sigma(\mathbf{s}_{v^*}) & \mathbf{0} \end{bmatrix}^\top.$$

Somit gilt $\mathbf{U}_{2_{v^*}}^\top \in \mathbb{R}^{n\times(n-1)}$ und $\mathrm{Rang}(\mathbf{U}_{2_{v^*}}^\top) = n-1$.[14] Die Matrix $\mathbf{S}_{v^*}^\perp$ hat dabei vollen Rang, d.h. $\mathrm{Rang}(\mathbf{S}_{v^*}^\perp) = n-1$. Dies ist ersichtlich aus

$$\begin{aligned} n-1 &= \min\{\mathrm{Rang}(\mathbf{H}), \mathrm{Rang}(\mathbf{U}_{2_{v^*}}^\top)\} \geq \mathrm{Rang}(\mathbf{S}_{v^*}^\perp) \\ &\geq \mathrm{Rang}(\mathbf{H}) + \mathrm{Rang}(\mathbf{U}_{2_{v^*}}^\top) - (n-1) = n-1. \end{aligned}$$

[11] Vgl. [77]. Diese Äquivalenzbedingung beruht auf einer verallgemeinerten $S$-Prozedur, welche in [38] eingeführt wurde. $\mathbf{A} \otimes \mathbf{B}$ bezeichnet dabei das Kronecker-Produkt.

[12] Vgl. [68, Satz 2.1.1].

[13] Da die Blockmatrix $\begin{bmatrix} \mathbf{U}_{1_{v^*}} & \mathbf{U}_{2_{v^*}} \end{bmatrix}$ unitär ist, folgt, dass $\mathbf{U}_{2_{v^*}}^\top \mathbf{U}_{1_{v^*}} = \mathbf{0}$ und daher, dass $\mathbf{S}_{v^*}^\perp \mathbf{s}_{v^*} = \mathbf{H}\mathbf{U}_{2_{v^*}}^\top \mathbf{U}_{1_{v^*}} \sigma(\mathbf{s}_{v^*}) = \mathbf{0}$. Dabei ist $\sigma(\mathbf{s}_{v^*})$ der Singulärwert von $\mathbf{s}_{v^*}$.

[14] Die Matrix $\mathbf{U}_{v^*} = [\mathbf{U}_{1_{v^*}} \;\; \mathbf{U}_{2_{v^*}}] \in \mathbb{R}^{n\times n}$ ist unitär. Diese hat folglich $n$ unabhängige Spalten. Somit folgt, dass $\mathrm{Rang}(\mathbf{U}_{2_{v^*}}^\top) = n-1$.

## 5.3 Maximierung des Einzugsgebiets der konvergenzoptimalen invers-polynomialen WSVR

Ein konvergenzoptimaler Regler wird im Allgemeinen mit Hilfe des Satzes 5.1 entworfen. Dabei kann als hinreichende Existenzbedingung die Existenz einer nichtsättigenden WSVR beispielsweise aus Satz 3.1 oder 4.1 festgelegt werden. In diesem Fall ist die Größe der erzielten verschachtelten und kontraktiv invarianten Gebiete durch bestimmte Entwurfsbedingungen beschränkt, welche wegen der Stabilität und des nichtsättigenden Charakters der unterlegten Regelgesetze entstanden sind. Diese Entwurfsbedingungen sind in Gl. (3.18)-(3.19) und Gl. (4.14) des Satzes 3.1 bzw. 4.1 enthalten.

Im Folgenden wird eine alternative Methode vorgestellt, die kontraktive Invarianz von Ellipsoiden im Falle des konvergenzoptimalen Regelgesetzes zu überprüfen. Diese Methode setzt zwar nicht mehr die Existenz einer nichtsättigenden WSVR voraus, ist jedoch nur hinreichend für die kontraktive Invarianz der analysierten Gebiete. Aufgrund ihrer Komplexität wird darüber hinaus nur den Fall einer vereinfachten Selektionssgleichung dargestellt. Der Vorteil dieser Methode liegt aber in der Tatsache, dass sie im Allgemeinen größere Gebiete als diejenigen aus dem Satz 4.1 erzielen kann.

### 5.3.1 Invers-polynomiale WSVR mit vereinfachter Selektionsgleichung

Angenommen, es existiert eine Matrixfunktion $\mathbf{Q}_v : [\varepsilon,1] \to \mathbb{P}^n$, sodass die Gebiete $\mathcal{E}(v) = \{\mathbf{x} \in \mathbb{R}^n | \mathbf{x}^\top \mathbf{Q}_v \mathbf{x} < 1\}$ für das System $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} - \mathbf{b}\,\mathrm{sgn}(\mathbf{b}^\top \mathbf{Q}_v \mathbf{x})$ verschachtelt und kontraktiv invariant sind. Die Existenz dieser Matrixfunktion kann beispielsweise mit Hilfe des Satzes 4.1 erfolgen. Fraglich ist, wie weit man diese Ellipsoide skalieren kann, sodass das erzielte Einzugsgebiet der Ruhelage vergrößert wird. Der im Folgenden vorgestellte Satz beantwortet zwar diese Frage, ist jedoch nur für den Fall einer vereinfachten Selektionsgleichung, d.h. für $\mathrm{M_l} = -1$ und $\mathrm{M_u} = 0$ gültig. Die Betrachtung einer beliebigen Selektionsgleichung ist möglich, erfordert jedoch, wie es im Weiteren gezeigt wird, einen größeren Rechenaufwand.

**Korollar 5.3.** *Die Ellipsoide*

$$\mathcal{E}_1(v) = \{\mathbf{x} \in \mathbb{R}^n | g^*(\mathbf{x},v) = \mathbf{x}^\top \mathbf{R}_v^{-1}\mathbf{x} - 1 < 0, \mathbf{R}_v = v^{-1}\mathbf{R}_{c_{-1}} + \mathbf{R}_{c_0}\}$$

*seien für das System*

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} - \mathbf{b}\,\mathrm{sgn}(\mathbf{b}^\top \mathbf{R}_v^{-1}\mathbf{x}) \tag{5.29}$$

*kontraktiv invariant. Die skalierten Ellipsoide*

$$\mathcal{E}_1^*(v) = \{\mathbf{x} \in \mathbb{R}^n | g^*(\mathbf{x},v) = \mathbf{x}^\top \mathbf{R}_v^{-1}\mathbf{x} - d < 0, \quad \mathbf{R}_v = v^{-1}\mathbf{R}_{c_{-1}} + \mathbf{R}_{c_0},\ d > 1\} \tag{5.30}$$

*sind für das System aus Gl.* (5.29) *genau dann kontraktiv invariant, wenn an der Stelle* $\tilde{v} := v^{-1} = 1$ *das Polynom*

$$G(\mathbf{x},\tilde{v}) := \sum_{i=0}^{n-1} \tilde{v}^{(n-1)-i} h_i(\mathbf{x}), \tag{5.31}$$

*mit*

$$h_i(\mathbf{x}) := \mathbf{x}^\top(\mathbf{A}^\top \mathbf{N}_i + \mathbf{N}_i\mathbf{A})\mathbf{x} - 2|\mathbf{x}^\top \mathbf{N}_i \mathbf{b}| \tag{5.32}$$

$$\mathbf{N}_i := \begin{cases} \mathbf{R}_{c_{-1}}^A, & i = 0, \\ \Gamma_{n-1}^{n-1-i}(\mathbf{R}_{c_0}/\mathbf{R}_{c_{-1}})^A, & i = 1,\ldots,n-2, \\ \mathbf{R}_{c_0}^A, & i = n-1, \end{cases} \tag{5.33}$$

*und seine sämtlichen partiellen Ableitungen* $\partial_{\tilde{v}}^j G(\mathbf{x},\tilde{v})$, $\forall j \in \{1,\ldots,n-1\}$, *strikt negativ bzw. nicht positiv für alle* $\mathbf{x} \in \partial\mathcal{E}_1^*(1)$ *sind, d.h. wenn*

$$\begin{aligned} &\max_{\mathbf{x}^\top \mathbf{R}_1^{-1}\mathbf{x} = d} G(\mathbf{x},1) < 0, \\ &\max_{\mathbf{x}^\top \mathbf{R}_1^{-1}\mathbf{x} = d} \partial_{\tilde{v}}^j G(\mathbf{x},1) \le 0, \quad j \in \{1,\ldots,n-1\} \end{aligned} \tag{5.34}$$

*gilt.*

*Bemerkung* 5.1. Der Skalar $d > 1$ wird verwendet, um die bereits erzielten verschachtelten und kontraktiv invarianten Gebiete zu skalieren. △

*Beweis.* Die Gebiete $\mathcal{E}_1^*(v)$ aus Gl. (5.30) sind für das System aus Gl. (5.29) definitionsgemäß dann und nur dann kontraktiv invariant, wenn

$$\partial_t g^*(\mathbf{x}(t),v) = \mathbf{x}^\top(\mathbf{A}^\top\mathbf{R}_v^{-1} + \mathbf{R}_v^{-1}\mathbf{A})\mathbf{x} - 2|\mathbf{x}^\top\mathbf{R}_v^{-1}\mathbf{b}| < 0, \quad \forall\,\mathbf{x} \in \partial\mathcal{E}_1^*(v), v \in [\varepsilon,1]$$

gilt, d.h. wenn

$$\det(\mathbf{R}_v)\partial_t g^*(\mathbf{x}(t),v) = \mathbf{x}^\top(\mathbf{A}^\top\mathbf{R}_v^A + \mathbf{R}_v^A\mathbf{A})\mathbf{x} - 2|\mathbf{x}^\top\mathbf{R}_v^A\mathbf{b}| < 0, \quad \forall\,\mathbf{x} \in \partial\mathcal{E}_1^*(v), v \in [\varepsilon,1] \tag{5.35}$$

gilt. Die Adjunkte der polynomialen Matrix $\mathbf{R}_v$ ist wiederum eine polynomiale Matrix in $v$ und ist gegeben durch[15)]

$$\begin{aligned}\mathbf{R}_v^A &= (v^{-1}\mathbf{R}_{c_{-1}} + \mathbf{R}_{c_0})^A \\ &= \left(\mathbf{R}_{c_0}^A + \sum_{i=1}^{n-2} v^{-i}\Gamma_{n-1}^i(\mathbf{R}_{c_0}/\mathbf{R}_{c_{-1}})^A + v^{-(n-1)}\mathbf{R}_{c_{-1}}^A\right) \\ &= \sum_{i=0}^{n-1} v^{i-(n-1)}\mathbf{N}_i,\end{aligned}$$

mit

$$\begin{aligned}\mathbf{N}_0 &:= \mathbf{R}_{c_{-1}}^A, \\ &\vdots \\ \mathbf{N}_i &:= \Gamma_{n-1}^{n-1-i}(\mathbf{R}_{c_0}/\mathbf{R}_{c_{-1}})^A, \quad i = 1,\ldots,n-2, \\ &\vdots \\ \mathbf{N}_{n-1} &:= \mathbf{R}_{c_0}^A.\end{aligned}$$

Dabei ist zu beachten, dass der Grad des Polynoms von der Systemordnung $n$ abhängig ist. Folglich gilt

$$\mathbf{x}^\top(\mathbf{A}^\top\mathbf{R}_v^A + \mathbf{R}_v^A\mathbf{A})\mathbf{x} - 2|\mathbf{x}^\top\mathbf{R}_v^A\mathbf{b}| \le \sum_{i=0}^{n-1} v^{i-(n-1)}h_i(\mathbf{x}),$$

[15)] Vgl. [77, Korollar 2.2] oder Lemma A.2 (Anhang).

mit

$$h_i(\mathbf{x}) = \mathbf{x}^\top(\mathbf{A}^\top\mathbf{N}_i + \mathbf{N}_i\mathbf{A})\mathbf{x} - 2|\mathbf{x}^\top\mathbf{N}_i\mathbf{b}|.$$

Die Bedingung aus Gl. (5.35) ist folglich erfüllt, wenn

$$G(\mathbf{x},v) := \sum_{i=0}^{n-1} v^{i-(n-1)} h_i(\mathbf{x}) < 0, \quad \forall\, \mathbf{x} \in \partial\mathcal{E}_1^*(v), v \in [\varepsilon,1] \tag{5.36}$$

gilt. Die Überprüfung der Bedingung aus Gl. (5.36) erfolgt im Weiteren mit Hilfe der Newton-Regel[16]. Für die Anwendung der Newton-Regel wird eine Variablensubstitution verwendet. Für $\tilde{v} := v^{-1}$ ist die Bedingung aus Gl. (5.36) äquivalent zu

$$G(\mathbf{x},\tilde{v}) = \sum_{i=0}^{n-1} \tilde{v}^{(n-1)-i} h_i(\mathbf{x}) < 0, \forall\, \mathbf{x} \in \partial\mathcal{E}_1^*\left(\frac{1}{\tilde{v}}\right), \tilde{v} \in \left[1,\frac{1}{\varepsilon}\right].$$

Aus der Anwendung der Newton-Regel ist das Polynom $G(\mathbf{x},\tilde{v})$ in $\tilde{v}$ negativ für alle $\tilde{v} > 1$, wenn an der Stelle $\tilde{v} = 1$ dieses und seine sämtlichen partiellen Ableitungen nicht positiv sind, d.h. wenn

$$\partial_{\tilde{v}}^j\, G(\mathbf{x},1) \leq 0, \quad \forall\, j \in \{0,\ldots,n-1\}, \quad \forall\, \mathbf{x} \in \mathcal{E}_1^*\,(1)\,. \tag{5.37}$$

Das Polynom ist negativ für alle $\tilde{v} \geq 1$ falls zusätzlich

$$G(\mathbf{x},1) < 0, \quad \forall\, \mathbf{x} \in \mathcal{E}_1^*(1) \tag{5.38}$$

gilt. Die Bedingungen sind also für beliebig kleine $\varepsilon \in (0,1)$ hinreichend.

Die Funktionen $\partial_{\tilde{v}}^j G(\mathbf{x},1)$, mit $j = 0,\ldots,n-1$, hängen dabei nicht mehr von $v$ ab und deren Maximum bezüglich $\mathbf{x}$ befindet sich auf dem Rand des äußersten Ellipsoids $\partial\mathcal{E}_1^*(1)$. Um dies zu zeigen, sei $\mathbf{x} \in \partial\mathcal{E}_1^*(1) = \{\mathbf{x} \in \mathbb{R}^n | \mathbf{x}^\top\mathbf{R}_1^{-1}\mathbf{x} = d,\, d > 1\}$ und eine Zahl $k \in (0,1]$. Es folgt

$$\begin{aligned}
&\partial_{\tilde{v}}^j G(k\mathbf{x},1) \\
&= \sum_{i=0}^{n-1-j} \left(\prod_{l=0}^{j-1} n-1-i-l\right) \left[k^2\mathbf{x}^\top(\mathbf{A}^\top\mathbf{N}_i + \mathbf{N}_i\mathbf{A})\mathbf{x} - 2k|\mathbf{x}^\top\mathbf{N}_i\mathbf{b}|\right] \\
&= k^2 \sum_{i=0}^{n-1-j} \left(\prod_{l=0}^{j-1} n-1-i-l\right) \left[\mathbf{x}^\top(\mathbf{A}^\top\mathbf{N}_i + \mathbf{N}_i\mathbf{A})\mathbf{x} - \frac{2}{k}|\mathbf{x}^\top\mathbf{N}_i\mathbf{b}|\right]
\end{aligned}$$

[16] Vgl. Lemma A.3 (Anhang).

$$\leq k^2 \sum_{i=0}^{n-1-j} \left( \prod_{l=0}^{j-1} n-1-i-l \right) \left[ \mathbf{x}^\top (\mathbf{A}^\top \mathbf{N}_i + \mathbf{N}_i \mathbf{A}) \mathbf{x} - 2|\mathbf{x}^\top \mathbf{N}_i \mathbf{b}| \right]$$
$$\leq \partial_v^j G(\mathbf{x},1).$$

Daraus folgt, dass die Bedingungen aus Gl. (5.37) und(5.38) äquivalent zu den Bedingungen aus Gl. (5.34) sind. □

*Bemerkung* 5.2. Aus Gl. (5.34) ist ersichtlich, dass, falls diese Bedingungen für ein $d \geq 1$ erfüllt sind, dann sind sie für alle $d^* \leq d$ erfüllt. Daraus folgt, dass die Bestimmung des maximalen Wertes von $d$ mittels des Bisektionsverfahrens erfolgen kann. △

Die Überprüfung der Bedingung aus Gl. (5.34) kann wie in [37, Theorem 1] durchgeführt werden. Folgender Satz verdeutlicht dies.

**Satz 5.4** [Nach [37, Theorem 1]] *Es seien* $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{b} \in \mathbb{R}^n$ *und* $\mathbf{L} \in \mathrm{Sym}^n$, *mit* $\mathbf{Lb} \neq \mathbf{0}$ *und* $\mathbf{A}^\top \mathbf{L} + \mathbf{LA} \neq \mathbf{0}$, *sowie* $\mathbf{R} \in \mathbb{P}^n$. *Des Weiteren seien die reellen Zahlen* $\lambda_1, \ldots, \lambda_J > 0$, *sodass*

$$\det \begin{bmatrix} \lambda_j \mathbf{R} - \mathbf{A}^\top \mathbf{L} - \mathbf{LA} & \mathbf{R} \\ d^{-1} \mathbf{Lbb}^\top \mathbf{L} & \lambda_j \mathbf{R} - \mathbf{A}^\top \mathbf{L} - \mathbf{PL} \end{bmatrix} = 0 \tag{5.39}$$

*und*

$$\mathbf{b}^\top \mathbf{L} (\mathbf{A}^\top \mathbf{L} + \mathbf{LA} - \lambda_j \mathbf{R})^{-1} \mathbf{Lb} > 0. \tag{5.40}$$

*Dann gilt für die Funktion*

$$g(\mathbf{x}) := \mathbf{x}^\top (\mathbf{A}^\top \mathbf{L} + \mathbf{LA}) \mathbf{x} - 2|\mathbf{x}^\top \mathbf{Lb}| \tag{5.41}$$

$$\max_{\mathbf{x}^\top \mathbf{R} \mathbf{x} = d} g(\mathbf{x}) < 0 \tag{5.42}$$

*dann und nur dann, wenn*

$$\lambda_{\max} \left[ \mathbf{U}_2^\top (\mathbf{A}^\top \mathbf{L} + \mathbf{LA}) \mathbf{U}_2 (\mathbf{U}_2^\top \mathbf{R} \mathbf{U}_2)^{-1} \right] < 0, \tag{5.43}$$

*mit der Matrix* $\mathbf{U}_2 \in \mathbb{R}^{n \times (n-1)}$ *aus der Singulärwertzerlegung des Vektors* $\mathbf{Lb}$, *d.h. aus*

$$\mathbf{Lb} = [\mathbf{u}_1 \;\; \mathbf{U}_2] \begin{bmatrix} \sigma(\mathbf{Lb}) \\ \mathbf{0} \end{bmatrix} v, \quad \mathbf{U}_2^\top \mathbf{u}_1 = \mathbf{0}, \tag{5.44}$$

*und*

$$\lambda_j d - \mathbf{b}^\top \mathbf{L}(\mathbf{A}^\top \mathbf{L} + \mathbf{L}\mathbf{A} - \lambda_j \mathbf{R})^{-1}\mathbf{L}\mathbf{b} < 0, \quad \forall\, j \in \{1, \ldots, J\}. \tag{5.45}$$

*Falls Gl.* (5.43) *erfüllt ist und keine reellen Zahlen* $\lambda_j > 0$ *existieren, welche Gl.* (5.39) *und* (5.40) *erfüllen, dann ist Gl.* (5.42) *erfüllt.*

*Bemerkung* 5.3. Die Lösungen der Gl. (5.39) sind gleichzeitig Eigenwerte der Matrix

$$\begin{bmatrix} \mathbf{R}^{-1/2}(\mathbf{A}^\top \mathbf{L} + \mathbf{L}\mathbf{A})\mathbf{R}^{-1/2} & -\mathbf{I} \\ -d^{-1}\mathbf{R}^{-1/2}\mathbf{L}\mathbf{b}\mathbf{b}^\top \mathbf{L}\mathbf{R}^{-1/2} & \mathbf{R}^{-1/2}(\mathbf{A}^\top \mathbf{L} + \mathbf{L}\mathbf{A})\mathbf{R}^{-1/2} \end{bmatrix}. \tag{5.46}$$

Dies folgt aus der Multiplikation der Matrix aus Gl. (5.39) links und rechts mit der nichtsingulären Matrix

$$\begin{bmatrix} \mathbf{R}^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}^{-1/2} \end{bmatrix},$$

wobei die Matrix $\mathbf{R}^{-1/2}$ die (eindeutig bestimmte) Quadratwurzel[17] der positiv definiten Matrix $\mathbf{R}^{-1}$ ist. △

*Beweis.* Für den Fall $\mathbf{x}^\top \mathbf{L}\mathbf{b} = 0$ gilt

$$g(\mathbf{x}) = \mathbf{x}^\top (\mathbf{A}^\top \mathbf{L} + \mathbf{L}\mathbf{A})\mathbf{x}.$$

In diesem Fall lassen sich die Nebenbedingungen

$$\mathbf{x}^\top \mathbf{R}\mathbf{x} = d \tag{5.47}$$

$$\mathbf{x}^\top \mathbf{L}\mathbf{b} = 0 \tag{5.48}$$

in eine einzige Nebenbedingung wie folgt umformen. Alle möglichen Lösungen der Gl. (5.48) können in der Form

$$\mathbf{x}^\top = \mathbf{h}^\top \mathbf{U}_2^\top \tag{5.49}$$

mit einem beliebigen Vektor $\mathbf{h} \in \mathbb{R}_*^{n-1}$ und der Matrix $\mathbf{U}_2 \in \mathbb{R}^{n \times (n-1)}$ aus der Singulärwertzerlegung des Vektors $\mathbf{L}\mathbf{b}$, d.h. aus Gl. (5.44), geschrieben werden. Dies kann man durch Einsetzen der Gl. (5.49) in Gl. (5.48) verifizieren. Durch Einsetzen dieser Lösungen in Gl. (5.47) folgt

$$\mathbf{h}^\top \mathbf{U}_2^\top \mathbf{R}\mathbf{U}_2 \mathbf{h} = d$$

[17] Vgl. [8, Theorem 10.6.1].

und somit

$$\max_{\substack{\mathbf{x}^\top\mathbf{R}\mathbf{x}=d\\ \mathbf{x}^\top\mathbf{L}\mathbf{b}=0}} g(\mathbf{x}) = \max_{\mathbf{h}^\top\mathbf{U}_2^\top\mathbf{R}\mathbf{U}_2\mathbf{h}=d} \mathbf{h}^\top\mathbf{U}_2^\top(\mathbf{A}^\top\mathbf{L}+\mathbf{L}\mathbf{A})\mathbf{U}_2\mathbf{h}$$
$$= \lambda_{\max}\left[\mathbf{U}_2^\top(\mathbf{A}^\top\mathbf{L}+\mathbf{L}\mathbf{A})\mathbf{U}_2(\mathbf{U}_2^\top\mathbf{R}\mathbf{U}_2)^{-1}\right] < 0$$

aufgrund der Bedingung aus Gl. (5.43). Wir betrachten im Weiteren nur noch den Fall $\mathbf{x}^\top\mathbf{L}\mathbf{b} > 0$, da auf Grund der quadratischen Form der Funktion $g(\mathbf{x})$ der Fall $\mathbf{x}^\top\mathbf{L}\mathbf{b} < 0$ zum gleichen maximalen Wert führt. Das Verfahren der Lagrange-Multiplikatoren liefert die Optimalitätsbedingungen

$$(\mathbf{A}^\top\mathbf{L}+\mathbf{L}\mathbf{A}-\lambda\mathbf{R})\mathbf{x} = \mathbf{L}\mathbf{b}, \quad \lambda\in\mathbb{R}, \tag{5.50}$$
$$\mathbf{x}^\top\mathbf{R}\mathbf{x} = d, \tag{5.51}$$

und das Maximum der Funktion $g(\mathbf{x})$ aus Gl. (5.41) lautet

$$g_{\max}(\mathbf{x}) = \lambda d - \mathbf{x}^\top\mathbf{L}\mathbf{b}. \tag{5.52}$$

Falls $\lambda \leq 0$, gilt $g_{\max}(\mathbf{x}) < 0$ da $\mathbf{x}^\top\mathbf{P}\mathbf{b} > 0$. Es wird also nur noch der Fall $\lambda > 0$ betrachtet.

Aus Gl. (5.43) folgt, dass[18)]

$$\lambda_{\max}[\mathbf{U}_2^\top(\mathbf{A}^\top\mathbf{L}+\mathbf{L}\mathbf{A})\mathbf{U}_2(\mathbf{U}_2^\top\mathbf{R}\mathbf{U}_2)^{-1}]$$
$$= \max\{\lambda\in\mathbb{R} : \det\left[\mathbf{U}_2^\top(\mathbf{A}^\top\mathbf{L}+\mathbf{L}\mathbf{A}-\lambda\mathbf{R})\mathbf{U}_2\right]=0\} < 0.$$

Dies bedeutet, dass das größte $\lambda\in\mathbb{R}$, sodass

$$\det\left[\mathbf{U}_2^\top(\mathbf{A}^\top\mathbf{L}+\mathbf{L}\mathbf{A}-\lambda\mathbf{R})\mathbf{U}_2\right]=0$$

gilt, negativ ist und damit, dass für $\lambda > 0$ die Matrix $\mathbf{A}^\top\mathbf{L}+\mathbf{L}\mathbf{A}-\lambda\mathbf{R}$ nichtsingulär ist. Dies kann man wie folgt erklären. Da die Matrixfunktion $f_1(\lambda) = \mathbf{A}^\top\mathbf{L}+\mathbf{L}\mathbf{A}-\lambda\mathbf{R}$ monoton fallend ist, ist die Matrixfunktion $f_2(\lambda) = \mathbf{U}_2^\top(\mathbf{A}^\top\mathbf{L}+\mathbf{L}\mathbf{A}-\lambda\mathbf{R})\mathbf{U}_2$ fallend[19)], und die Matrixfunktion $f_3(\lambda) = \det\left[\mathbf{U}_2^\top(\mathbf{A}^\top\mathbf{L}+\mathbf{L}\mathbf{A}-\lambda\mathbf{R})\mathbf{U}_2\right]$ monoton fallend[20)]. Für $\lambda$ größer als $\lambda_{\max}[\mathbf{U}_2^\top(\mathbf{A}^\top\mathbf{L}+\mathbf{L}\mathbf{A})\mathbf{U}_2(\mathbf{U}_2^\top\mathbf{R}\mathbf{U}_2)^{-1}]$, z.B. für $\lambda > 0$, ist die Determinante kleiner null und die Matrix $\mathbf{A}^\top\mathbf{L}+\mathbf{L}\mathbf{A}-\lambda\mathbf{R}$ daher nichtsingulär.

[18)] Vgl. [8, Fact 8.15.21].
[19)] Vgl. [8, Prop. 8.6.13, xv)].
[20)] Vgl. [8, Prop. 8.6.13, xxii)].

Lösen der Gl. (5.50) ergibt

$$\mathbf{x} = (\mathbf{A}^\top\mathbf{L} + \mathbf{L}\mathbf{A} - \lambda\mathbf{R})^{-1}\mathbf{L}\mathbf{b}, \quad \lambda > 0, \tag{5.53}$$

und Einsetzen in Gl. (5.51) ergibt

$$\mathbf{b}^\top\mathbf{L}(\mathbf{A}^\top\mathbf{L} + \mathbf{L}\mathbf{A} - \lambda\mathbf{R})^{-1}\mathbf{R}(\mathbf{A}^\top\mathbf{L} + \mathbf{L}\mathbf{A} - \lambda\mathbf{R})^{-1}\mathbf{L}\mathbf{b} = d.$$

Dies ist äquivalent zu[21)]

$$\det\begin{bmatrix} d & -\mathbf{b}^\top\mathbf{L}\Psi^{-1} \\ -\Psi^{-1}\mathbf{L}\mathbf{b} & \mathbf{R}^{-1} \end{bmatrix} = 0, \tag{5.54}$$

mit $\Psi = \lambda\mathbf{R} - \mathbf{A}^\top\mathbf{L} - \mathbf{L}\mathbf{A}$. Im Weiteren folgt aus Gl. (5.54), dass

$$\begin{aligned}
\det\left(\begin{bmatrix} d & \mathbf{0}^\top \\ \mathbf{0} & \mathbf{R}^{-1} \end{bmatrix} - \begin{bmatrix} \mathbf{b}^\top\mathbf{L} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}\begin{bmatrix} \Psi^{-1} & \mathbf{0} \\ \mathbf{0} & \Psi^{-1} \end{bmatrix}\begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{L}\mathbf{b} & \mathbf{0} \end{bmatrix}\right) &= 0, \\
\det\left(\begin{bmatrix} \Psi & \mathbf{0} \\ \mathbf{0} & \Psi \end{bmatrix} - \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{L}\mathbf{b} & \mathbf{0} \end{bmatrix}\begin{bmatrix} d^{-1} & \mathbf{0}^\top \\ \mathbf{0} & \mathbf{R} \end{bmatrix}\begin{bmatrix} \mathbf{b}^\top\mathbf{L} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}\right) &= 0, \\
\det\begin{bmatrix} \lambda\mathbf{R} - \mathbf{A}^\top\mathbf{L} - \mathbf{L}\mathbf{A} & -\mathbf{R} \\ -d^{-1}\mathbf{L}\mathbf{b}\mathbf{b}^\top\mathbf{L} & \lambda\mathbf{R} - \mathbf{A}^\top\mathbf{L} - \mathbf{L}\mathbf{A} \end{bmatrix} &= 0, \\
\det\begin{bmatrix} \lambda\mathbf{R} - \mathbf{A}^\top\mathbf{L} - \mathbf{L}\mathbf{A} & \mathbf{R} \\ d^{-1}\mathbf{L}\mathbf{b}\mathbf{b}^\top\mathbf{L} & \lambda\mathbf{R} - \mathbf{A}^\top\mathbf{L} - \mathbf{L}\mathbf{A} \end{bmatrix} &= 0.
\end{aligned}$$

Die letzte Gleichung ist Gl. (5.39). Aus dieser Gleichung kann $\lambda$ bestimmt werden. Wie in der Bemerkung 5.3 gezeigt, sind die Lösungen gleichzeitig Eigenwerte der Matrix aus Gl. (5.46), d.h. deren Berechnung kann numerisch erfolgen. Darüber hinaus muss für ein erzieltes $\lambda > 0$ aufgrund $\mathbf{x}^\top\mathbf{L}\mathbf{b} > 0$ auch

$$\mathbf{b}^\top\mathbf{L}(\mathbf{A}^\top\mathbf{L} + \mathbf{L}\mathbf{A} - \lambda\mathbf{R})^{-1}\mathbf{L}\mathbf{b} > 0$$

gelten, d.h., Gl. (5.40) muss auch erfüllt sein. Schließlich ist der maximale Wert der Funktion $g(\mathbf{x})$ gegeben in Gl. (5.52) negativ dann und nur dann, wenn

$$\lambda_j d - \mathbf{b}^\top\mathbf{L}(\mathbf{A}^\top\mathbf{L} + \mathbf{L}\mathbf{A} - \lambda_j\mathbf{R})^{-1}\mathbf{L}\mathbf{b} < 0, \quad \forall\, j \in \{1, \ldots, J\},$$

gilt, d.h. wenn Gl. (5.45) erfüllt ist, wobei die positiven Zahlen $\lambda_j$ Gl. (5.39)-(5.40) erfüllen müssen. □

[21)] Dies folgt aus der Anwendung von [8, Prop. 8.2.3].

Eine Erweiterung des Korollars 5.3 für polynomiale Selektionsgleichungen mit einem beliebigen Grad ist ebenfalls möglich, da auch in diesem Fall die Adjunkte der Matrix $\mathbf{R}_v$ eine polynomiale Matrix ist. Jedoch erhöhen sich der Grad dieses Polynoms[22)] und der Aufwand in der Berechnung der Adjunkten der Matrix $\mathbf{R}_v$.

## 5.4 Regelungsentwurf

Die *Bang-Bang* WSVR kann mit Hilfe der nicht-konservativen *klassischen* WSVR oder der *invers-polynomialen* WSVR wie folgt entworfen werden:

### 5.4.1 *Klassische* WSVR mittels iLF

**Schritt 5a** Nach den Entwurfsschritten 1a-4a aus Abschnitt 3.3 verwende alternativ das konvergenzoptimale Regelgesetz $u_s$ aus Gl. (5.6).

**Schritt 6a** Alternativ, verwende den *High-Gain* Regler $u_{\text{sat}}$ aus Gl. (5.8) mit dem Parameter $\kappa_v = 0.5\mu_v$ und $\mu_v$ aus Gl. (5.11) und (5.19).

### 5.4.2 *Invers-polynomiale* WSVR

**Schritt 3b** Nach den Entwurfsschritten 1b-2b aus Abschnitt 4.3 verwende alternativ das konvergenzoptimale Regelgesetz $u_s$ aus Gl. (5.6) mit der Matrix $\mathbf{Q}_v$ aus Gl. (4.13).

**Schritt 4b** Alternativ, verwende den *High-Gain*-Regler $u_{\text{sat}}$ aus Gl. (5.8) mit dem Parameter $\kappa_v = 0.5\mu_v$ und $\mu_v$ aus Gl. (5.11), (5.24) und (5.26).

**Schritt 5b** Berechne mit Hilfe des Korollars 5.3 und des Satzes 5.4 das maximale Einzugsgebiet der Ruhelage des Systems, das durch das Regelgesetz $u_s$ aus Gl. (5.6) erzielbar ist.

Tabelle 5.1 zeigt eine Zusammenfassung der hier entwickelten nicht-konservativen *klassischen* und *invers-polynomialen* WSVR zusammen mit den dazu verwendeten Sätzen. Im nächsten Abschnitt werden diese Regler anhand von zwei Beispielen veranschaulicht und mit anderen nichtlinearen Reglern verglichen.

[22)] Vgl. [39, Lemma A.1] oder Lemma A.1 (Anhang).

**Tabelle 5.1:** Zusammenfassung der nichtkonservativen *klassischen* und *invers-polynomialen* WSVR.

| Strecke: $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b}u, \quad \mathbf{x} \in \mathbb{R}^n, u \in \mathbb{R}, \lvert u\rvert \leq 1$ | |
|---|---|
| Die *klassische* WSVR mittels iLF | |
| Aus Satz 3.1 zusammen mit Satz 5.1 | |
| Regelgesetze | Selektionsgleichung |
| $u(\mathbf{x}) = -(\mathbf{D}_v^{-r}\hat{\mathbf{a}} - \mathbf{a})\mathbf{x}$ | |
| $\hat{\mathbf{a}} = \mathbf{a} + c\mathbf{P}^{-1}\mathbf{b}, c > 1/2$ | $\mathbf{x}^\top\mathbf{Q}_v\mathbf{x} - d = 0, d \geq c^2(\mathbf{b}^\top\mathbf{P}^{-1}\mathbf{b})$ |
| $u_{\mathrm{s}}(\mathbf{x}) = -\operatorname{sgn}(\mathbf{b}^\top\mathbf{Q}_v\mathbf{x})$ | $\mathbf{Q}_v = \mathbf{D}_v^{-r}\mathbf{P}_1\mathbf{D}_v, \mathbf{P}_1 = d\mathbf{P}^{-1}$ |
| $u_{\mathrm{sat}}(\mathbf{x}) = -\operatorname{sat}(\kappa_v\mathbf{b}^\top\mathbf{Q}_v\mathbf{x})$ | |
| Die invers-polynomiale WSVR | |
| Aus Satz 4.2 zusammen mit Satz 5.1 | |
| Regelgesetze | Selektionsgleichung |
| $u(\mathbf{x}) = -c\mathbf{b}^\top\mathbf{Q}_v\mathbf{x}, c > 1/2$ | $\mathbf{x}^\top\mathbf{Q}_v\mathbf{x} - d = 0, d \geq c^2(\mathbf{b}^\top\mathbf{R}_\varepsilon^{-1}\mathbf{b})$ |
| $u_{\mathrm{s}}(\mathbf{x}) = -\operatorname{sgn}(\mathbf{b}^\top\mathbf{Q}_v\mathbf{x})$ | $\mathbf{Q}_v = \mathbf{R}_v^{-1}, \mathbf{R}_v = \sum_{i=\mathrm{M_l}}^{\mathrm{M_u}} v^i\mathbf{R}_{\mathrm{c}_i}$ |
| $u_{\mathrm{sat}}(\mathbf{x}) = -\operatorname{sat}(\kappa_v\mathbf{b}^\top\mathbf{Q}_v\mathbf{x})$ | |
| Aus Satz 5.1 zusammen mit Korollar 5.3 | |
| Regelgesetze | Selektionsgleichung |
| $u_{\mathrm{s}}^*(\mathbf{x}) = -\operatorname{sgn}(\mathbf{b}^\top\mathbf{Q}_v\mathbf{x})$ | $\mathbf{x}^\top\mathbf{Q}_v\mathbf{x} - d = 0, d > 1$ |
| $u_{\mathrm{sat}}^*(\mathbf{x}) = -\operatorname{sat}(\kappa_v\mathbf{b}^\top\mathbf{Q}_v\mathbf{x})$ | $\mathbf{Q}_v = \mathbf{R}_v^{-1}, \mathbf{R}_v = \sum_{i=\mathrm{M_l}}^{\mathrm{M_u}} v^i\mathbf{R}_{\mathrm{c}_i}$ |

## 5.5 Beispiele

### 5.5.1 Allgemeine Strecke zweiter Ordnung

Betrachtet wird folgende allgemeine Strecke zweiter Ordnung mit einem stabilen Eigenwert $\lambda = -1$ und einem Eigenwert bei null

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u, \quad |u| \leq 1.$$

Die *null-steuerbare* Region ist der gesamte Zustandsraum. Die Lösung des Validierungsproblems (3.14)-(3.16) lautet

$$\mathbf{P} = \begin{bmatrix} 0.7027 & -0.2610 \\ -0.2610 & 0.5354 \end{bmatrix}.$$

Für ein gegebenes $\varepsilon = 0.1$ und eine (für $v = 1$) mittels $d = 2.1$ skalierte Ellipse $\mathcal{E}_\Delta(1)$ beeinflusst der Reglerparameter $\nu$ aus Gl. (3.19) des Satzes 3.1 die Größe des Gebiets $\mathcal{L}(1)$. Zwei verschiedene Werte illustrieren den Einfluss dieses Parameters. Der erste Wert $\nu = 0.99$ ist sehr nah an der oberen Grenze des erlaubten Intervalls und erzielt ein Gebiet, dessen Rand fast tangential zu dem Rand der Ellipse $\mathcal{E}_\Delta(1)$ ist. Der kleinste Wert $\nu = 0.5212$ erzielt ein viel größeres Gebiet. Abbildung 5.1 veranschaulicht die Ellipse $\mathcal{E}_\Delta(1)$, die Anfangsauslenkung $\mathbf{x}_0 = [0.0594 \quad -0.4766]^\top \in \partial\mathcal{E}_\Delta(1)$ auf derem Rand und die Gebiete $\mathcal{L}(1)$ für die verschiedenen Werte von $\nu$. In beiden Fällen hält die nichtsättigende *klassische* WSVR die Stellgrößenbeschränkung ein.

Tabelle 5.2 zeigt die Regelparameter $\varepsilon$, $d$ und $\nu$, sowie den jeweils erzielten maximalen Wert von $r$. Es ist ersichtlich, dass dieser vom Wert des Parameters $\nu$ beeinflusst wird. Dies liegt daran, dass ein größerer Wert von $\nu$ einen größeren Wert auf der linken Seite der Gl. (3.20) für gegebene $w$ und $\varepsilon$ erzielt, mit der Folge, dass das Intervall $w \in [\varepsilon^r,1]$ verkleinert wird und, somit, dass ein kleineres $r$ erzielt wird. Darüber hinaus beeinflusst der Wert des Parameters $r \in (0,1]$ das Ausregelverhalten. Für $r \to 0$ handelt es sich bei der Regelung um eine nichtsättigende lineare Zustandsrückführung. Für $r \in (0,1]$ ist die nichtsättigende WSVR schneller aufgrund ihrer variablen Struktur. Schließlich zeigt die Tabelle auch den Faktor $\eta$, der die Verstärkung des *High-Gain*-Reglers über den notwendigen minimalen Wert $0.5\mu_v$ erhöht. Je größer $\eta$ ist, desto besser approximiert der *High-Gain*-Regler den konvergenzoptimalen Regler.

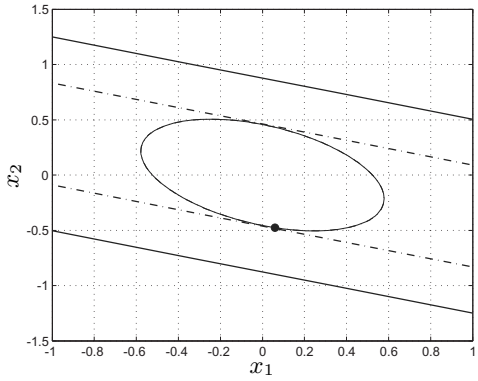


**Bild 5.1:** Das kontraktiv invariante Gebiet für $v = 1$ und $d = 2.1$ ($\mathcal{G}_\Delta(1)$, $-$), sowie die Gebiete $\mathcal{L}(1)$ für $\nu = 0.5212$ ($-$) und $\nu = 0.99$ ($-.$).

**Tabelle 5.2:** Regelparameter für die nichtsättigende *klassische* WSVR mittels iLF ($u_f$) und für den konvergenzoptimalen Regler ($u_s$).

| Regler | $\varepsilon$ | $d$ | $\nu$ | $r$ | $\eta$ | Symbol |
|---|---|---|---|---|---|---|
| $u_f$ | 0.1 | 2.1 | 0.5212 | 1 | - | $(-)$ |
| $u_s$ | 0.1 | 2.1 | 0.5212 | 1 | - | $(-\circ)$ |
| $u_{\text{sat}}$ | 0.1 | 2.1 | 0.5212 | 1 | 16 | $(-+)$ |
| $u_f$ | 0.1 | 2.1 | 0.99 | 0.0048 | - | $(-.)$ |
| $u_s$ | 0.1 | 2.1 | 0.99 | 0.0048 | - | $(-.\square)$ |
| $u_{\text{sat}}$ | 0.1 | 2.1 | 0.99 | 0.0048 | 160 | $(-.\times)$ |

Abbildung 5.2 zeigt das Ausregelverhalten für die verschiedenen Reglerparameter. Dabei wird die gleiche Anfangsauslenkung $\mathbf{x}_0$ gewählt. Verglichen werden die nichtsättigende *klassische* WSVR mit den jeweiligen konvergenzoptimalen Reglern und den High-Gain Reglern, sowie mit dem zeitoptimalen Regler. Der High-Gain Regler aus Satz 5.1 approximiert dabei sehr gut den jeweiligen konvergenzoptimalen Regler. Die gewählte Verstärkung ist gegeben durch $\kappa_v = 0.5\eta\mu_v$, wobei $\eta = 160$ für $\nu = 0.99$ und $\eta = 16$ für $\nu = 0.5212$ gewählt wurde, und der Parameter $\mu_v$ aus Gl. (5.11) mit $\mathbf{S}_v^\perp$ aus Gl. (5.19) und $\mathbf{S}_1^\perp = [-0.9374 \;\; 0.3482]^\top$ berechnet wurde. Schließlich werden darin die Stellgrößenverläufe der nichtsättigenden *klassischen* WSVR, der High-Gain Regler und des zeitoptimalen Reglers.

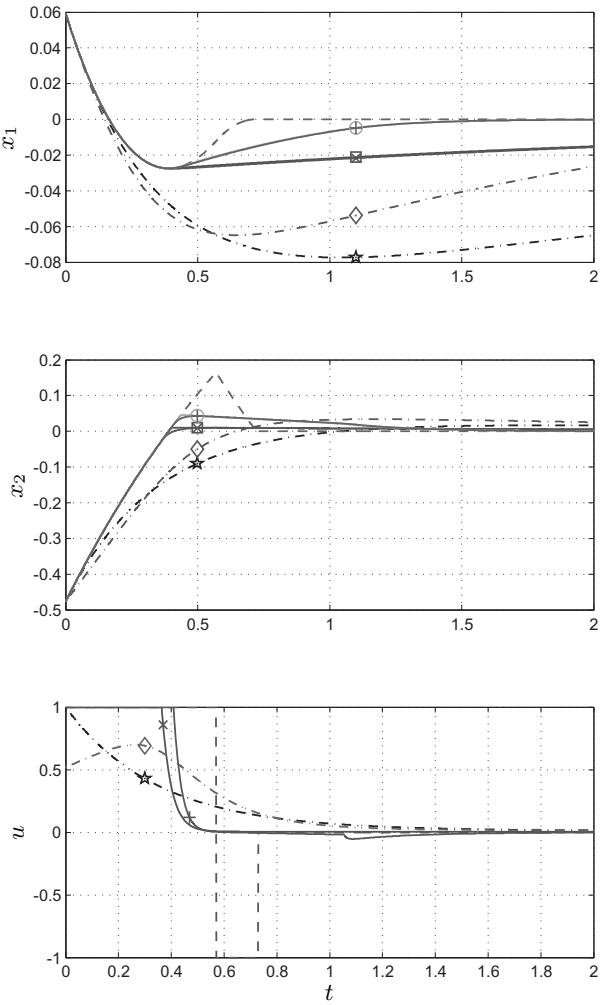


**Bild 5.2:** Simulation der Zustände und Stellgrößen für die nichtsättigende *klassische* WSVR ($\nu = 0.5212$ $(-.\diamond)$ und $\nu = 0.99$ $(-.\star)$), für den konvergenzoptimalen Regler ($\nu = 0.5212$ $(-\circ)$ und $\nu = 0.99$ $(-.\square)$), für den High-Gain Regler ($\nu = 0.5212$ $(-+)$ und $\nu = 0.99$ $(-.\times)$) und für den zeitoptimalen Regler $(--)$.

Aus Übersichtlichkeitsgründen werden die Stellgrößenverläufe der konvergenzoptimalen Regler nicht gezeigt, welche hochfrequent zwischen 1 und $-1$ schalten. Es ist ersichtlich, dass die nichtsättigende *klassische* WSVR für $\nu = 0.99$ nah an der Stellgrößenbeschränkung ist. Dies liegt daran, dass der Rand des entsprechenden Gebiets $\mathcal{L}(v)$ fast tangential zu dem Rand der Ellipse $\mathcal{E}_{\Delta}(1)$ ist, wie man in Abbildung 5.1 sehen kann. Der zeitoptimale Regler schaltet wie erwartet ein Mal zwischen 1 und $-1$. Der High-Gain Regler hat jeweils einen sättigenden aber stetigen Verlauf und erzielt, wie erwartet, eine sehr gute Approximation des Ausregelverhaltens, welches durch den entsprechenden konvergenzoptimalen Regler erzielt wird.

### 5.5.2 Fusionsreaktor

Betrachtet wird folgende lineare Strecke mit einem instabilen und einem stabilen Eigenwert

$$\dot{\mathbf{x}} = \begin{bmatrix} 1 & 0 \\ 0 & -0.5 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 1 \\ -0.5 \end{bmatrix} u, \quad |u| \leq 1.$$

Das Modell beschreibt die wesentlichen physikalischen Eigenschaften von felderzeugenden Strömen, wobei die erzeugten Magnetfelder die Position des Plasmas in einem Fusionsreaktor halten.[23)] Bei dieser Strecke ist zu beachten, dass die *null-steuerbare* Region $\mathbf{x} \in [-1{,}1] \times \mathbb{R}$ ist. Diese ist die Region im Zustandsraum welche durch eine beschränkte Stellgröße $|u| \leq 1$ überhaupt ausregelbar ist. Sie ist beschränkt, da die Strecke einen instabilen Eigenwert bei $\lambda_2 = 1$ besitzt.[24)]

In diesem Beispiel werden die konvergenzoptimale WSVR mit inverspolynomialer Selektionsgleichung und die nichtlineare Regelung aus [28] verglichen. Als Anfangsauslenkung wird $\mathbf{x}_0 = [0.7 \;\; 2.8]^\top$ gewählt. Die Koeffizienten $\mathbf{P}_{c_0}$ und $\mathbf{P}_{c_{-1}}$ des Matrixpolynoms $\mathbf{P}_v$ aus dem Validierungsproblem (4.10)-(4.12) mit $\mathrm{M_l} = -1$ und $\mathrm{M_u} = 0$, sowie $\varepsilon = 0.01$, sind im Abschnitt C.1 (Anhang) angegeben.

In Abbildung 5.3 wird die erzielte Ellipse $\mathcal{E}_{\mathrm{P}}(1) = \{\mathbf{x} \in \mathbb{R}^n | \mathbf{x}^\top \mathbf{P}_1^{-1} \mathbf{x} - 1 < 0\}$ (-) für $v = 1$ aus dem Validierungsproblem gezeigt. Sie enthält zwar nicht die gewünschte Anfangsauslenkung, sie kann aber skaliert werden, sodass sich die Anfangsauslenkung auf derem Rand befindet. Eine Skalierung mit $d^* = (\mathbf{x}_0^\top \mathbf{P}_1^{-1} \mathbf{x}_0)^{-1} = 1.1058$ erzielt die Ellipse

[23)] Vgl. [28] und die Referenzen darin.

[24)] Vgl. [36, Proposition 2.2.1] und Gl. (2.4.2) für die Analyse von *null-steuerbaren* Regionen von Systemen mit reellen Eigenwerten.

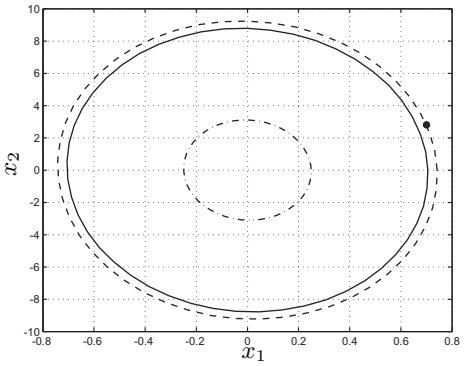


**Bild 5.3:** Anfangszustand $\mathbf{x}_0 = [0.7{,}2.8]^\top$ (•) und erzielte Ellipsen im Zustandsraum für $v = 1$: die aus dem Validierungsproblem (4.10)-(4.12) erzielte Ellipse $\mathcal{E}_\mathrm{P}(1)$ (-), die skalierte Ellipse $\mathcal{E}_\mathrm{P}^u(1)$ (- .) für die Einhaltung der Stellgrößenbeschränkung durch den Regler $u(\mathbf{x}) = -\mathbf{b}^\top\mathbf{P}_v^{-1}\mathbf{x}$, und die skalierte Ellipse $\mathcal{E}_\mathrm{P}^*(1)$ (- -) für den Regler $u_s(\mathbf{x}) = -\operatorname{sgn}(\mathbf{b}^\top\mathbf{P}_v^{-1}\mathbf{x})$.

$\mathcal{E}_\mathrm{P}^*(1) = \{\mathbf{x} \in \mathbb{R}^n | \mathbf{x}^\top\mathbf{P}_1^{-1}\mathbf{x} - d^* < 0\}$ (- -), mit $\mathbf{x}_0 \in \partial\mathcal{E}_\mathrm{P}^*(1)$ (•). Mit Hilfe des Korollars 5.3 muss jedoch überprüft werden, ob die skalierten Ellipsen $\mathcal{E}_\mathrm{P}^*(v)$, mit $v \in [\varepsilon,1]$, ebenfalls kontraktiv invariant für das System mit der konvergenzoptimalen WSVR $u_\mathrm{s}(\mathbf{x}) = -\operatorname{sgn}(\mathbf{b}^\top\mathbf{P}_v^{-1}\mathbf{x})$ sind. Dies wird ebenfalls im Abschnitt C.1 (Anhang) gezeigt. Die dritte Ellipse $\mathcal{E}_\mathrm{P}^u(1) = \{\mathbf{x} \in \mathbb{R}^n | \mathbf{x}^\top\mathbf{P}_1^{-1}\mathbf{x} - d < 0\}$ (- .) mit $d = 0.1249$ zeigt schließlich das Einzugsgebiet der Ruhelage für das System mit der nichtsättigenden invers-polynomialen WSVR $u(\mathbf{x}) = -c\mathbf{b}^\top\mathbf{P}_v^{-1}\mathbf{x}$ und $c = 1$ aus Satz 4.1. Es ist ersichtlich, dass die nichtsättigende invers-polynomiale WSVR ein kleineres Einzugsgebiet der Ruhelage erzielt.

Abbildung 5.4 zeigt die Simulationsergebnisse für den konvergenzoptimalen Regler aus Gl. (5.6), den dazugehörigen *High-Gain* Regler aus Gl. (5.8) und den in [28] vorgestellten nichtlinearen Regler

$$u_\mathrm{nonlin} = \operatorname{sat}(-6x_1 - 3x_2(1 - |x_1|)),$$

der speziell für die Regelung linearer Strecken mit einem einzigen instabilen Pol und einer Stellgrößenbeschränkung entwickelt wurde. Die Zeitverläufe des konvergenzoptimalen und des *High-Gain* Reglers sind deckungsgleich, d.h. der *High-Gain* Regler erzielt fast die gleiche Performance mit einem stetigen Regelgesetz. Lediglich die Stellgrößen unterscheiden sich.

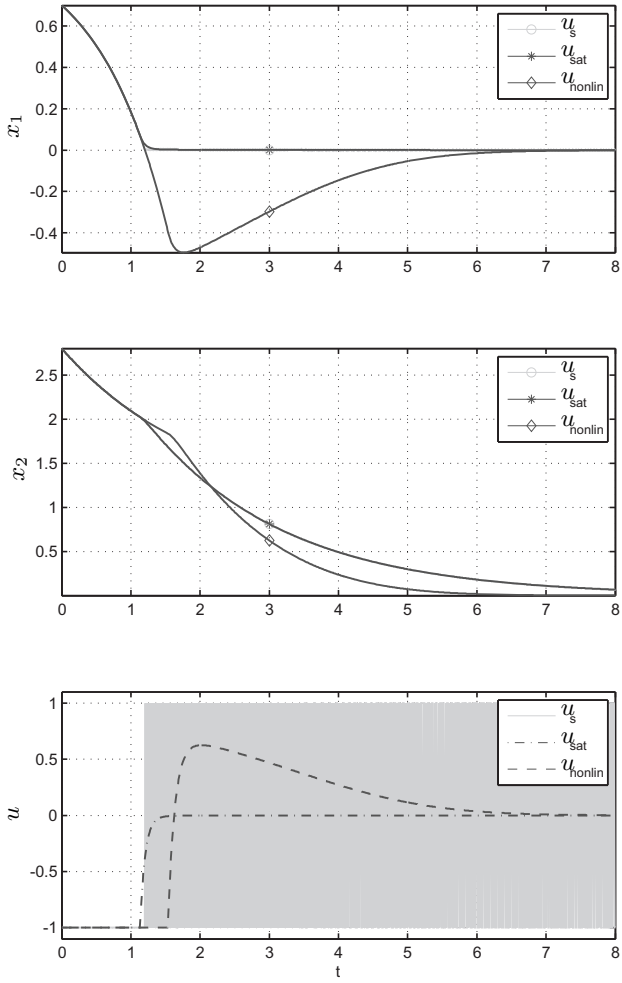


**Bild 5.4:** Simulation des Fusionsreaktor-Modells. Die ersten zwei Abbildungen zeigen die Zeitverläufe der Zustände im Falle des konvergenzoptimalen Reglers $u_s(\mathbf{x})$ ($\circ$) und des dazugehörigen *High-Gain*-Reglers $u_{sat}(\mathbf{x})$ ($*$), welche deckungsgleich sind, sowie des nichtlinearen Reglers aus [28] ($\Diamond$). Die dritte Abbildung zeigt die Zeitverläufe der jeweiligen Stellgrößen: des *High-Gain*-Reglers mit der *invers-polynomialen* Selektionsstrategie (-.), des nichtlinearen Reglers aus [28] (- -), sowie des konvergenzoptimalen Reglers mit der *invers-polynomialen* Selektionsstrategie (-). Letzterer schaltet hochfrequent zwischen $-1$ und 1 und entspricht dem grauen Bereich.

# 6 WSVR-Synthese in Regelstrecken-Ensembles

Um die Performance-Analyse einer nichtlinearen Regelungsmethode über eine einzelne Regelstrecke hinaus zu untersuchen, konzentriert sich Kapitel 8 auf die Entwicklung einer Methode zur Performance-Analyse in Regelstreckenensembles. Die Performance einer Regelungsmethode für ein Regelstreckenensemble kann beispielsweise an einzelnen Strecken aus dem Ensemble exakt überprüft und im übrigen Bereich interpoliert werden. Dabei muss aber garantiert werden, dass im ganzen Interpolationsbereich auch Regler existieren. Dies wird in diesem Kapitel untersucht. Die jeweiligen Bedingungen für die *klassische* WSVR mittels iLFs werden im Abschnitt 6.1 und diejenigen für die *invers-polynomiale* WSVR im Abschnitt 6.2 vorgestellt.

Es werden Regelstreckenensembles betrachtet, welche von einem Parameter aus einer kompakten Menge, vgl. Def. 11 (Anhang), abhängen. Die Abhängigkeit wird als polynomial angenommen. Das hier analysierte Regelstreckenensemble ist wie folgt definiert:

$$\dot{\mathbf{x}} = \mathbf{A}(\theta)\mathbf{x} + \mathbf{b}(\theta)u, \quad \mathbf{x} \in \mathbb{R}^n, u \in \mathbb{R}^n, |u| \leq 1, \theta \in \Theta, \tag{6.1}$$

mit

$$\mathbf{A}(\theta) := \sum_{i=0}^{n_a} \theta^i \mathbf{A}_{c_i}, \quad \mathbf{A}_{c_i} \in \mathbb{R}^{n\times n}, i = 0, \ldots, n_a,\ n_a \in \mathbb{N}, \tag{6.2}$$

$$\mathbf{b}(\theta) := \sum_{i=0}^{n_b} \theta^i \mathbf{b}_{c_i}, \quad \mathbf{b}_{c_i} \in \mathbb{R}^n, i = 0, \ldots, n_b,\ n_b \in \mathbb{N}. \tag{6.3}$$

Es wird darüber hinaus angenommen, dass $n_a, n_b \in \mathbb{N}$ und $\Theta$ bekannt sind. Der Wert von $\theta \in \Theta$ variiert während eines Ausregelvorgangs nicht und ist erst bei der Reglerimplementierung bekannt. Ein solches Problem wird auch *in-situ* Regler-Tuning[1)] genannt.

[1)] Vgl. [39].

Um zu garantieren, dass für jedes $\theta \in \Theta$ ein WSVR Regelgesetz existiert, kann eine bereits existierende Methode aus dem Bereich der robusten Regelung verwendet werden, welche im Zusammenhang mit der WSVR mittels iLF und polynomialer Selektionsstrategien in [40] vorgestellt wurde. Die Methode garantiert die Existenz eines einzigen Reglers für das gesamte Regelstreckenensemble. Dies ist zwar hinreichend aber nicht notwendig und kann entsprechend konservativ sein.

Eine alternative Methode besteht darin, Existenzbedingungen für Regler zu formulieren, die sich an das Regelstreckenensemble anpassen. Sind diese Bedingungen darüber hinaus sowohl hinreichend als auch notwendig, so ist die Untersuchung nicht mehr konservativ. Ein derart entworfener Regler ist dann zwar nicht bezüglich eines Gütemaßes optimiert, er garantiert aber die Stabilisierbarkeit des Regelstreckenensembles durch die untersuchte Regelungsmethode, sodass eine Performance-Analyse für das gesamte Ensemble erfolgen kann. Die Untersuchung dieser Bedingungen wird, wie in den vorherigen Abschnitten, für die *klassische* WSVR mittels iLF und für die invers-polynomiale WSVR vorgestellt.

Dieser Abschnitt ist wie folgt gegliedert. Abschnitt 6.1 stellt die notwendigen und hinreichenden Existenzbedingungen einer *klassischen* WSVR mittels iLF für das Regelstreckenensemble aus Gl. (6.4)-(6.6) vor. In Abschnitt 6.2 werden die notwendigen und hinreichenden Existenzbedingungen einer invers-polynomialen WSVR für das Regelstreckenensemble aus Gl. (6.1)-(6.3) hergeleitet.

## 6.1 Die *klassische* WSVR mittels iLF

Bei der *klassischen* WSVR mittels iLF beschränkt sich diese Untersuchung auf parametrische Reglerstrecken, welche bereits in Steuerungsnormalform vorliegen. Diese haben also die Form

$$\dot{\mathbf{x}} = \mathbf{A}(\theta)\mathbf{x} + \mathbf{b}u, \quad \mathbf{x} \in \mathbb{R}^n, |u| \leq 1, \quad \theta \in \Theta \subset \mathbb{R}, \tag{6.4}$$

mit

$$\mathbf{A}(\theta) := \sum_{i=0}^{n_a} \theta^i \mathbf{A}_{c_i}, \quad \mathbf{A}_i \in \mathbb{R}^{n\times n}, n_a \in \mathbb{N}, \tag{6.5}$$

und

$$\mathbf{A}_{c_i} := \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_{i_0} & -a_{i_1} & -a_{i_2} & \cdots & -a_{i_{n-1}} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}. \tag{6.6}$$

Im Folgenden werden die Existenzbedingungen eines Regelgesetzes der Form

$$u = -\mathbf{k}^\top(v;\theta)\mathbf{x}, \quad u \in \mathbb{R},\ |u| \leq 1, \tag{6.7}$$

mit der Reglerverstärkung

$$\mathbf{k}(v;\theta) := \mathbf{D}^{-1}(v)\hat{\mathbf{a}}(\theta) - \mathbf{a}(\theta), \tag{6.8}$$

bestimmt, wobei der noch zu ermittelnde parameterabhängige Vektor $\hat{\mathbf{a}}(\theta)$ in polynomialer Form angenommen wird, d.h.

$$\hat{\mathbf{a}}(\theta) := \sum_{i=0}^{\hat{n}_a} \theta^i \hat{\mathbf{a}}_{c_i}, \quad \hat{\mathbf{a}}_{c_i} \in \mathbb{R}^n, \tag{6.9}$$

und

$$\mathbf{a}^\top(\theta) := -[0 \ \cdots \ 1]\mathbf{A}(\theta) = \sum_{i=0}^{n_a} \theta^i \mathbf{a}_i^\top, \quad \mathbf{a}_{c_i}^\top = -[0 \ \cdots \ 1]\mathbf{A}_{c_i}. \tag{6.10}$$

Während des Ausregelvorgangs wird der Parameter $v$ durch die Selektionsstrategie

$$\mathbf{x}^\top\mathbf{P}(v;\theta)\mathbf{x} - 1 = 0, \tag{6.11}$$

bestimmt, wobei

$$\begin{aligned} \mathbf{P}(v;\theta) &:= \mathbf{D}^{-1}(v)\mathbf{P}(1;\theta)\mathbf{D}^{-1}(v), \quad \mathbf{P}(1;\theta) \in \mathbb{P}^n, \forall\, \theta \in \Theta \\ \mathbf{D}(v) &:= \operatorname{diag}(v^n, \ldots, v). \end{aligned} \tag{6.12}$$

Gesucht sind ein von $\theta \in \Theta$ abhängiger polynomialer Vektor $\hat{\mathbf{a}}(\theta)$ und eine polynomiale Matrix $\mathbf{P}(1;\theta)$, sodass für alle $\theta \in \Theta$ die Selektionsstrategie eine eindeutige Lösung im Intervall $v \in (0,1]$ hat, das Regelgesetz $u$ beschränkt ist, d.h. $|u| \leq 1$, und der geschlossene Regelkreis stabil ist. Folgender Satz stellt die notwendigen und hinreichenden Bedingungen für

die Existenz einer stabilisierenden WSVR dieser Klasse für das Streckenensemble aus Gl. (6.4)-(6.6) vor. Der Satz baut auf dem Satz 3.1, Seite 17, aus Abschnitt 3 auf und erweitert diesen zur Berücksichtigung von Streckenensembles durch den konstanten Parameter $\theta \in \Theta$.

**Satz 6.1** *Gegeben sei das Ensemble von LTI-Systemen mit einer Eingangsgröße und Stellgrößenbeschränkung aus Gl.* (6.4)-(6.6)*, sowie eine Zahl* $\varepsilon \in (0,1)$*. Folgende Aussagen sind äquivalent:*

*i) Es existieren die Zahlen* $n_{\hat{a}} \in \mathbb{N}$ *und* $n_r \in \mathbb{N}$*, die Vektoren* $\hat{\mathbf{a}}_{c_i} \in \mathbb{R}^n$ *mit* $i = 1,\ldots,n_{\hat{a}}$ *sowie die Matrizen* $\mathbf{Q}_{c_i}$*,* $i = 0,\ldots,n_r$*, sodass für jedes* $\theta \in \Theta$ *die Gebiete*

$$\mathcal{G}_\Delta(v;\theta) := \{\mathbf{x} \in \mathbb{R}^n | g_\Delta(\mathbf{x},v;\theta) = \mathbf{x}^\top \mathbf{Q}(v;\theta)\mathbf{x} - 1 < 0\} \tag{6.13}$$

*mit*

$$\mathbf{Q}(v;\theta) := \mathbf{D}^{-1}(v)\,\mathbf{Q}(1;\theta)\,\mathbf{D}^{-1}(v), \tag{6.14}$$

$$\mathbf{Q}(1;\theta) := \sum_{i=0}^{n_r} \theta^i \mathbf{Q}_{c_i} \succ \mathbf{0} \tag{6.15}$$

*verschachtelt und kontraktiv invariant für alle* $v \in (0,1]$ *für das System aus Gl.* (6.4)-(6.6) *mit dem Regelgesetz*

$$u = -\mathbf{k}^\top(v;\theta)\mathbf{x}, \tag{6.16}$$

$$\mathbf{k}(v;\theta) := \mathbf{D}^{-1}(v)\hat{\mathbf{a}}(\theta) - \mathbf{a}(\theta),\ \mathbf{D}(v) = \mathrm{diag}(v^n,\ldots,v),$$

$$\hat{\mathbf{a}}(\theta) := \sum_{i=0}^{\hat{n}_a} \theta^i \hat{\mathbf{a}}_{c_i}, \quad \hat{\mathbf{a}}_{c_i} \in \mathbb{R}^n, \quad \mathbf{a}^\top(\theta) := -[0 \ \cdots \ 1]\mathbf{A}(\theta), \tag{6.17}$$

*sind. Darüber hinaus gilt*

$$|u| < 1, \quad \forall\, \mathbf{x} \in \mathcal{G}_\Delta(v;\theta), \quad v \in [\varepsilon,1]. \tag{6.18}$$

*ii) Es existieren die Zahl* $n_p \in \mathbb{N}$ *und die Matrizen* $\mathbf{P}_{c_i} \in \mathrm{Sym}^n$*, sodass für alle* $\theta \in \Theta$

$$\mathbf{P}(\theta) := \sum_{i=0}^{n_p} \theta^i \mathbf{P}_{c_i} \succ \mathbf{0}, \tag{6.19}$$

$$\mathbf{A}(\theta)\mathbf{P}(\theta) + \mathbf{P}(\theta)\mathbf{A}^\top(\theta) \prec \mathbf{b}\mathbf{b}^\top, \tag{6.20}$$

$$\mathbf{N}\mathbf{P}(\theta) + \mathbf{P}(\theta)\mathbf{N} \prec \mathbf{0}, \quad \mathbf{N} := \mathrm{diag}(-n, \ldots, -1). \tag{6.21}$$

*Falls ii) gilt, dann ist ein stabilisierendes Regelgesetz gegeben durch*

$$\hat{\mathbf{a}}(\theta) = \mathbf{a}(\theta) + \frac{\det(\mathbf{P}(\theta))}{\delta}\mathbf{P}^{-1}(\theta)\mathbf{b}, \quad \delta := \min_{\theta\in\Theta} \det(\mathbf{P}(\theta)), \tag{6.22}$$

$$\mathbf{Q}(1;\theta) = d(\theta)\det(\mathbf{P}(\theta))\mathbf{P}^{-1}(\theta), \tag{6.23}$$

*wobei der Skalierungsfaktor* $d(\theta)$ *für ein gegebenes* $\theta \in \Theta$ *aus dem Optimierungsproblem*

$$\min\, d(\theta), \ \textit{sodass}$$

$$d(\theta) > 0, \tag{6.24}$$

$$\begin{bmatrix} \det(\mathbf{P}(\theta))\mathbf{P}(\theta)^{-1} & \hat{\mathbf{a}}(\theta) - \mathbf{D}(v)\mathbf{a}(\theta) \\ (\hat{\mathbf{a}}(\theta) - \mathbf{D}(v)\mathbf{a}(\theta))^\top & d(\theta) \end{bmatrix} \succ \mathbf{0}, \quad \forall\, v \in [\varepsilon,1], \tag{6.25}$$

*berechnet wird.*

*Bemerkung* 6.1. Da die Matrix $\mathbf{P}(\theta)$ in polynomieller Form vorliegt, können Gl. (6.19)-(6.21) in äquivalente parameterunabhängige LMIs transformiert werden. Die Transformation basiert auf [78, Lemma 4.4], welches eine Generalisierung der in [38] vorgestellten $S$-Prozedur verwendet. △

*Bemerkung* 6.2. Für ein gegebenes $\theta \in \Theta$ ist dieser Satz sehr ähnlich zu dem Satz 3.1 aus Abschnitt 3. Der einzige Unterschied besteht in der Form des vorgegebenen Regelgesetzes in Gl. (6.22)-(6.25). △

*Beweis.* Der Beweis wird auf die gleiche Weise wie im Satz 3.1 aus Abschnitt 3 ausgeführt. Durch Einsetzen des vorgegebenen Regelgesetzes aus Gl. (6.16), mit dem Vektor $\hat{\mathbf{a}}(\theta)$ aus Gl. (6.22), und der vorgegebenen Matrix $\mathbf{Q}(1;\theta)$ aus Gl. (6.23) wird gezeigt, dass die Bedingungen aus Gl. (6.19)-(6.21) hinreichend für $i$) sind. Deren Notwendigkeit wird mit Hilfe von Finsler's Lemma gezeigt. Da die Beweise in diesem Punkt deckungs-

gleich sind, wird hier nur auf den Unterschied zwischen den beiden Regelgesetzen eingegangen.

Das Regelgesetz aus Gl. (6.16), mit dem Vektor $\hat{\mathbf{a}}(\theta)$ aus Gl. (6.22), wird so bestimmt, dass es polynomial in $\theta$ ist. Dies wird durch den zusätzlichen Skalierungsfaktor $\det(\mathbf{P}(\theta))$ in Gl. (6.22) erreicht, da

$$\det(\mathbf{P}(\theta))\mathbf{P}(\theta)^{-1} = \mathbf{P}(\theta)^{A} \tag{6.26}$$

gilt, und die Adjunkte $\mathbf{P}(\theta)^{A}$ der Matrix $\mathbf{P}(\theta)$ wiederum polynomial in $\theta$ ist. Dies gilt auch für die Matrix $\mathbf{Q}(1;\theta)$. Darüber hinaus wird der konstante Skalierungsfaktor $\delta$ eingeführt, sodass $\det(\mathbf{P}(\theta))/\delta \geq 1$, für alle $\theta \in \Theta$ ist. Dies muss erfüllt sein, damit das Regelgesetz kontraktiv invariante Gebiete $\mathcal{G}_{\Delta}(v;\theta)$ für alle $v \in (0,1]$ erzeugt, vgl. Satz 3.1, bzw. dessen Beweis. Schließlich wird der Skalierungsfaktor $d(\theta)$ in Gl. (6.23) eingeführt, um sicherzustellen, dass das Regelgesetz $u$ beschränkt ist, d.h. dass Gl. (6.18) gilt. Jeder Skalierungsfaktor $d(\theta)$, der die Bedingungen (6.24) und (6.25) erfüllt, ist zulässig. Der kleinste Faktor $d(\theta)$ ergibt dabei das maximale Einzugsgebiet $\mathcal{G}_{\Delta}(1;\theta)$. Für ein gegebenes $\theta \in \Theta$ ist die linke Seite der Bedingung (6.25) eine polynomiale Matrix in $v \in [\varepsilon,1]$, welche, wie im Satz 3.1, in eine äquivalente LMI transformiert werden kann. □

## 6.2 *Invers-polynomiale* WSVR für Regelstreckenensembles

Wie im vorigen Abschnitt baut der nächste Satz auf dem Satz 4.1, Seite 28, auf und erweitert ihn zur Berücksichtigung von Streckenensembles durch den konstanten Parameter $\theta \in \Theta$. Dieser Satz wurde zum ersten Mal im [59] vorgestellt.

**Satz 6.2** *Gegeben seien die Zahlen $a,b \in \mathbb{N}$, die Matrizen $\mathbf{A}_{c_i} \in \mathbb{R}^{n\times n}$, mit $i = 0,\ldots,a$ und die Vektoren $\mathbf{b}_{c_i} \in \mathbb{R}^n$, mit $i = 0,\ldots,b$, für das folgende LTI-System mit einer Eingangsgröße und Stellgrößenbeschränkung*

$$\dot{\mathbf{x}} = \mathbf{A}(\theta)\mathbf{x} + \mathbf{b}(\theta)u, \quad u \in \mathbb{R}, |u| \leq 1, \tag{6.27}$$

*wobei*

$$\mathbf{A}(\theta) := \sum_{i=0}^{a} \theta^i \mathbf{A}_{c_i}, \quad \mathbf{b}(\theta) := \sum_{i=0}^{b} \theta^i \mathbf{b}_{c_i}, \tag{6.28}$$

*sowie eine reelle Zahl* $\varepsilon \in (0,1)$. *Folgende Aussagen sind äquivalent:*

*i) Es existieren die Zahlen* $R \in \mathbb{N}$, $\mathrm{M_l}, \mathrm{M_u} \in \mathbb{Z}$, *mit* $\mathrm{M_l} < 0$, $\mathrm{M_l} \leq \mathrm{M_u}$, *die Matrizen* $\mathbf{R}_{c_{ij}} \in \mathrm{Sym}^n$, *mit* $i = \mathrm{M_l}, \ldots, \mathrm{M_u}$, $j = 0, \ldots, R$, *sowie die Vektorfunktion* $k(v;\theta) : \mathcal{V} \to \mathbb{R}^n$, *mit* $\mathcal{V} := \{(v,\theta) | v \in [\varepsilon,1],\, \theta \in \Theta\}$, *sodass für jedes gegebene* $\theta \in \Theta$ *die Gebiete*

$$\mathcal{G}_\Lambda(v;\theta) := \left\{ \mathbf{x} \in \mathbb{R}^n \,\middle|\, g_\Lambda(\mathbf{x},v;\theta) = \mathbf{x}^\top \mathbf{Q}(v;\theta)\mathbf{x} - 1 < 0 \right\}, \tag{6.29}$$

*mit*

$$\mathbf{Q}(v;\theta) := \mathbf{R}(v;\theta)^{-1}, \tag{6.30}$$

$$\mathbf{R}(v;\theta) := \sum_{i=\mathrm{M_l}}^{\mathrm{M_u}} v^i \left( \sum_{j=0}^{R} \theta^j \mathbf{R}_{c_{ij}} \right) \succ \mathbf{0}, \quad \forall\, (v,\theta) \in \mathcal{V}, \tag{6.31}$$

*für alle* $v \in [\varepsilon,1]$ *verschachtelt und kontraktiv invariant für das System* (6.27) *mit dem Regelgesetz*

$$u = -\mathbf{k}(v;\theta)^\top \mathbf{x}, \quad \mathbf{k}(v;\theta) : \mathcal{V} \to \mathbb{R}^n \tag{6.32}$$

*sind, und*

$$|u(\mathbf{x})| \leq 1, \quad \forall \mathbf{x} \in \mathcal{G}_\Lambda(v;\theta),\ v \in [\varepsilon,1]. \tag{6.33}$$

*ii) Es existieren die Zahlen* $\rho \in \mathbb{N}$, $\mathrm{m_l}, \mathrm{m_u} \in \mathbb{Z}$, *mit* $\mathrm{m_l} < 0$, $\mathrm{m_l} \leq \mathrm{m_u}$ *und die Matrizen* $\mathbf{P}_{c_{ij}} \in \mathrm{Sym}^n$, *mit* $i = \mathrm{m_l}, \ldots, \mathrm{m_u}$, $j = 0, \ldots, r$, *sodass für jedes Tupel* $(v;\theta) \in \mathcal{V}$ *gilt*

$$\mathbf{P}(v;\theta) := \sum_{i=\mathrm{m_l}}^{\mathrm{m_u}} v^i \left( \sum_{j=0}^{r} \theta^j \mathbf{P}_{c_{ij}} \right) \succ \mathbf{0}, \tag{6.34}$$

$$\partial_v \mathbf{P}(v;\theta) \succ \mathbf{0}, \tag{6.35}$$

$$\mathbf{A}(\theta)\mathbf{P}(v;\theta) + \mathbf{P}(v;\theta)\mathbf{A}(\theta)^\top \prec \mathbf{b}(\theta)\mathbf{b}(\theta)^\top. \tag{6.36}$$

*Falls ii) gilt, dann ist ein stabilisierendes Regelgesetz in i) gegeben durch*

$$\mathbf{k}(v;\theta)^\top = \mathbf{b}(\theta)^\top \mathbf{P}(v;\theta)^{-1}, \quad \mathbf{Q}(v;\theta) = d(\theta)\, \mathbf{P}(v;\theta)^{-1}, \tag{6.37}$$

*wobei*

$$d(\theta) \geq \mathbf{b}(\theta)^\top \mathbf{P}(\varepsilon;\theta)^{-1} \mathbf{b}(\theta) > 0. \tag{6.38}$$

*Bemerkung* 6.3. Die parameterabhängigen Bedingungen aus Gl. (6.34)-(6.36) können in äquivalente parameterunabhängige LMIs transformiert werden. Die Transformation beruht auf der allgemeinen $S$-Prozedur aus [38] und wurde in [77] vorgestellt, vgl. Abschnitt A.3 (Anhang). Diese kann als Verallgemeinerung der im Abschnitt 4 verwendeten Transformation für polynomiale Matrizen mit nur einem Parameter gesehen werden. Für den Fall mehrparametriger polynomialer Matrizen werden im folgenden Abschnitt die LMIs vorgestellt. △

*Beweis.* Für ein vorgegebenes $\theta \in \Theta$ sind die Punkte *i*) und *ii*) äquivalent zu den gleichnamigen Punkten aus Theorem 4.1. Daher wird hier der Beweis nicht wiederholt aufgeführt. □

### 6.2.1 Umwandlung der Stabilitätsbedingungen in LMIs

Gl. (6.34)-(6.36) können in äquivalente parameterunabhängige LMIs transformiert werden. Die Transformation beruht auf [77, Theorem 6.3], und wird im Folgenden beschrieben.

**Satz 6.3** [Vgl. [77, Theorem 6.3]] *Folgende Aussagen sind äquivalent:*
*i) Für alle* $(\rho_1,\rho_2) \in [-1,1] \times [-1,1]$ *gilt für die polynomiell parameterabhängige quadratische Matrix*[a)]

$$\mathbf{P}(\rho_1,\rho_2) := \left(\boldsymbol{\rho}_2^{[\bar{\alpha}_2]} \otimes \boldsymbol{\rho}_1^{[\bar{\alpha}_1]} \otimes \mathbf{I}_n\right)^\top \mathbf{P}_\Sigma \left(\boldsymbol{\rho}_2^{[\bar{\alpha}_2]} \otimes \boldsymbol{\rho}_1^{[\bar{\alpha}_1]} \otimes \mathbf{I}_n\right) \succ \mathbf{0}, \quad (6.39)$$

*mit* $\boldsymbol{\rho}_1^{[\bar{\alpha}_1]} = [1 \;\; \rho_1 \;\; \rho_1^2 \;\; \cdots \;\; \rho_1^{\bar{\alpha}_1-1}]^\top$ *und* $\boldsymbol{\rho}_2^{[\bar{\alpha}_2]} = [1 \;\; \rho_2 \;\; \rho_2^2 \;\; \cdots \;\; \rho_2^{\bar{\alpha}_2-1}]^\top$.
*ii) Es existieren die Matrizen* $\mathbf{D}_1,\mathbf{D}_2 \in \mathrm{Sym}^q$, *mit* $\mathbf{D}_1,\mathbf{D}_2 \succ \mathbf{0}$ *und* $q = 2 \cdot (\bar{\alpha}_1 - 1) \cdot (\bar{\alpha}_2 - 1) \cdot n$, *sodass*

$$-\mathbf{P}_\Sigma + \begin{bmatrix}\mathbf{J}_K \\ \mathbf{C}_K\end{bmatrix}^\top \begin{bmatrix}-\mathbf{D}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{D}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{D}_2\end{bmatrix} \begin{bmatrix}\mathbf{J}_K \\ \mathbf{C}_K\end{bmatrix} \prec \mathbf{0}, \quad (6.40)$$

*wobei*

$$\mathbf{J}_K := \begin{bmatrix} \hat{\mathbf{J}}_{\alpha_2-1} \otimes \check{\mathbf{J}}_{\alpha_1-1} \otimes \mathbf{I}_n \\ \check{\mathbf{J}}_{\alpha_2-1} \otimes \check{\mathbf{J}}_{\alpha_1-1} \otimes \mathbf{I}_n \\ \hat{\mathbf{J}}_{\alpha_2-1} \otimes \hat{\mathbf{J}}_{\alpha_1-1} \otimes \mathbf{I}_n \\ \check{\mathbf{J}}_{\alpha_2-1} \otimes \check{\mathbf{J}}_{\alpha_1-1} \otimes \mathbf{I}_n \end{bmatrix}, \quad \mathbf{C}_K := \begin{bmatrix} \hat{\mathbf{J}}_{\alpha_2-1} \otimes \hat{\mathbf{J}}_{\alpha_1-1} \otimes \mathbf{I}_n \\ \check{\mathbf{J}}_{\alpha_2-1} \otimes \hat{\mathbf{J}}_{\alpha_1-1} \otimes \mathbf{I}_n \\ \hat{\mathbf{J}}_{\alpha_2-1} \otimes \hat{\mathbf{J}}_{\alpha_1-1} \otimes \mathbf{I}_n \\ \hat{\mathbf{J}}_{\alpha_2-1} \otimes \check{\mathbf{J}}_{\alpha_1-1} \otimes \mathbf{I}_n \end{bmatrix}, \quad (6.41)$$

*und*

$$\hat{\mathbf{J}}_k = \begin{bmatrix} \mathbf{I}_k & \mathbf{0}_{k,1} \end{bmatrix}, \quad \check{\mathbf{J}}_k = \begin{bmatrix} \mathbf{0}_{k,1} & \mathbf{I}_k \end{bmatrix}.$$

[a] Vgl. Def. 27 (Anhang) für die Definition einer polynomiell parameterabhängigen quadratischen (matrixwertigen) Funktion.

### Negative Exponenten

Für die Umwandlung der Polynome aus Gl. (6.34)-(6.36) in die Form eines Polynoms aus Gl. (6.39) müssen diese erstens so transformiert werden, dass die Exponenten von $v$ positiv sind. Dies ist möglich durch eine mehrmalige Multiplikation der Gleichungen mit dem Parameter $v$. Da

$$\begin{aligned} \mathbf{P}(v,\theta) = \sum_{i=\mathrm{m_l}}^{\mathrm{m_u}} v^i \left( \sum_{j=0}^{r} \theta^j \mathbf{P}_{c_{ij}} \right) &= v^{\mathrm{m_l}} \sum_{i=\mathrm{m_l}}^{\mathrm{m_u}} v^{i-\mathrm{m_l}} \left( \sum_{j=0}^{r} \theta^j \mathbf{P}_{c_{ij}} \right) \\ &= v^{\mathrm{m_l}} \sum_{k=0}^{\mathrm{m_u}-\mathrm{m_l}} v^k \left( \sum_{j=0}^{r} \theta^j \mathbf{P}_{c_{k+\mathrm{m_l},j}} \right) \end{aligned}$$

gilt, ist Gl. (6.34) äquivalent zu

$$\mathbf{S}_1(v,\theta) := \sum_{k=0}^{\mathrm{m_u}-\mathrm{m_l}} v^k \left( \sum_{j=0}^{r} \theta^j \mathbf{P}_{c_{k+\mathrm{m_l},j}} \right) \succ \mathbf{0}, \quad \forall\, (v,\theta) \in [\varepsilon,1] \times \Theta, \quad (6.42)$$

und, da

$$\begin{aligned}\partial_v \mathbf{P}(v,\theta) &= \sum_{i=\mathrm{m_l}}^{\mathrm{m_u}} i v^{i-1} \left( \sum_{j=0}^{r} \theta^j \mathbf{P}_{c_{ij}} \right) \\ &= v^{\mathrm{m_l}-1} \sum_{i=\mathrm{m_l}}^{\mathrm{m_u}} i v^{i-\mathrm{m_l}} \left( \sum_{j=0}^{r} \theta^j \mathbf{P}_{c_{ij}} \right) \\ &= v^{\mathrm{m_l}-1} \sum_{k=0}^{\mathrm{m_u}-\mathrm{m_l}} (k+\mathrm{m_l}) v^k \left( \sum_{j=0}^{r} \theta^j \mathbf{P}_{c_{k+\mathrm{m_l},j}} \right)\end{aligned}$$

gilt, ist Gl. (6.35) äquivalent zu

$$\mathbf{S}_2(v,\theta) := \sum_{k=0}^{\mathrm{m_u}-\mathrm{m_l}} (k+\mathrm{m_l}) v^k \left( \sum_{j=0}^{r} \theta^j \mathbf{P}_{c_{k+\mathrm{m_l},j}} \right) \succ \mathbf{0}, \quad \forall\, (v,\theta) \in [\varepsilon,1] \times \Theta. \tag{6.43}$$

Schließlich ist Gl. (6.36) äquivalent zu

$$\begin{aligned}\mathbf{S}_3(v,\theta) &:= \mathbf{A}(\theta)\mathbf{S}_1(v,\theta) + \mathbf{S}_1(v,\theta)\mathbf{A}(\theta)^\top - v^{-\mathrm{m_l}} \mathbf{b}(\theta)\mathbf{b}(\theta)^\top \\ &= \sum_{k=0}^{\mathrm{m_u}-\mathrm{m_l}} v^k \left( \sum_{i=0}^{n_a} \sum_{j=0}^{r} \theta^{i+j} (\mathbf{A}_i \mathbf{P}_{c_{k+\mathrm{m_l},j}} + \mathbf{P}_{c_{k+\mathrm{m_l},j}} \mathbf{A}_i^\top) \right) \\ &\quad - v^{-\mathrm{m_l}} \sum_{i=0}^{n_b} \sum_{i=0}^{n_b} \theta^{i+j} \mathbf{b}_i \mathbf{b}_j^\top \prec \mathbf{0}, \quad \forall\, (v,\theta) \in [\varepsilon,1] \times \Theta.\end{aligned} \tag{6.44}$$

Gl.(6.42)-(6.44) können in der allgemeinen Form

$$\mathbf{S}_\iota(v,\theta) = \sum_{k=0}^{n_v} v^k \sum_{i=0}^{n_\theta} \theta^i \mathbf{S}_{c_{k,i}} \succ \mathbf{0}, \quad \forall\, (v,\theta) \in [\varepsilon,1] \times \Theta, \quad \iota = 1,2,3. \tag{6.45}$$

geschrieben werden. Gl. (6.45) wird im Folgenden mit Hilfe des Theorems 6.3 in eine parameterunabhängige LMI transformiert. Dazu werden erstens zwei Intervalltransformationen durchgeführt, sodass beide Parameter im Intervall $[-1,1]$ liegen und anschließend die polynomiale Matrix umgeformt, sodass diese in der Form aus Punkt *i*) des Theorems 6.3 vorliegt.

**Intervalltransformationen**

Da der Parameter $v$ aus Gl. (6.45) im Intervall $v \in [\varepsilon,1]$ liegt, wird stattdessen der transformierte Parameter $\tilde{v} := 2/\alpha v - \beta/\alpha$, mit $\alpha = 1-\varepsilon$ und $\beta = 1+\varepsilon$ verwendet, wobei $\tilde{v} \in [-1,1]$ gilt. Für Gl. (6.45) ergibt sich nach einigen Umformungen

$$\mathbf{S}_\iota(v,\theta) = \sum_{k=0}^{n_v} \tilde{v}^k \alpha^k \sum_{j=k}^{n_v} 2^{-j} \binom{j}{k} \beta^{j-k} \left( \sum_{i=0}^{n_\theta} \theta^i \mathbf{S}_{c_{j,i}} \right) =: \tilde{\mathbf{S}}_\iota(\tilde{v},\theta). \tag{6.46}$$

Wir nehmen an, dass das Intervall $\Theta$ in der Form $\Theta = [\theta_l,\theta_u]$ vorliegt. Eine zweite Intervalltransformation wird durch $\tilde{\theta} := 2/\gamma\theta - \delta/\gamma$, mit $\gamma = \theta_u - \theta_l$ und $\delta = \theta_u + \theta_l$ vorgenommen, sodass $\tilde{\theta} \in [-1,1]$ gilt. Folglich ist Gl. (6.46) äquivalent zu

$$\begin{aligned}\tilde{\mathbf{S}}_\iota(\tilde{v},\theta) = &\sum_{k=0}^{n_v} \tilde{v}^k \alpha^k \sum_{j=k}^{n_v} 2^{-j} \binom{j}{k} \beta^{j-k} \left( \sum_{i=0}^{n_\theta} \tilde{\theta}^i \gamma^i \sum_{l=i}^{n_\theta} 2^{-l} \binom{l}{i} \delta^{l-i} \mathbf{S}_{c_{jl}} \right) \\ &=: \tilde{\tilde{\mathbf{S}}}_\iota(\tilde{v},\tilde{\theta}).\end{aligned}$$

Mit der Notation $\tilde{v} =: \rho_1$, $\tilde{\theta} =: \rho_2$ kann die Matrix $\tilde{\tilde{\mathbf{S}}}_\iota(\tilde{v},\tilde{\theta})$ in der allgemeinen Form

$$\mathbf{T}_\iota(\rho_1,\rho_2) := \tilde{\tilde{\mathbf{S}}}_\iota(\rho_1,\rho_2) = \sum_{k_1=k_2=0}^{\bar{\rho}_1,\bar{\rho}_2} \rho_1^{k_1} \rho_2^{k_2} \mathbf{T}_{\iota c_{k_1,k_2}}, \tag{6.47}$$

mit $\bar{\rho}_1 := n_v$, $\bar{\rho}_2 := n_\theta$, sowie

$$\mathbf{T}_{\iota c_{k_1,k_2}} := \alpha^{k_1} \sum_{j=k_1}^{\bar{\rho}_1} 2^{-j} \binom{j}{k_1} \beta^{j-k_1} \gamma^{k_2} \sum_{l=k_2}^{\bar{\rho}_2} 2^{-l} \binom{l}{k_2} \delta^{l-k_2} \mathbf{S}_{c_{jl}}$$

und $\alpha := 1-\varepsilon$, $\beta := 1+\varepsilon$, $\gamma := \theta_u - \theta_l$, $\delta := \theta_u + \theta_l$ geschrieben werden.

**Umformung der polynomialen Matrix $\mathbf{S}(\tilde{v},\tilde{\theta})$**

Gl. (6.47) kann darüber hinaus in der Form

$$\mathbf{T}_\iota(\rho_1,\rho_2) = \left( \boldsymbol{\rho}_2^{[\bar{\alpha}_2]} \otimes \boldsymbol{\rho}_1^{[\bar{\alpha}_1]} \otimes \mathbf{I}_n \right)^\top \mathbf{T}_{\iota_\Sigma} \left( \boldsymbol{\rho}_2^{[\bar{\alpha}_2]} \otimes \boldsymbol{\rho}_1^{[\bar{\alpha}_1]} \otimes \mathbf{I}_n \right), \tag{6.48}$$

mit $\boldsymbol{\rho}_1^{[\alpha_1]} = [1 \;\; \rho_1 \;\; \rho_1^2 \;\; \cdots \;\; \rho_1^{\bar{\alpha}_1-1}]^\top$ und $\boldsymbol{\rho}_2^{[\bar{\alpha}_2]} = [1 \;\; \rho_2 \;\; \rho_2^2 \;\; \cdots \;\; \rho_2^{\alpha_2-1}]^\top$, sowie $\bar{\alpha}_i := \lceil \bar{\rho}_i/2 \rceil + 1$, $i = 1{,}2$, geschrieben werden, wobei die konstante Block-Matrix $\mathbf{T}_{\iota_\Sigma} \in \mathrm{Sym}^{\alpha_1 \cdot \alpha_2 \cdot n}$ aus

$$\mathbf{T}_{\iota_\Sigma} := \frac{1}{2}(\bar{\mathbf{T}}_{\iota_\Sigma} + \bar{\mathbf{T}}_{\iota_\Sigma}^\top)$$

gebildet wird. Dabei wird der $(f_1,f_2)$-te Block der Matrix $\bar{\mathbf{T}}_{\iota_\Sigma}$ durch

$$\bar{\mathbf{T}}_{\iota_{\Sigma_{(f_1,f_2)}}} := \mathbf{T}_{\iota_{c_{k_1,k_2}}}$$

gebildet, wobei die Indizes $f_1$ und $f_2$ aus

$$\begin{aligned} f_1 &:= f_K(\alpha_1,\alpha_2) \\ f_2 &:= f_K(\beta_1,\beta_2) \end{aligned}$$

mit

$$f_K(c_1,c_2) := \bar{\alpha}_1 c_2 + c_1 + 1$$

und

$$\alpha_i := \left\lceil \frac{k_i}{2} \right\rceil, \quad \beta_i := \left\lfloor \frac{k_i}{2} \right\rfloor, \quad i = 1{,}2,$$

berechnet werden. Die restlichen Komponenten der Blockmatrix, welche durch die Indizes $f_1$ und $f_2$ nicht festgelegt wurden, werden zu $\mathbf{0}_{n,n}$ gesetzt. Die Umformung aus Gl. (6.48) ist allerdings nicht eindeutig.

Die Sätze 6.1 und 6.2 beinhalten die notwendigen und hinreichenden Bedingungen für die Stabilisierbarkeit eines Regelstreckenensembles durch *klassische* bzw. *invers-polynomiale* WSVRs. Sie können als Erweiterungen der im Kapitel 3 und 4 vorgestellten Regelungen gesehen werden.

Abschließend werden hier noch die Hauptbeiträge des ersten Teils der Arbeit zusammengefasst. Dieser stellt mehrere Weiterentwicklungen weicher strukturvariabler Regelungen mittels impliziter Ljapunov-Funktionen vor, deren Hauptaugenmerk die Nicht-Konservativität der Regelgesetze bildet. Nach einem einleitenden Kapitel über die Stabilisierung linearer Systeme mit Stellgrößenbeschränkung werden im Kapitel 3 die hinreichenden und notwendigen Bedingungen der *klassischen* WSVR mittels iLF vorgestellt. Daran anschließend wird die hier neu-entwickelte *invers-polynomiale* WSVR in Kapitel 4 eingeführt. Beide Kapitel haben dabei eine ähnliche Struktur. Sie beginnen mit der Definition der jeweiligen Regelung,

werden dann mit den nicht-konservativen Stabilitätsbedingungen fortgeführt und enden mit einem Abschnitt über einen möglichen Regelungsentwurf. Kapitel 5 stellt die konvergenzoptimale (*Bang-Bang*) WSVR dar. Dabei werden ebenfalls nicht-konservative Stabilitätsbedingungen vorgestellt. Darüber hinaus wird der Aufbau stetiger Approximationen der vorgestellten *Bang-Bang* Regelgesetze und, abschließend, ein möglicher Regelungsentwurf vorgestellt. Der letzte Abschnitt des Kapitels veranschaulicht die oben vorgestellten Regelgesetze anhand von zwei Beispielen mit Reglerstrecken zweiter Ordnung. Der erste Teil der Arbeit endet mit der im Kapitel 6 vorgestellten WSVR-Synthese für Regelstreckenensembles. Diese werden im zweiten Teil der Arbeit im Rahmen der Performance-Analyse in nichtlinearen Regelkreisensembles verwendet.

# Teil II

# Performance-Analyse nichtlinearer Regelkreise

# 7 Performance-Maße in nichtlinearen Regelkreisen

Lineare und nichtlineare Regelkreise unterscheiden sich wesentlich in ihren Eigenschaften. Das Superpositionsprinzip, die Darstellbarkeit jedes linearen Systems mit Hilfe einer endlichen Anzahl von Parametern, die Anwendbarkeit des Satzes von Cayley-Hamilton mit den jeweiligen Folgen für den linearen Regelkreis, z.B. bzgl. Steuerbarkeit und Beobachtbarkeit, haben zu einer vollständingen Theorie linearer Systeme geführt. Diese Eigenschaften sind im Allgemeinen bei nichtlinearen Systemen nicht vorhanden. Dies hat zur Folge, dass nur bestimmte Klassen nichtlinearer Systeme, welche besondere Eigenschaften haben, einheitlich untersucht werden können. Eine Klassifizierung nichtlinearer Systeme kann man z.B. in [65] finden.

Die Eigenschaften eines nichtlinearen Systems können unter bestimmten Bedingungen mit Hilfe der linearen Systemtheorie untersucht werden. Diese wird zur asymptotischen Analyse nichtlinearer Zustandslösungen in der Nähe linearer Lösungen erweitert. Mit dieser Thematik beschäftigen sich beispielsweise [11, 31, 43, 48, 56, 67, 76]. Ein Beispiel einer solchen Methode ist die Analyse von Dauerschwingungen in der Nähe linearer harmonischer Schwingungen, vgl. dazu [31]. Die Erweiterung der linearen Systemtheorie in dieser Richtung basiert auf der asymptotischen Methode von Krylov und Bogoliubov, vgl. [43]. In der Arbeit von [56] wurde diese Methode auf gedämpfte nichtlineare Schwingungen erweitert. Die Voraussetzungen für diese Untersuchung war, dass die Dämpfung und *Frequenz* der nichtlinearer Schwingung nur langsam variieren. Diese Voraussetzung wurde in der Arbeit von [76] gemildert, in der die nichtlineare Schwingung in der Nähe einer linearer Schwingung mit zeitvarianter Dämpfung und Frequenz analysiert wurde. Die Anwendung dieser Methode auf den Fall weicher strukturvariabler Regelungen wurde in [57] gezeigt. Dabei wurde eine Lösungsmethode unter Verwendung von Potenzreihen vorgestellt, um die meistens sehr komplexe Form der angenäherten Zeitlösung numerisch effizient zu berechnen. Zwei weitere Anwendungen von solchen asymptotischen Methoden können in [15], sowie [20] gefunden werden.

Die Performance-Analyse nichtlinearer Systeme wird mit Hilfe verschiedener Performance-Maße (Gütemaße) durchgeführt, welche exakt berechnet oder angenähert werden können, oder wofür bestimmte Ober- und/oder Untergrenzen berechnet werden können. In diesem Kapitel werden mehrere Performance-Maße sowie deren Einsatz zur Performance-Analyse in nichtlinearen Regelkreisen analysiert. Neben der Konvergenzrate eines exponentiell stabilen nichtlinearen Systems, welche vielfach sowohl für die Performance-Analyse als auch für den Reglerentwurf verwendet wird, wird in dieser Arbeit - auf Basis eines neu entwickelten Zweipunktreglers mit einer parameterabhängigen Schaltfunktion - auch ein neues Performance-Maß vorgestellt, das den *Fehlklassifikationsanteil* einer zeitsuboptimalen Regelung mit Schaltfunktion mißt. Unabhängig von der Zeitlösung des Systems, quantifiziert dieses Maß den *Abstand* zwischen der Schaltfunktion eines zeitoptimalen Reglers und einer Schaltfunktion eines zeitsuboptimalen Reglers. In dem gleichen Zusammenhang wird der Quotient zwischen der *Einschwingzeit* des neuen Reglers und der *Einschwingzeit* des zeitoptimalen Reglers analysiert.

Das Kapitel ist wie folgt gegliedert. Abschnitt 7.1 stellt die wesentlichen Unterschiede zwischen linearen und nichtlinearen Regelkreisen dar. Abschnitt 7.2 stellt eine Klassifizierung von Performance-Maßen dar. Abschnitt 7.3 stellt ein neues Performance-Maß dar, das geeignet für umschaltende Regler ist und den Fehlklassifikationsanteil einer zeitsuboptimalen Regelung mit Schaltfunktion misst. Abschnitt 7.4 stellt ein weiteres Performance-Maß dar, das das Verhältnis zwischen der *Einschwingzeit* des zeitoptimalen Reglers und der des konvergenzoptimalen Reglers mißt. Schließlich beschreibt Abschnitt 7.5 die Konvergenzrate eines exponentiell stabilen nichtlinearen Systems und gibt einen theoretischen Rahmen für dessen Bestimmung an.

## 7.1 Lineare und nichtlineare Regelkreise

Wie bereits erwähnt, unterscheiden sich die linearen und die nichtlinearen Regelkreise wesentlich in ihren Eigenschaften. Das Fehlen des Superpositionsprinzips im Fall nichtlinearer Regelkreise führt beispielsweise dazu, dass der Zusammenhang zwischen Ein- und Ausgangsgröße nicht durch eine komplexe Übertragungsfunktion darstellbar ist, und somit keine frequenzbasierten Untersuchungsmethoden anwendbar sind. Darüber hinaus gelten die Ergebnisse, die für einen Unterraum des $\mathbb{R}^n$ erzeugt wurden,

nicht im gesamten Zustandsraum. Die Ergebnisse sind also nur lokal gültig.

Darüber hinaus ist die Darstellbarkeit eines nichtlinearen Systems mit Hilfe einer endlichen Anzahl an Parametern nicht mehr möglich, da auch eine genügend glatte Funktion $f(\mathbf{x})$ höchstens durch eine unendliche Reihe, z.B. in Form einer Taylorreihe in der Umgebung eines Arbeitspunktes $\mathbf{x}(0)$, darstellbar ist. Dies führt dazu, dass die Zeitlösung allein bei bestimmten nichtlinearen Systemen exakt bestimmbar ist, im Allgemeinen aber bestenfalls nur approximiert werden kann.

Sind Dauerschwingungen bei linearen Systemen theoretische Phänomene, die aufgrund unvermeidbaren Rauschens oder Parameterschwankung in einem realen System nicht vorkommen können, so sind diese - zumindest die stabilen Dauerschwingungen - in nichtlinearen Regelkreisen reale Phänomene, wobei deren Amplituden und Frequenzen unabhängig von den Anfangsauslenkungen sind. Diese bilden eine besondere Eigenschaft eines nichtlinearen Systems und werden daher oft als *Selbstschwingungen*[1)] bezeichnet.

Weitere Eigenschaften nichtlinearer Regelkreise, wie z.B. Bifurkationen, Chaos, sowie - bezogen auf die Instabilität - Unbeschränktheit des Ausgangssignals eines instabilen nichtlinearen Systems (nach endlicher Zeit), oder die mögliche Inexistenz eines mathematischen Modells für die Beschreibung nichtlinearer Dynamiken werden beispielsweise in [75] behandelt. Bei der in dieser Arbeit untersuchten Spezialklasse nichtlinearer Systeme sind jedoch solche Eigenschaften nicht vorhanden, sodass sie hier keine weitere Beachtung finden.

Bezüglich der Performance erfahren auf der anderen Seite die linearen Systeme fundamentale Grenzen, welche beispielsweise durch das *Gleichgewichtstheorem*[2)] anschaulich sind. Diese sind zwar prinzipiell mit Hilfe nichtlinearer Regler überwindbar, dies jedoch auf Kosten einer Erhöhung der Komplexität des Reglers.

Die Performance-Maße für die nichtlinearen Regelkreise sind folglich angepasst an die besonderen Eigenschaften verschiedener Systemklassen. Eine besondere Klasse nichtlinearer Regelkreise, die in dieser Arbeit untersucht wird, bilden die linearen Strecken mit Stellgrößenbegrenzung, welche durch lineare oder nichtlineare Regler geregelt sind. Allgemein lässt sich der Regelkreis wie in Abbildung 7.1 darstellen. Ist das Regelgesetz linear,

[1)] Vgl. [29].
[2)] Vgl. [47, Abschnitt 7.4.4].

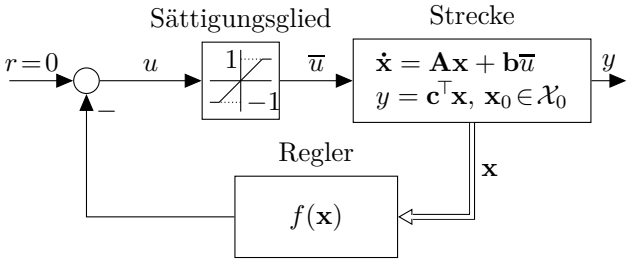


**Bild 7.1:** Nichtlinearer Regelkreis.

so bildet die Sättigungsfunktion die einzige Nichtlinearität im Regelkreis. Ein solches Modell ist in der Literatur auch als *Hammerstein*-Systemmodell bekannt.[3)] Dabei ist die Sättigungsfunktion als symmetrisch und, ohne Beschränkung der Allgemeinheit, normiert angenommen, d.h.

$$\overline{u} = \mathrm{sat}(u) = \mathrm{sign}(u)\min\{1,|u|\}.$$

Eine fehlende Symmetrie der Sättigungsfunktion ändert wesentlich die Eigenschaften des Systems. Diese Art von Systemen wird hier jedoch nicht behandelt.

Ist der Regler nichtlinear, so muss dieser nicht notwendigerweise analytisch sein. In dieser Arbeit werden Regelgesetze analysiert, welche nichtlinear in den Systemzuständen sind und außerdem unstetige Komponenten beinhalten können. Ein nichtlinearer Regler dieser Art ist beispielsweise ein *Bang-Bang*-Regler ($u(\mathbf{x}) = \mathrm{sgn}(s(\mathbf{x}))$).

Im Allgemeinen lässt sich obiger nichtlineare Regelkreis in der Form

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} - \mathbf{b}\,\mathrm{sat}(f(\mathbf{x})), \quad \mathbf{x} \in \mathbb{R}^n, f(\mathbf{0}) = 0 \tag{7.1}$$

beschreiben. Die Ruhelage des Systems ist dabei $\mathbf{x}_\mathrm{R} = \mathbf{0}$. Solche Regelungen für lineare Systeme mit Stellgrößenbegrenzung wurden vielfach in der Literatur untersucht, vgl. z.B. [35] für lineare und *Bang-Bang*-artige Regelgesetze mit **zustandslinearen** Umschaltstrategien.

Im Folgenden wird eine Klassifizierung von Performance-Maßen für nichtlineare Regelkreise mit besonderer Berücksichtigung der oben vorgestellten Systemklasse vorgenommen.

[3)] Vgl. [65, Abschnitt 1.2].

## 7.2 Klassifizierung von Performance-Maßen

Im Allgemeinen werden Performance-Maße verwendet, um Aussagen über das dynamische Verhalten des nichtlinearen Regelkreises zu machen. Sie basieren oft auf der Antwortfunktion der homogenen Zustandsgleichung[4)] im Fall stabiler Systeme, im Weiteren als *transientes Verhalten* bezeichnet, also auf der Antwort ausschließlich zu verschiedenen Anfangsauslenkungen, ohne weitere aktive externe Störungen. Diese unterscheidet sich von dem transienten Verhalten in einem linearen Regelkreis, welches das Einschwingverhalten des Systems und nicht die Antwortfunktion der homogenen Zustandsgleichung darstellt. Letztere beinhaltet bei linearen Systemen auch das stationäre Verhalten.

Das transiente Verhalten in nichtlinearen Regelkreisen, also die Antwort ausschließlich zu verschiedenen Anfangsauslenkungen, wird dabei genutzt, um verschiedene **zustandsbezogene** Performance-Maße zu formulieren. Hierfür relevante Performance-Maße sind beispielsweise die *Zeitkonstante*, d.h. die Zeit, in der die Zustandsnorm auf weniger als $35\%\,(\approx e^{-1})$ ihres Anfangswertes $\|\mathbf{x}(0)\|$ sinkt und die *Einschwingzeit* der Zustandsnorm, d.h. die Zeit, wann die Zustandsnorm auf weniger als $5\%\,(\approx e^{-3})$ ihres Anfangswertes sinkt. Im Fall exponentiell stabiler Systemen können diese Maße aus der *Konvergenzrate* des Systems, vgl. [69], berechnet werden, welche sich leichter als die Zeitlösung des Systems ermitteln lässt und geeignet für die Reglersynthese über optimierungsbasierte Ansätze ist. Diese wird im nächsten Abschnitt vorgestellt.

Auch relativ zur zeitoptimaler Regelung ist es relevant, wie schnell eine (nichtlineare) Regelungsmethode ist. Ist die zeitoptimale Schaltfunktion bekannt, so kann man das Verhältnis der beiden Einschwingzeiten untersuchen. Dies wird im Weiteren *relative Einschwingzeit* genannt. Zudem kann man den Fehlklassifikationsanteil einer zeitsuboptimalen Regelung mit Schaltfunktion als Performance-Maß verwenden. Diese Maße werden in den nächsten Abschnitten vorgestellt.

Ein weiteres zustandsbezogenes Performance-Maß ist die Größe des Einzugsgebiets der Ruhelage, welche beispielsweise durch das *Volumen des Einzugsgebiets* quantifiziert werden kann, denn je größer das Einzugsgebiet ist, desto mehr Anfangsauslenkungen können ausgeregelt werden. Das maximale Einzugsgebiet ist dabei im Fall linearer Strecken mit Stellgrößenbegrenzungen auf die asymptotisch-null-steuerbare Region beschränkt,

[4)] Der entsprechende englische Begriff lautet *zero-input response*.

die im Fall instabiler Strecken begrenzt ist.[5)] Im Fall ellipsoidaler Gebiete lässt sich, wie im vorigen Teil der Arbeit gezeigt, das Volumen des Einzugsgebiets mittels LMIs optimieren.

Im Fall schwingungsfähiger nichtlinearer Systeme kann, wie bereits erwähnt, auch eine zeitveränderliche *Frequenz* der nichtlinearen Schwingung ermittelt werden. Diese ist im Fall nichtlinearer Regelkreise, ebenso wie die *Dämpfung*, eine Funktion der Amplitude der nichtlinearen Schwingung. Mit Hilfe der sogenannten *amplitudenabhängigen Frequenz* und *Dämpfung* kann die Zeitlösung des Systems approximiert werden und damit die oben genannten Performance-Maße, sowie die *Anzahl der Schwingungsperioden* der nichtlinearen Schwingung approximiert werden.

Andere Performance-Maße können die Zustandseigenschaften eines nichtlinearen Systems quantifizieren, welche durch besondere Steuergrößen entstehen. Diese werden ***input-to-state***-bezogene Performance-Maße genannt. Beispielsweise ist in [16] die *Erreichbarkeitsmenge* von Zuständen mittels einer Steuergröße mit einer Gesamtenergie von eins, oder einer Steuergröße mit einer Obergrenze von eins von Interesse. Die *Erreichbarkeitsmenge* gruppiert die Zustände eines nichtlinearen Systems, welche von einem bestimmten Zustand aus mit Hilfe einer beliebigen Steuergröße in einer maximalen Zeitspanne von $T$ Sekunden erreicht werden können, vgl. [3]. Beispielsweise ist die Erreichbarkeitsmenge (von jedem Zustandspunkt aus) eines steuerbaren linearen Systems (ohne Stellgrößenbeschränkungen) der gesamte Zustandsraum. Im Fall steuerbarer linearer Systeme mit Stellgrößenbeschränkungen ist dies nur bei stabilen oder semi-stabilen Systemen der Fall, vgl. [35, Kapitel 2]. Auch bei nichtlinearen Systemen ist dies nicht immer der Fall, wie das Beispiel aus [3, S. 163] zeigt.

Auch die Eigenschaften der Ausgangsgröße eines nichtlinearen Systems, welche durch die Zustände determiniert werden, können durch Performance-Maße quantifiziert werden. Diese werden ***state-to-output***-bezogene Performance-Maße genannt. Beispielsweise gehört die *Ausgangsenergie* eines (nichtlinearen) Systems dazu, welche für Reglersynthese mit optimierungsbasierten Ansätzen geeignet sind, vgl. z.B. [30]. Auch die *Überschwingweite*, d.h. die maximale Amplitude der Ausgangsgröße, kann zu solchen Performance-Maßen gezählt werden.

Schließlich können die Eigenschaften der Ausgangsgröße eines nichtlinearen Systems, welche durch die Steuergrößen determiniert werden, mit Hilfe von sogenannten ***input-to-output***-bezogenen Performance-Maßen

[5)] Vgl. [35, Proposition 2.2.1].

quantifiziert werden. Beispielsweise quantifiziert die $L_2$-*Verstärkung* die kleinste obere Schranke des Verhältnises zwischen der Ausgangs- und Eingangsenergie. Dieses ist definiert als

$$L_2 := \sup_{\|\mathbf{u}\|_{\mathrm{L2}} \neq 0} \frac{\|\mathbf{y}\|_{\mathrm{L2}}}{\|\mathbf{u}\|_{\mathrm{L2}}},$$

mit

$$\|\mathbf{w}\|_{\mathrm{L2}}^2 := \int_0^\infty \mathbf{w}^\top \mathbf{w} \mathrm{d}t.$$

Eine Anwendung dieses Maßes kann z.B. in [53] gefunden werden. Ein ähnliches Maß ist die *RMS-Verstärkung*[6], welche als

$$\mathrm{RMS} := \sup_{\mathrm{RMS}(\mathbf{u}) \neq 0} \frac{\mathrm{RMS}(\mathbf{y})}{\mathrm{RMS}(\mathbf{u})},$$

mit[7]

$$\mathrm{RMS}(\mathbf{w})^2 := \limsup_{T \to \infty} \int_0^T \mathbf{w}^\top \mathbf{w} \mathrm{d}t,$$

definiert ist.

Die in diesem Abschnitt beschriebene Klassifizierung von Performance-Maßen stammt aus [16] und ist nicht vollständig. Beispielsweise werden Kombinationen der o.g. Maße, wie das *Lagrange'sche*, das *Mayer'sche*, sowie das *Bolza'sche Gütemaß* in der Arbeit von [30] beschrieben und analysiert. Das *Bolza'sche Gütemaß* bezeichnet dabei die Addition der beiden anderen Gütemaße und hat die Form

$$J = h(\mathbf{x}(T),T) + \int_0^T f_0(\mathbf{x}(t),\mathbf{u}(t),t)\mathrm{d}t,$$

wobei $\mathbf{x}(t)$ den Zustandsvektor, $\mathbf{u}(t)$ den Steuervektor, $T$ eine bestimmte Zeit, $h(\,\cdot\,)$ und $f_0(\,\cdot\,)$ vorgegebene Funktionen bezeichnen. Im Weiteren beschäftigt sich die Arbeit nur mit *zustandsbezogenen* Performance-Maßen, der Konvergenzrate, dem Fehlklassifikationsanteil der zeitsuboptimalen Schaltfunktion, sowie der relativen Einschwingzeit. Die im letzten Kapitel dieser Arbeit beschriebenen Methoden zur Performance-Analyse mittels *Computerexperimenten* lassen sich jedoch auf beliebige Performance-Maße übertragen.

[6] RMS ist eine Abkürzung für *Root Mean Square*.

[7] Vgl. Def. 16 (Anhang) für die Definition der Limes superior einer Funktion.

## 7.3 Der Fehlklassifikationsanteil einer zeitsuboptimalen Regelung mit Schaltfunktion

Im Abschnitt 5 wurde durch die Maximierung einer $V$-induzierten Matrixnorm ein *Bang-Bang*-Regler erzielt, der im Fall linearer Systeme mit Stellgrößenbeschränkung die gleiche Struktur wie der zeitoptimale Regler besitzt, jedoch eine parameterabhängige Schaltfunktion aufweist, wobei der Parameter zustandsabhängig ist. Um die Performance dieses zeitsuboptimalen Reglers zu quantifizieren, können die beiden Schaltfunktionen verglichen werden. Dies wird im Folgenden gezeigt.

Das Ziel der zeitoptimalen Regelung ist es, den Zustandsvektor $\mathbf{x}(t)$ aus einem beliebigen Punkt $\mathbf{x}_0$ der *null-steuerbaren* Region in kürzester Zeit in die Ruhelage $\mathbf{x}_\mathrm{R} = \mathbf{0}$ des Systems zu bewegen. Mathematisch formulieren lässt sich dies durch

$$\min_{u\in\mathbb{R}} J = \int_0^T 1\mathrm{d}t = T,$$

unter den Nebenbedingungen

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{A}\mathbf{x} + \mathbf{b}u, \\ \mathbf{x}(0) = \mathbf{x}_0, &\quad \mathbf{x}(T) = 0, \\ |u| &\leq 1. \end{aligned}$$

Dieses Optimierungsproblem mit Nebenbedingungen kann durch den Lagrange-Multiplikatorenansatz in ein Optimierungsproblem ohne Nebenbedingungen transformiert werden. Dabei ist der Multiplikator (im Folgenden mit $\boldsymbol{\psi}(t)$ bezeichnet) aufgrund der Differentialgleichung aus der ersten Nebenbedingung eine zeitabhängige Vektorfunktion. Diese Transformation und die anschließende Lösung des Problems lassen sich durch Einführen der *Hamilton-Funktion* leichter darstellen, vgl. [30]. In diesem Fall hat Letztere die Form

$$\begin{aligned} \mathrm{H}(\mathbf{x}, \boldsymbol{\psi}, u) :&= -1 + \boldsymbol{\psi}^\top(\mathbf{A}\mathbf{x} + \mathbf{b}u) \\ &= -1 + \boldsymbol{\psi}^\top\mathbf{A}\mathbf{x} + \boldsymbol{\psi}^\top\mathbf{b}u, \end{aligned} \tag{7.2}$$

wobei die Lösung des obigen Optimierungsproblems die Bedingungen

$$\dot{\mathbf{x}} = \frac{\partial H(\mathbf{x}, \boldsymbol{\psi}, u)}{\partial \boldsymbol{\psi}} = \mathbf{A}\mathbf{x} + \mathbf{b}u, \tag{7.3}$$

$$\dot{\boldsymbol{\psi}} = -\frac{\partial H(\mathbf{x}, \boldsymbol{\psi}, u)}{\partial \mathbf{x}} = -\mathbf{A}^\top \boldsymbol{\psi}, \tag{7.4}$$

$$u_{\text{opt}} = \arg\max_u \mathrm{H}(\mathbf{x}, \boldsymbol{\psi}, u) \tag{7.5}$$

erfüllen muss. Bedingung (7.5) stellt den wesentlichen Aspekt des *Maximumprinzips von Pontrjagin*[8] dar, das aufgrund der Beschränkung der Stellgröße $|u| \leq 1$ zur Anwendung kommt. Aus der Bedingung (7.5) folgt, dass die Lösung durch

$$u_{\text{opt}}(t) = \operatorname{sgn}(\boldsymbol{\psi}_{\text{opt}}^\top(t)\mathbf{b}) \tag{7.6}$$

gegeben sein muss. Die Vektorfunktion $\boldsymbol{\psi}_{\text{opt}}(t)$ folgt dabei aus Bedingung (7.4), d.h., es gilt

$$\boldsymbol{\psi}_{\text{opt}}(t) = e^{-\mathbf{A}^\top t}\boldsymbol{\psi}_{\text{opt}}(0), \tag{7.7}$$

wobei der unbekannte Parametervektor $\boldsymbol{\psi}_{\text{opt}}(0)$ unter Nutzung der Endbedingung $\mathbf{x}(T) = 0$ mit Hilfe der Zeitlösung des Gesamtsystems berechnet werden muss. Es ergibt sich dabei mit der Lösung von Gl. (7.3) die Vektorgleichung

$$\int_0^T e^{-\mathbf{A}\tau}\mathbf{b} \operatorname{sgn}\left(\mathbf{b}^\top e^{-\mathbf{A}^\top\tau}\boldsymbol{\psi}_{\text{opt}}(0)\right) \mathrm{d}\tau = -\mathbf{x}_0. \tag{7.8}$$

Diese ist besonders im Fall von Strecken höherer Ordnung nicht mehr analytisch lösbar, sodass die zeitoptimale Schaltfunktion aus Gl. (7.6) nicht mehr in einer geschlossenen Form angegeben werden kann.

Die regelgesetzabhängige Komponente der Hamilton-Funktion ist im Optimum gegeben durch

$$\mathrm{H}(\mathbf{x}, \boldsymbol{\psi}_{\text{opt}}, u_{\text{opt}}) = \boldsymbol{\psi}_{\text{opt}}^\top(t)\mathbf{b}u_{\text{opt}} = |\boldsymbol{\psi}_{\text{opt}}^\top(t)\mathbf{b}|. \tag{7.9}$$

Sei ein zeitsuboptimales Regelgesetz gegeben durch

$$u_{\text{subopt}}(t) := \operatorname{sgn}(\boldsymbol{\psi}_{\text{subopt}}^\top(t)\mathbf{b}).$$

[8] Vgl. [30].

Beide Regelgesetze unterscheiden sich also durch die Vektorfunktion $\boldsymbol{\psi}(t)$. Dies hat zur Folge, dass durch den zeitsuboptimalen Regler die Endbedingung $\mathbf{x}(T) = \mathbf{0}$ nicht mehr erfüllt ist. Nehmen wir an, dass im **suboptimalen** Fall stattdessen gilt

$$\mathbf{x}(T) = e^{\mathbf{A}T}\boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \neq \mathbf{0}.$$

Gl. (7.8) wird in diesem Fall

$$\int_0^T e^{-\mathbf{A}\tau}\mathbf{b}\,\mathrm{sgn}\left(\mathbf{b}^\top e^{-\mathbf{A}^\top\tau}\boldsymbol{\psi}_{\mathrm{subopt}}(0)\right)\mathrm{d}\tau = \boldsymbol{\epsilon} - \mathbf{x}_0. \tag{7.10}$$

Durch Subtrahieren der beiden Gleichungen (7.8) und (7.10) ergibt sich

$$\begin{aligned}\boldsymbol{\epsilon} &= \int_0^T e^{-\mathbf{A}\tau}\mathbf{b}\left[\mathrm{sgn}\left(\mathbf{b}^\top e^{-\mathbf{A}^\top\tau}\boldsymbol{\psi}_{\mathrm{subopt}}(0)\right) - \mathrm{sgn}\left(\mathbf{b}^\top e^{-\mathbf{A}^\top\tau}\boldsymbol{\psi}_{\mathrm{opt}}(0)\right)\right]\mathrm{d}\tau \\ &= \int_0^T e^{-\mathbf{A}\tau}\mathbf{b}\left[\mathrm{sgn}\left(\mathbf{b}^\top\boldsymbol{\psi}_{\mathrm{subopt}}(\tau)\right) - \mathrm{sgn}\left(\mathbf{b}^\top\boldsymbol{\psi}_{\mathrm{opt}}(\tau)\right)\right]\mathrm{d}\tau\end{aligned}$$

und somit[9)]

$$\|\boldsymbol{\epsilon}\| \leq \int_0^T \|e^{-\mathbf{A}\tau}\mathbf{b}\| \cdot \left|\mathrm{sgn}\left(\mathbf{b}^\top\boldsymbol{\psi}_{\mathrm{subopt}}(\tau)\right) - \mathrm{sgn}\left(\mathbf{b}^\top\boldsymbol{\psi}_{\mathrm{opt}}(\tau)\right)\right|\mathrm{d}\tau. \tag{7.11}$$

Abgesehen von der *Gewichtung* $\|e^{-\mathbf{A}\tau}\mathbf{b}\|$, je kleiner die Zeitspanne ist, wo die Regelgesetze unterschiedliche Vorzeichen aufweisen, desto kleiner ist $\|\boldsymbol{\epsilon}\|$. Für ein gegebenes Einzugsgebiet $\mathcal{G}$ der Ruhelage wäre es dann denkbar, das Maß

$$J_{\mathrm{H}} := \frac{1}{2}\int\cdots\int_{\mathcal{G}} |\,\mathrm{sgn}(\boldsymbol{\psi}_{\mathrm{opt}}^\top(\mathbf{x})\mathbf{b}) - \mathrm{sgn}(\boldsymbol{\psi}_{\mathrm{subopt}}^\top(\mathbf{x})\mathbf{b})|\mathrm{d}\mathbf{x} \tag{7.12}$$

zu verwenden, welches die Fläche zwischen den beiden Schaltfunktionen berechnet, wo die Regelgesetze unterschiedliche Vorzeichen haben. Um den jeweiligen Wert im Verhältnis zum gesamten Einzugsgebiet der Ruhelage

[9)] Dies kann man wie folgt erklären: Sei $\mathbf{s} := \int_a^b \mathbf{f}(t)\mathrm{d}t \neq \mathbf{0}$, mit $\mathbf{f} : [a,b] \to \mathbb{R}^n$. Es gilt $\|\mathbf{s}\|^2 = \mathbf{s}^\top\mathbf{s} = \mathbf{s}^\top\int_a^b \mathbf{f}(t)\mathrm{d}t = \int_a^b \mathbf{s}^\top\mathbf{f}(t)\mathrm{d}t \leq \int_a^b \|\mathbf{s}\|\|\mathbf{f}(t)\|\mathrm{d}t = \|\mathbf{s}\|\int_a^b \|\mathbf{f}(t)\|\mathrm{d}t$. Durch Teilen mit $\|\mathbf{s}\| \neq 0$ folgt $\|\int_a^b \mathbf{f}(t)\mathrm{d}t\| \leq \int_a^b \|\mathbf{f}(t)\|\mathrm{d}t$.

zu setzen, wird das Performance-Maß zum Volumen des Einzugsgebiets normiert, d.h.

$$J_{\mathrm{H}}^{\mathrm{n}} := \frac{J_{\mathrm{H}}}{\mathrm{Vol}(\mathcal{G})}. \tag{7.13}$$

Dieses Maß wird normierter *Fehlklassifikationsanteil einer zeitsuboptimalen Regelung mit Schaltfunktion*[10] genannt.

## 7.4 Relative *Einschwingzeit*

Als alternativer Vergleich der Regelgüte zwischen einer zeitsuboptimalen und einer zeitoptimalen Regelung wird die relative *Einschwingzeit* verwendet. Diese ist definiert als das Verhältnis zwischen der mittleren *Einschwingzeit* einer zeitsuboptimalen Regelung ($u_{\mathrm{subopt}}$) und der mittleren *Einschwingzeit* einer zeitoptimalen Regelung ($u_{\mathrm{opt}}$), d.h.

$$J_{\mathrm{t_a}} := \frac{\frac{1}{m}\sum_{k=1}^{m} t_{\mathrm{a}_{5\%}}^{u_{\mathrm{subopt}}}(\mathbf{x}_k(0))}{\frac{1}{m}\sum_{k=1}^{m} t_{\mathrm{a}_{5\%}}^{u_{\mathrm{opt}}}(\mathbf{x}_k(0))}, \quad \mathbf{x}_k(0) \in \partial\mathcal{G}, k = 1,\ldots,m, \tag{7.14}$$

wobei $\mathbf{x}_k(0) \in \partial\mathcal{G}$, mit $k = 1,\ldots,m$, äquidistante Anfangsauslenkungen auf dem Rand des Einzugsgebietes $\mathcal{G}$ sind, und die *Einschwingzeit* als

$$t_{\mathrm{a}_{5\%}}^{u}(\mathbf{x}_k(0)) := \max_{\substack{t>0 \\ \|\mathbf{x}_k(t)\| \geq 0.05\|\mathbf{x}_k(0)\|}} t \tag{7.15}$$

definiert ist. Dabei stellt $\mathbf{x}_k(t)$ die Trajekorie des Systems $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b}u$ dar.

Offensichtlich können die letzten zwei Performance-Maße nicht analytisch berechnet werden. Diese werden im nächsten Kapitel für ein Streckenensemble empirisch geschätzt.

## 7.5 Konvergenzrate

Im Fall exponentiell stabiler Systeme kann das Konvergenzverhalten der Zeitlösung durch die Konvergenzrate (od. *Rate der exponentiellen Konver-*

[10] Der Begriff Fehlklassifikationsanteil (engl. *misclassification ratio*) stammt aus dem Bereich des maschinellen Lernens. Darin bezeichnet dieses Maß den Anteil der falsch klassifizierten (Test-)Daten an den gesamten (Test-)Daten.

*genz*) des Systems analysiert werden, ohne dabei die exakte Zeitlösung zu kennen. Die Konvergenzrate ist wie folgt definiert:

**Definition 1** [Konvergenzrate eines exponentiell stabilen Systems] *Gegeben sei das Differentialgleichungssystem* $\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t))$*, mit* $\mathbf{x}(t) \in \mathbb{R}^n$ *und* $f(\mathbf{0}) = \mathbf{0}$*, wobei die Ruhelage* $\mathbf{x}_\mathrm{R} = \mathbf{0}$ *exponentiell stabil ist. Der maximale Abklingfaktor* $\alpha > 0$*, wofür eine positive Zahl* $\gamma > 0$ *existiert, sodass*

$$\|\mathbf{x}(t)\| \leq \gamma \|\mathbf{x}(0)\| e^{-\alpha t}, \forall t > 0 \text{ und } \forall\, \mathbf{x}(0) \in \mathcal{B}_\epsilon(\mathbf{x}_\mathrm{R})$$

*gilt, heißt* Konvergenzrate des Systems.

Dieses Maß stellt eine obere Grenze der Zustandsnorm zu jedem Zeitpunkt $t > 0$ dar. Diese Grenze kann verwendet werden, um den jeweiligen Wert der Zustandsnorm zum Zeitpunkt $t$ im Verhältnis zum Anfangswert $\|\mathbf{x}(0)\|$ zu setzen. Dies ergibt sich aus der äquivalenten Darstellung

$$\|\mathbf{x}(t)\| \leq \|\mathbf{x}(0)\| e^{\ln\gamma - \alpha t}.$$

Beispielsweise kann die *Zeitkonstante* approximiert werden, d.h. die Zeit, in der die Zustandsnorm auf weniger als $35\%\,(\approx e^{-1})$ ihres Anfangswertes $\|\mathbf{x}(0)\|$ sinkt. Auch die *Einschwingzeit* der Zustandsnorm kann damit approximiert werden, d.h. die Zeit, wann die Zustandsnorm auf weniger als $5\%\,(\approx e^{-3})$ ihres Anfangswertes sinkt. Daraus folgt eine Approximation der *Einschwingzeit* durch

$$t_{\mathrm{a}_{5\%}} \approx \frac{(\ln\gamma + 3)}{\alpha}. \tag{7.16}$$

Es ist ersichtlich, dass je größer die Konvergenzrate $\alpha$ ist, desto kleiner ist die *Einschwingzeit*. Im Fall linearer Systeme entspricht die Konvergenzrate betragsmäßig dem Realteil des Eigenwertes/Eigenwertpaares der Systemmatrix, der/das am nächsten zur imaginären Achse liegt. Dies ist aus der Zeitlösung eines linearen Systems ersichtlich. Im Fall nichtlinearer Systeme ist die Konvergenzrate im Allgemeinen nicht mehr exakt bestimmbar.

### 7.5.1 Formulierung mittels Matrixnormen

In der Literatur existieren eine Reihe von Methoden, welche eine untere Grenze der Konvergenzrate liefern. Sie basieren im Allgemeinen auf Matrixnormen. Diese, sowie ein dazu passender theoretischer Rahmen werden im Folgenden erläutert. Dazu werden zwei Matrixnormen vorgestellt,

die *logarithmische Matrixnorm* und die *V-induzierte logarithmische Matrixnorm*.

**Definition 2** [Logarithmische Matrixnorm, vgl. [73, Section 2.2.2]] *Gegeben sei eine quadratische Matrix* $\mathbf{A} \in \mathbb{R}^{n\times n}$. *Der Skalar* $\mu(\mathbf{A}) \in \mathbb{R}$, *definiert als*

$$\mu(\mathbf{A}) := \lim_{h\to 0^+} \frac{\|\mathbf{I}_n + h\mathbf{A}\|_i - 1}{h} \tag{7.17}$$

*wird* logarithmische Matrixnorm *genannt. Dabei bezeichnet* $\|\mathbf{A}\|_i$ *eine induzierte Matrixnorm, vgl. Def. 24 (Anhang).*

*Bemerkung* 7.1 (Abgrenzung zur induzierten Matrixnorm). Obwohl hier der Begriff *logarithmische Matrixnorm* verwendet wird, handelt es sich bei $\mu(\mathbf{A})$ nicht um eine induzierte Matrixnorm im Sinne der Def. 24 (Anhang). Ein wesentlicher Unterschied bildet die Tatsache, dass die logarithmische Matrixnorm auch negative Werte annehmen kann, wobei die induzierte Matrixnorm im eigentlichen Sinn nur positive Werte annimmt. Der Begriff *logarithmische Matrixnorm* wurde zum ersten Mal in [23] verwendet. △

*Bemerkung* 7.2 (Anschauliche Darstellung einer logarithmischen Matrixnorm, vgl. [72]). Für eine kontinuierliche Funktion $F : \mathbb{R}^n \to \mathbb{R}$ heißt der Skalar

$$\partial F(\mathbf{x};\mathbf{y}) := \lim_{h\to 0^+} \frac{F(\mathbf{x}+h\mathbf{y}) - F(\mathbf{x})}{h} \tag{7.18}$$

falls er existiert, *Richtungsableitung von* $F$ *an der Stelle* $\mathbf{x}$ *in Richtung* $\mathbf{y}$. Es ist ersichtlich, dass die logarithmische Matrixnorm aus Gl. (7.17) die Richtungsableitung der induzierten Matrixnorm $\|\cdot\|_i$ an der Stelle $\mathbf{I}_n$ in die Richtung der Matrix $\mathbf{A}$ ist. △

*Bemerkung* 7.3 (Ausgewählte Eigenschaften der logarithmischen Matrixnorm, siehe [24]).

(*i*) Die Grenze $\mu(\mathbf{A})$ existiert für jede Matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$.

(*ii*) $\mu(\mathbf{I}_n) = 1$, $\mu(-\mathbf{I}_n) = -1$,

(*iii*) $-\|\mathbf{A}\|_i \le -\mu(-\mathbf{A}) \le \operatorname{Re}\lambda(\mathbf{A}) \le \mu(\mathbf{A}) \le \|\mathbf{A}\|_i$,

(*iv*) $\mu(\alpha\mathbf{A}) = \alpha\mu(\mathbf{A}), \forall\, \alpha \ge 0$,

(*v*) $\mu[\nu\mathbf{A} + (1-\nu)\mathbf{B}] \le \nu\mu(\mathbf{A}) + (1-\nu)\mu(\mathbf{B}), \forall\, \nu \in [0,1], \mathbf{A},\mathbf{B} \in \mathbb{R}^{n\times n}$.

△

**Tabelle 7.1:** Ausgewählte Vektornormen, induzierte Matrixnormen und logarithmische Matrixnormen.

| | | |
|---|---|---|
| $\|\mathbf{x}\|_\infty := \max_i \|x_{(i)}\|$ | $\|\mathbf{A}\|_{i_\infty} := \max_i \sum_{j=1}^{n} \|a_{(ij)}\|$ | $\mu_\infty(\mathbf{A}) = \max_i [a_{(ij)} + \sum_{j\neq i} \|a_{(ij)}\|]$ |
| $\|\mathbf{x}\|_1 := \sum_{i=1}^{n} \|x_{(i)}\|$ | $\|\mathbf{A}\|_{i_1} := \max_j \sum_{i=1}^{n} \|a_{(ij)}\|$ | $\mu_1(\mathbf{A}) = \max_j [a_{(jj)} + \sum_{i\neq j} \|a_{(ij)}\|]$ |
| $\|\mathbf{x}\|_2 := \sqrt{\sum_{i=1}^{n} \|x_{(i)}\|^2}$ | $\|\mathbf{A}\|_{i_2} := \sqrt{\lambda_{\max}(\mathbf{A}^\top \mathbf{A})}$ | $\mu_2(\mathbf{A}) = \lambda_{\max}(\mathbf{A}^\top + \mathbf{A})/2$ |

Tabelle 7.1 gibt mehrere Normen und die entsprechenden logarithmischen Matrixnormen wieder. Für LTI und LTV-Systeme kann eine untere und obere Grenze der Zustandsnorm mit Hilfe logarithmischer Matrixnormen angegeben werden. Folgendes Lemma verdeutlicht dies.

**Lemma 7.1 [Grenzen der Zustandsnorm basierend auf logaritmischen Matrixnormen, vgl. [73, Section 2.5]].** *Für das LTV-System* $\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t)$*, mit* $t \geq 0$*,* $\mathbf{x} \in \mathbb{R}^n$ *und der stetigen matrixwertigen Funktion* $\mathbf{A}(t) : [0,\infty) \to \mathbb{R}^{n\times n}$*, vgl. Def. 25 (Anhang), gilt für jedes* $t \geq t_0 \geq 0$*,*

$$\begin{aligned} \|\mathbf{x}(t_0)\| \exp\left\{ \int_{t_0}^{t} -\mu(-\mathbf{A}(\tau))\,\mathrm{d}\,\tau \right\} \\ \leq \|\mathbf{x}(t)\| \\ \leq \|\mathbf{x}(t_0)\| \exp\left\{ \int_{t_0}^{t} \mu(\mathbf{A}(\tau))\,\mathrm{d}\,\tau \right\}. \end{aligned}$$

*Bemerkung* 7.4. Mit Hilfe der oberen Grenze aus Lemma 7.1 kann die Konvergenzrate jedoch nur in Spezialfällen, z.B. für stabile LTI-Systeme mit *normalen* Systemmatrizen, d.h. mit Systemmatrizen bei denen $\mathbf{A}^\top\mathbf{A} = \mathbf{A}\mathbf{A}^\top$ gilt, bestimmt werden. △

Diese Einschränkung gilt, weil für allgemeine stabile LTI-Systeme die logarithmische Matrixnorm sowohl positive als auch negative Werte annehmen kann. Jedoch sind im Fall von *normalen* Matrizen $\mathbf{A}$ mit negativen Eigenwerten die Eigenwerte der Matrix $\mathbf{A} + \mathbf{A}^\top$ ebenfalls negativ. Dies ist

ersichtlich aus der Zeitlösung eines LTI-Systems $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$,

$$\mathbf{x}(t) = e^{\mathbf{A}t}\mathbf{x}(0),$$

und aus der Identität

$$\|\mathbf{x}(t)\|^2 = \|e^{\mathbf{A}t}\mathbf{x}(0)\|^2 = \mathbf{x}(0)^\top e^{\mathbf{A}^\top t} e^{\mathbf{A}t}\mathbf{x}(0).$$

Dabei gilt

$$\mathbf{x}(0)^\top e^{\mathbf{A}^\top t} e^{\mathbf{A}t}\mathbf{x}(0) = \mathbf{x}(0)^\top e^{(\mathbf{A}+\mathbf{A}^\top)t}\mathbf{x}(0),$$

dann und nur dann, wenn die Matrix $\mathbf{A}$ *normal* ist.[11)] Somit ergibt sich in diesem Fall

$$\|e^{\mathbf{A}t}\mathbf{x}(0)\|^2 = \mathbf{x}(0)^\top e^{(\mathbf{A}+\mathbf{A}^\top)t}\mathbf{x}(0),$$

wobei beide Seiten der oberen Identität dann und nur dann für alle $\mathbf{x}(0) \in \mathbb{R}^n$ gegen null konvergieren, wenn die Matrizen $\mathbf{A}$ bzw. $\mathbf{A} + \mathbf{A}^\top$ nur Eigenwerte mit negativen Realteilen besitzen. Ist im Fall einer nicht *normalen* Systemmatrix der maximale Eigenwert der Matrix $\mathbf{A}^\top + \mathbf{A}$ positiv, so ist die obere Grenze aus Lemma 7.1 unbrauchbar für die Berechnung der Konvergenzrate. Folgendes Beispiel verdeutlicht dies.

**Beispiel 7.1.** *Für das LTI-System* $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$ *mit der (nicht* normalen*) Systemmatrix*

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}$$

*und Eigenwerten* $\lambda_1 = -1$ *und* $\lambda_2 = -2$*, entspricht die Konvergenzrate* $\alpha = 1$ *betragsmäßig dem Eigenwert, der am nächsten zur imaginären Achse liegt. Die obere Grenze aus Lemma 7.1 ergibt aber*

$$\|\mathbf{x}\| \leq \|\mathbf{x}_0\| \exp\{\mu_2(\mathbf{A})(t - t_0)\},$$

*wobei für die logarithmische Matrixnorm* $\mu_2(\mathbf{A}) = 0.5\lambda_{\max}(\mathbf{A} + \mathbf{A}^\top) = 0.0811 > 0$ *gilt. Folglich ist diese obere Grenze für die Bestimmung der Konvergenzrate in diesem Fall unbrauchbar.*

[11)] Vgl. [8, Fakt 11.1.5].

Eine modifizierte Form der logarithmischen Matrixnorm, welche für jedes stabile System einen negativen Wert einnimmt, kann für jede konvexe und positiv definite Funktion $V(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}_+$ wie folgt definiert werden.

**Definition 3** [$V$-induzierte logarithmische Matrixnorm, vgl. [72]] *Gegeben sei die stetige, konvexe und positiv definite Funktion $V(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}_+$. Die* $V$-induzierte logarithmische Matrixnorm *ist definiert als*[a)]

$$\mu_V(\mathbf{A}) := \sup_{\mathbf{x}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \left[\frac{\partial V(\mathbf{x};\mathbf{A}\mathbf{x})}{V(\mathbf{x})}\right]. \tag{7.19}$$

[a)] Wie in Gl. (7.18) definiert, bezeichnet $\partial V(\mathbf{x};\mathbf{A}\mathbf{x})$ die Richtungsableitung der Funktion $V$ an der Stelle $\mathbf{x}$ in Richtung $\mathbf{A}\mathbf{x}$.

*Bemerkung* 7.5 (Äquivalenz zur logarithmischen Matrixnorm). Die $V$-induzierte logarithmische Matrixnorm ist äquivalent zur logarithmischen Matrixnorm falls die Funktion $V$ eine Vektornorm darstellt, vgl. [72, Lemma 1] für den Beweis. △

*Bemerkung* 7.6 (Eigenschaften der $V$-induzierten logarithmischen Matrixnorm, vgl. [72]).

$$\begin{aligned}
&(i) \quad \mu_V(\alpha\mathbf{A}) = \alpha\mu_V(\mathbf{A}), \quad \forall\,\alpha \geq 0,\\
&(ii) \quad \mu_V(\mathbf{A}+\mathbf{B}) \leq \mu_V(\mathbf{A}) + \mu_V(\mathbf{B}), \quad \forall\,\mathbf{A},\mathbf{B} \in \mathbb{R}^{n\times n},\\
&(iii) \quad \mu_V(\nu\mathbf{A} + (1-\nu)\mathbf{B}) \leq \nu\mu_V(\mathbf{A}) + (1-\nu)\mu_V(\mathbf{B}), \quad \forall\nu \in [0,1],\\
&\qquad\quad \forall\,\mathbf{A},\mathbf{B} \in \mathbb{R}^{n\times n}.
\end{aligned}$$

△

Für LTI-Systeme ist die Existenz einer $V$-induzierten logarithmischen Matrixnorm mit negativem Wert durch die Stabilität des Systems gewährleistet. Die Existenzbedingung ist sowohl notwendig als auch hinreichend. Folgendes Lemma verdeutlicht dies.

**Lemma 7.2 [Vgl. [72, Theorem 5]].** *Gegeben sei die Matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$. Dann existiert eine konvexe, positiv definite Funktion $V$, sodass $\mu_V(\mathbf{A}) < 0$ dann und nur dann wenn $\mathbf{A}$ Hurwitz ist.*

*Beweis.* Falls die Matrix $\mathbf{A}$ Hurwitz ist, dann existiert eine positiv definite und konvexe Ljapunov-Funktion der Form $V(\mathbf{x}) := \mathbf{x}^\top\mathbf{P}\mathbf{x}$, wobei die

Matrix $\mathbf{P} \succ \mathbf{0}$ die Ljapunov-Gleichung

$$\mathbf{A}^\top \mathbf{P} + \mathbf{P}\mathbf{A} = -\mathbf{Q} \tag{7.20}$$

mit $\mathbf{Q} \succ \mathbf{0}$ löst, siehe dazu Lemma A.4 (Anhang). Umgekehrt, falls es eine konvexe und positiv definite Funktion $V(\mathbf{x})$ existiert, sodass $\mu_V(\mathbf{A}) < 0$, dann folgt aus Gl. (7.19), dass

$$\mu_V(\mathbf{A}) := \sup_{\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} \left[ \frac{\partial V(\mathbf{x}; \mathbf{A}\mathbf{x})}{V(\mathbf{x})} \right] < 0,$$

was äquivalent ist zu

$$\partial V(\mathbf{x}; \mathbf{A}\mathbf{x}) < 0, \forall\, \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}.$$

Daraus folgt, dass die Funktion $V$ eine gültige Ljapunov-Funktion des Systems $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$ ist und somit, dass die Matrix $\mathbf{A}$ Hurwitz ist. □

Mit Hilfe der modifizierten logarithmischen Matrixnorm kann eine untere und obere Grenze der Funktion $V(\mathbf{x})$ angegeben werden. Folgendes Lemma verdeutlicht dies.

**Lemma 7.3 [Vgl. [72, Theorem 1, (iv)]].** *Für das LTV-System* $\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t)$*, mit* $t \geq 0$*,* $\mathbf{x} \in \mathbb{R}^n$ *und der stetigen matrixwertigen Funktion* $\mathbf{A}(t) : [0,\infty) \to \mathbb{R}^{n \times n}$*, vgl. Def. 25 (Anhang), gilt für jedes* $t \geq 0$

$$\begin{aligned} V(\mathbf{x}(0)) \exp\left\{ \int_0^t -\mu_V(-\mathbf{A}(\tau))\, d\tau \right\} \\ \leq V(\mathbf{x}(t)) \\ \leq V(\mathbf{x}(0)) \exp\left\{ \int_0^t \mu_V(\mathbf{A}(\tau))\, d\tau \right\}. \end{aligned}$$

Für quadratische Funktionen der Form $V(\mathbf{x}) = \mathbf{x}^\top \mathbf{P}\mathbf{x}$, mit $\mathbf{P} \succ \mathbf{0}$, kann eine obere und untere Grenze der Zustandsnorm angegeben werden. Folgendes Lemma verdeutlicht dies.

**Lemma 7.4.** *Für das LTV-System $\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t)$, mit $t \geq 0$, $\mathbf{x} \in \mathbb{R}^n$ und der stetigen matrixwertigen Funktion $\mathbf{A}(t) : [0,\infty) \to \mathbb{R}^{n\times n}$, gilt für jedes $t \geq 0$*

$$\sqrt{\frac{1}{\kappa(\mathbf{P})}}\|\mathbf{x}(0)\| \exp\left\{\frac{1}{2}\int_0^t -\mu_V(-\mathbf{A}(\tau))d\tau\right\} \leq \|\mathbf{x}(t)\| \leq \sqrt{\kappa(\mathbf{P})}\|\mathbf{x}(0)\| \exp\left\{\frac{1}{2}\int_0^t \mu_V(\mathbf{A}(\tau))d\tau\right\},$$

*wobei $\kappa(\mathbf{P}) := \lambda_{\max}(\mathbf{P})/\lambda_{\min}(\mathbf{P})$ die Konditionszahl der Matrix $\mathbf{P}$ ist.*

*Beweis.* Für $\mathbf{P} \succ \mathbf{0}$ gilt

$$\lambda_{\min}(\mathbf{P})\|\mathbf{x}(t)\|^2 = \mathbf{x}(t)^\top \lambda_{\min}(\mathbf{P})\mathbf{I}\mathbf{x}(t) \leq \mathbf{x}(t)^\top \mathbf{P}\mathbf{x}(t),\ \forall\, t \geq 0, \mathbf{x}(t) \neq \mathbf{0}, \tag{7.21}$$

folgt aus Lemma 7.3 für $V(\mathbf{x}) = \mathbf{x}^\top\mathbf{P}\mathbf{x} > 0$, $\forall\, \mathbf{x} \in \mathbb{R}^n\backslash\{\mathbf{0}\}$, dass

$$\lambda_{\min}(\mathbf{P})\|\mathbf{x}(t)\|^2 \leq \mathbf{x}(t)^\top\mathbf{P}\mathbf{x}(t) \leq V(\mathbf{x}(0))e^{\mu_V(\mathbf{A})t}.$$

Mit

$$\mathbf{x}(0)^\top\mathbf{P}\mathbf{x}(0) \leq \mathbf{x}(0)^\top\lambda_{\max}(\mathbf{P})\mathbf{I}\mathbf{x}(0) = \lambda_{\max}(\mathbf{P})\|\mathbf{x}(0)\|^2, \quad \mathbf{x}(0) \neq \mathbf{0},$$

folgt, dass

$$\|\mathbf{x}(t)\| \leq \sqrt{\frac{\lambda_{\max}(\mathbf{P})}{\lambda_{\min}(\mathbf{P})}}\|\mathbf{x}(0)\|e^{\frac{1}{2}\mu_V(\mathbf{A})t},\ \forall \mathbf{x} \in \mathbb{R}^n,$$

d.h. der Abklingfaktor $\alpha$ ist mindestens $\frac{1}{2}\mu_V(\mathbf{A})$. Die untere Grenze wird ähnlich zur oberen Grenze hergeleitet. □

In diesem Fall kann die modifizierte Matrixnorm auch in analytischer Form berechnet werden.

*Bemerkung* 7.7 (Berechnung von $\mu_V(\mathbf{A})$ für stabile LTI-Systeme). Mit Hilfe der Ljapunov-Gleichung aus Gl. (7.20) kann für stabile LTI-Systeme und quadratische Ljapunov-Funktionen die Matrixnorm $\mu_V(\mathbf{A})$ relativ einfach berechnet werden. Für eine beliebige symmetrisch und positiv definite

Matrix $\mathbf{Q} \succ \mathbf{0}$ ist die Lösung der Ljapunov-Gleichung $\mathbf{A}^\top\mathbf{P} + \mathbf{P}\mathbf{A} = -\mathbf{Q}$ eine positiv definite Matrix $\mathbf{P} \succ \mathbf{0}$, und $\mu_V(\mathbf{A})$ wird

$$\begin{aligned}\mu_V(\mathbf{A}) &:= \sup_{\mathbf{x}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \left[\frac{\partial V(\mathbf{x};\mathbf{A}\mathbf{x})}{V(\mathbf{x})}\right] \\ &= \sup_{\mathbf{x}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \left[-\frac{\mathbf{x}^\top\mathbf{Q}\mathbf{x}}{\mathbf{x}^\top\mathbf{P}\mathbf{x}}\right] = -\inf_{\mathbf{x}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \left[\frac{\mathbf{x}^\top\mathbf{Q}\mathbf{x}}{\mathbf{x}^\top\mathbf{P}\mathbf{x}}\right].\end{aligned}$$

Folglich gilt

$$-\mu_V(\mathbf{A}) \leq \frac{\mathbf{x}^\top\mathbf{Q}\mathbf{x}}{\mathbf{x}^\top\mathbf{P}\mathbf{x}}, \quad \forall\,\mathbf{x} \in \mathbb{R}^n\setminus\{\mathbf{0}\}. \tag{7.22}$$

Da $\mathbf{P} \succ \mathbf{0}$ und, somit, $\mathbf{x}^\top\mathbf{P}\mathbf{x} > 0$, folgt, dass $-\mu_V(\mathbf{A})\mathbf{x}^\top\mathbf{P}\mathbf{x} \leq \mathbf{x}^\top\mathbf{Q}\mathbf{x}$, d.h., dass

$$-\mu_V(\mathbf{A})\mathbf{P} - \mathbf{Q} \preceq \mathbf{0}. \tag{7.23}$$

Durch Links- und Rechtsmultiplizieren der Gl. (7.23) mit der symmetrischen und nichtsingulären Matrix $\mathbf{P}^{-1/2}$ folgt[12)]

$$-\mu_V(\mathbf{A})\mathbf{I} - \mathbf{P}^{-1/2}\mathbf{Q}\mathbf{P}^{-1/2} \preceq \mathbf{0}. \tag{7.24}$$

Dies ist äquivalent zu[13)]

$$-\mu_V(\mathbf{A}) \leq \lambda_{\min}(\mathbf{P}^{-1/2}\mathbf{Q}\mathbf{P}^{-1/2}) = \lambda_{\min}(\mathbf{Q}\mathbf{P}^{-1}). \tag{7.25}$$

Der Wert $\lambda_{\min}(\mathbf{Q}\mathbf{P}^{-1})$ wird erreicht falls $\mathbf{x}$ der zugehörige Eigenvektor zum Eigenwert $\lambda_{\min}(\mathbf{Q}\mathbf{P}^{-1})$ ist. Es handelt sich folglich um ein Minimum. Dann folgt, dass

$$\mu_V(\mathbf{A}) = -\lambda_{\min}(\mathbf{Q}\mathbf{P}^{-1}). \tag{7.26}$$

△

[12)] Die Matrix $\mathbf{P}^{-1/2}$ stellt die Quadratwurzel der Matrix $\mathbf{P}$ dar. Da die Matrix $\mathbf{P}^{-1}$ positiv definit ist, ist die Matrix $\mathbf{P}^{-1/2}$ eindeutig durch $\mathbf{P}^{-1/2}\mathbf{P}^{-1/2} = \mathbf{P}^{-1}$ definiert. Darüber hinaus ist diese Matrix ebenfalls positiv definit, und folglich, symmetrisch. In den Gl. (7.23) und (7.24) sind beide Matrizen somit *kongruent*, vgl. Def. 19 (Anhang).

[13)] Vgl. [8, Lemma 8.4.1]. Dabei wird auch die Tatsache verwendet, dass die Matrizen $\mathbf{P}^{-1/2}\mathbf{Q}\mathbf{P}^{-1/2}$ und $\mathbf{Q}\mathbf{P}^{-1}$ *ähnlich* sind, vgl. Def. 18 (Anhang), und somit gleiche Eigenwerte haben. Dies lässt sich durch Links- und Rechtsmultiplizieren der ersten Matrix mit den nichtsingulären Matrizen $\mathbf{P}^{1/2}$ bzw. $\mathbf{P}^{-1/2}$ zeigen.

Da die Matrizen $\mathbf{Q}$ und $\mathbf{P}^{-1}$ positiv definit sind, kann man noch feststellen, dass alle Eigenwerte der Matrix $\mathbf{Q}\mathbf{P}^{-1}$ positiv sind[14)] und somit, dass, $\mu_V(\mathbf{A})$ negativ ist. Der Wert $\mu_V(\mathbf{A})$ stellt folglich eine untere Grenze der Konvergenzrate des Systems aus Def. 1 dar, d.h.

$$\underline{\alpha}_V := -\frac{1}{2}\mu_V(\mathbf{A}) \leq \alpha. \tag{7.27}$$

Je nach Wahl der Funktion $V(\mathbf{x})$, kann diese untere Grenze maximiert werden. Durch Einsetzen der Gl. (7.27) in Gl. (7.22) ergibt sich als Bedingung für die untere Grenze der Konvergenzrate

$$2\underline{\alpha}_V \leq \frac{\mathbf{x}^\top \mathbf{Q}\mathbf{x}}{\mathbf{x}^\top \mathbf{P}\mathbf{x}}, \quad \forall\, \mathbf{x} \neq \mathbf{0}$$

und somit die Bedingung

$$-\mathbf{x}^\top(\mathbf{A}^\top\mathbf{P} + \mathbf{P}\mathbf{A})\mathbf{x} \geq 2\underline{\alpha}_V \mathbf{x}^\top \mathbf{P}\mathbf{x}, \quad \forall\, \mathbf{x} \neq \mathbf{0}.$$

Die maximale untere Grenze $\alpha_V^*$, d.h. die Konvergenzrate des LTI-Systems, kann folglich durch Lösen des konvexen Optimierungsproblems

$$\max_{\mathbf{P},\,\alpha_V} \alpha_V, \text{ sodass}$$

$$\mathbf{P} \succ \mathbf{0},\ \alpha_V > 0, \tag{7.28}$$

$$\mathbf{A}^\top\mathbf{P} + \mathbf{P}\mathbf{A} + 2\alpha_V \mathbf{P} \preceq \mathbf{0}. \tag{7.29}$$

berechnet werden. Die maximale Grenze $\alpha_V^*$ entspricht betragsmäßig dem Realteil des Eigenwertes/Eigenwertpaares, der/das am nächsten zur imaginären Achse liegt, d.h.

$$\alpha = \alpha_V^* = -\max_i \{\operatorname{Re}\lambda_i(\mathbf{A})\}.$$

Auch für exponentiell stabile **nichtlineare** Systeme kann man mit Hilfe quadratischer Ljapunov-Funktionen und der dazugehörigen kontraktiv invarianten Gebiete, vgl. Def. 13 (Anhang), eine untere Grenze der Konvergenzrate angeben. Dies wird in Lemma 7.5 gezeigt. Darauf basierend, stellt folgendes Lemma eine untere Grenze der Konvergenzrate eines allgemeinen exponentiell stabilen nichtlinearen Systems dar.

[14)] Vgl. [8, Korollar 8.3.7].

**Lemma 7.5.** *Gegeben sei ein nichtlineares System*

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t)), \quad \mathbf{x} \in \mathbb{R}^n, \mathbf{f}(\mathbf{0}) = \mathbf{0}, \tag{7.30}$$

*und das kontraktiv invariante Gebiet*

$$\mathcal{E}(\mathbf{P},c) := \{\mathbf{x} \in \mathbb{R}^n | \, \mathbf{x}^\top \mathbf{P}\mathbf{x} \leq c\},$$

*mit* $\mathbf{P} \succ \mathbf{0}$. *Es sei darüber hinaus angenommen, dass*

$$\mathbf{x}^\top \mathbf{P}\mathbf{f}(\mathbf{x}) \leq -\mathbf{x}^\top \mathbf{Q}\mathbf{x}, \quad \forall\, \mathbf{x} \in \mathcal{E}(\mathbf{P},c)\backslash\{\mathbf{0}\}, \tag{7.31}$$

*gilt, wobei* $\mathbf{Q} \succ \mathbf{0}$. *Dann gelten folgende Aussagen:*

*a)* $\|\mathbf{x}(t)\| \leq \sqrt{\kappa(\mathbf{P})}\|\mathbf{x}(0)\| \exp\left\{-\lambda_{\min}(\mathbf{Q}\mathbf{P}^{-1})t\right\}$, *für jedes* $\mathbf{x}(0) \in \mathcal{E}(\mathbf{P},c)\backslash\{\mathbf{0}\}$ *und* $t \geq 0$.

*b) Die Ruhelage* $\mathbf{x}_\mathrm{R} = \mathbf{0}$ *ist exponentiell stabil.*

*c) Die quadratische Ljapunov-Funktion* $V(\mathbf{x})$ *nimmt entlang jeder Trajektorien des Systems exponentiell ab, und die Hälfte der größten unteren Schranke der Menge aller Abklingfaktoren*

$$\underline{\alpha}_V(c) := \frac{1}{2} \cdot \inf_{\mathbf{x} \in \mathcal{E}(\mathbf{P},c)\backslash\{\mathbf{0}\}} \left( -\frac{\partial V(\mathbf{x}; f(\mathbf{x}))}{V(\mathbf{x})} \right) = \lambda_{\min}(\mathbf{Q}\mathbf{P}^{-1}), \tag{7.32}$$

*stellt eine Untergrenze der Konvergenzrate* $\alpha$ *des Systems innerhalb des ellipsoidalen Gebietes* $\mathcal{E}(\mathbf{P},c)\backslash\{\mathbf{0}\}$ *dar, d.h. es gilt*

$$0 < \underline{\alpha}_V(c) \leq \alpha, \quad \forall\, c > 0. \tag{7.33}$$

*Beweis.* a) Da das Gebiet $\mathcal{E}(\mathbf{P},c)$ kontraktiv invariant für das System aus Gl. (7.30) ist, ist die quadratische Funktion $V(\mathbf{x}) = \mathbf{x}^\top \mathbf{P}\mathbf{x}$ eine gültige Ljapunov-Funktion des Systems. Darüber hinaus gilt

$$\|\mathbf{x}(t)\|^2 \lambda_{\min}(\mathbf{P}) \leq \mathbf{x}(t)^\top \mathbf{P}\mathbf{x}(t) \leq \lambda_{\max}(\mathbf{P})\|\mathbf{x}(t)\|^2, \quad \forall\, \mathbf{x} \neq \mathbf{0}, \forall\, t \geq 0.$$

Aus Gl. (7.31) folgt, dass

$$\frac{\dot{V}(\mathbf{x})}{V(\mathbf{x})} = \frac{2\mathbf{x}^\top \mathbf{P}\mathbf{f}(\mathbf{x})}{\mathbf{x}^\top \mathbf{P}\mathbf{x}} \leq -\frac{2\mathbf{x}^\top \mathbf{Q}\mathbf{x}}{\mathbf{x}^\top \mathbf{P}\mathbf{x}} \leq \sup_{\mathbf{x} \in \mathcal{E}(\mathbf{P},c)} \left( -\frac{2\mathbf{x}^\top \mathbf{Q}\mathbf{x}}{\mathbf{x}^\top \mathbf{P}\mathbf{x}} \right) = -2\lambda_{\min}(\mathbf{Q}\mathbf{P}^{-1}).$$

Integriert man die obige Ungleichung entlang einer Trajektorie vom Zeitpunkt $t_0 = 0$ zum Zeitpunkt $t$, erhält man

$$V(\mathbf{x}(t)) \leq V(\mathbf{x}(0)) \exp\{-2\lambda_{\min}(\mathbf{Q}\mathbf{P}^{-1})t\}.$$

Schließlich folgt, dass

$$\|\mathbf{x}(t)\|^2 \lambda_{\min}(\mathbf{P}) \leq \|\mathbf{x}(0)\|^2 \lambda_{\max}(\mathbf{P}) \exp\{-2\lambda_{\min}(\mathbf{Q}\mathbf{P}^{-1})t\}$$

und somit, dass

$$\|\mathbf{x}(t)\| \leq \sqrt{\kappa(\mathbf{P})}\|\mathbf{x}(0)\| \exp\left\{-\lambda_{\min}(\mathbf{Q}\mathbf{P}^{-1})t\right\}, \forall \mathbf{x}(0) \in \mathcal{E}(\mathbf{P},c)\backslash\{\mathbf{0}\}, t \geq 0.$$

b) Dies folgt unmittelbar aus a).

c) Für jeden Zustand $\mathbf{x} \in \mathcal{E}(\mathbf{P},c)\backslash\{\mathbf{0}\}$ existiert für das exponentiell stabile nichtlineare System $\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t))$ ein $\alpha_{\mathbf{x}} > 0$, sodass $\dot{V}(\mathbf{x})/V(\mathbf{x}) \leq -2\alpha_{\mathbf{x}}$, d.h. $\dot{V}(\mathbf{x}) \leq -2\alpha_{\mathbf{x}} V(\mathbf{x})$. Da aus Gl. (7.32) folgt, dass $\underline{\alpha}_V(c) \leq \alpha_{\mathbf{x}}, \forall \mathbf{x} \in \mathcal{E}(\mathbf{P},c)\backslash\{\mathbf{0}\}$, folgt im Weiteren, dass $\dot{V}(\mathbf{x}) \leq -2\alpha_{\mathbf{x}} V(\mathbf{x}) \leq -2\underline{\alpha}_V(c) V(\mathbf{x})$. Daraus folgt dann, dass $V(\mathbf{x}) \leq V(\mathbf{x}_0) e^{-2\underline{\alpha}_V(c)t}, \forall \mathbf{x} \in \mathcal{E}(\mathbf{P},c)\backslash\{\mathbf{0}\}$ und schließlich aus Gl. (7.21), dass

$$\|\mathbf{x}(t)\| \leq \sqrt{\lambda_{\max}(\mathbf{P})/\lambda_{\min}(\mathbf{P})}\|\mathbf{x}(0)\| e^{-\underline{\alpha}_V(c)t}, \quad \forall \mathbf{x}(0) \in \mathcal{E}(\mathbf{P},c)\backslash\{\mathbf{0}\},$$

d.h. dass $\underline{\alpha}_V(c)$ eine untere Grenze der Konvergenzrate des Systems ist. □

**Korollar 7.6.** *Die größte untere Schranke der variablen Ljapunov-Funktion-basierten Konvergenzrate eines exponentiell stabilen nichtlinearen Systems $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}(t))$, $\mathbf{x} \in \mathbb{R}^n$ innerhalb des Einzugsgebietes der Ruhelage $\mathbf{x}_{\mathrm{R}} = \mathbf{0}$,*

$$\underline{\alpha}_V = \inf_{c>0}\left\{\underline{\alpha}_V(c) = \frac{1}{2} \cdot \inf_{\mathbf{x}\in\mathcal{E}(\mathbf{P},c)\backslash\{\mathbf{0}\}} \left(-\frac{\partial V(\mathbf{x}; f(\mathbf{x}))}{V(\mathbf{x})}\right)\right\}, \tag{7.34}$$

*stellt eine untere Grenze der Konvergenzrate $\alpha$ des nichtlinearen Systems dar.*

*Beweis.* Der Beweis folgt unmittelbar aus dem Beweis des Lemmas 7.5 und wird hier weggelassen. □

Auch mittels impliziter Ljapunov-Funktionen können Bedingungen für die exponentielle Stabilität eines nichtlinearen Systems gestellt werden. Folgender Satz zeigt in diesem Zusammenhang unter welchen Bedingungen die Ruhelage eines nichtlinearen Systems in **zustandsabhängiger Koeffizientenform** exponentiell stabil ist. Darüber hinaus wird eine obere Grenze der Zustandsnorm angegeben. Der Satz kann als Erweiterung von

[5, Satz 5] für den Fall nichtlinearer Systeme in zustandsabhängiger Koeffizientenform mit ellipsoidalen kontraktiv invarianten Gebieten gesehen werden. Der Beweis basiert auf dem Nachweis der asymptotischen Stabilität eines nichtlinearen Systems mittels impliziter Ljapunov-Funktionen aus [2, Satz 4].

Ein ähnliches Theorem, das auch dynamische Systeme in zustandsabhängiger Koeffizientenform analysiert und ebenfalls Bedingungen für exponentielle Stabilität untersucht, ist in [54, Theorem 1] angegeben. Darin wird der Zustandsraum in endlich viele Gebiete aufgeteilt, die sich überlappen können und in denen ein bestimmtes lineares System aktiv ist. Die Gesamtdynamik des darin analysierten Systems ergibt sich als gewichteter Durchschnitt der lokal aktiven linearen Dynamiken. Eine obere Grenze der Zustandsnorm in einer der o.g. Regionen ist auch angegeben.

**Satz 7.7** *In einer Menge*

$$\mathcal{V}_0 := \{(v,\mathbf{x}) | 0 < v < 1, \mathbf{x} \in \mathcal{U}_0 \backslash \{\mathbf{0}\} \subset \mathcal{B}_\epsilon(0)\},$$

*sei eine stetige und differenzierbare Funktion* $g(v,\mathbf{x})$ *gegeben durch*

$$g(v,\mathbf{x}) := \mathbf{x}^\top \mathbf{P}(v(\mathbf{x}))\mathbf{x} - 1, \quad \mathbf{P}(v) : (0,1) \to \mathbb{P}^n, \tag{7.35}$$

*welche folgende Bedingungen erfüllt:*

***i)*** *Für* $\mathbf{x} \to 0$ *resultiert aus* $g(v,\mathbf{x}) = 0$ *der Grenzübergang* $v \to 0^+$,

***ii)*** $\lim\limits_{v \to 0^+} g(v,\mathbf{x}) > 0$ *und* $\lim\limits_{v \to 1^-} g(v,\mathbf{x}) < 0$ *für alle* $\mathbf{x} \in \mathcal{U}_0 \backslash \{\mathbf{0}\}$.

*Seien darüber hinaus die durch die Funktion* $g(v,\mathbf{x})$ *bestimmten Gebiete bezeichnet durch*

$$\mathcal{E}(v) := \left\{ \mathbf{x} \,\middle|\, \mathbf{x}^\top \mathbf{P}(v(\mathbf{x}))\mathbf{x} < 1 \right\} \subseteq \mathcal{U}_0.$$

*Das autonome nichtlineare System*

$$\dot{\mathbf{x}}(t) = \mathbf{A}(\nu(\mathbf{x}(t)))\mathbf{x}(t), \quad \mathbf{x} \in \mathbb{R}^n, \, \mathbf{x}_\mathrm{R} = \mathbf{0}, \tag{7.36}$$

*mit*

$$\nu(\mathbf{x}) := \begin{cases} v(\mathbf{x}), \ \textit{mit } g(v,\mathbf{x}) = 0, & \mathbf{x} \in \mathcal{E}(1)\backslash\mathcal{E}(v_{\min}) \\ v_{\min}, & \mathbf{x} \in \mathcal{E}(v_{\min}), \end{cases} \tag{7.37}$$

*und der stetigen matrixwertigen Funktion*[a)] $\mathbf{A}(v) : [v_{\min},1) \to \mathbb{R}^{n\times n}$*, wobei* $v_{\min} \in (0,1)$ *gegeben ist, besitze im Weiteren eine eindeutige Lösung für jeden Anfangswert* $\mathbf{x}(0) \in \mathcal{B}_\epsilon(0)$.

*Sind die Bedingungen*

***iii)*** $-\infty < \frac{\partial g(v,\mathbf{x})}{\partial v} < 0$ *für alle* $(v,\mathbf{x}) \in \mathcal{V}_0$,

***iv)*** $\frac{\partial g(v,\mathbf{x}(t))}{\partial t} < 0$ *für alle* $(v,\mathbf{x}) \in \mathcal{V}_0$, *mit* $g(v,\mathbf{x}) = 0$

*erfüllt, dann gilt:*

*a) Für jedes* $\mathbf{x} \in \mathcal{U}_0\backslash\{\mathbf{0}\}$ *besitzt die Gleichung* $g(v,\mathbf{x}) = 0$ *eine eindeutige Lösung* $v = v(\mathbf{x})$, *mit* $v \in (0,1)$.

*b) Die stetige Funktion*

$$V(\mathbf{x}) := \begin{cases} v(\mathbf{x}), \ \textit{mit } g(v,\mathbf{x}) = 0, & \mathbf{x} \in \mathcal{E}(1)\backslash\mathcal{E}(v_{\min}), \\ v_{\min}\mathbf{x}^\top\mathbf{P}(v_{\min})\mathbf{x}, & \mathbf{x} \in \mathcal{E}(v_{\min}) \end{cases} \tag{7.38}$$

*stellt in* $\mathcal{U}_0$ *eine Ljapunov-Funktion des Systems dar.*

*c) Jedes abgeschlossene Gebiet* $\mathcal{E}(c)$, *mit* $c \in [v_{\min},1]$, *ist ein kontraktiv invariantes Gebiet der Ruhelage.*

*d) Für alle* $c \in [v_{\min},1]$ *sind die Ränder der kontraktiv invarianten Gebiete* $\mathcal{E}(c)$ *disjunkt, d.h. es gilt*

$$\partial\mathcal{E}(c_i) \cap \partial\mathcal{E}(c_j) = \varnothing, \quad \forall 0 < c_i, c_j < 1, c_i \neq c_j.$$

*e) Für alle* $c \in [v_{\min},1]$ *sind die Gebiete* $\mathcal{E}(c)$ *ineinander verschachtelt, d.h. es gilt*

$$\mathcal{E}(c_i) \subset \mathcal{E}(c_j), \quad \forall\, v_{\min} \leq c_i < c_j < 1.$$

*f) Für jeden Zeitpunkt* $t \geq 0$ *ist eine obere Grenze der Zustandsnorm* $\|\mathbf{x}(t)\|$ *gegeben durch*

$$\|\mathbf{x}(t)\| \leq \left(\frac{\gamma_{\max}}{v_{\min}}\right)^{1/2} \|\mathbf{x}(0)\| \exp\left\{-\frac{1}{2}\alpha_{\min}t\right\}, \quad \forall \mathbf{x}(0) \in \partial\mathcal{E}(1),$$

*wobei* $\gamma_{\max}$ *durch*

$$\gamma_{\max} := \max_{v\in[v_{\min},1]} \kappa(\mathbf{P}_v) = \frac{\lambda_{\max}(\mathbf{P}_{v_{\min}})}{\lambda_{\min}(\mathbf{P}_1)} \tag{7.39}$$

*und* $\alpha_{\min}$ *durch*

$$\alpha_{\min} := \frac{1}{v_{\min}\lambda_{\min}(\mathbf{P}_1)} \frac{\min\limits_{v\in[v_{\min},1]} \lambda_{\min}(\mathbf{A}_v^\top\mathbf{P}_v + \mathbf{P}_v\mathbf{A}_v)}{\min\limits_{v\in[v_{\min},1]} \lambda_{\min}(\mathbf{P}'_v\mathbf{P}_v^{-1})} > 0 \tag{7.40}$$

*gegeben sind.*

*g) Die Ruhelage* $\mathbf{x}_{\mathrm{R}} = \mathbf{0}$ *ist* **exponentiell stabil**.

[a)] Vgl. Def. 25 (Anhang).

*Beweis.* a) Dies wurde in [2, Theorem 4] bewiesen.

b) Für die Anwendung der direkten Methode von Ljapunov wird aufgrund der besonderen Definition der Ljapunov-Funktion der Zustandsraum in drei Gebiete eingeteilt. In dem Gebiet $\mathbf{x} \in \mathcal{E}(1)\backslash(\mathcal{E}\,(v_{\min}) \cup \partial\mathcal{E}(v_{\min}))$, sowie in dem Gebiet $\mathbf{x} \in \mathcal{E}(v_{\min})$ ist die Funktion $V(\mathbf{x})$ stetig und differenzierbar. In einem dritten Gebiet $\mathbf{x} \in \partial\mathcal{E}(v_{\min})$ ist diese nicht differenzierbar und die direkte Methode von Ljapunov wird mit Hilfe von Dini-Derivierten angewandt. Im ersten Gebiet gilt

$$V(\mathbf{x}) = v(\mathbf{x}) > 0, \quad \text{mit } g(v,\mathbf{x}) = 0, \quad \forall\,\mathbf{x} \in \mathcal{E}(1)\backslash(\mathcal{E}\,(v_{\min}) \cup \partial\mathcal{E}(v_{\min}))\,,$$

sowie, aufgrund der Bedingungen iii) und iv),

$$\dot{V}(\mathbf{x}) = \dot{v}(\mathbf{x}) = -\frac{\partial g(v,\mathbf{x}(t))/\partial t}{\partial g(v,\mathbf{x})/\partial v} < 0, \quad \forall\,\mathbf{x} \in \mathcal{E}(1)\backslash(\mathcal{E}\,(v_{\min}) \cup \partial\mathcal{E}(v_{\min}))\,.$$

Im zweiten Gebiet gilt

$$\begin{aligned} V(\mathbf{x}) &= v_{\min}\mathbf{x}^\top\mathbf{P}(v_{\min})\mathbf{x} > 0, \quad \forall\,\mathbf{x} \in \mathcal{E}(v_{\min})\backslash\{\mathbf{0}\}, \\ V(\mathbf{0}) &= 0, \end{aligned}$$

sowie aufgrund der Bedingung iv)

$$\begin{aligned} \dot{V}(\mathbf{x}) &= v_{\min}\mathbf{x}^\top\left(\mathbf{A}(v_{\min})^\top\mathbf{P}(v_{\min}) + \mathbf{P}(v_{\min})\mathbf{A}(v_{\min})\right)\mathbf{x} \\ &< 0, \quad \forall\,\mathbf{x} \in \mathcal{E}(v_{\min})\backslash\{\mathbf{0}\}. \end{aligned} \tag{7.41}$$

Im dritten Gebiet, $\mathbf{x} \in \partial\mathcal{E}(v_{\min})$, ist die Funktion $V(\mathbf{x}) = v(\mathbf{x})$ nicht differenzierbar. Daher bedarf es bei der Anwendung der direkten Methode von Ljapunov der rechten oberen Dini-Derivierten der stetigen Funktion $V(\mathbf{x})$ entlang einer Trajektorie $\mathbf{x}(t)$.[15] Diese ist definiert als[16]

$$D^+V(\mathbf{x}(t)) := \limsup_{h\to 0^+} \frac{V(\mathbf{x}(t+h)) - V(\mathbf{x}(t))}{h}.$$

Aus Bedingung iv) folgt für $\mathbf{x} \in \partial\mathcal{E}(v_{\min})$ und jedes beliebig kleine $h > 0$, dass

$$g(v,\mathbf{x}(t+h)) < g(v,\mathbf{x}(t)) = 0,$$

d.h. $\mathbf{x}(t+h) \in \mathcal{E}(v_{\min})$. Somit ist die Ljapunov-Funktion in diesem Punkt $V(\mathbf{x}(t+h)) = v_{\min}\mathbf{x}(t+h)^\top\mathbf{P}(v_{\min})\mathbf{x}(t+h)$. Folglich gilt

$$\begin{aligned} D^+V(\mathbf{x}(t)) &= \limsup_{h\to 0^+} \frac{v_{\min}\mathbf{x}(t+h)^\top\mathbf{P}(v_{\min})\mathbf{x}(t+h) - v_{\min}}{h} \\ &= \dot{V}_l(\mathbf{x}(t)), \quad \forall\, \mathbf{x} \in \partial\mathcal{E}(v_{\min}), \end{aligned}$$

wobei $V_l(\mathbf{x}(t)) = v_{\min}\mathbf{x}(t)^\top\mathbf{P}(v_{\min})\mathbf{x}(t)$. Aus Gl. (7.41) folgt aber, dass $\dot{V}_l(\mathbf{x}(t)) < 0$, für jedes $\mathbf{x} \in \mathbb{R}^n\backslash\{\mathbf{0}\}$. Daher gilt

$$D^+V(\mathbf{x}(t)) < 0, \quad \forall\, \mathbf{x} \in \partial\mathcal{E}(v_{\min}).$$

Daraus folgt,[17] dass die Funktion $V(\mathbf{x})$ eine gültige Ljapunov-Funktion des Systems ist.

c) Da die Funktion $V(\mathbf{x})$ eine gültige Ljapunov-Funktion des Systems ist, folgt, dass für jedes $c > 0$ das Gebiet

$$G := \{\mathbf{x}\,|V(\mathbf{x}) < c\}$$

kontraktiv invariant ist. Das Gebiet ist dabei gleich mit dem Gebiet $\mathcal{E}(c)$.[18]

d-e) Dies wurde in [2, Abschnitt III] bewiesen.

[15] Vgl. Def. 17.

[16] Vgl. Def. 16 (Anhang) für die Definition der Limes superior einer Funktion und Def. 17 (Anhang) für die Definition der rechten oberen Dini-Derivierten einer stetigen Funktion.

[17] Vgl. [62, Theorem 6.3].

[18] Vgl. [2, Theorem 5].

f) In jedem Gebiet $\mathbf{x} \in \partial\mathcal{E}(v) \subseteq \mathcal{E}(1)\backslash\mathcal{E}(v_{\min})$ gilt für die zeitliche Ableitung der Ljapunov-Funktion $V(\mathbf{x})$

$$\dot{V}(\mathbf{x}) = \dot{v}(\mathbf{x}) = -\frac{\partial g(v,\mathbf{x}(t))/\partial t}{\partial g(v,\mathbf{x})/\partial v} = -\frac{\mathbf{x}^\top(\mathbf{A}_v^\top\mathbf{P}_v + \mathbf{P}_v\mathbf{A}_v)\mathbf{x}}{\mathbf{x}^\top\mathbf{P}'_v\mathbf{x}}, \tag{7.42}$$

wobei $\mathbf{P}'_v = \frac{\partial\mathbf{P}(v)}{\partial v} \prec \mathbf{0}$. Dabei gilt[19] für jedes $v \in [v_{\min},1]$ und $\mathbf{x} \in \partial\mathcal{E}(v) \subseteq \mathcal{E}(1)\backslash\mathcal{E}(v_{\min})$

$$\frac{\mathbf{x}^\top\mathbf{P}'_v\mathbf{x}}{\mathbf{x}^\top\mathbf{P}_v\mathbf{x}} \geq \min_{\mathbf{x}\in\partial\mathcal{E}(v)} \frac{\mathbf{x}^\top\mathbf{P}'_v\mathbf{x}}{\mathbf{x}^\top\mathbf{P}_v\mathbf{x}} = \lambda_{\min}(\mathbf{P}'_v\mathbf{P}_v^{-1}) < 0.$$

Daraus folgt, dass

$$\frac{\mathbf{x}^\top\mathbf{P}'_v\mathbf{x}}{\mathbf{x}^\top\mathbf{P}_v\mathbf{x}} \geq \min_{v\in[v_{\min},1]} \lambda_{\min}(\mathbf{P}'_v\mathbf{P}_v^{-1}) =: k_1 < 0, \quad \forall\,\mathbf{x} \in \mathcal{E}(1)\backslash\mathcal{E}(v_{\min}).$$

Darüber hinaus gilt für jedes $v \in [v_{\min},1]$ und $\mathbf{x} \in \partial\mathcal{E}(v)$

$$\lambda_{\min}(\mathbf{A}_v^\top\mathbf{P}_v + \mathbf{P}_v\mathbf{A}_v)\|\mathbf{x}\|^2 \leq \mathbf{x}^\top(\mathbf{A}_v^\top\mathbf{P}_v + \mathbf{P}_v\mathbf{A}_v)\mathbf{x},$$

wobei für

$$k_2 := \min_{v\in[v_{\min},1]} \lambda_{\min}(\mathbf{A}_v^\top\mathbf{P}_v + \mathbf{P}_v\mathbf{A}_v) < 0,$$

folgt

$$k_2\|\mathbf{x}\|^2 \leq \mathbf{x}^\top(\mathbf{A}_v^\top\mathbf{P}_v + \mathbf{P}_v\mathbf{A}_v)\mathbf{x}, \quad \forall\,\mathbf{x} \in \partial\mathcal{E}(1)\backslash\mathcal{E}(v_{\min}).$$

Da $\mathbf{x}^\top\mathbf{P}_v\mathbf{x} = 1$, $\forall\,\mathbf{x} \in \partial\mathcal{E}(v) \subseteq \mathcal{E}(1)\backslash\mathcal{E}(v_{\min})$, folgt aus Gl. (7.42), dass

$$\dot{v}(\mathbf{x}) = -\frac{\mathbf{x}^\top(\mathbf{A}_v^\top\mathbf{P}_v + \mathbf{P}_v\mathbf{A}_v)\mathbf{x}}{\mathbf{x}^\top\mathbf{P}'_v\mathbf{x}} \leq -\frac{\mathbf{x}^\top(\mathbf{A}_v^\top\mathbf{P}_v + \mathbf{P}_v\mathbf{A}_v)\mathbf{x}}{k_1} \leq -\frac{k_2}{k_1}\|\mathbf{x}\|^2. \tag{7.43}$$

Darüber hinaus gilt

$$v\lambda_{\min}(\mathbf{P}_v)\|\mathbf{x}\|^2 \leq v(\mathbf{x}) \leq v\lambda_{\max}(\mathbf{P}_v)\|\mathbf{x}\|^2, \quad \forall\,\mathbf{x} \in \partial\mathcal{E}(v),$$

[19] Vgl. [8, Fakt 8.15.21] und [9]. Voraussetzungen dafür sind, dass die Matrix $\mathbf{P}'_v$ symmetrisch ist und die Matrix $\mathbf{P}_v$ positiv definit ist. Die letzte Beziehung $\lambda_{\min}(\mathbf{P}'_v\mathbf{P}_v^{-1}) < 0$ folgt aus [8, Korollar 8.3.7].

und, somit,

$$\begin{aligned}\min_{v\in[v_{\min},1]} v\lambda_{\min}(\mathbf{P}_v)\|\mathbf{x}\|^2 & \\ \leq v(\mathbf{x}) & \\ \leq \max_{v\in[v_{\min},1]} v\lambda_{\max}(\mathbf{P}_v)\|\mathbf{x}\|^2, & \quad \forall\, \mathbf{x}\in\mathcal{E}(1)\backslash\mathcal{E}(v_{\min}).\end{aligned}$$

Da die matrixwertige Funktion[20] $\mathbf{P}_v$ monoton fallend ist, gilt noch

$$\begin{aligned}\min_{v\in[v_{\min},1]} v\lambda_{\min}(\mathbf{P}_v) &= v_{\min}\lambda_{\min}(\mathbf{P}_1) > 0\\ \max_{v\in[v_{\min},1]} v\lambda_{\max}(\mathbf{P}_v) &= \lambda_{\max}(\mathbf{P}_{v_{\min}}) > 0.\end{aligned}$$

Aus Gl. (7.43) folgt dann, dass

$$\begin{aligned}\dot{v}(\mathbf{x}(t)) &\leq -\frac{k_2}{k_1}\|\mathbf{x}(t)\|^2\\ &\leq -\frac{k_2}{k_1}\frac{1}{v_{\min}\lambda_{\min}(\mathbf{P}_1)}v(\mathbf{x}(t)), \quad \forall\, \mathbf{x}(t)\in\partial\mathcal{E}(v)\backslash\mathcal{E}(v_{\min}).\end{aligned}$$

Integriert man die obige Ungleichung entlang einer Trajektorie des Systems vom Zeitpunkt $t_0 = 0$ zum Zeitpunkt $t$, erhält man

$$v(\mathbf{x}(t)) \leq v(\mathbf{x}(0))\exp\left(-\frac{k_2}{k_1}\frac{1}{v_{\min}\lambda_{\min}(\mathbf{P}_1)}t\right), \forall\, \mathbf{x}(t)\in\partial\mathcal{E}(v)\backslash\mathcal{E}(v_{\min}), t\geq 0.$$

Schließlich folgt, dass

$$\begin{aligned}v_{\min}\lambda_{\min}(\mathbf{P}_1)\|\mathbf{x}(t)\|^2 & \\ \leq v(\mathbf{x}(t)) & \\ \leq v(\mathbf{x}(0))\exp\left(-\frac{k_2}{k_1}\frac{1}{v_{\min}\lambda_{\min}(\mathbf{P}_1)}t\right) & \\ \leq \lambda_{\max}(\mathbf{P}_{v_{\min}})\|\mathbf{x}(0)\|^2\exp\left(-\frac{k_2}{k_1}\frac{1}{v_{\min}\lambda_{\min}(\mathbf{P}_1)}t\right), &\end{aligned}$$

[20] Vgl. Def. 25 (Anhang).

und, somit, dass

$$\|\mathbf{x}\| \leq \sqrt{\frac{\lambda_{\max}(\mathbf{P}_{v_{\min}})}{v_{\min}\lambda_{\min}(\mathbf{P}_1)}}\|\mathbf{x}_0\| \cdot \exp\left(-\frac{1}{2}\frac{\min\limits_{v\in[v_{\min},1]}\lambda_{\min}(\mathbf{A}_v^\top\mathbf{P}_v+\mathbf{P}_v\mathbf{A}_v)}{\min\limits_{v\in[v_{\min},1]}\lambda_{\min}(\mathbf{P}_v'\mathbf{P}_v^{-1})}\frac{1}{v_{\min}\lambda_{\min}(\mathbf{P}_1)}t\right).$$

g) Diese Aussage folgt unmittelbar aus e). Dabei gilt noch $\gamma_{\max} > 0$ und $\alpha_{\min} > 0$. Dies folgt aus der Tatsache, dass für jedes $v \in [v_{\min},1]$ die Matrix $\mathbf{P}_v$ positiv definit und die Matrix $\mathbf{A}_v^\top\mathbf{P}_v + \mathbf{P}_v\mathbf{A}_v$ negativ definit sind. □

Der Wert $\alpha_{\min}$ aus Gl. (7.40) stellt eine untere Grenze der *Konvergenzrate* des Systems im gesamten Bereich $\mathcal{E}(1)$ dar. Eine gebietsabhängige Konvergenzrate kann darüber hinaus wie folgt formuliert werden.

**Korollar 7.8 [Gebietsabhängige Konvergenzrate des Systems].** *Für das System aus Satz 7.7 ist für jeden Zeitpunkt $t \geq t_0$, mit $\mathbf{x}(t_0) \in \partial\mathcal{E}(v^*)$ und $v^* \in (0,1]$, eine obere Grenze der Zustandsnorm $\|\mathbf{x}(t)\|$ gegeben durch*

$$\|\mathbf{x}(t)\| \leq (\gamma_{\max}(v^*))^{1/2}\,\|\mathbf{x}(t_0)\|\exp\left\{-\frac{1}{2}\alpha_{\min}(v^*)t\right\}, \quad \mathbf{x}(t_0) \in \partial\mathcal{E}(v^*),$$

*wobei $\gamma_{\max}(v^*)$ durch*

$$\gamma_{\max}(v^*) := \max_{v\in[v_{\min},v^*]}\kappa(\mathbf{P}_v) = \frac{\lambda_{\max}(\mathbf{P}_{v_{\min}})}{\lambda_{\min}(\mathbf{P}_{v^*})}$$

*und $\alpha_{\min}(v^*)$ durch*

$$\alpha_{\min}(v^*) := \min_{v\in[v_{\min},v^*]}\left\{-\lambda_{\max}[(\mathbf{A}_v^\top\mathbf{P}_v+\mathbf{P}_v\mathbf{A}_v)\mathbf{P}_v^{-1}]\right\} > 0 \qquad (7.44)$$

*definiert sind.*

Der Wert $\alpha_{\min}(v^*)$, mit $v^* \in (0,1]$, wird *gebietsabhängige Konvergenzrate* des Systems genannt. Ein Vorteil der gebietsabhängigen Konvergenzrate ist, dass sie unabhängig von der Systemordnung eine skalare Funktion ist. Darüber hinaus stellt diese eine untere Grenze der Konvergenzrate

des Systems innerhalb des jeweiligen Gebietes dar. Diese gebietsabhängige Konvergenzrate kann exakt bestimmt, approximiert oder interpoliert werden. Dies wird in den Kapiteln 7.5.2 und 8 gezeigt.

### 7.5.2 Analyse der Konvergenzrate

Eine exakte Performance-Analyse in einem nichtlinearen Regelkreis setzt die Zeitlösung des Systems voraus. Weil diese nur in seltenen Fällen zur Verfügung steht, wird im Folgenden die *gebietsabhängige* Konvergenzrate analysiert, welche im Korrolar 7.8 definiert wurde, und verwendet, um eine obere Grenze der Zustandsnorm anzugeben.

Im Unterschied zu den linearen Systemen ist die Konvergenzrate eines nichtlinearen Systems nicht konstant, sondern abhängig von dem Abstand zur Ruhelage und in den meisten Fällen nicht exakt bestimmbar. Eine untere Grenze der Konvergenzrate kann jedoch mittels (impliziter) Ljapunov-Funktionen angegeben werden. Diese ist ebenfalls im Abschnitt 7.5 vorgestellt worden. Deren Berechnung wird im Fall der nicht-sättigenden und konvergenzoptimalen *klassischen* WSVR mittels iLF und *invers-polynomialen* WSVR vorgestellt.

#### Nicht-sättigende WSVR

Mit Hilfe des Satzes 7.7 kann die exponentielle Stabilität des geschlossenen Kreises aus Punkt $i$) des Satzes 3.1 und aus Punkt $i$) des Satzes 4.1 unter Verwendung der (Ljapunov-)Funktion $V_\star(\mathbf{x})$

$$V_\star(\mathbf{x}) := \begin{cases} v(\mathbf{x}), \text{ mit } g_\star(\mathbf{x},v) = 0, & \mathbf{x} \in \mathcal{G}_\star(1) \backslash \mathcal{G}_\star(\varepsilon), \\ \varepsilon \mathbf{x}^\top \mathbf{Q}_\varepsilon \mathbf{x}, & \mathbf{x} \in \mathcal{G}_\star(\varepsilon), \end{cases} \tag{7.45}$$

mit $\star = \Delta$ für die *klassische* WSVR und $\star = \mathrm{P}$ für die *invers-polynomiale* WSVR, nachgewiesen werden. Der geschlossene Regelkreis hat dabei die allgemeine Form $\dot{\mathbf{x}} = \hat{\mathbf{A}}_{\star_v} \mathbf{x}$. Wie im Korollar 7.8 gezeigt, bildet der kleinste Abklingfaktor der Funktion $V_\star(\mathbf{x})$ entlang der Trajektorien des Systems $\dot{\mathbf{x}} = \hat{\mathbf{A}}_{\star_v} \mathbf{x}$ eine obere Grenze der Zustandsnorm $\|\mathbf{x}\|$ für alle Anfangsauslenkungen, die auf dem Rand eines Gebietes $\mathbf{x} \in \partial\mathcal{G}_\star(v)$ starten. Dieser Faktor ist

$$\underline{\alpha}_\star(v) := \frac{1}{2} \min_{\mathbf{x} \in \partial\mathcal{G}_\star(v)} \left( -\frac{\dot{V}_\star(\mathbf{x})}{V_\star(\mathbf{x})} \right).$$

Für alle $\mathbf{x} \in \partial\mathcal{G}_\star(v) \subseteq \mathcal{G}_\star(1)\backslash\mathcal{G}_\star(\varepsilon)$ gilt des Weiteren

$$\begin{aligned}\underline{\alpha}_\star(v) &= \frac{1}{2v} \min_{\mathbf{x}\in\partial\mathcal{G}_\star(v)} (-\dot{v}) \\ &= \frac{1}{2v} \min_{\mathbf{x}\in\partial\mathcal{G}_\star(v)} \left[\frac{\mathbf{x}^\top(\hat{\mathbf{A}}_{\star_v}^\top \mathbf{Q}_v + \mathbf{Q}_v\hat{\mathbf{A}}_{\star_v})\mathbf{x}}{\mathbf{x}^\top\mathbf{Q}_v'\mathbf{x}}\right] \\ &= \frac{1}{2v}\lambda_{\min}\left[(\hat{\mathbf{A}}_{\star_v}^\top \mathbf{Q}_v + \mathbf{Q}_v\hat{\mathbf{A}}_{\star_v})(\mathbf{Q}_v')^{-1}\right], \quad \forall\, v \in (\varepsilon,1]. \end{aligned} \tag{7.46}$$

Da die Matrix $\hat{\mathbf{A}}_{\star_v}^\top \mathbf{Q}_v + \mathbf{Q}_v\hat{\mathbf{A}}_{\star_v}$ symmetrisch ist und die Matrix $\mathbf{Q}_v' \prec 0$ negativ definit ist, folgt[21], dass $\mathrm{Spec}[(\hat{\mathbf{A}}_{\star_v}^\top \mathbf{Q}_v + \mathbf{Q}_v\hat{\mathbf{A}}_{\star_v})(\mathbf{Q}_v')^{-1}] \subset \mathbb{R}$. Da noch $\hat{\mathbf{A}}_v^\top\mathbf{Q}_v + \mathbf{Q}_v\hat{\mathbf{A}}_v \prec 0$, folgt[22] darüber hinaus, dass $\lambda[(\hat{\mathbf{A}}_v^\top\mathbf{Q}_v + \mathbf{Q}_v\hat{\mathbf{A}}_v)\mathbf{Q}_v^{-1}] > 0$. Somit ist $\underline{\alpha}_\star(v)$ für alle $v \in [\varepsilon,1]$ und jedes $\varepsilon \in (0,1)$ positiv.

Für alle $\mathbf{x} \in \partial\mathcal{G}_\star(\varepsilon)$ gilt

$$\begin{aligned}\underline{\alpha}_\star(\varepsilon) &= \frac{1}{2} \min_{\mathbf{x}\in\partial\mathcal{G}_\star(\varepsilon)} \left[-\frac{\mathbf{x}^\top(\hat{\mathbf{A}}_{\star_\varepsilon}^\top \mathbf{Q}_\varepsilon + \mathbf{Q}_\varepsilon\hat{\mathbf{A}}_{\star_\varepsilon})\mathbf{x}}{\mathbf{x}^\top\mathbf{Q}_\varepsilon\mathbf{x}}\right] \\ &= \frac{1}{2}\lambda_{\min}\left[-(\hat{\mathbf{A}}_{\star_\varepsilon}^\top \mathbf{Q}_\varepsilon + \mathbf{Q}_\varepsilon\hat{\mathbf{A}}_{\star_\varepsilon})\mathbf{Q}_\varepsilon^{-1}\right] \end{aligned} \tag{7.47}$$

Ebenfalls gilt $\underline{\alpha}_\star(\varepsilon) > 0$ da die Matrizen $-(\hat{\mathbf{A}}_{\star_\varepsilon}^\top \mathbf{Q}_\varepsilon + \mathbf{Q}_\varepsilon\hat{\mathbf{A}}_{\star_\varepsilon})$ und $\mathbf{Q}_\varepsilon$ positiv definit sind. Darüber hinaus ist der geschlossene Regelkreis für alle $\mathbf{x} \in \mathcal{G}_\star(\varepsilon)$ linear, sodass die Konvergenzrate des Systems innerhalb des Gebietes $\mathbf{x} \in \mathcal{G}_\star(\varepsilon)$ konstant bleibt. Dies kann man wie folgt veranschaulichen. Jeder Zustand $\mathbf{y} \in \mathcal{G}_\star(\varepsilon)$ kann in der Form $\mathbf{y} = k\mathbf{x}$, mit $k \in (0,1)$ und $\mathbf{x} \in \partial\mathcal{G}_\star(\varepsilon)$ geschrieben werden, und es gilt

$$\begin{aligned}\min_{\mathbf{y}\in\mathcal{G}_\star(\varepsilon)} &- \frac{\mathbf{y}^\top(\hat{\mathbf{A}}_{\star_\varepsilon}^\top \mathbf{Q}_\varepsilon + \mathbf{Q}_\varepsilon\hat{\mathbf{A}}_{\star_\varepsilon})\mathbf{y}}{\mathbf{y}^\top\mathbf{Q}_\varepsilon\mathbf{y}} \\ &= \min_{\substack{\mathbf{x}\in\partial\mathcal{G}_\star(\varepsilon)\\ k\in(0,1]}} -\frac{(k\mathbf{x})^\top(\hat{\mathbf{A}}_{\star_\varepsilon}^\top \mathbf{Q}_\varepsilon + \mathbf{Q}_\varepsilon\hat{\mathbf{A}}_{\star_\varepsilon})(k\mathbf{x})}{(k\mathbf{x})^\top\mathbf{Q}_\varepsilon(k\mathbf{x})} \\ &= \min_{\mathbf{x}\in\partial\mathcal{G}_\star(\varepsilon)} -\frac{\mathbf{x}^\top(\hat{\mathbf{A}}_{\star_\varepsilon}^\top \mathbf{Q}_\varepsilon + \mathbf{Q}_\varepsilon\hat{\mathbf{A}}_{\star_\varepsilon})\mathbf{x}}{\mathbf{x}^\top\mathbf{Q}_\varepsilon\mathbf{x}} = 2\underline{\alpha}_\star(\varepsilon). \end{aligned}$$

[21] Vgl. [66], Fakt 6.52.
[22] Vgl. [8], Fakt 8.3.7.

Es ist dabei ersichtlich, dass aufgrund der Tatsache, dass die Funktion $V_\star(\mathbf{x})$ für $\mathbf{x} \in \partial\mathcal{G}_\star(\varepsilon)$ nicht differenzierbar ist, die Funktion $\alpha_\star(v)$ für $v = \varepsilon$ nicht stetig ist. Gl. (7.46) und (7.47) nehmen im Fall der *klassischen* und der *invers-polynomialen* WSVR besondere Formen ein. Dies wird im Folgenden gezeigt.

### Die *klassische* WSVR mittels iLF

In diesem Fall lautet die Systemmatrix des geschlossenen Regelkreises

$$\hat{\mathbf{A}}_{\Delta_v} = \frac{1}{v^r}\mathbf{D}_v^r\hat{\mathbf{A}}_{\Delta_1}\mathbf{D}_v^{-r}, \quad \hat{\mathbf{A}}_{\Delta_1} := \mathbf{A} - c\mathbf{b}\mathbf{b}^\top\mathbf{P}^{-1},$$

mit $\mathbf{D}_v = \mathrm{diag}(v^n,v^{n-1},\ldots,v)$. Für $\mathbf{Q}_v = \mathbf{D}_v^{-r}(d\cdot\mathbf{P}^{-1})\mathbf{D}_v^{-r}$ folgt, dass für $\mathbf{x} \in \partial\mathcal{G}_\Delta(v) \subseteq \mathcal{G}_\Delta(1)\backslash\mathcal{G}_\Delta(\varepsilon)$ gilt

$$\begin{aligned}\underline{\alpha}_\Delta(v) &= \frac{1}{2v}\lambda_{\min}[\mathbf{D}_v^{-r}(\hat{\mathbf{A}}_{\Delta_1}^\top\mathbf{P}^{-1} + \mathbf{P}^{-1}\hat{\mathbf{A}}_{\Delta_1})(\mathbf{N}\mathbf{P}^{-1} + \mathbf{P}^{-1}\mathbf{N})^{-1}\mathbf{D}_v^r] \\ &= \frac{1}{2v}\lambda_{\min}[(\hat{\mathbf{A}}_{\Delta_1}^\top\mathbf{P}^{-1} + \mathbf{P}^{-1}\hat{\mathbf{A}}_{\Delta_1})(\mathbf{N}\mathbf{P}^{-1} + \mathbf{P}^{-1}\mathbf{N})^{-1}], \quad v \in (\varepsilon,1].\end{aligned} \tag{7.48}$$

Dies beruht auf der Tatsache, dass beide Matrizen *ähnlich* sind, vgl. Def. 18 (Anhang), und folglich gleiche Eigenwerte haben. Aus Gl. (7.48) ist ersichtlich, dass mit kleiner werdendem $v$ die LF-basierte Konvergenzrate $\underline{\alpha}_\Delta(v)$ steigt. Da die Gebiete ineinander verschachtelt sind, folgt auch, dass innerhalb eines Einzugsgebietes $\mathcal{G}_\Delta(v)$ die definitionsgemäß minimale Konvergenzrate auf dem Rand des Gebietes erzielt wird.

Für alle $\mathbf{x} \in \partial\mathcal{G}_\Delta(\varepsilon)$ gilt

$$\begin{aligned}\underline{\alpha}_\Delta(\varepsilon) &= \frac{1}{2}\lambda_{\min}\left[-(\hat{\mathbf{A}}_{\Delta_\varepsilon}^\top\mathbf{Q}_\varepsilon + \mathbf{Q}_\varepsilon\hat{\mathbf{A}}_{\Delta_\varepsilon})\mathbf{Q}_\varepsilon^{-1}\right] \\ &= \frac{1}{2}\lambda_{\min}\left[-(\hat{\mathbf{A}}_{\Delta_1}^\top\mathbf{P}^{-1} + \mathbf{P}^{-1}\hat{\mathbf{A}}_{\Delta_1})\mathbf{P}\right].\end{aligned}$$

Folgender Satz fasst diese Ergebnisse zusammen:

**Satz 7.9** [Konvergenzrate der nicht-sättigenden *klassischen* WSVR mittels iLF] *Für den geschlossenen Regelkreis aus Punkt i) des Satzes 3.1*

*gilt:*

$$(a)\ \underline{\alpha}_{\Delta}(v)=\begin{cases}\frac{1}{2v}\lambda_{\min}[(\hat{\mathbf{A}}_{\Delta_1}^{\top}\mathbf{P}^{-1}+\mathbf{P}^{-1}\hat{\mathbf{A}}_{\Delta_1})(\mathbf{N}\mathbf{P}^{-1}+\mathbf{P}^{-1}\mathbf{N})^{-1}], & v\in(\varepsilon,1],\\ \frac{1}{2}\lambda_{\min}\left[-(\hat{\mathbf{A}}_{\Delta_1}^{\top}\mathbf{P}^{-1}+\mathbf{P}^{-1}\hat{\mathbf{A}}_{\Delta_1})\mathbf{P}\right], & v=\varepsilon,\end{cases}$$

$(b)$ $\underline{\alpha}_{\Delta}(v)$ *ist streng monoton steigend mit sinkendem* $v$.

**Die *invers-polynomiale* WSVR**

In diesem Fall lautet die Systemmatrix des geschlossenen Regelkreises

$$\hat{\mathbf{A}}_{\mathrm{P}_v} := \mathbf{A} - \mathbf{b}\mathbf{b}^{\top}\mathbf{P}_v^{-1}$$

wobei

$$\mathbf{P}_v = \sum_{i=\mathrm{m_l}}^{\mathrm{m_u}} v^i \mathbf{P}_{c_i}.$$

Mit $\mathbf{Q}_v = d\cdot\mathbf{P}_v^{-1}$ ergibt sich für $\mathbf{x}\in\partial\mathcal{G}_{\mathrm{P}}(v)\subseteq\mathcal{G}_{\mathrm{P}}(1)\backslash\mathcal{G}_{\mathrm{P}}(\varepsilon)$

$$\begin{aligned}\underline{\alpha}_{\mathrm{P}}(v) &= \frac{1}{2v}\lambda_{\min}[(\hat{\mathbf{A}}_{\mathrm{P}_v}^{\top}\mathbf{Q}_v+\mathbf{Q}_v\hat{\mathbf{A}}_{\mathrm{P}_v})(\mathbf{Q}_v')^{-1}]\\ &= \frac{1}{2v}\lambda_{\min}[(\hat{\mathbf{A}}_{\mathrm{P}_v}^{\top}\mathbf{P}_v^{-1}+\mathbf{P}_v^{-1}\hat{\mathbf{A}}_{\mathrm{P}_v})\mathbf{P}_v(\mathbf{P}_v')^{-1}\mathbf{P}_v]\\ &= \frac{1}{2v}\lambda_{\min}[\mathbf{P}_v^{-1}(\mathbf{P}_v\hat{\mathbf{A}}_{\mathrm{P}_v}^{\top}\mathbf{P}_v^{-1}+\hat{\mathbf{A}}_{\mathrm{P}_v})\mathbf{P}_v(\mathbf{P}_v')^{-1}\mathbf{P}_v]\\ &= \frac{1}{2v}\lambda_{\min}[(\mathbf{P}_v\hat{\mathbf{A}}_{\mathrm{P}_v}^{\top}\mathbf{P}_v^{-1}+\hat{\mathbf{A}}_{\mathrm{P}_v})\mathbf{P}_v(\mathbf{P}_v')^{-1}]\\ &= \frac{1}{2v}\lambda_{\min}[(\mathbf{P}_v\hat{\mathbf{A}}_{\mathrm{P}_v}^{\top}+\hat{\mathbf{A}}_{\mathrm{P}_v}\mathbf{P}_v)(\mathbf{P}_v')^{-1}].\end{aligned}$$

Für $\mathbf{x}\in\mathcal{G}_{\mathrm{P}}(\varepsilon)$ ergibt sich

$$\begin{aligned}\underline{\alpha}_{\mathrm{P}}(\varepsilon) &= \frac{1}{2}\lambda_{\min}\left[-(\hat{\mathbf{A}}_{\mathrm{P}_\varepsilon}^{\top}\mathbf{Q}_\varepsilon+\mathbf{Q}_\varepsilon\hat{\mathbf{A}}_{\mathrm{P}_\varepsilon})\mathbf{Q}_\varepsilon^{-1}\right]\\ &= \frac{1}{2}\lambda_{\min}\left[-(\hat{\mathbf{A}}_{\mathrm{P}_\varepsilon}^{\top}\mathbf{P}_\varepsilon^{-1}+\mathbf{P}_\varepsilon^{-1}\hat{\mathbf{A}}_{\mathrm{P}_\varepsilon})\mathbf{P}_\varepsilon\right].\end{aligned}$$

Folgender Satz fasst die Ergebnisse zusammen:

**Satz 7.10** [Konvergenzrate der nicht-sättigenden *invers-polynomialen*-WSVR] *Für den geschlossenen Regelkreis aus Punkt i) des Satzes 4.1,*

*Seite 28, gilt:*

$$\underline{\alpha}_{\mathrm{P}}(v) = \begin{cases} \frac{1}{2v}\lambda_{\min}[(\mathbf{P}_v\hat{\mathbf{A}}_{\mathrm{P}_v}^{\top} + \hat{\mathbf{A}}_{\mathrm{P}_v}\mathbf{P}_v)(\mathbf{P}_v')^{-1}], & v \in (\varepsilon,1], \\ \frac{1}{2}\lambda_{\min}\left[-(\hat{\mathbf{A}}_{\mathrm{P}_\varepsilon}^{\top}\mathbf{P}_\varepsilon^{-1} + \mathbf{P}_\varepsilon^{-1}\hat{\mathbf{A}}_{\mathrm{P}_\varepsilon})\mathbf{P}_\varepsilon\right], & v = \varepsilon, \end{cases}$$

### Konvergenzoptimale WSVR

Die Konvergenzrate des Systems aus Punkt *b*) des Satzes 5.1 kann man - wie auch in [37] dargestellt - mit Hilfe des Lagrange-Multiplikatoren-Ansatzes analysieren. Untersucht wird der kleinste Abklingfaktor der Funktion

$$V_{\mathrm{s}}(\mathbf{x}) := \begin{cases} v(\mathbf{x}), \text{ mit } g_{\mathrm{s}}(\mathbf{x},v) = 0, & \mathbf{x} \in \mathcal{G}_{\mathrm{s}}(1)\backslash\mathcal{G}_{\mathrm{s}}(\varepsilon), \\ \varepsilon\mathbf{x}^{\top}\mathbf{Q}_\varepsilon\mathbf{x}, & \mathbf{x} \in \mathcal{G}_{\mathrm{s}}(\varepsilon), \end{cases} \tag{7.49}$$

entlang der Trajektorien des Systems $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} - \mathbf{b}\,\mathrm{sgn}(\mathbf{b}^{\top}\mathbf{Q}_v\mathbf{x})$. Da die Funktion $V_{\mathrm{s}}(\mathbf{x})$ gerade ist, d.h. $V_{\mathrm{s}}(\mathbf{x}) = V_{\mathrm{s}}(-\mathbf{x})$, reicht es aus die Zustandspunkte zu betrachten, wofür $\mathbf{x}^{\top}\mathbf{Q}_v\mathbf{b} \geq 0$ gilt. Für $\mathbf{x} \in \partial\mathcal{G}_{\mathrm{s}}(v) \subseteq \mathcal{G}_{\mathrm{s}}(1)\backslash\mathcal{G}_{\mathrm{s}}(\varepsilon)$ ergibt sich

$$\begin{aligned} 2\underline{\alpha}_{\mathrm{s}}(v) &:= \min_{\mathbf{x}\in\partial\mathcal{G}_{\mathrm{s}}(v)} \left(-\frac{\dot{V}_{\mathrm{s}}(\mathbf{x})}{V_{\mathrm{s}}(\mathbf{x})}\right) \\ &= \frac{1}{v}\min_{\mathbf{x}\in\partial\mathcal{G}_{\mathrm{s}}(v)} \left(-\dot{V}_{\mathrm{s}}(\mathbf{x})\right) \\ &= \frac{1}{v}\min_{\mathbf{x}\in\partial\mathcal{G}_{\mathrm{s}}(v)} \frac{\mathbf{x}^{\top}(\mathbf{A}^{\top}\mathbf{Q}_v + \mathbf{Q}_v\mathbf{A})\mathbf{x} - 2\mathbf{b}^{\top}\mathbf{Q}_v\mathbf{x}}{\mathbf{x}^{\top}\mathbf{Q}_v'\mathbf{x}}. \end{aligned} \tag{7.50}$$

Schließlich ergibt sich für jeden Zustand $\mathbf{x} \in \partial\mathcal{G}_{\mathrm{s}}(\varepsilon)$

$$\begin{aligned} 2\underline{\alpha}_{\mathrm{s}}(\varepsilon) &:= \min_{\mathbf{x}\in\partial\mathcal{G}_{\mathrm{s}}(\varepsilon)} \left(-\frac{\dot{V}_{\mathrm{s}}(\mathbf{x})}{V_{\mathrm{s}}(\mathbf{x})}\right) \\ &= \min_{\mathbf{x}\in\partial\mathcal{G}_{\mathrm{s}}(\varepsilon)} \left[-\frac{\mathbf{x}^{\top}(\mathbf{A}^{\top}\mathbf{Q}_\varepsilon + \mathbf{Q}_\varepsilon\mathbf{A})\mathbf{x} - 2\mathbf{b}^{\top}\mathbf{Q}_\varepsilon\mathbf{x}}{\mathbf{x}^{\top}\mathbf{Q}_\varepsilon\mathbf{x}}\right] \\ &= \min_{\mathbf{x}\in\partial\mathcal{G}_{\mathrm{s}}(\varepsilon)} \left[-\mathbf{x}^{\top}(\mathbf{A}^{\top}\mathbf{Q}_\varepsilon + \mathbf{Q}_\varepsilon\mathbf{A})\mathbf{x} - 2\mathbf{b}^{\top}\mathbf{Q}_\varepsilon\mathbf{x}\right]. \end{aligned} \tag{7.51}$$

Aufgrund der Komplexität der Bestimmungsfunktion, aber auch der komplexeren Form der Matrix $\mathbf{Q}_v$ lässt sich die untere Grenze $\underline{\alpha}_{\mathrm{s}}(v)$ nicht

mehr analytisch berechnen. Diese wird nicht mehr exakt gelöst sondern interpoliert. Dies geschieht im nächsten Kapitel.

Zusammenfassend lässt sich feststellen, dass unter den Gütemaßen für nichtlineare Regelkreise die Konvergenzrate eins der vorteilhaftesten ist. Dies liegt an seiner möglichen Verwendung zur Approximation der Güte des Ausregelverhaltens aber auch zur Optimierung des Ausregelverhaltens. Letztere basiert auf einer unteren Grenze der Konvergenzrate, die im Falle der hier analysierten nichtlinearen Regelungsmethoden mit Hilfe einer impliziten Ljapunov-Funktion angegeben werden konnte. Im nächsten Kapitel wird eine Methode zur Performance-Analyse vorgestellt, welche die bereits existierende Theorie über den Design und Analyse von *Computerexperimenten* anwendet. Diese analysiert die Performance einer (nichtlinearen) Regelungsmethode in Ensembles von nichtlinearen Regelkreisen.

# 8 *Computerexperimente* unter Einsatz Bayes'scher Methoden

*Computerexperimente* bilden neben physikalischen Experimenten (Versuche an elektromechanischen Versuchsständen, klinische Versuche, landwirtschaftliche Feldversuche etc.) eine Methode zur Generierung von Beobachtungen über die Eigenschaften eines Versuchsobjekts infolge der Variation verschiedener Faktoren, vgl. z.B. [27, 64]. Diese Faktoren sind die Eingangsvariablen, und die Eigenschaften des Versuchsobjekts sind die Ausgangsvariablen des Experiments. Im Fall von *Computerexperimenten* wird der Zusammenhang zwischen den Eingangs- und den Ausgangsvariablen in Form eines Rechnercodes basierend auf einem mathematischen Modell beschrieben, dessen Komplexität im Allgemeinen sehr hoch ist. In vielen Fällen wären die entsprechenden physikalischen Experimente sogar nicht durchführbar - z.B. aus ethischen, ökonomischen oder zeitlichen Gründen - sodass das *Computerexperiment* die einzige Alternative bietet. Unter Verwendung numerischer Verfahren können dabei simulierte Beobachtungen generiert werden, welche für eine Prädiktion des Verhaltens des Versuchsobjekts verwendet werden können. Beispiele technologischer und wissenschaftlicher Entwicklungen basierend auf *Computerexperimenten* beinhalten die Untersuchung des Verhaltens von Fusionsreaktoren, künstlichen Prothesen, integrierten Schaltungen, thermischen Energiespeichern und vielen anderen Objekten aus nahezu allen natur- und ingenieurwissenschaftlichen Bereichen.

In dieser Arbeit wird zum ersten Mal das theoretische Konzept der *Computerexperimente* auf die Performance-Analyse von Regelmethoden übertragen. Das Versuchsobjekt ist ein Ensemble nichtlinearer Regelkreise, dessen Eigenschaften infolge der Variation eines oder mehrerer Faktoren analysiert werden. Bild 8.1 zeigt das untersuchte Regelkreisensemble. Wie man im Bild sehen kann, beeinflussen die Faktoren $\boldsymbol{\zeta}$, welche beliebige Werte aus einer kompakten Menge annehmen können, die Dynamik der Regelstrecke und des Reglers. Das *Computerexperiment* besteht aus mehreren Versuchen mit jeweils verschiedenen Faktoren. In jedem Versuch werden irgendwelche Faktorenwerte $\boldsymbol{\zeta}$ vorgegeben und für die erzeugte Re-

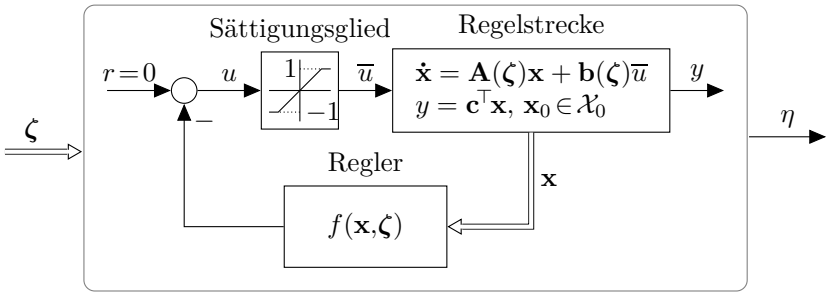


**Bild 8.1:** Aufbau eines *Computerexperiments*.

gelstrecke automatisch ein Regler entworfen. Die erzielte Performance im geschlossenen Regelkreis bildet eine *Beobachtung*. Es handelt sich hier um ein deterministisches Ergebnis, d.h. zwei Versuche mit gleichen Faktoren bestimmen gleiche *Beobachtungen*. Anschließend kann anhand der *Beobachtungen* eine Performanceprädiktion für alle restlichen Faktorenwerte gemacht werden. Ziel kann also die Prädiktion der Performance einer Regelmethode für eine neue Strecke aus dem gegebenen Regelstreckenensemble sein, ohne dabei einen Regler entwerfen zu müssen. Dabei wird auch eine Quantifizierung der **erwarteten** Prädiktionssicherheit von Interesse sein. Weitere Ziele können sein, den Einfluss verschiedener Faktoren auf die Performance der Regelmethode zu analysieren oder die erwartete Performance für eine zufällig gewählte Strecke aus einem Streckenensemble zu bestimmen.

Die Faktoren sind in einem Vektor $\boldsymbol{\zeta} := [\boldsymbol{\zeta}_k \quad \boldsymbol{\zeta}_e \quad \boldsymbol{\zeta}_m]^\top$ gruppiert. Die Einteilung der Eingangsvariablen spiegelt die allgemeine Klassifikation im Rahmen von *Computerexperimenten* wider, vgl. [64]. Diese werden in drei große Kategorien aufgeteilt: die *Kontrollvariablen* ($\boldsymbol{\zeta}_k = [\zeta_{k_1} \ \cdots \ \zeta_{k_l}]^\top$), welche vom Experimentator gezielt geändert werden können, wie z.B. die Spezifikationen eines eletromechanischen Aufbaus (z.B. die Motorspezifikation für die Laufkatze bei einer Verladebrücke), die *Umgebungsvariablen* ($\boldsymbol{\zeta}_e := [\zeta_{e_1} \ \cdots \ \zeta_{e_p}]^\top$), manchmal auch *Rauschvariablen* genannt, welche stochastischer Natur und unbeeinflußbar sind, und *Modellvariablen* ($\boldsymbol{\zeta}_m := [\zeta_{m_1} \ \cdots \ \zeta_{m_c}]^\top$), welche Unsicherheiten in der mathematischen Modellierung beschreiben und entweder unbekannt oder stochastischer Natur sind.

Je nach Typ der Eingangsvariablen werden verschiedene Ziele des Experiments formuliert. Es wird zwischen einem homogenen Eingang (Vorhandensein von nur einem Typ von Eingangsvariablen) und einem gemischten Eingang (Vorhandensein von zwei Typen von Eingangsvariablen) unterschieden. In dieser Arbeit wird darüber hinaus zwischen dem (einfacheren) eindimensionalen (eine einzige Eingangsvariable) und dem mehrdimensionalen Fall (mehrere Eingangsvariablen) unterschieden. Darauf basierend können folgende allgemeine Probleme behandelt werden:

- **Prädiktion** des Ausgangs $\eta = h(\boldsymbol{\zeta})$ in einem Bereich $\boldsymbol{\zeta} \in \mathcal{D}_{\boldsymbol{\zeta}}$, wobei der Zusammenhang zwischen Eingangs- und Ausgangsvariablen durch die unbekannte Funktion $\eta = h(\boldsymbol{\zeta})$ beschrieben wird, welche als Realisierung des skalaren Zufallsfeldes $\mathcal{H}(\boldsymbol{\zeta})$ angenommen wird,
- **Optimierung** - d.h. Bestimmung des Eingangs $\boldsymbol{\zeta}$, welcher einen optimalen Ausgangwert $\eta$ erzielt,
- **Nullstellensuche** - d.h. Bestimmung eines Eingangs $\boldsymbol{\zeta} \in \mathcal{L}(\eta_0)$ für ein vorgegebenes Niveau des Ausgangs $\mathcal{L}(\eta_0) := \{\boldsymbol{\zeta} \in \mathcal{D}_{\boldsymbol{\zeta}} | h(\boldsymbol{\zeta}) = \eta_0\}$,
- **Unsicherheitsanalyse** - d.h. Bestimmung wesentlicher und unwesentlicher Eingänge $\boldsymbol{\zeta} \in \mathcal{D}_{\boldsymbol{\zeta}}$ für die Variation des Ausgangs $\eta = h(\boldsymbol{\zeta})$,
- **Sensitivitätsanalyse**, als Verallgemeinerung der Unsicherheitsanalyse - d.h. Bestimmung der Art, wie Unsicherheit über den Eingang $\boldsymbol{\zeta}$ den Ausgang des Systems $\eta$ beeinflusst,
- **Integration des Ausgangs** - d.h. die Bestimmung des Erwartungswertes des skalaren Zufallfeldes $\mathcal{H}(\boldsymbol{\zeta})$, wobei der Eingang $\boldsymbol{\zeta}$ als Realisierung eines Zufallsvektors mit einer bestimmten Verteilung $\mathcal{F}$, d.h. $\mathbf{Z} \sim \mathcal{F}$, betrachtet wird.
- **Kalibrierung** - d.h. Anpassung verschiedener Modellvariablen $\boldsymbol{\zeta}_m$, sodass die beobachteten Ausgangswerte eines physikalischen Experiments den Ausgangswerten des Modells eines Rechnercodes entsprechen.

Tabelle 8.1 zeigt eine typische Klassifizierung verschiedener Ziele in Abhängigkeit der Eingangsvariablen, vgl. [64].

Im Rahmen der hier untersuchten Performance-Analyse von Regelmethoden gruppiert der Eingangsvektor mehrere Parameter einer Regelstrecke. Alle möglichen Werte des Parametervektors aus einer vorgegebenen

**Tabelle 8.1:** Zielklassifizierung je nach Art der Eingangsvariablen.

| Ziel | Kontroll- | Umgebungs- | Modellvar. |
|---|---|---|---|
| Prädiktion, Optimierung, Nullstellensuche | x<br>x | x | |
| Kalibrierung | | | x |
| Unsicherheits- und Sensitivitätsanalyse<br>Integration des Ausgangs | | x | |

(kompakten) Menge generieren zusammen ein Regelstreckenensemble. Wie bereits erwähnt, umfasst das *Computerexperiment* eine Serie von Versuchen mit verschiedenen Parametervektoren. In jedem Versuch wird eine neue Strecke durch einen bestimmten Wert des Parametervektors $\boldsymbol{\zeta}$ (Design-Punkt) generiert und dafür automatisch ein Regler $u = -f(\mathbf{x},\boldsymbol{\zeta})$ entworfen. Der Wert eines Performance-Maßes des jeweiligen Regelkreises bildet dabei den Ausgang $\eta = h(\boldsymbol{\zeta})$. Ein Performance-Maß ist z.B.

- der Fehlklassifikationsanteil einer zeitsuboptimalen Regelung mit Schaltfunktion aus Gl. (7.13)
- die relative *Einschwingzeit* $J_{\mathrm{t_a}}$ aus Gl. (7.14),
- die gebietsabhängige Konvergenzrate des Systems aus Gl. (7.50)- (7.51).

Jedes untersuchte Performance-Maß erzielt für einen geschlossenen Regelkreis mit dem Eingangsvektor $\boldsymbol{\zeta}_i$ einen skalaren Wert $\eta_i := h(\boldsymbol{\zeta}_i)$. Die skalare Funktion $h(\,\cdot\,)$ ist dabei unbekannt, auch wenn es sich um eine deterministische Funktion handelt, da jeder Funktionswert erst durch die Ausführung des Rechnercodes bekannt ist. Eine Serie von $N$ Versuchen ergibt entsprechend eine Menge von $N$ Werten des jeweiligen Performance-Maßes. Die Werte $\boldsymbol{\zeta}_i$, mit $i = 1,\ldots,N$, werden *Design-Punkte* und die Werte $\eta_i$, mit $i = 1,\ldots,N$, *Trainingsdaten* genannt. Letztere werden in einem Vektor $\boldsymbol{\eta} := [\eta_1 \;\; \cdots \;\; \eta_N]^\top$ zusammengefasst. Anhand des erzielten Vektors $\boldsymbol{\eta}$ werden die oben genannten Problemstellungen gelöst. Beispielsweise können für die Prädiktion Neuronale Netze, *Splines* oder Prädikto-

ren basierend auf Gauß'schen Zufallsfeldern[1)] verwendet werden. Letztere haben den Vorteil, dass sie auch die erwartete Unsicherheit der Schätzung quantifizieren können. Die Minimierung der erwarteten Unsicherheit kann beispielsweise verwendet werden, um die $N$ Parametervektoren $\boldsymbol{\zeta}_i$, mit $i = 1, \ldots, N$ zu wählen. In dieser Arbeit werden diese Schätzer unter Berücksichtigung der Bayes'schen Methodik verwendet.

Im Fall der Prädiktion wird eine Bayes'sche Interpolationsmethode für deterministische Funktionen, vgl. [21, 22], angewandt, welche in Abschnitt 8.2 beschrieben wird. Diese Methode betrachtet den unbekannten Ausgang des *Computerexperiments* als ein skalares Zufallsfeld $\mathcal{H}(\cdot)$ (meistens *Zufallsfunktion* genannt, vgl. [64, S. 24-25]), dessen Verteilung noch bestimmt werden muss. Als Prädiktionswert für einen neuen Punkt $\boldsymbol{\zeta}_0$ wird der Erwartungswert des skalaren Zufallsfeldes $\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta}_0)|\mathbf{H} = \boldsymbol{\eta}\}$ an dem Punkt $\boldsymbol{\zeta}_0$ verwendet. Die **Optimierung** und **Nullstellensuche** sind mit der vorigen Problemstellung eng verwandt und werden mit Hilfe der prädiktiven Verteilung des Ausgangs berechnet. Die **Sensitivitätsanalyse** wird ebenfalls unter Anwendung des Bayes'schen Ansatzes durchgeführt. Unter Verwendung der *A-posteriori*-Verteilung des skalaren Zufallsfeldes $\mathcal{H}(\boldsymbol{\zeta}_e)$ wird eine Inferenz für verschiedene Sensitivitätsmaße durchgeführt, vgl. [51]. Die in dieser Arbeit vorgestellten Sensitivitätsmaße sind die *Haupteffekte*, die *Interaktionen*, sowie die varianzbasierten *Haupteffekt- und Interaktionenindizes*. Diese und die dafür verwendete Bayes'sche Inferenz werden in Abschnitt 8.3 vorgestellt.

Das Kapitel ist wie folgt gegliedert. Der einleitende Abschnitt 8.1 enthält die Grundidee des Bayes'schen Ansatzes, eine Einführung über Gauß'sche Zufallsfelder und über den hier verwendeten besten linearen erwartungstreuen Prädiktor. Abschnitt 8.2 zeigt den Aufbau von prädiktiven *A-posteriori*-Verteilungen über den partiellen und den vollständigen Bayes'schen Ansatz, das Design-Problem, die Maße zur Analyse der Prädiktionsgenauigkeit, sowie ein Beispiel einer Funktion mit einer Variablen. Abschnitt 8.3 zeigt die Sensitivitätsanalyse im Rahmen von *Computerexperimenten* und die jeweilige Bayes'sche Inferenz. Der Abschnitt endet mit einem Beispiel einer Funktion mit zwei Variablen und der Veranschaulichung der Prädiktion, sowie der Sensitivitätsanalyse. Der letzte Abschnitt, Abschnitt 8.4, wendet die oben vorgestellte Theorie über das Design von *Computerexpe-*

[1)] Wie im nächsten Abschnitt beschrieben wird, stellen Zufallsfelder (*engl.*: *random fields*, vgl. [6]) eine Verallgemeinerung stochastischer Prozesse beliebiger Dimension dar, wobei der Parameter *Zeit* durch einen beliebigen anderen Parameter ersetzt werden kann.

*rimenten* auf die Performance-Analyse von nichtlinearen Regelungsmethoden an. Der Abschnitt enthält das Prädiktionsbeispiel einer nicht-simulierten Strecke, ein Beispiel für die Sensitivitätsanalyse und einen empirischen Vergleich zwischen den Prädiktoren.

# 8.1 Vorbemerkungen

## 8.1.1 Notationen

In diesem Kapitel bezeichnen $\mathcal{H}(\cdot)$, $\mathcal{Z}(\cdot)$ und $\boldsymbol{\mathcal{C}}(\cdot)$ skalare bzw. mehrdimensionale Zufallsfelder mit den Realisierungen (auch Pfade genannt) $h(\cdot)$, $z(\cdot)$ bzw. $\mathbf{c}(\cdot)$; $H$ und $\mathbf{H}$ bezeichnen Zufallsvariablen bzw. Zufallsvektoren mit den Realisierungen $\eta$ bzw. $\boldsymbol{\eta}$; $[\cdot]$ bezeichnet die Wahrscheinlichkeitsverteilung (oder kurz Verteilung) einer Zufallsvariable, die durch ihre Wahrscheinlichkeitsdichte (oder kurz Dichte) gegeben ist. Die Multiplikation $[X]\cdot[Y]$ bezeichnet dabei die Multiplikation von zwei Wahrscheinlichkeitsdichten $f_X(x)\cdot f_Y(y)$. Das Symbol $\propto$ bezeichnet proportionale Verteilungen, d.h. Verteilungen, welche sich voneinander nur durch eine Konstante unterscheiden. Schließlich bezeichnet $\mathcal{D}_{\boldsymbol{\zeta}_x}$ den Definitionsbereich der Variable $\boldsymbol{\zeta}_x$.

## 8.1.2 Grundidee des Bayes'schen Ansatzes

Im untersuchten Bereich $\boldsymbol{\zeta} \in \mathcal{D}_{\boldsymbol{\zeta}}$ bildet $h(\cdot) : \mathcal{D}_{\boldsymbol{\zeta}} \to \mathbb{R}$ eine (deterministische) reellwertige Funktion. Diese ist als unbekannt angenommen, obwohl die Funktionswerte anhand eines bekannten Rechnercodes generiert werden. Da die Komponenten des Eingangsvektors $\boldsymbol{\zeta}$ kontinuierliche Größen sind, wäre die Funktion erst durch unendlich viele Simulationen des Rechnercodes vollständig determiniert. In dieser Arbeit wird die Funktion $h(\cdot)$ als Realisierung eines skalaren Zufallsfeldes betrachtet, dessen Verteilung noch bestimmt werden muss. Das skalare Zufallsfeld - auch *Zufallsfunktion* genannt - wird mit $\mathcal{H}(\cdot)$ bezeichnet, um es von seiner Realisierung $h(\cdot)$ zu unterscheiden. Im Rahmen des Bayes'schen Ansatzes wird eine Verteilung für die Zufallsfunktion $\mathcal{H}(\cdot)$ formuliert - die sogenannte *A-priori*-Verteilung - welche anhand der *Trainingsdaten* $\boldsymbol{\eta}$ aktualisiert wird. Letztere wird *A-posteriori*-Wahrscheinlichkeitsdichte genannt und wird verwendet, um Inferenzen über die unbekannte Funktion $h(\cdot)$ (das gewählte Performance-

Maß für die Regelmethode) oder über verschiedene Sensitivitätsmaße zu machen. Dies wird in den nächsten Abschnitten dieses Kapitels gezeigt.

Die Wahl der *A-priori*-Verteilung der Zufallsfunktion basiert auf den vorhandenen Informationen des Experimentators über die unbekannte Funktion $h(\cdot)$. Wie in [50] erläutert, kann man vorerst allgemeine Fragen über $h(\cdot)$ stellen, wie z.B.

1. Ist die Funktion $h(\cdot)$ stetig im gesamten Bereich $\boldsymbol{\zeta} \in \mathcal{D}_{\boldsymbol{\zeta}}$?
2. Kann das Wissen über ein bestimmtes $h(\boldsymbol{\zeta}_1)$ Informationen über $h(\boldsymbol{\zeta}_2)$ liefern, wenn $\boldsymbol{\zeta}_2$ nah an $\boldsymbol{\zeta}_1$ ist?

Können beide Fragen mit 'ja' beantwortet werden, so kann als *A-priori*-Verteilung beispielsweise ein skalares Gauß'sches Zufallsfeld (Gauß'sche Zufallsfunktion) verwendet werden. In vielen Fällen wird diese Wahl aufgrund der Flexibilität des Gauß'schen Zufallsfeldes getroffen. Gauß'sche Zufallsfelder werden kurz in Abschnitt 8.1.3 eingeführt. Sie sind die meist verwendeten Modelle im Rahmen von *Computerexperimenten*. Dies bedeutet aber nicht, dass ein skalares Gauß'sches Zufallsfeld die Unsicherheit über die unbekannte Funktion perfekt modellieren kann. Vielmehr können die Parameter des Gauß'schen Zufallsfeldes an die schon vorhandenen Informationen angepasst werden, und man spricht in diesem Fall von einer ***angepassten*** *A-priori*-Verteilung.

### 8.1.3 Skalare Gauß'sche Zufallsfelder

Ein skalares Gauß'sches Zufallsfeld (GRF, ***G**aussian **R**andom **F**ield*) ist wie folgt definiert:

**Definition 4** [Gauß'sches Zufallsfeld, [64, S.27]] *Gegeben sei die Menge $\mathcal{D}_{\boldsymbol{\zeta}} \subset \mathbb{R}^m$ mit einem positiven m-dimensionalen Volumen. Das Zufallsfeld $\mathcal{H}(\boldsymbol{\zeta})$, mit $\boldsymbol{\zeta} \in \mathcal{D}_{\boldsymbol{\zeta}}$, heißt ein* Gauß'sches Zufallsfeld *falls für $N \geq 1$ und jede Wahl von $\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_N$ aus $\mathcal{D}_{\boldsymbol{\zeta}}$, der Zufallsvektor $(\mathcal{H}(\boldsymbol{\zeta}_1), \ldots, \mathcal{H}(\boldsymbol{\zeta}_N))$ eine multivariate Gauß'sche Verteilung hat.*

Die im Rahmen von *Computerexperimenten* verwendeten GRFs haben darüber hinaus folgende Eigenschaften:

- **Nichtsingularität**: Das skalare Zufallsfeld $\mathcal{H}(\boldsymbol{\zeta})$ heißt *nichtsingulär* falls für jedes $N \geq 1$ die mit jeder Wahl von Eingangsvektoren $\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_N$ verbundene Kovarianzmatrix der multivariaten

Gauß'schen Verteilung von $(\mathcal{H}(\boldsymbol{\zeta}_1),\ldots,\mathcal{H}(\boldsymbol{\zeta}_N))$ nichtsingulär ist. Diese Eigenschaft ist u.a. notwendig für den Aufbau des Prädiktors.

- **Trennbarkeit**: Diese Eigenschaft[2)] stellt sicher, dass die endlich dimensionale Verteilung des Zufallsvektors $(\mathcal{H}(\boldsymbol{\zeta}_1),\ldots,\mathcal{H}(\boldsymbol{\zeta}_N))$ die Eigenschaften einer Realisierung $h(\boldsymbol{\zeta})$, wie beispielsweise deren Stetigkeit und Differenzierbarkeit, bestimmt.

- **Starke/schwache Stationarität und Ergodizität**: Wie im Fall von Zeitreihen, verfügt man bei den hier analysierten Zufallsfeldern über jeweils nur eine Beobachtung $h(\boldsymbol{\zeta})$ für ein bestimmtes $\boldsymbol{\zeta} \in \mathcal{D}_{\boldsymbol{\zeta}}$. Ein zweiter Versuch mit demselben Parametervektor $\boldsymbol{\zeta} \in \mathcal{D}_{\boldsymbol{\zeta}}$ würde die gleiche Beobachtung $h(\boldsymbol{\zeta})$ generieren. Man verfügt also nicht über eine Stichprobe von Werten. Um eine Prädiktion des (unbekannten) Wertes $h(\boldsymbol{\zeta}_{\text{neu}})$ auf Basis der Erwartungswertfunktion $\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}$ machen zu können, muss das Zufallsfeld *stark stationär* und *ergodisch* sein, vgl. z.B. [64, Abschnitt 2.3.2].
  - **starke Stationarität**: Das Zufallsfeld $\mathcal{H}(\boldsymbol{\zeta})$ heißt *stark stationär* (oder einfach *stationär*) falls für jedes $\mathbf{d} \in \mathbb{R}^m$ und $N \geq 1$, und alle $\boldsymbol{\zeta}_i \in \mathcal{D}_{\boldsymbol{\zeta}}$, $i = 1,\ldots,N$, sodass $\boldsymbol{\zeta}_i + \mathbf{d} \in \mathcal{D}_{\boldsymbol{\zeta}}$, $i = 1,\ldots,N$, die Zufallsvektoren $(\mathcal{H}(\boldsymbol{\zeta}_1),\ldots,\mathcal{H}(\boldsymbol{\zeta}_N))$ und $(\mathcal{H}(\boldsymbol{\zeta}_1 + \mathbf{d}),\ldots,\mathcal{H}(\boldsymbol{\zeta}_N + \mathbf{d}))$ die **gleiche** Verteilung besitzen. Dies gilt auch für $N = 1$, d.h. alle Zufallsvariablen $\mathcal{H}(\boldsymbol{\zeta})$, mit $\boldsymbol{\zeta} \in \mathcal{D}_{\boldsymbol{\zeta}}$, haben die gleiche Verteilung. Folglich haben sie einen konstanten Erwartungswert und eine konstante Varianz.
  - **schwache Stationarität**: Das Zufallsfeld $\mathcal{H}(\boldsymbol{\zeta})$ heißt *schwach stationär* falls

$$\mu(\boldsymbol{\zeta}) := \mathrm{E}\{\mathcal{H}(\zeta)\} = \mu, \text{ und} \tag{8.1}$$

$$c(\boldsymbol{\zeta}_1,\boldsymbol{\zeta}_2) := \mathrm{Cov}\{\mathcal{H}(\boldsymbol{\zeta}_1),\mathcal{H}(\boldsymbol{\zeta}_2)\} = C(\boldsymbol{\zeta}_1 - \boldsymbol{\zeta}_2) \tag{8.2}$$

gilt, wobei $C(\boldsymbol{\zeta}_1 - \boldsymbol{\zeta}_2)$ *Kovarianzfunktion* des Zufallsfeldes genannt wird. Gl. (8.1) bedeutet, dass die Erwartungswertfunktion des skalaren Zufallsfeldes für alle $\boldsymbol{\zeta} \in \mathcal{D}_{\boldsymbol{\zeta}}$ konstant ist. Gl. (8.2) bedeutet, dass die Beobachtungen generiert von zwei Parametervektoren mit dem gleichen *Abstand* und gleiche *Orientierung* den gleichen Autokovarianzfunktionswert haben. *Star-*

[2)] Eine formale Definition der Trennbarkeit wird in [6] angegeben.

*ke Stationarität* impliziert *schwache Stationarität*. Der Umkehrschluß ist nicht notwendigerweise wahr.

– **Ergodizität**: Ein schwach stationärer Zufallsprozess $\mathcal{H}(\zeta)$ mit der konstanten Erwartungswertfunktion $\mathrm{E}\{\mathcal{H}(\zeta)\} = \mu$ heißt *ergodisch* falls

$$\lim_{N\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}\mathcal{H}(\boldsymbol{\zeta}_i)\right) = \mu$$

gilt. Dies bedeutet, dass je mehr Beobachtungen verwendet werden, desto besser wird die Schäzung des Erwartungswertes des Zufallsfeldes. Eine wichtige Voraussetzung dafür ist, dass $\lim_{\mathbf{d}\to\infty} C(\mathbf{d}) = 0$. Das bedeutet, dass weit auseinander liegende Beobachtungen nicht zu stark zusammenhängen dürfen.

- **Isotropie** Ein stark-stationäres Zufallsfeld $\mathcal{H}(\boldsymbol{\zeta})$ heißt *isotropisch* falls

$$\mathrm{Cov}\{H(\boldsymbol{\zeta}_1),H(\boldsymbol{\zeta}_2)\} = C(\|\boldsymbol{\zeta}_1 - \boldsymbol{\zeta}_2\|) \tag{8.3}$$

gilt. Gl. (8.3) bedeutet, dass die Beobachtungen generiert von zwei Parametervektoren mit dem gleichen *Abstand* (unabhängig von der *Orientierung*) den gleichen Autokovarianzfunktionswert haben.

Die Voraussetzung der Stationarität des Zufallsfeldes $\mathcal{H}(\boldsymbol{\zeta})$ kann durch ein Modell der Form

$$\mathcal{H}(\boldsymbol{\zeta}) = \mathbf{f}(\boldsymbol{\zeta})^{\top}\boldsymbol{\beta} + \mathcal{Z}(\boldsymbol{\zeta}),$$

umgangen werden, wobei $\mathbf{f}(\boldsymbol{\zeta})$ bekannte, bzw. vorher festgelegte, Regressionsfunktionen, $\boldsymbol{\beta} := [\beta_1,\ldots,\beta_p]^{\top}$ ein Vektor von unbekannten Regressionskoeffizienten, und $\mathcal{Z}(\boldsymbol{\zeta})$ ein mittelwertfreies skalares Zufallsfeld mit der konstanten Varianz $\mathrm{Var}\{\mathcal{Z}(\boldsymbol{\zeta})\} = \sigma^2_{\mathcal{Z}}$ darstellen, welches alle obigen Eigenschaften besitzt. Bei einer solchen Modellierung ist $\mathcal{H}(\boldsymbol{\zeta})$ nicht mehr stationär. Weitere Modellierungen, die die Flexibilität des Zufallsfeldes anstreben, werden in [64] diskutiert.

Ein skalares Gauß'sches Zufallsfeld ist durch dessen Erwartungswertfunktion $\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}$ und Autokovarianz-Funktion $\mathrm{Cov}\{\mathcal{H}(\boldsymbol{\zeta}_1),\mathcal{H}(\boldsymbol{\zeta}_2)\}$ vollständig spezifiziert. Alternativ zur Autokovarianz-Funktion wird die *Korrelationsfunktion* spezifiziert. Diese ist definiert als

$$R(\mathbf{d}) := \frac{\mathrm{Cov}\{\mathcal{H}(\boldsymbol{\zeta}_1),\mathcal{H}(\boldsymbol{\zeta}_2)\}}{\sqrt{\mathrm{Var}\{\mathcal{H}(\boldsymbol{\zeta}_1)\}\cdot\mathrm{Var}\{\mathcal{H}(\boldsymbol{\zeta}_2)\}}} = \frac{C(\mathbf{d})}{\sigma^2_{\mathcal{Z}}}, \quad \mathbf{d}\in\mathbb{R}^m,$$

wobei $\mathrm{Var}\{\mathcal{H}(\boldsymbol{\zeta}_1)\} = \mathrm{Var}\{\mathcal{H}(\boldsymbol{\zeta}_2)\} = \mathrm{Cov}\{\mathcal{H}(\boldsymbol{\zeta}),\mathcal{H}(\boldsymbol{\zeta})\} = C(\mathbf{0}) =: \sigma_{\mathcal{Z}}^2$ die konstante Varianz des stationären skalaren Zufallsfeldes $\mathcal{H}(\boldsymbol{\zeta})$ ist. Die Korrelationsfunktion $R(\mathbf{d})$, mit $R(\mathbf{0}) = 1$, muss positiv semidefinit und symmetrisch um den Ursprung sein, d.h. $R(\mathbf{d}) = R(-\mathbf{d})$. Eine weitere wichtige Eigenschaft der Korrelationsfunktion betrifft die Stetigkeit der Realisierungen (auch *Pfade* genannt) eines skalaren Zufallsfeldes (die sogenannten *Stichprobenfunktionen*). Nach [6] hat ein stationäres skalares Zufallsfeld $\mathcal{H}(\,\cdot\,)$ mit der Korrelationsfunktion $R(\,\cdot\,)$ *fast sicher*, d.h. mit Wahrscheinlichkeit eins, stetige Pfade falls die Korrelationsfunktion $R(\mathbf{d})$

1. stetig im Ursprung ist, d.h. $\lim\limits_{\mathbf{d}\to\mathbf{0}} R(\mathbf{d}) = 1$, und
2. für $\mathbf{d} \to \mathbf{0}$ *schnell genug* gegen eins konvergiert. Dies ist z.B. der Fall, wenn [64, S. 38]

$$1 - R(\mathbf{d}) \leq \frac{c}{|\log(\|\mathbf{d}\|)|^{1+\epsilon}}, \quad \forall\, \|\mathbf{d}\| < \delta,$$

für irgendein $c > 0$, $\epsilon > 0$ und $\delta < 1$ gilt.

Jede Korrelationsfunktion der (eindimensionalen) *Potenz-Exponential-Familie*

$$R(\,\cdot\,) : \mathbb{R} \to (0,\infty), \quad R(d;\psi,p) := \exp\{-|d/\psi|^p\}, \quad \psi > 0,\, 0 < p \leq 2 \tag{8.4}$$

erfüllt diese Bedingungen und generiert folglich stetige *Stichprobenfunktionen* mit der Probabilität eins. Der Fall $p = 2$ entspricht der sogenannten *Gauß'schen* Korrelationsfunktion. Im mehrdimensionalen Fall hat die Korrelationsfunktion der *Potenz-Exponential-Familie* die Form

$$R(\,\cdot\,) : \mathbb{R}^m \to (0,\infty),$$
$$R(\mathbf{d};\psi,p) := \exp\left\{-\sum_{i=1}^{m} |d_i/\psi_i|^{p_i}\right\}, \quad \psi_i > 0,\, 0 < p_i \leq 2, \forall i = 1,\ldots,m. \tag{8.5}$$

Die Korrelationsfunktion aus Gl. (8.5) wurde durch Multiplikation der jeweiligen eindimensionalen Korrelationsfunktionen aus derselben Familie gebildet. Dies ist möglich, weil das Produkt mehrerer Korrelationsfunktionen wiederum eine Korrelationsfunktion darstellt.

### 8.1.4 Bester linearer erwartungstreuer Prädiktor (BLUP)

Das Problem der **Prädiktion** wird als Prädiktion einer Zufallsvariable $H_0$ basierend auf den (bekannten) Trainingsdaten $\mathbf{H} = [H_1 \quad \cdots \quad H_N]^\top$ formuliert. Wir unterscheiden dabei die Zufallsvariable $H_i$, mit $i = 0,\ldots,N$, von ihrer Realisierung $\eta_i$. Der Prädiktor wird mit $\hat{H}_0$ bezeichnet. In dieser Arbeit wird nur der beste lineare erwartungtreue Prädiktor (BLUP, ***B**est **L**inear **U**nbiased **P**redictor*) analysiert. Dieser hat die allgemeine Form

$$\hat{H}_0 := c_0 + \mathbf{c}^\top \mathbf{H}, \quad c_0 \in \mathbb{R}, \mathbf{c} \in \mathbb{R}^N,$$

wobei die Parameter $c_0$ und $\mathbf{c}$ noch bestimmt werden müssen. Der Prädiktor heißt *erwartungstreu* bezüglich einer Verteilungsfamilie $\mathcal{F}$ für $(H_0,\mathbf{H})$, falls $E_F\{\hat{H}_0\} = E_F\{H_0\}$, für $(H_0,\mathbf{H}) \sim F$ und $F \in \mathcal{F}$. Der *beste lineare* Prädiktor heißt einer, der unter allen möglichen linearen erwartungstreuen Prädiktoren den mittleren quadratischen Prädiktionsfehler (MSPE, ***M**ean **S**quared **P**rediction **E**rror*)

$$\mathrm{MSPE}(\hat{H}_0,F) := E_F\{(\hat{H}_0 - H_0)^2\},$$

minimiert. In [64, Theorem 3.2.1] wird gezeigt, dass im Fall eines Zufallsvektors $(H_0,\mathbf{H})$ mit einer vorgegebenen Verteilung $F \in \mathcal{F}$, der Erwartungswert der bedingten Verteilung $[H_0|\mathbf{H}]$ den besten MSPE-Prädiktor für $H_0$ bildet, d.h.

$$\hat{H}_0 := \mathrm{E}\{H_0|\mathbf{H} = \boldsymbol{\eta}\},$$

unter der Annahme, dass dieser bedingte Erwartungswert von $H_0$, gegeben $\mathbf{H} = \boldsymbol{\eta}$, $\mathrm{E}\{H_0|\mathbf{H} = \boldsymbol{\eta}\}$, existiert.

## 8.2 Prädiktive Verteilungen

In dieser Arbeit wird der Ausgang des *Computerexperiments* als skalares Zufallsfeld modelliert. Dieses hat die Form

$$\mathcal{H}(\boldsymbol{\zeta}) = \mathbf{f}(\boldsymbol{\zeta})^\top \boldsymbol{\beta} + \mathcal{Z}(\boldsymbol{\zeta}), \tag{8.6}$$

wobei $\mathbf{f}(\boldsymbol{\zeta}) := [f_1(\boldsymbol{\zeta}) \quad \cdots \quad f_m(\boldsymbol{\zeta})]^\top$ bekannte, d.h. vorher festgelegte, Regressionsfunktionen gruppiert, $\boldsymbol{\beta} := [\beta_1 \quad \cdots \quad \beta_m]^\top$ einen Vektor von **unbekannten** (konstanten) Regressionskoeffizienten darstellt, und $\mathcal{Z}(\boldsymbol{\zeta})$

ein mittelwertfreies stationäres GRF mit der **unbekannten** konstanten Varianz $\sigma_{\mathcal{Z}}^2$ und **unbekannten** Korrelationsfunktion $R(\cdot)$ ist. Die Korrelationsfunktion wird in parametrischer Form angenommen, d.h. es gilt $R(\cdot) = R(\cdot|\boldsymbol{\psi})$, wobei der Vektor $\boldsymbol{\psi}$ die unbekannten Parameter der Korrelationsfunktion gruppiert. Anhand eines gegebenen Vektors $\boldsymbol{\eta} := [h(\boldsymbol{\zeta}_1) \quad \cdots \quad h(\boldsymbol{\zeta}_N)]^\top$ von Trainingsdaten wird eine Prädiktion des Ausgangs für einen neuen Datenpunkt $\boldsymbol{\zeta}_0$ gesucht.

Das angenommene Modell aus Gl. (8.6) impliziert,[3)] dass die Zufallsvariable $H_0 := \mathcal{H}(\boldsymbol{\zeta}_0)$ und der Zufallsvektor $\mathbf{H} := [H(\zeta_1) \quad \cdots \quad H(\zeta_N)]^\top$ eine gemeinsame multivariate Gauß'sche Verteilung haben, d.h.

$$\begin{bmatrix} H_0 \\ \mathbf{H} \end{bmatrix} \sim \mathcal{N}_{1+N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right), \tag{8.7}$$

mit Erwartungswert und Kovarianzmatrix

$$\boldsymbol{\mu} := \begin{bmatrix} \mathbf{f}_0^\top \\ \mathbf{F} \end{bmatrix} \boldsymbol{\beta}, \tag{8.8}$$

$$\boldsymbol{\Sigma} := \sigma_{\mathcal{Z}}^2 \begin{bmatrix} 1 & \mathbf{r}_0^\top \\ \mathbf{r}_0 & \mathbf{R} \end{bmatrix}. \tag{8.9}$$

Dabei bezeichnet $\mathbf{f}_0 := \mathbf{f}_0(\boldsymbol{\zeta}_0) = [f_1(\boldsymbol{\zeta}_0) \quad \cdots \quad f_m(\boldsymbol{\zeta}_0)]^\top$ einen Vektor von bekannten Regressionsfunktionswerten, $\mathbf{F} := [\mathbf{f}_1(\boldsymbol{\zeta}_1) \quad \cdots \quad \mathbf{f}_N(\boldsymbol{\zeta}_N)]^\top$ eine $N \times m$-Matrix von bekannten Regressionsfunktionswerten, mit $\mathbf{f}_i(\boldsymbol{\zeta}_i) := [f_1(\boldsymbol{\zeta}_i) \quad \cdots \quad f_m(\boldsymbol{\zeta}_i)]^\top$, $i = 1,\ldots,N$, $\mathbf{r}_0^\top := [R(\boldsymbol{\zeta}_0 - \boldsymbol{\zeta}_1|\boldsymbol{\psi}) \quad \cdots \quad R(\boldsymbol{\zeta}_0 - \boldsymbol{\zeta}_N|\boldsymbol{\psi})]^\top$ einen Vektor von Korrelationen zwischen dem Zufallsvektor $\mathbf{H}$ und der Zufallsvariable $H_0$, und, schließlich, $\mathbf{R}$ eine Matrix von Korrelationen zwischen den Zufallsvariablen $H_1,\ldots,H_N$ mit den Elementen $\mathbf{R}_{(i,j)} := R(\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_j|\boldsymbol{\psi})$, mit $i,j = 1,\ldots,N$.

Der BLUP-Prädiktor von $H_0$ ist, wie im vorherigen Abschnitt erwähnt, $\hat{H}_0 := \mathrm{E}\{H_0|\mathbf{H} = \boldsymbol{\eta}\}$, d.h. der Erwartungswert der Verteilung $[H_0|\mathbf{H}]$. Um diesen Erwartungswert zu berechnen, wird der Bayes'sche Ansatz angewandt. Dieser basiert auf dem Satz von Bayes.

*Bemerkung* 8.1 (Anwendung des Satzes von Bayes). Über den Parameter $\omega$ eines Modells seien - beispielsweise aufgrund vorheriger Erfahrungen mit ähnlichen Modellen oder Daten - irgendwelche probabilistische Annahmen in der Form einer *A-priori*-Wahrscheinlichkeitsdichte getroffen worden. Sei diese mit $f_\Omega(\omega)$ bezeichnet. Mit Hilfe von neuen Informationen in der Form

[3)] Dies folgt aus der Definition eines Gauß'schen Zufallsfeldes, vgl. Def. 4, Seite 119.

eines neu erhobenen Datensatzes $\boldsymbol{\eta}$ kann die Wahrscheinlichkeitsdichte des Parameters $\omega$ aktualisiert werden. Diese wird *A-posteriori*-Dichte genannt und wird mit $f_{\Omega|\mathbf{H}=\eta}(\omega)$ bezeichnet. Der Zusammenhang zwischen den beiden Wahrscheinlichkeitsdichten kann durch den Satz von Bayes in der Form

$$f_{\Omega|\mathbf{H}=\boldsymbol{\eta}}(\omega) = \frac{f_{\mathbf{H}|\Omega=\omega}(\boldsymbol{\eta}) f_{\Omega}(\omega)}{f_{\mathbf{H}}(\boldsymbol{\eta})} \tag{8.10}$$

geschrieben werden. Dabei sind $f_{\mathbf{H}|\Omega=\omega}(\boldsymbol{\eta})$ die *Likelihood* (auch *inverse Wahrscheinlichkeitsdichte* genannt), welche anhand der erhobenen Daten $\boldsymbol{\eta}$ gewählt wird, $f_{\Omega}(\omega)$ die *A-priori*-Dichte von $\Omega$, und $f_{\mathbf{H}}(\boldsymbol{\eta})$ die *marginale Likelihood* (oder *Evidenz*), welche den Erwartungswert der Likelihood bezüglich der *A-priori*-Dichte des Parameters $\omega$, d.h.

$$f_{\mathbf{H}}(\boldsymbol{\eta}) = \int_{-\infty}^{\infty} f_{\mathbf{H}|\Omega=\omega}(\boldsymbol{\eta}) f_{\Omega}(\omega) \mathrm{d}\omega,$$

darstellt und folglich eine Konstante ist. △

Für die Bildung der prädiktiven Verteilung $[H_0|\mathbf{H}]$ ist es weiterhin von Bedeutung, welche Parameter des Modells aus Gl. (8.6) bekannt sind. Diese Parameter sind der Koeffizientenvektor $\boldsymbol{\beta}$, die Varianz $\sigma_{\mathcal{Z}}^2$ und der Parametervektor $\boldsymbol{\psi}$ der Korrelationsfunktion $R(\,\cdot\,|\boldsymbol{\psi})$.[4)] In dieser Arbeit wird davon ausgegangen, dass keiner dieser Parameter bekannt ist. Ist einer der Parameter bekannt, so vereinfacht sich die Untersuchung, vgl. [64] für solche Fälle. Im Folgenden werden zwei Methoden vorgestellt. In der ersten Methode - *partieller Bayes'scher Ansatz* - wird die prädiktive Verteilung $[H_0|\mathbf{H}]$ unter Verwendung des Bayes'schen Ansatzes für die Parameter $\boldsymbol{\beta}$ und $\sigma_{\mathcal{Z}}^2$ berechnet, wobei der Parametervektor $\boldsymbol{\psi}$ vorerst als bekannt angenommen und anschließend empirisch, z.B. mit Hilfe der Maximum Likelihood Methode, geschätzt wird. Diese prädiktive Verteilung heißt *plug-in* Verteilung. Die zweite Methode - *vollständiger Bayes'scher Ansatz* - berechnet eine prädiktive Verteilung $[H_0|\mathbf{H}]$ unter Verwendung des Bayes'schen Ansatzes für alle Parameter $\boldsymbol{\beta}$, $\sigma_{\mathcal{Z}}^2$ und $\boldsymbol{\psi}$. Beide Ansätze sind in zwei Etappen gegliedert. In der ersten Etappe wird der Parametervektor $\boldsymbol{\psi}$ als bekannt angenommen, und in der zweiten Etappe wird dieser empirisch geschätzt bzw. durch die Bayes'sche Methode inferiert. Die erste Etappe enthält die folgenden Schritte:

[4)] Es wird dabei angenommen, dass die Korrelationsfunktion bis auf den Parametervektor $\boldsymbol{\psi}$ bekannt ist.

**Schritt 1** Wähle beliebige *A-priori*-Verteilungen $[\boldsymbol{\beta}|\sigma_{\mathcal{Z}}^2]$ und $[\sigma_{\mathcal{Z}}^2]$, um die gemeinsame *A-priori*-Verteilung $[\boldsymbol{\beta},\sigma_{\mathcal{Z}}^2] = [\boldsymbol{\beta}|\sigma_{\mathcal{Z}}^2] \cdot [\sigma_{\mathcal{Z}}^2]$ zu bilden.

**Schritt 2** Wähle eine Verteilung $[\mathbf{H}|\boldsymbol{\beta},\sigma_{\mathcal{Z}}^2]$ für den Zufallsvektor $\mathbf{H}$ und eine Verteilung $\left[(H_0,\mathbf{H})|\,\boldsymbol{\beta},\sigma_{\mathcal{Z}}^2\right]$ für den Zufallsvektor $(H_0,\mathbf{H})$.

**Schritt 3** Unter Verwendung des Satzes von Bayes und der Verteilungen aus Schritt 1 und 2, berechne die gemeinsame *A-posteriori*-Wahrscheinlichkeitsdichte $[(\boldsymbol{\beta},\sigma_{\mathcal{Z}}^2)|\mathbf{H}]$ mit Hilfe der Regel

$$[(\boldsymbol{\beta},\sigma_{\mathcal{Z}}^2)|\mathbf{H}] = \frac{[\mathbf{H}|(\boldsymbol{\beta},\sigma_{\mathcal{Z}}^2)] \cdot [\boldsymbol{\beta},\sigma_{\mathcal{Z}}^2]}{[\mathbf{H}]}. \tag{8.11}$$

**Schritt 4** Berechne aus der Verteilung $[(\boldsymbol{\beta},\sigma_{\mathcal{Z}}^2)|\mathbf{H}]$ die *A-posteriori*-Randverteilungen $[\boldsymbol{\beta}|\mathbf{H}]$ und $[\sigma_{\mathcal{Z}}^2|\mathbf{H}]$. Es folgt

$$[\boldsymbol{\beta}|\mathbf{H}] = \int_{\mathcal{D}_{\sigma_{\mathcal{Z}}^2}} \left[(\boldsymbol{\beta},\sigma_{\mathcal{Z}}^2)|\mathbf{H}\right] \mathrm{d}\sigma_{\mathcal{Z}}^2, \tag{8.12}$$

$$[\sigma_{\mathcal{Z}}^2|\mathbf{H}] = \int\cdots\int_{\mathcal{D}_{\boldsymbol{\beta}}} \left[(\boldsymbol{\beta},\sigma_{\mathcal{Z}}^2)|\mathbf{H}\right] \mathrm{d}\boldsymbol{\beta}. \tag{8.13}$$

*Bemerkung* 8.2 (Zu Schritt 1 - Wahl der *A-priori*-Verteilungen $[\boldsymbol{\beta}|\sigma_{\mathcal{Z}}^2]$ und $[\sigma_{\mathcal{Z}}^2]$). Die Wahl dieser Verteilungen kann einen großen Einfluß auf die *A-posteriori*-Wahrscheinlichkeitsdichte $[(\boldsymbol{\beta},\sigma_{\mathcal{Z}}^2)|\mathbf{H}]$ aus Schritt 3 und auf die *prädiktive* Verteilung $[H_0|\mathbf{H}]$ haben. Diese können nur im Fall bestimmter *A-priori*-Verteilungen analytisch angegeben werden. Einen solchen einfachen Fall bilden die *nicht-informativen* und die ***konjugierten*** *A-priori*-Verteilungen. Die *nicht-informative* Verteilung hat definitionsgemäß keinen Einfluß auf die *A-posteriori*-Verteilung. Die ***konjugierten*** *A-priori*-Verteilungen sind solche *A-priori*-Verteilungen, die - durch Multiplikation mit der *Likelihood*, wie in Gl. (8.10) - *A-posteriori*-Wahrscheinlichkeitsdichten aus derselben Klasse erzeugen. Ein Beispiel einer ***konjugierten*** *A-priori*-Verteilung ist die multivariate Normalverteilung für $\boldsymbol{\beta}$ mit einer multivariaten Normalverteilung als *Likelihood* von $\mathbf{H}$ und mit bekannter Kovarianzmatrix $\sigma_{\mathcal{Z}}^2\mathbf{R}$.

Tabelle 8.2 zeigt die hier untersuchte Wahl von *A-priori*-Verteilungen. Für jeden Parameter wird jeweils eine *informative* und eine *nicht-informative A-priori*-Verteilung gewählt. Für den Zufallsvektor $\boldsymbol{\beta}|\sigma_{\mathcal{Z}}^2$ wird als ***informative*** *A-priori*-Verteilung die Normalverteilung gewählt, d.h.

**Tabelle 8.2:** Die untersuchten *A-priori*-Verteilungen $[(\boldsymbol{\beta},\sigma_{\mathcal{Z}}^2)]$.

| | $[\sigma_{\mathcal{Z}}^2]$ | |
|---|---|---|
| $[\boldsymbol{\beta}\vert\sigma_{\mathcal{Z}}^2]$ | $c_0/\chi_{\nu_0}^2$ | $1/\sigma_{\mathcal{Z}}^2$ |
| $\mathcal{N}_m(\boldsymbol{\beta}_0,\sigma_{\mathcal{Z}}^2\boldsymbol{\Sigma}_0)$ | (1) | (2) |
| 1 | (3) | (4) |

$\boldsymbol{\beta}|\sigma_{\mathcal{Z}}^2 \sim \mathcal{N}_m(\boldsymbol{\beta}_0,\sigma_{\mathcal{Z}}^2\boldsymbol{\Sigma}_0)$, wobei $\boldsymbol{\beta}_0$ und $\boldsymbol{\Sigma}_0$ vorher festgelegt werden. Oft wird die Matrix $\boldsymbol{\Sigma}_0$ in Diagonalform gewählt. Die *nicht-informative A-priori*-Verteilung ist $[\boldsymbol{\beta}|\sigma_{\mathcal{Z}}^2] \propto 1$. Für die Varianz $\sigma_{\mathcal{Z}}^2$ des skalaren Zufallsfeldes $\mathcal{Z}$ wird als *informative*-Verteilung nicht die Normalverteilung angenommen, da diese auch negative Realisierungen erlaubt, sondern die Verteilung einer Konstanten $c_0 > 0$ geteilt durch eine $\mathcal{X}_{\nu_0}^2$ (Chi-Quadrat)-verteilte Zufallsvariable mit $\nu_0$ Freiheitsgraden.[5)] Die Konstanten $c_0$ und $\nu_0$ werden als bekannt angenommen, d.h. vorher festgelegt. Die *nicht-informative A-priori*-Verteilung ist $[\sigma_{\mathcal{Z}}^2] \sim 1/\sigma_{\mathcal{Z}}^2$, die sogenannte *Jeffreys A-priori*-Verteilung. △

*Bemerkung* 8.3 (Zu Schritt 2 - Wahl der Verteilungen $[\mathbf{H}|\boldsymbol{\beta},\sigma_{\mathcal{Z}}^2]$ und $[(H_0,\mathbf{H})|\boldsymbol{\beta},\sigma_{\mathcal{Z}}^2]$). In dieser Arbeit werden ausschließlich Normalverteilungen betrachtet, da angenommen wird, dass der Ausgang des *Computerexperiments* als skalares Gauß'sches Zufallsfeld modelliert wird, d.h., es gilt

$$[\mathbf{H}|(\boldsymbol{\beta},\sigma_{\mathcal{Z}}^2)] \sim \mathcal{N}_N(\mathbf{F}\boldsymbol{\beta},\sigma_{\mathcal{Z}}^2\mathbf{R}), \tag{8.14}$$

$$\left[\begin{bmatrix} H_0 \\ \mathbf{H} \end{bmatrix}\middle|(\boldsymbol{\beta},\sigma_{\mathcal{Z}}^2)\right] \sim \mathcal{N}_{N+1}\left(\begin{bmatrix} \mathbf{f}_0^\top \\ \mathbf{F} \end{bmatrix}\boldsymbol{\beta},\sigma_{\mathcal{Z}}^2\begin{bmatrix} 1 & \mathbf{r}_0^\top \\ \mathbf{r}_0 & \mathbf{R} \end{bmatrix}\right). \tag{8.15}$$

△

*Bemerkung* 8.4 (Zu Schritt 3). Die *Evidenz* $[\mathbf{H}]$ stellt eine Konstante dar, die für Schätzungszwecke keine Rolle spielt. Somit kann man Gl. (8.11) vereinfacht in der Form

$$[(\boldsymbol{\beta},\sigma_{\mathcal{Z}}^2)|\mathbf{H}] \propto [\mathbf{H}|(\boldsymbol{\beta},\sigma_{\mathcal{Z}}^2)] \cdot [\boldsymbol{\beta},\sigma_{\mathcal{Z}}^2].$$

schreiben. △

[5)] Falls eine Zufallsvariable $X$ Chi-Quadrat verteilt ist, d.h. $X \sim \chi_{\nu_0}^2$, dann besitzt die Zufallsvariable $Y := c_0/X$ eine *inverse* Chi-Quadrat-Verteilung, d.h. $Y \sim c_0\chi_{\nu_0}^{-2}$. Die inverse Chi-Quadrat-Verteilung wird auch als Inv-$\chi_{\nu_0}^{-2}$ bezeichnet.

*Bemerkung* 8.5 (Zu Schritt 4). Die Erwartungswerte der beiden *A-posteriori*-Randverteilungen werden als Schätzer des unbekannten Koeffizientenvektors $\boldsymbol{\beta}$ bzw. der unbekannten Varianz $\sigma^2_{\mathcal{Z}}$ verwendet, d.h.

$$\hat{\boldsymbol{\beta}} := \mathrm{E}\{\boldsymbol{\beta}|\mathbf{H}\},$$
$$\hat{\sigma}^2_{\mathcal{Z}} := \mathrm{E}\{\sigma^2_{\mathcal{Z}}|\mathbf{H}\}.$$

△

### 8.2.1 Der partielle Bayes'sche Ansatz

Bei dem partiellen Bayes'schen Ansatz wird der Parametervektor $\boldsymbol{\psi}$ empirisch geschätzt, vgl. [64, Abschnitt 3.3]. Der Ansatz enthält die Schritte 1-4 und

**Schritt 5a** Berechne die *prädiktive* Verteilung $[H_0|\mathbf{H}]$. Dies wird in mehreren Schritten gemacht. Nach einigen Umformungen ergibt sich

$$[H_0|\mathbf{H}] = \int_{\mathcal{D}_{\sigma^2_{\mathcal{Z}}}} \int \cdots \int_{\mathcal{D}_{\boldsymbol{\beta}}} [H_0|\mathbf{H},\boldsymbol{\beta},\sigma^2_{\mathcal{Z}}][\mathbf{H}|\boldsymbol{\beta},\sigma^2_{\mathcal{Z}}][\boldsymbol{\beta}|\sigma^2_{\mathcal{Z}}]\mathrm{d}\boldsymbol{\beta}\mathrm{d}\sigma^2_{\mathcal{Z}}. \quad (8.16)$$

Die einzelnen Schritte werden im Anhang B.4.1 gezeigt.

**Schritt 6a** Berechne einen Schätzer $\hat{\boldsymbol{\psi}}$ für den Parametervektor $\boldsymbol{\psi}$ aus Gl. (B.26) und (B.27) mit Hilfe einer der im Folgenden vorgestellten empirischen Schätzmethoden.

**Schritt 7a** Berechne den Prädiktor $\hat{H}_0 := \mathrm{E}\{H_0|\mathbf{H} = \boldsymbol{\eta}\}$ und seine Varianz $\mathrm{Var}\{H_0|\mathbf{H} = \boldsymbol{\eta}\}$ aus der Verteilung $[H_0|\mathbf{H}]$ aus Schritt 5a mit dem Parametervektor $\hat{\boldsymbol{\psi}}$ aus Schritt 6a. Es ergibt sich

$$\hat{\mathcal{H}}(\boldsymbol{\zeta}_0) =: \hat{H}_0 = \mathbf{f}_0^\top \hat{\boldsymbol{\beta}} + \hat{\mathbf{r}}_0^\top \hat{\mathbf{R}}^{-1}(\boldsymbol{\eta} - \mathbf{F}\hat{\boldsymbol{\beta}}), \quad (8.17)$$

wobei $\hat{\mathbf{r}}_0$ und $\hat{\mathbf{R}}^{-1}$ vom geschätzten Parametervektor $\hat{\boldsymbol{\psi}}$ aus Schritt 6a abhängen.

*Bemerkung* 8.6 (Zu Schritt 5a - Berechnung der *A-posteriori*-Wahrscheinlichkeitsdichte $[H_0|\mathbf{H}]$). Die Berechnung aus Schritt 5a wurde im Fall der *A-priori*-Verteilungen aus Tabelle 8.2 in [64, Theorem 4.1.2] angegeben.

Der Satz besagt, dass im Fall der *A-priori*-Verteilungen (1) – (4) aus Tabelle 8.2 und der Verteilung $[H_0,\mathbf{H}|(\boldsymbol{\beta},\sigma_{\mathcal{Z}}^2)]$ aus Gl. (8.15), die *A-posteriori*-Wahrscheinlichkeitsdichte $[H_0|\mathbf{H}]$ eine eindimensionale *nicht-zentrale* $t$-Verteilung ist, d.h.

$$H_0|\mathbf{H} \sim \mathcal{T}_1(\nu_i,\mu_i,\sigma_i^2) \tag{8.18}$$

mit $\nu_i$ Freiheitsgraden, *Nichtzentralitätsparameter* $\mu_i$ und Skalierungsparameter $\sigma_i^2$, $i = (1),\ldots,(4)$. Diese Parameter werden im Theorem B.1 (Anhang) angegeben. Sie hängen vom Parametervektor $\boldsymbol{\psi}$ der Korrelationsfunktion $R(\cdot\,|\boldsymbol{\psi})$ ab. △

Dabei wird angenommen, dass die Trainingsdaten $\boldsymbol{\eta}$ die Realisierungen einer Gauß'schen (bedingten) Verteilung sind, d.h.

$$[\mathbf{H}|\boldsymbol{\beta},\sigma_{\mathcal{Z}}^2,\boldsymbol{\psi}] \sim \mathcal{N}_N(\mathbf{F}\boldsymbol{\beta},\sigma_{\mathcal{Z}}^2\mathbf{R}). \tag{8.19}$$

Gl. (8.19) ist dabei sehr ähnlich zu Gl. (8.14). Letztere enthält der Einfachheit halber den Parametervektor $\boldsymbol{\psi}$ nicht ausdrücklich, welcher in jenem Schritt als bekannt angenommen wurde.

### Maximum-Likelihood-Methode (MLE)

Die *Log-Likelihood*-Funktion der Verteilung $[\mathbf{H}|\boldsymbol{\beta},\sigma_{\mathcal{Z}}^2,\boldsymbol{\psi}]$ hat folglich bis auf einen konstanten Term die Form

$$\begin{aligned}\mathcal{L}(\boldsymbol{\eta};\boldsymbol{\beta},\sigma_{\mathcal{Z}}^2,\boldsymbol{\psi}) = -\frac{1}{2}\Big[&N\log\sigma_{\mathcal{Z}}^2 + \log(\det(\mathbf{R}(\boldsymbol{\psi})))\\ &+\frac{1}{\sigma_{\mathcal{Z}}^2}(\boldsymbol{\eta}-\mathbf{F}\boldsymbol{\beta})^\top\mathbf{R}(\boldsymbol{\psi})^{-1}(\boldsymbol{\eta}-\mathbf{F}\boldsymbol{\beta})\Big]\end{aligned}$$

Im Fall eines bekannten Parametervektors $\boldsymbol{\psi}$ ergibt die Maximierung der *Log-Likelihood*-Funktion bezüglich $\boldsymbol{\beta}$ und $\sigma_{\mathcal{Z}}^2$

$$\begin{aligned}\beta_{\mathrm{MLE}}(\boldsymbol{\psi}) &:= \arg\max_{\boldsymbol{\beta}} \mathcal{L}(\boldsymbol{\eta};\boldsymbol{\beta},\sigma_{\mathcal{Z}}^2,\boldsymbol{\psi}) = (\mathbf{F}^\top\mathbf{R}(\boldsymbol{\psi})^{-1}\mathbf{F})^{-1}\mathbf{F}^\top\mathbf{R}(\boldsymbol{\psi})^{-1}\boldsymbol{\eta},\\ \sigma_{\mathcal{Z}_{\mathrm{MLE}}}^2(\boldsymbol{\psi}) &:= \arg\max_{\sigma_{\mathcal{Z}}^2} \mathcal{L}(\boldsymbol{\eta};\boldsymbol{\beta},\sigma_{\mathcal{Z}}^2,\boldsymbol{\psi})\\ &= \frac{1}{N}\left(\boldsymbol{\eta}-\mathbf{F}\boldsymbol{\beta}_{\mathrm{MLE}}(\boldsymbol{\psi})\right)^\top\mathbf{R}(\boldsymbol{\psi})^{-1}(\boldsymbol{\eta}-\mathbf{F}\boldsymbol{\beta}_{\mathrm{MLE}}(\boldsymbol{\psi})),\end{aligned}$$

sowie ein Maximum von

$$\begin{aligned}\mathcal{L}^*(\beta_{\mathrm{MLE}}(\boldsymbol{\psi}),\sigma^2_{\mathcal{Z}_{\mathrm{MLE}}}(\boldsymbol{\psi}),\boldsymbol{\psi}) &\qquad (8.20)\\ := \max_{\boldsymbol{\beta},\sigma^2_{\mathcal{Z}}} \mathcal{L}(\boldsymbol{\eta};\boldsymbol{\beta},\sigma^2_{\mathcal{Z}},\boldsymbol{\psi})&\\ = -\frac{1}{2}\left[N\log\sigma^2_{\mathcal{Z}_{\mathrm{MLE}}}(\boldsymbol{\psi}) + \log(\det(\mathbf{R}(\boldsymbol{\psi}))) + N\right], &\qquad (8.21)\end{aligned}$$

das vom Parametervektor $\boldsymbol{\psi}$ abhängt. Eine (numerische) Maximierung bezüglich $\boldsymbol{\psi}$ ergibt schließlich einen empirischen Schätzer des unbekannten Parametervektors, d.h.

$$\hat{\boldsymbol{\psi}} =: \boldsymbol{\psi}_{\mathrm{MLE}} = \arg\max_{\boldsymbol{\psi}} \mathcal{L}^*(\beta_{\mathrm{MLE}}(\boldsymbol{\psi}),\sigma^2_{\mathcal{Z}_{\mathrm{MLE}}}(\boldsymbol{\psi}),\boldsymbol{\psi}). \qquad (8.22)$$

Alternativ zur *Maximum-Likelihood*-Methode kann die sogenannte ***begrenzte*** *Maximum-Likelihood-Methode*, vgl. [64, Abschnitt 3.3] verwendet werden. Die *begrenzte* Maximum-Likelihood-Methode basiert auf der gleichen Schätzmethode aus dem vorherigen Abschnitt, jedoch unter Verwendung einer kleineren Menge von Trainingsdaten. Diese ergibt sich aus $\boldsymbol{\eta}_b := \mathbf{C}\boldsymbol{\eta}$, wobei die Matrix $\mathbf{C} \in \mathbb{R}^{(N-m)\times N}$, die Bedingung $\mathbf{CF} = \mathbf{0}$ erfüllt. Die Matrix $\mathbf{C}^\top$ bildet also eine Basis des Nullraumes der Matrix $\mathbf{F}^\top$. Die Menge $\boldsymbol{\eta}_b$ beinhaltet weniger Trainingsdaten als $\boldsymbol{\eta}$, die Verteilung der Trainingsdaten enthält aber den Parametervektor $\boldsymbol{\beta}$ nicht mehr, d.h. $\mathbf{H}_b := \mathbf{CH} \sim \mathcal{N}_{N-m}(\mathbf{CF}\boldsymbol{\beta} = \mathbf{0},\sigma^2_{\mathcal{Z}}\mathbf{CR}(\boldsymbol{\psi})\mathbf{C}^\top)$.

### Kreuzvalidierungsmethode (XVal)

Diese Methode generiert $N$ Prädiktionen $\hat{\mathcal{H}}_{\overline{i}}(\boldsymbol{\psi})$, mit $i = 1,\ldots,N$, basierend jeweils auf einem Vektor $\boldsymbol{\eta}_{\overline{i}}$ von Trainingsdaten, der den i-ten Datenpunkt $\eta(\zeta_i)$ nicht enthält. Da die Datenpunkte $\eta(\boldsymbol{\zeta}_i)$, mit $i = 1,\ldots,N$, aber bekannt sind, kann man einen empirischen quadratischen Prädiktionsfehler (E-MSPE) berechnen, d.h.

$$\text{E-MSPE}(\boldsymbol{\psi}) := \sum_{i=1}^{N}\left(\hat{H}_{\overline{i}}(\boldsymbol{\psi}) - \eta(\boldsymbol{\zeta}_i)\right)^2, \qquad (8.23)$$

und zur Schätzung des Parametervektors $\boldsymbol{\psi}$ verwenden. Der *Kreuzvalidierung*-Schätzer $\hat{\boldsymbol{\psi}}$ minimiert den empirischen quadratischen Prädiktionsfehler aus Gl. (8.23).

### 8.2.2 Der vollständige Bayes'sche Ansatz

Bei dem vollständigen Bayes'schen Ansatz wird der Parametervektor $\boldsymbol{\psi}$ ebenfalls mit Hilfe des Bayes'schen Ansatzes berechnet. Es wird davon ausgegangen, dass der Zufallsvektor $\boldsymbol{\psi}$ unabhängig vom Zufallsvektor $(\boldsymbol{\beta},\sigma_Z^2)$ ist, d.h.

$$[\boldsymbol{\beta},\sigma_Z^2,\boldsymbol{\psi}] = [\boldsymbol{\beta},\sigma_Z^2]\cdot[\boldsymbol{\psi}]. \tag{8.24}$$

Die prädiktive Verteilung $[H_0|\mathbf{H}]$ aus Gl. (8.18) kann dabei als eine bedingte Verteilung $[H_0|\mathbf{H},\boldsymbol{\psi}]$ angesehen werden. Diese Methode besteht dann aus den Schritten 1-4 und:

**Schritt 5b** Wähle eine beliebige *A-priori*-Verteilung $[\boldsymbol{\psi}]$.

**Schritt 6b** Berechne die *A-posteriori*-Verteilung $[\boldsymbol{\psi}|\mathbf{H}]$ mit Hilfe des Bayes'schen Ansatzes und der Verteilungen aus den Schritten 1-3. Es folgt

$$[\boldsymbol{\psi}|\mathbf{H}] = \int_{\mathcal{D}_{\sigma_Z^2}} \cdots \int_{\mathcal{D}_{\boldsymbol{\beta}}} [\boldsymbol{\beta},\sigma_Z^2,\boldsymbol{\psi}|\mathbf{H}]\mathrm{d}\boldsymbol{\beta}\,\mathrm{d}\sigma_Z^2, \tag{8.25}$$

wobei der Integrand aus Gl. (8.25) mit Hilfe des Bayes'schen Ansatzes aus

$$[\boldsymbol{\beta},\sigma_Z^2,\boldsymbol{\psi}|\mathbf{H}] = [\mathbf{H}|\boldsymbol{\beta},\sigma_Z^2,\boldsymbol{\psi}]\cdot[\boldsymbol{\beta},\sigma_Z^2,\boldsymbol{\psi}] \tag{8.26}$$

berechnet werden kann.

**Schritt 7b** Berechne die *prädiktive* Verteilung $[H_0|\mathbf{H}]$ aus

$$[H_0|\mathbf{H}] = \int \cdots \int_{\mathcal{D}_{\boldsymbol{\psi}}} [H_0,\boldsymbol{\psi}|\mathbf{H}]\mathrm{d}\boldsymbol{\psi} = \int \cdots \int_{\mathcal{D}_{\boldsymbol{\psi}}} [H_0|\mathbf{H},\boldsymbol{\psi}][\boldsymbol{\psi}|\mathbf{H}]\mathrm{d}\boldsymbol{\psi}, \tag{8.27}$$

wobei $[H_0|\mathbf{H},\boldsymbol{\psi}]$ die gleiche Verteilung wie die Zufallsvariable $H_0|\mathbf{H}$ aus Gl. (8.18) hat, und die Verteilung $[\boldsymbol{\psi}|\mathbf{H}]$ in Schritt 6b berechnet wurde.

**Schritt 8b** Berechne den Prädiktor $\hat{H}_0 := \mathrm{E}\{H_0|\mathbf{H}=\boldsymbol{\eta}\}$ und seine Varianz $\mathrm{Var}\{H_0|\mathbf{H}=\boldsymbol{\eta}\}$ aus der Verteilung $[H_0|\mathbf{H}]$ aus Schritt 7b.

*Bemerkung* 8.7 (Zu Schritt 6b und 7b). Für einfache *A-priori*-Verteilungen $[\boldsymbol{\beta},\sigma_Z^2,\boldsymbol{\psi}]$ kann die $(m+1)$-dimensionale Integration aus Gl. (8.26) analytisch berechnet werden. Die Integration aus Gl. (8.27) kann nur in seltenen Fällen analytisch berechnet werden, vgl. z.B. [33] für den Fall $m=$

2, unter Verwendung von isotropischen Korrelationsfunktionen (*Potenz-Exponential-Familie* und *Matérn*-Korrelationsfunktionen).

Kann die analytische Berechnung nicht erfolgen, so besteht die Möglichkeit, die *A-posteriori*-Verteilung $[\boldsymbol{\psi}|\mathbf{H}]$ und/oder die *prädiktive* Verteilung $[H_0|\mathbf{H}]$ unter Verwendung der Gibbs Sampling-Methode oder des Metropolis-Hastings-Algorithmus numerisch zu approximieren, vgl. z.B. [12]. Beide Verfahren werden zur numerischen Approximation von Randverteilungen bei gegebener gemeinsamer Verteilung verwendet. Die gemeinsame Verteilung kann dabei auch eine *A-posteriori*-Verteilung, wie z.B. $[\boldsymbol{\beta},\sigma_Z^2,\boldsymbol{\psi}|\mathbf{H}]$, sein. Bei der Gibbs Sampling-Methode werden alle bedingten Verteilungen als bekannt vorausgesetzt. Bei dem Metropolis-Hastings-Algorithmus entfällt auch diese Voraussetzung. △

### 8.2.3 Das *Design-Problem*

Das *Design-Problem* beschäftigt sich mit der Frage, welche *Design*-Punkte $\boldsymbol{\zeta}_i$, mit $i = 1,\ldots,N$, gewählt werden sollten, sodass ein vordefiniertes Ziel des *Computerexperiments* erreicht wird. Ein solches Ziel ist beispielsweise

- eine möglichst gute Abdeckung des untersuchten Bereichs $\boldsymbol{\zeta} \in \mathcal{D}_{\boldsymbol{\zeta}}$ - *raumausschöpfendes Design*,
- die Optimierung eines bestimmten statistischen Gütemaßes - *optimales-Design* - wie z.B.
  - die Minimierung einer bestimmten Funktion der Kovarianzmatrix der Parameter des Modells ($\boldsymbol{\beta}$, $\sigma_Z^2$, $\boldsymbol{\psi}$):
    - die Determinante der Kovarianzmatrix - *D-optimales*-Design,
    - die Spur der Kovarianzmatrix - *A-optimales*-Design,
  - die Minimierung der erwarteten Varianz des prädizierten Wertes im ganzen Untersuchungsbereich - *I-optimales*-Design.

Die optimalen Designs setzen jedoch voraus, dass der jeweilige Parameter des Modells bekannt ist. Daher wird das *Design-Problem* manchmal in zwei Stufen gelöst, in einer ersten Stufe wird der jeweilige Parameter des Modells durch ein raumausschöpfendes Design bestimmt, und in einer zweiten Stufe, beispielsweise, dessen Varianz durch ein optimales Design minimiert. In dieser Arbeit wird die erste Stufe verwendet.

**Raumausschöpfende Designs**

Für raumausschöpfende Designs können sowohl deterministische als auch statistische Auswahlstrategien der Design-Punkte verwendet werden. Eine sehr einfache deterministische Strategie wäre, die Design-Punkte basierend auf einem Grid zu wählen. Statistische Strategien basieren auf einfachen oder stratifizierten Stichproben. Die raumausschöpfenden Designs werden in drei Klassen unterteilt:

- Stichprobenbasierte Designs
  - *einfache Zufallsstichprobe* generiert aus einer bestimmten Verteilung,
  - *geschichtete Zufallsstichprobe*, d.h. einfache Zufallsstichproben generiert in jeder Schicht aus der vorher geschichteten Design-Region,
  - *Latin-Hypercube-Sampling* (LHS): diese Methode generiert (marginal) gleichmäßig verteilte Punkte über jede Dimension des Parameterbereichs, vgl. Abschnitt B.5 (Anhang) für eine Beschreibung im zweidimensionalen Fall. Für den höherdimensionalen Fall siehe [64].
- Designs basierend auf Entfernungsmaßen, wie z.B.
  - Abstand $p$-ter Ordnung: $\rho_{\mathrm{p}}(\boldsymbol{\zeta}_1,\boldsymbol{\zeta}_2) = \left[\sum_{j=1}^{d} |\boldsymbol{\zeta}_1 - \boldsymbol{\zeta}_2|^p\right]^{1/p}$
    - Das *maximin* Design $\mathcal{X}_{\mathcal{D}_{\boldsymbol{\zeta}_{\mathrm{Mm}}}}$: Maximiert unter allen Design-Mengen $\mathcal{X}_{\mathcal{D}} \subset \mathcal{D}_{\boldsymbol{\zeta}}$ den kleinsten Abstand zwischen jeweils zwei Design-Punkten aus einer Design-Menge, d.h.

$$\mathcal{X}_{\mathcal{D}_{\boldsymbol{\zeta}_{\mathrm{Mm}}}} := \arg \max_{\mathcal{X}_{\mathcal{D}} \subset \mathcal{D}_{\boldsymbol{\zeta}}} \min_{\boldsymbol{\zeta}_1,\boldsymbol{\zeta}_2 \in \mathcal{X}_{\mathcal{D}}} \rho_{\mathrm{p}}(\boldsymbol{\zeta}_1,\boldsymbol{\zeta}_2).$$

    - Das *minimax* Design $\mathcal{D}_{\boldsymbol{\zeta}_{\mathrm{mM}}}$: Verwendet den Abstand zwischen einem beliebigen vorgegebenen Design-Punkt $\boldsymbol{\zeta}$ und einer Design-Menge $\mathcal{X}_{\mathcal{D}}$, welcher als

$$\rho_{\mathrm{d}}(\boldsymbol{\zeta},\mathcal{X}_{\mathcal{D}}) := \min_{\boldsymbol{\zeta}_i \in \mathcal{X}_{\mathcal{D}}} \rho_{\mathrm{p}}(\boldsymbol{\zeta},\boldsymbol{\zeta}_i)$$

      definiert wird, und minimiert unter allen Design-Mengen $\mathcal{X}_{\mathcal{D}} \subset \mathcal{D}_{\boldsymbol{\zeta}}$ den maximalen Abstand zwischen allen möglichen Design-Punkten $\boldsymbol{\zeta} \in \mathcal{D}_{\boldsymbol{\zeta}}$ und einer Design-Menge $\mathcal{X}_{\mathcal{D}}$,

d.h.

$$\mathcal{X}_{\mathcal{D}_{\mathrm{mM}}} := \arg \min_{\mathcal{X}_{\mathcal{D}} \subset \mathcal{D}_{\boldsymbol{\zeta}}} \max_{\boldsymbol{\zeta} \in \mathcal{D}_{\boldsymbol{\zeta}}} \rho_{\mathrm{d}}(\boldsymbol{\zeta}, \mathcal{X}_{\mathcal{D}}).$$

– Durchschnittlicher (funktionaler) Abstand zwischen zwei beliebigen Punkten aus einer Design-Menge (bei normierten Eingangsvariablen, beispielsweise $\mathcal{D}_{\boldsymbol{\zeta}} = [0,1]^d$):

$$m_{(\mathrm{p},\lambda)}(\mathcal{X}_{\mathcal{D}}) := \left( \frac{1}{\binom{N}{2}} \sum_{\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_j \in \mathcal{X}_{\mathcal{D}}} \left[ \frac{d^{1/p}}{\rho_{\mathrm{p}}(\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2)} \right]^{\lambda} \right)^{1/\lambda}, \quad \lambda \geq 1,$$

wobei $0 < \rho_{\mathrm{p}}(\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2) \leq d^{1/p}$, $\forall\, \boldsymbol{\zeta}_1 \neq \boldsymbol{\zeta}_2 \in [0,1]^d$ gilt.

∗ Design basierend auf optimalen durchschnittlichen (funktionalen) Abständen $\mathcal{X}_{\mathcal{D}_{\mathrm{av}}}$, z.B.

$$\mathcal{X}_{\mathcal{D}_{\mathrm{av}}} := \min_{\mathcal{X}_{\mathcal{D}} \subset \mathcal{D}_{\boldsymbol{\zeta}}} m_{(\mathrm{p},\lambda)}(\mathcal{X}_{\mathcal{D}}).$$

- Designs mit gleichverteilten Design-Punkten: Unter der Annahme von normierten Eingangsvariablen, d.h. $\mathcal{D}_{\boldsymbol{\zeta}} = \bigotimes_{i=1}^{d} [a_i, b_i]$, minimieren diese Designs die Abweichung zwischen der empirischen Verteilung der Design-Punkte $F_N(\boldsymbol{\zeta}) := \frac{1}{N} \sum_{i=1}^{N} I\{\mathbf{Z}_i \leq \boldsymbol{\zeta}_i\}$ aus einer Design-Menge $\boldsymbol{\zeta} \in \mathcal{X}_{\mathcal{D}}$ und der mehrdimensionalen Gleichverteilung $F(\boldsymbol{\zeta}) := \prod_{i=1}^{d} \left( \frac{\zeta_i - a_i}{b_i - a_i} \right)$. Die Abweichung ist definiert als

$$D_{\mathrm{p}}(\mathcal{X}_{\mathcal{D}}) := \left[ \int \cdots \int_{\mathcal{D}_{\boldsymbol{\zeta}}} |F_N(\boldsymbol{\zeta}) - F(\boldsymbol{\zeta})|^p \mathrm{d}\boldsymbol{\zeta} \right]^{1/p}$$

und das optimale Design $\mathcal{X}_{\mathcal{D}_{\mathrm{u}}}$ als

$$\mathcal{X}_{\mathcal{D}_{\mathrm{u}}} := \arg \min_{\mathcal{X}_{\mathcal{D}} \subset \mathcal{D}_{\boldsymbol{\zeta}}} D_{\mathrm{p}}(\mathcal{X}_{\mathcal{D}}).$$

### 8.2.4 Prädiktionsgenauigkeit

Um die Genauigkeit der Prädiktion zu quantifizieren werden zwei Maße eingeführt. Das erste Maß ist der empirische quadratische Mittelwert des

Prädiktionsfehlers (ERMSPE ***E**mpirical **R**oot **M**ean **S**quared **P**rediction **E**rror*)

$$\text{ERMSPE} = \sqrt{\frac{1}{T_n}\sum_{i=1}^{T_n}\left(h(\boldsymbol{\zeta}_i) - \hat{\mathcal{H}}(\boldsymbol{\zeta}_i)\right)^2} \tag{8.28}$$

wobei $T_n$ die Anzahl der Test-Punkte ist. Das zweite Maß ist die erzielte Deckung des wahren Funktionswertes durch das $1-\alpha$ Konfidenzintervall (wobei z.B. $\alpha = 0.05$) des Prädiktors. Für das Konfidenzintervall gilt

$$\text{P}\left(H_0 \in \text{E}\{H_0|\mathbf{H}=\boldsymbol{\eta}\} \pm \sqrt{\text{Var}\{H_0|\mathbf{H}=\boldsymbol{\eta}\}}z_{1-\alpha/2}\right) = 1-\alpha, \tag{8.29}$$

wobei $z_{1-\alpha/2}$ das $(1-\alpha/2)$-Quantil der *A-posteriori*-Dichte $[H_0|\mathbf{H}]$ ist. Die erzielte Deckung ist definitionsgemäß der Anteil der Test-Punkte AC $\in$ [0,1] (AC = ***A**chieved **C**overage*), deren wahren Funktionswerte innerhalb des $1-\alpha$ Konfidenzintervalls liegen.

### 8.2.5 Beispiel: Prädiktion einer Funktion mit einer Variablen

Folgendes Beispiel illustriert den partiellen Bayes'schen Ansatz anhand der folgenden Funktion aus [64, Beispiel 4.1]

$$\eta = h(\zeta) = e^{-1.4\zeta}\cos(7\pi\zeta/2), \quad 0 \le \zeta \le 1. \tag{8.30}$$

Für die Interpolation werden $N = 7$ *Design-Punkte* $\mathcal{X}_\mathcal{D} := \{\zeta_1, \ldots, \zeta_N\}$ durch *Latin-Hypercube-Sampling* (vgl. Abschnitt B.5 (Anhang)) erzeugt, welche die Trainingsdaten $\boldsymbol{\eta} := [\eta_1 \ \cdots \ \eta_N]$ generieren. Das angenommene Modell des Ausgangs lautet

$$\mathcal{H}(\zeta) = \beta + \mathcal{Z}(\zeta), \tag{8.31}$$

wobei der freie Koeffizient $\beta \in \mathbb{R}$ unbekannt und $\mathcal{Z}(\zeta)$ ein mittelwertfreies skalares Zufallsfeld mit einer noch unbekannten konstanten Varianz $\sigma_\mathcal{Z}^2 > 0$ sind. Als *A-priori*-Verteilung $[\boldsymbol{\beta}, \sigma_\mathcal{Z}^2]$ wird die *nicht-informative* Verteilung (4) aus Tabelle 8.2 verwendet. Der Ausgang $\mathcal{H}(\zeta)$ des Modells stellt - im Unterschied zum wahren und deterministischen Ausgangswert $\eta \in \mathbb{R}$ aus Gl. (8.30) - ein skalares Gauß'sches Zufallsfeld dar, dessen Realisierungen Funktionen von $\zeta$ sind. Ebenso wird der Vektor von Trainingsdaten als Zufallsvektor $\mathbf{H}$ (mit einer multivariaten Normalverteilung)

angenommen. Der Wert des skalaren Zufallsfeldes an einem Test-Punkt $\zeta_0 \notin \mathcal{X}_\mathcal{D}$ wird durch $H_0 := \mathcal{H}(\zeta_0)$ bezeichnet.

Unter diesen Annahmen besitzt die Zufallsvariable $H_0|\mathbf{H}$ eine *A-posteriori* nichtzentrale $t$-Verteilung (vgl. Theorem B.1 (Anhang)) mit $\nu_4 = 6$ Freiheitsgraden, sowie mit Nichtzentralitätsparameter $\mu_4(\zeta_0,\psi)$ und Skalierungsparameter $\sigma_4^2(\zeta_0,\psi)$. Die letzten Parameter können aus Theorem B.1 (Anhang) entnommen werden. Es gilt folglich

$$\frac{H_0|\mathbf{H} - \mu_4(\zeta_0,\psi)}{\sigma_4(\zeta_0,\psi)} \sim \mathcal{T}_1(\nu_4,0,1), \tag{8.32}$$

wobei $\psi$ den noch unbekannten Parameter der Korrelationsfunktion $R(\,\cdot\,|\psi)$ darstellt. Dafür wird die Gauß'sche Korrelationsfunktion gewählt, welche zu der *Potenz-Exponential-Familie* aus Gl. (8.4), mit $p = 2$, gehört. Der Parameter $\psi$ wurde mit Hilfe der Maximum Likelihood Methode aus Abschnitt 8.2.1 auf $\hat{\psi} = 0.2965$ geschätzt.

Bild 8.2 zeigt den Verlauf der wahren Funktion $h(\zeta)$, sowie den Prädiktor $\hat{H}_0 = \mathrm{E}\{H_0|\mathbf{H} = \boldsymbol{\eta}\} = \mu_4(\zeta_0,\hat{\psi})$ und die punktweise entsprechenden 95% Prädiktionsintervallgrenzen $\hat{H}_0 \pm \sigma_4(\zeta_0,\hat{\psi})t_{\nu_4}^{\alpha/2}$ des Prädiktors für 200 äquidistante Test-Punkte $\zeta_0$ aus dem Intervall $\zeta \in [0,1]$. Dabei stellt $t_{\nu_4}^{\alpha/2} = F_{H_0|\mathbf{H}}^{-1}(1-\alpha/2|\nu_4 = 6) = 2.4469$ den oberen $\alpha/2$ kritischen Punkt der Verteilung $\mathcal{T}_1(\nu_4,0,1)$, mit $\alpha = 0.05$, dar. Dies bedeutet, dass

$$\mathrm{P}\left\{ H_0 \in \mu_4(\zeta_0,\hat{\psi}) \pm \sigma_4(\zeta_0,\hat{\psi})t_{\nu_4}^{\alpha/2} \,\middle|\, \mathbf{H} \right\} = 1 - \alpha. \tag{8.33}$$

Die Untersuchung der Prädiktionsgenauigkeit, welche anhand der Maße aus Abschnitt 8.2.4 quantifiziert wird, ergibt einen empirischen quadratischen Mittelwert des Prädiktionsfehlers von ERMSPE = 0.034 und einen Anteil der erzielten Deckung von AC = 1.

## 8.3 Sensitivitätsanalyse

Wie in den vorherigen Abschnitten beschrieben, bildet der Ausgang eines *Computerexperiments* eine unbekannte deterministische Funktion $\eta = h(\boldsymbol{\zeta}) = h(\zeta_1,\ldots,\zeta_d)$. Die Sensitivitätsanalyse beschäftigt sich mit der Untersuchung des Einflusses eines oder mehrerer Eingänge $\zeta_i, i = 1,\ldots,d$, auf den Ausgang $\eta$. Haben beispielsweise bestimmte Eingänge einen nur sehr kleinen Einfluß auf den Ausgang, so können sie bei der Prädiktion

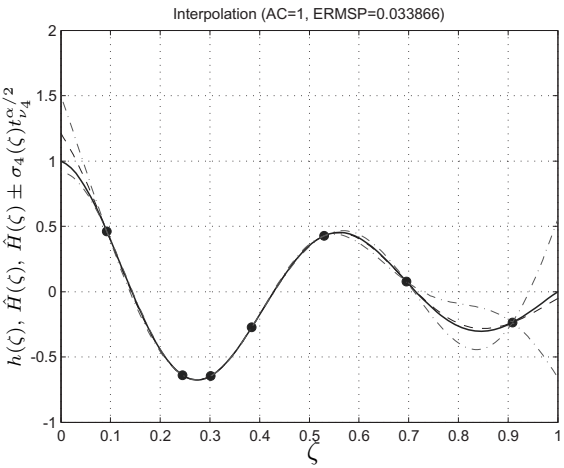


**Bild 8.2:** Verlauf der Funktion $h(\zeta)$ (-), Menge der Trainingsdaten $\boldsymbol{\eta}$ (•), Prädiktor $\hat{H}_0 = \mu_4(\zeta_0)$ (- -), sowie Prädiktionsintervallgrenzen $\mu_4(\zeta_0,\hat{\psi}) \pm \sigma_4(\zeta_0,\hat{\psi}) t_{\nu_4}^{\alpha/2}$ (-.) für 200 äquidistante Test-Punkte $\zeta_0$ aus dem Intervall [0,1].

des Ausgangs vernachlässigt werden. Dies würde zu einer Vereinfachung der Untersuchung führen. Darüber hinaus hilft die Sensitivitätsanalyse Interaktionen zwischen den Eingängen $\zeta_i$, $i = 1,\ldots,d$, zu identifizieren. Existieren solche Interaktionen nicht, dann ist der Einfluß eines bestimmten Eingangs unabhängig von den anderen Eingängen. Dies ist vor allem im Rahmen des *Designs* (der Auswahl der Design-Punkte) von Bedeutung. Die Änderung der Eingänge kann dabei infinitesimal klein (*lokale* Sensitivitätsanalyse) oder groß (*globale* Sensitivitätsanalyse) sein.

Die Sensitivitätsanalyse kann auf den Ausgang $\eta = h(\boldsymbol{\zeta})$ des Versuchs oder auf den Prädiktor $\hat{\mathcal{H}}(\boldsymbol{\zeta})$ angewandt werden. Der wesentliche Unterschied besteht darin, dass bei der Anwendung auf den Prädiktor, die Schlußfolgerungen das vorgeschlagene Modell, und nicht das *Computerexperiment* - wie im ersten Fall - betreffen. In Abschnitt 8.3.1 werden die Sensitivitätsmaße und in Abschnitt 8.3.2 die Bayes'sche Inferenz vorgestellt.

### 8.3.1 Sensitivitätsmaße

Die in dieser Arbeit untersuchten Sensitivitätsmaße, welche in [51] eingeführt wurden, basieren entweder auf der Zerlegung des Ausgangs des Modells $h(\boldsymbol{\zeta})$ oder $\mathcal{H}(\boldsymbol{\zeta})$ in *Haupteffekte* und *Interaktionen* - je nachdem ob einzelne Eingänge oder mehrere Eingänge gleichzeitig betrachtet werden - oder aber auf der Reduktion der Varianz des Ausgangs infolge Fixierung eines oder mehrerer Eingänge.

#### Ausgangszerlegung in Haupteffekte und Interaktionen

Die Haupteffekte und Interaktionen werden durch eine Zerlegung des Ausgangs des Modells konstruiert. Die Komponenten der Zerlegung beinhalten jeweils einen oder mehrere Eingänge aus $\mathbf{Z} := [Z_1 \quad \cdots \quad Z_d]$.[6)] Die Komponenten, welche einen einzigen Eingang betrachten, heißen *Haupteffekte*, die restlichen heißen *Interaktionen*. Alle Eingänge werden dabei als unabhängige Zufallsvariablen mit jeweils vorgegebener Verteilung betrachtet. Die Zerlegung des Modells hat die Form

$$\mathcal{H}(\boldsymbol{\zeta}) = \mathrm{E}\{\mathcal{H}\} + \sum_{i=1}^{d} \mathcal{C}_i(\zeta_i) + \sum_{i<j} \mathcal{C}_{i,j}(\boldsymbol{\zeta}_{i,j}) + \sum_{i<j<k} \mathcal{C}_{i,j,k}(\boldsymbol{\zeta}_{i,j,k}) + \\ \ldots + \mathcal{C}_{1,2,\ldots,d}(\boldsymbol{\zeta}) \tag{8.34}$$

mit $\boldsymbol{\zeta}_{i,j} := [\zeta_i \quad \zeta_j]$ und $\boldsymbol{\zeta}_{i,j,k} := [\zeta_i \quad \zeta_j \quad \zeta_k]$. Die einzelnen Komponenten sind definitionsgemäß

$$\mathcal{C}_i(\zeta_i) := \mathrm{E}_{\mathbf{Z}_{\overline{i}}}\{\mathcal{H}(\boldsymbol{\zeta})|Z_i\} - \mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}, \tag{8.35}$$

$$\boldsymbol{\mathcal{C}_{i,j}}(\boldsymbol{\zeta}_{i,j}) := \mathrm{E}_{\mathbf{Z}_{\overline{(i,j)}}}\{\mathcal{H}(\boldsymbol{\zeta})|\mathbf{Z}_{i,j}\} - \mathcal{C}_i(\zeta_i) - \mathcal{C}_j(\zeta_j) - \mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}, \tag{8.36}$$

$$\boldsymbol{\mathcal{C}}_{i,j,k}(\boldsymbol{\zeta}_{i,j,k}) := \mathrm{E}_{\mathbf{Z}_{\overline{(i,j,k)}}}\{\mathcal{H}(\boldsymbol{\zeta})|\mathbf{Z}_{i,j,k}\} - \boldsymbol{\mathcal{C}}_{i,j}(\boldsymbol{\zeta}_{i,j}) - \boldsymbol{\mathcal{C}}_{i,k}(\boldsymbol{\zeta}_{i,k}) - \boldsymbol{\mathcal{C}}_{j,k}(\boldsymbol{\zeta}_{j,k}) \\ - \mathcal{C}_i(\zeta_i) - \mathcal{C}_j(\zeta_j) - \mathcal{C}_k(\zeta_k) - \mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}, \tag{8.37}$$

$$\vdots$$

wobei der letzte Term $\boldsymbol{\mathcal{C}}_{1,2,\ldots,d}(\boldsymbol{\zeta})$ als Differenz zwischen der Summe aller anderen Komponenten und dem Ausgang des Modells $\mathcal{H}(\boldsymbol{\zeta})$ definiert wird.

[6)] Wir unterscheiden dabei durch diese Notation den Zufallsvektor $\mathbf{Z}$ von seiner Realisierung $\boldsymbol{\zeta}$.

Zusammengefasst kann ein solcher Effekt beschrieben werden als

$$\boldsymbol{\mathcal{C}}_p(\boldsymbol{\zeta}_p) := \mathrm{E}_{\mathbf{Z}_{\overline{p}}}\{H|\mathbf{Z}_p\} - \sum_{k=1}^{\nu_p} \sum_{\pi_k \subseteq p} \boldsymbol{\mathcal{C}}_{\pi_k}(\boldsymbol{\zeta}_{\pi_k}) - \mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}, \tag{8.38}$$

wobei $p$ eine Menge von $\nu_p$ Indizes, $\pi_k \subseteq p$ eine Untermenge von $k$ Indizes aus der Menge $p$ und $\overline{p}$ die Menge der $d - \nu_p$ Indizes darstellt, welche sich nicht in der Menge $p$ befinden. Die Zerlegung aus Gl. (8.34) wird in der Literatur als *höherdimensionale Modellbeschreibung* (HDMR ***H**igh **D**imensional **M**odel **R**epresentation*) bezeichnet. Dabei sind $\mathcal{C}_i(\zeta_i)$ die *Haupteffekte* und $\boldsymbol{\mathcal{C}}_p(\boldsymbol{\zeta}_p)$ die *Interaktionen p-ter Ordnung*.

### Ausgangsvarianzreduktion

Ein Sensitivitätsmaß kann die Reduktion der Varianz $\mathrm{Var}\{\mathcal{H}(\boldsymbol{\zeta})\}$ des Zufallsfeldes $\mathcal{H}(\boldsymbol{\zeta})$ in Abhängigkeit von einem oder mehreren Eingängen quantifizieren. Im Fall eines einzigen Eingangs hat dieses die Form[7)]

$$V_i := \mathrm{Var}_{Z_i}\{\mathrm{E}_{\mathbf{Z}_{\overline{i}}}\{\mathcal{H}(\boldsymbol{\zeta})|Z_i\}\}. \tag{8.39}$$

Dessen Wahl kann man wie folgt erklären. Sei $\mathrm{Var}_{\mathbf{Z}_{\overline{i}}}\{\mathcal{H}(\boldsymbol{\zeta})|Z_i = \zeta_i^*\}$ die resultierende Varianz von $\mathcal{H}(\boldsymbol{\zeta})$ infolge der Unsicherheit über alle Eingänge $\mathbf{Z}$ außer $Z_i$. Diese wird *bedingte Varianz* genannt. Eine intuitive Folge der Fixierung eines Eingangs $Z_i = \zeta_i^*$ ist, dass die resultierende Varianz $\mathrm{Var}_{\mathbf{Z}_{\overline{i}}}\{\mathcal{H}(\boldsymbol{\zeta})|Z_i = \zeta_i^*\}$ kleiner als die *unbedingte Varianz* $\mathrm{Var}\{\mathcal{H}(\boldsymbol{\zeta})\}$ ist. Je kleiner dieser Wert ist, desto wichtiger ist der Faktor $Z_i$. Dieser Wert ist jedoch abhängig von dem festgelegten Wert von $\zeta_i^*$. Daher wird noch der Mittelwert über alle möglichen Werte von $\zeta_i^*$ angewendet, d.h. $\mathrm{E}_{Z_i}\{\mathrm{Var}_{\mathbf{Z}_{\overline{i}}}\{\mathcal{H}(\boldsymbol{\zeta})|Z_i = \zeta_i\}\}$, welches immer kleiner oder gleich der unbedingten Varianz $\mathrm{Var}\{\mathcal{H}(\boldsymbol{\zeta})\}$ ist. Aus dem Satz von der totalen Varianz, folgt darüber hinaus

$$\mathrm{E}_{Z_i}\{\mathrm{Var}_{\mathbf{Z}_{\overline{i}}}\{\mathcal{H}(\boldsymbol{\zeta})|Z_i\}\} + \mathrm{Var}_{Z_i}\{\mathrm{E}_{\mathbf{Z}_{\overline{i}}}\{\mathcal{H}(\boldsymbol{\zeta})|Z_i\}\} = \mathrm{Var}\{\mathcal{H}(\boldsymbol{\zeta})\}. \tag{8.40}$$

Aus Gl. (8.40) ist ersichtlich, dass der Wert $\mathrm{Var}_{Z_i}\{\mathrm{E}_{\mathbf{Z}_{\overline{i}}}\{\mathcal{H}(\boldsymbol{\zeta})|Z_i\}\}$ das Ausmaß der Varianzreduktion des Ausgangs durch die Fixierung des Eingangs $Z_i$ quantifiziert. Durch eine Normierung mit $\mathrm{Var}\{\mathcal{H}(\boldsymbol{\zeta})\}$ erzielt man das Sensitivitätsmaß

$$S_i := \frac{\mathrm{Var}_{Z_i}\{\mathrm{E}_{\mathbf{Z}_{\overline{i}}}\{\mathcal{H}(\boldsymbol{\zeta})|Z_i\}\}}{\mathrm{Var}\{\mathcal{H}(\boldsymbol{\zeta})\}}, \tag{8.41}$$

[7)] Vgl. [63, Kapitel 1].

welches varianzbasierter *Haupteffektindex* des Eingangs $Z_i$ genannt wird. Zusammengefasst können die varianzbasierten Haupteffekt- und Interaktionenindizes beschrieben werden als

$$S_p := \frac{\mathrm{Var}_{\mathbf{Z}_p}\{\mathrm{E}_{\mathbf{Z}_{\overline{p}}}\{\mathcal{H}(\boldsymbol{\zeta})|\mathbf{Z}_p\}\}}{\mathrm{Var}\{\mathcal{H}(\boldsymbol{\zeta})\}}. \tag{8.42}$$

## 8.3.2 Bayes'sche Inferenz

Da der Ausgang des *Computerexperiments* eine unbekannte Funktion darstellt, können diese Sensitivitätsmaße nicht exakt berechnet werden. Sie werden daher, wie der Ausgang selbst, als Zufallsfelder (im Fall der Haupteffekte und Interaktionen) bzw. Zufallsvariablen (im Fall der Haupteffekt- und Interaktionenindizes) angenommen. Mittels Bayes'scher Inferenz werden dann für diese Sensitivitätsmaße *A-posteriori*-Erwartungswerte berechnet.

### Haupteffekte und Interaktionen

Die Haupteffekte und Interaktionen aus Gl. (8.35)-(8.37) hängen von dem Erwartungswert bzw. den Erwartungswertfunktionen der jeweiligen bedingten Zufallsfelder ab. Diese sind gegeben durch

$$\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\} = \int_{\mathcal{D}_{\zeta_1}} \cdots \int_{\mathcal{D}_{\zeta_d}} \mathcal{H}(\boldsymbol{\zeta}) \cdot f_1(\zeta_1) \cdot \ldots \cdot f_d(\zeta_d) \mathrm{d}\zeta_1 \cdots \mathrm{d}\zeta_d, \tag{8.43}$$

$$\begin{aligned} \mathrm{E}_{\mathbf{Z}_{\overline{i}}}\{\mathcal{H}(\boldsymbol{\zeta})|Z_i\} = \int_{\mathcal{D}_{\zeta_1}} \cdots \int_{\mathcal{D}_{\zeta_{i-1}}} \int_{\mathcal{D}_{\zeta_{i+1}}} \cdots \int_{\mathcal{D}_{\zeta_d}} \mathcal{H}(\boldsymbol{\zeta}) \\ \cdot f_1(\zeta_1) \cdot \ldots \cdot f_{i-1}(\zeta_{i-1}) \\ \cdot f_{i+1}(\zeta_{i+1}) \cdot \ldots \cdot f_d(\zeta_d) \mathrm{d}\zeta_1 \cdots \mathrm{d}\zeta_{i-1} \mathrm{d}\zeta_{i+1} \cdots \mathrm{d}\zeta_d, \end{aligned} \tag{8.44}$$

wobei $\mathcal{D}_{\boldsymbol{\zeta}_i}$, $i = 1, \ldots, d$, den Definitionsbereich der Zufallsvariable $Z_i$ darstellt, sowie - der Einfachheit halber in Vektorschreibweise dargestellt -

$$\mathrm{E}_{\mathbf{Z}_{\overline{(i,j)}}}\{\mathcal{H}(\boldsymbol{\zeta})|\mathbf{Z}_{i,j}\} = \int\cdots\int_{\substack{\bigotimes_{l=1}^{d} \mathcal{D}_{\boldsymbol{\zeta}_l} \\ l\neq i,j}} \mathcal{H}(\boldsymbol{\zeta}) \cdot \prod_{\substack{l=1 \\ l\neq i,j}}^{d} f_l(\zeta_l)\,\mathrm{d}\boldsymbol{\zeta}_{\overline{(i,j)}}, \tag{8.45}$$

$$\mathrm{E}_{\mathbf{Z}_{\overline{(i,j,k)}}}\{\mathcal{H}(\boldsymbol{\zeta})|\mathbf{Z}_{i,j,k}\} = \int\cdots\int_{\substack{\bigotimes_{l=1}^{d} \mathcal{D}_{\boldsymbol{\zeta}_l} \\ l\neq i,j,k}} \mathcal{H}(\boldsymbol{\zeta}) \cdot \prod_{\substack{l=1 \\ l\neq i,j,k}}^{d} f_l(\zeta_l)\,\mathrm{d}\boldsymbol{\zeta}_{\overline{(i,j,k)}}, \tag{8.46}$$

$$\vdots$$

Zusammengefasst können die Komponenten aus Gl. (8.43)-(8.46) für einen beliebigen Vektor $\boldsymbol{\zeta}_p$, mit einer Menge $p$ von $\nu_p$ Indizes, geschrieben werden als

$$\mathrm{E}_{\mathbf{Z}_{\overline{p}}}\{\mathcal{H}(\boldsymbol{\zeta})|\mathbf{Z}_p\} = \int\cdots\int_{\mathcal{D}_{\boldsymbol{\zeta}_{\overline{p}}}} \mathcal{H}(\boldsymbol{\zeta})\mathrm{d}\mathbf{G}_{\overline{p}}(\boldsymbol{\zeta}_{\overline{p}}), \tag{8.47}$$

mit

$$\mathrm{d}\mathbf{G}_{\overline{p}}(\boldsymbol{\zeta}_{\overline{p}}) := \prod_{\substack{l=1 \\ l\neq p_{(i)} \\ i=1,\ldots,\nu_p}}^{d} f_l(\zeta_l)\,\mathrm{d}\boldsymbol{\zeta}_{\overline{p}}$$

und

$$\mathcal{D}_{\boldsymbol{\zeta}_{\overline{p}}} := \bigotimes_{\substack{l=1 \\ l\neq p_{(i)} \\ i=1,\ldots,\nu_p}}^{d} \mathcal{D}_{\boldsymbol{\zeta}_l}.$$

Der Term $\mathrm{E}_{\mathbf{Z}_{\overline{p}}}\{\mathcal{H}(\boldsymbol{\zeta})|\mathbf{Z}_p\}$ aus Gl. (8.47) ist ein lineares Funktional des skalaren Zufallsfeldes $\mathcal{H}(\boldsymbol{\zeta})$. Somit wird das Zufallsfeld (oder die Zufallsvariable im Fall von $p = \emptyset$) $\mathrm{E}_{\mathbf{Z}_{\overline{p}}}\{\mathcal{H}(\boldsymbol{\zeta})|\mathbf{Z}_p\}$ für jeden Punkt $\boldsymbol{\zeta}_p$ eine nichtzentrale $t$-Verteilung wie in Gl. (8.18), Seite 129, haben, jedoch mit anderen Parametern. Der *A-posteriori*-Erwartungswert ist

$$\mathrm{E}^*\{\mathrm{E}_{\mathbf{Z}_{\overline{p}}}\{\mathcal{H}(\boldsymbol{\zeta})|\mathbf{Z}_p\}\} = \int\cdots\int_{\mathcal{D}_{\boldsymbol{\zeta}_{\overline{p}}}} \mathrm{E}^*\{\mathcal{H}(\boldsymbol{\zeta})\}\mathrm{d}\mathbf{G}_{\overline{p}}(\boldsymbol{\zeta}_{\overline{p}}), \tag{8.48}$$

wobei das Symbol $*$ die Berechnung des Erwartungswertes relativ zur *A-posteriori*-Wahrscheinlichkeitsdichte von $\mathrm{E}_{\mathbf{Z}_{\overline{p}}}\{\mathcal{H}(\boldsymbol{\zeta})|\mathbf{Z}_p\}$ bezeichnet. Aus

Gl. (8.17), Seite 128, folgt darüber hinaus, dass

$$\mathrm{E}^*\{\mathcal{H}(\boldsymbol{\zeta})\} := \mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})|\mathbf{H}\} = \mathbf{f}(\boldsymbol{\zeta})^\top\hat{\boldsymbol{\beta}} + \hat{\mathbf{r}}(\boldsymbol{\zeta})^\top\hat{\mathbf{R}}^{-1}(\boldsymbol{\eta} - \mathbf{F}\hat{\boldsymbol{\beta}}). \tag{8.49}$$

Durch Einsetzen der Gl. (8.49) in Gl. (8.48) folgt

$$\mathrm{E}^*\{\mathrm{E}_{\mathbf{Z}_{\overline{p}}}\{\mathcal{H}(\boldsymbol{\zeta})|\mathbf{Z}_p\}\} = \mathbf{s}_p^\top(\boldsymbol{\zeta}_p)\hat{\boldsymbol{\beta}} + \mathbf{t}_p^\top(\boldsymbol{\zeta}_p)\hat{\mathbf{R}}^{-1}(\boldsymbol{\eta} - \mathbf{F}\hat{\boldsymbol{\beta}}), \tag{8.50}$$

mit

$$\mathbf{s}_p^\top(\boldsymbol{\zeta}_p) := \int\cdots\int_{\mathcal{D}_{\boldsymbol{\zeta}_{\overline{p}}}} \mathbf{f}(\boldsymbol{\zeta})^\top \mathrm{d}\mathbf{G}_{\overline{p}}(\boldsymbol{\zeta}_{\overline{p}}), \tag{8.51}$$

$$\mathbf{t}_p^\top(\boldsymbol{\zeta}_p) := \int\cdots\int_{\mathcal{D}_{\boldsymbol{\zeta}_{\overline{p}}}} \hat{\mathbf{r}}(\boldsymbol{\zeta})^\top \mathrm{d}\mathbf{G}_{\overline{p}}(\boldsymbol{\zeta}_{\overline{p}}). \tag{8.52}$$

Die Integrale aus Gl. (8.51) und (8.52) können nur in wenigen Fällen analytisch berechnet werden. Eine numerische Lösung ist jedoch unproblematisch. Auf die gleiche Weise kann auch der *A-posteriori*-Erwartungswert $\mathrm{E}^*\{\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}\}$ für den Erwartungswert aus Gl. (8.43) berechnet werden. Es folgt

$$\mathrm{E}^*\{\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}\} = \mathbf{s}^\top\hat{\boldsymbol{\beta}} + \mathbf{t}^\top\hat{\mathbf{R}}^{-1}(\boldsymbol{\eta} - \mathbf{F}\hat{\boldsymbol{\beta}}), \tag{8.53}$$

mit $\mathbf{s}^\top := \mathbf{s}_p^\top(\boldsymbol{\zeta}_p)$ und $\mathbf{t}^\top := \mathbf{t}_p^\top(\boldsymbol{\zeta}_p)$ für $p = \emptyset$. Somit können die *A-posteriori*-Erwartungswerte der Haupteffekte und Interaktionen aus Gl. (8.38) unter Verwendung der Gl. (8.50) und Gl. (8.53) berechnet werden.

Für die Inferenz über die bedingte Ausgangsvarianzreduktion wird noch die *A-posteriori*-Kovarianzfunktion des skalaren Zufallsfeldes $\mathrm{E}_{\mathbf{Z}_{\overline{p}}}\{\mathcal{H}(\boldsymbol{\zeta})|\mathbf{Z}_p\}$ benötigt. Diese ist

$$\begin{aligned}
&\mathrm{Cov}^*\{\mathrm{E}_{\mathbf{Z}_{\overline{p}}}\{\mathcal{H}(\boldsymbol{\zeta})|\mathbf{Z}_p\}, \mathrm{E}_{\mathbf{Z}'_{\overline{q}}}\{\mathcal{H}(\boldsymbol{\zeta})|\mathbf{Z}'_q\}\} \\
&\quad = \int\cdots\int_{\mathcal{D}_{\boldsymbol{\zeta}_{\overline{p}}}}\int\cdots\int_{\mathcal{D}_{\boldsymbol{\zeta}'_{\overline{q}}}} \mathrm{Cov}^*\{\mathcal{H}(\boldsymbol{\zeta})\}\mathrm{d}\mathbf{G}_{\overline{p}}(\boldsymbol{\zeta}_{\overline{p}})\mathrm{d}\mathbf{G}_{\overline{q}}(\boldsymbol{\zeta}'_{\overline{q}}) \\
&\quad = \int\cdots\int_{\mathcal{D}_{\boldsymbol{\zeta}_{\overline{p}}}}\int\cdots\int_{\mathcal{D}_{\boldsymbol{\zeta}'_{\overline{q}}}} \hat{\sigma}^2_{\mathcal{Z}} r^*(\boldsymbol{\zeta},\boldsymbol{\zeta}')\mathrm{d}\mathbf{G}_{\overline{p}}(\boldsymbol{\zeta}_{\overline{p}})\mathrm{d}\mathbf{G}_{\overline{q}}(\boldsymbol{\zeta}'_{\overline{q}}),
\end{aligned} \tag{8.54}$$

wobei $\boldsymbol{\zeta} := (\boldsymbol{\zeta}_p,\boldsymbol{\zeta}_{\overline{p}})$, $\boldsymbol{\zeta}' := (\boldsymbol{\zeta}_q,\boldsymbol{\zeta}_{\overline{q}})$, der Wert $\hat{\sigma}^2_{\mathcal{Z}}$ eine Schätzung der *A-priori*-Varianz und die Funktion $r^*(\boldsymbol{\zeta},\boldsymbol{\zeta}')$ die *A-posteriori*-Korrelationsfunktion des skalaren Zufallsfeldes $\mathcal{Z}$ sind. Im Fall der *A-priori*-Verteilungen aus dem vorherigen Abschnitt werden diese Größen in Gl. (B.18) und (B.19) des Satzes B.1 (Anhang) angegeben.

**Bedingte Ausgangsvarianzreduktion**

Schließlich wird der *A-posteriori*-Erwartungswert der Sensitivitätsmaße $S_p$ aus Gl. (8.42) berechnet. Dieser wird durch

$$\mathrm{E}^*\{S_p\} \approx \frac{\mathrm{E}^*\{\mathrm{Var}_{\mathbf{Z}_p}\{\mathrm{E}_{\mathbf{Z}_{\overline{p}}}\{\mathcal{H}(\boldsymbol{\zeta})|\mathbf{Z}_p\}\}\}}{\mathrm{E}^*\{\mathrm{Var}\{\mathcal{H}(\boldsymbol{\zeta})\}\}} \tag{8.55}$$

angenähert, vgl. [51], da der exakte Erwartungswert nicht analytisch berechnet werden kann. Dabei gilt

$$\begin{aligned}\mathrm{Var}_{\mathbf{Z}_p}\{\mathrm{E}_{\mathbf{Z}_{\overline{p}}}\{\mathcal{H}(\boldsymbol{\zeta})|\mathbf{Z}_p\}\} &= \mathrm{E}_{\mathbf{Z}_p}\{\mathrm{E}_{\mathbf{Z}_{\overline{p}}}\{\mathcal{H}(\boldsymbol{\zeta})|\mathbf{Z}_p\}^2\} - \mathrm{E}_{\mathbf{Z}_p}\{\mathrm{E}_{\mathbf{Z}_{\overline{p}}}\{\mathcal{H}(\boldsymbol{\zeta})|\mathbf{Z}_p\}\}^2 \\ &= \mathrm{E}_{\mathbf{Z}_p}\{\mathrm{E}_{\mathbf{Z}_{\overline{p}}}\{\mathcal{H}(\boldsymbol{\zeta})|\mathbf{Z}_p\}^2\} - \mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}^2. \end{aligned} \tag{8.56}$$

Der *A-posteriori*-Erwartungswert des zweiten Terms auf der rechten Seite von Gl. (8.56) ist[8)]

$$\begin{aligned}\mathrm{E}^*\{\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}^2\} &:= \mathrm{E}\{\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}^2|\mathbf{H}\} \\ &= \mathrm{Var}\{\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}|\mathbf{H}\} + \mathrm{E}\{\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}|\mathbf{H}\}^2 \\ &= \mathrm{Var}^*\{\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}\} + (\mathrm{E}^*\{\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}\})^2, \end{aligned} \tag{8.57}$$

wobei der erste Term aus Gl. (8.54), mit $\mathbf{Z}_p = \mathbf{Z}'_q$ und $q = p = \emptyset$, und der zweite Term aus Gl. (8.53) entnommen werden können.

Der *A-posteriori*-Erwartungswert des ersten Terms auf der rechten Seite von Gl. (8.56), $\mathrm{E}_{\mathbf{Z}_p}\{\mathrm{E}_{\mathbf{Z}_{\overline{p}}}\{\mathcal{H}(\boldsymbol{\zeta})|\mathbf{Z}_p\}^2\}$, lautet

$$\begin{aligned}&\mathrm{E}^*\left\{\mathrm{E}_{\mathbf{Z}_p}\left\{\mathrm{E}_{\mathbf{Z}_{\overline{p}}}\{\mathcal{H}(\boldsymbol{\zeta})|\mathbf{Z}_p\}^2\right\}\right\} \\ &\qquad = \mathrm{E}^*\left\{\mathrm{E}_{\mathbf{Z}_p}\left\{\mathrm{E}_{\mathbf{Z}_{\overline{p}}}\{\mathcal{H}(\boldsymbol{\zeta})|\mathbf{Z}_p\} \cdot \mathrm{E}_{\mathbf{Z}_{\overline{p}}}\{\mathcal{H}(\boldsymbol{\zeta})|\mathbf{Z}_p\}\right\}\right\}. \end{aligned}$$

Unter Verwendung der Gl. (8.47), Seite 141, ist dieser weiterhin äquivalent zu

$$\begin{aligned}&\mathrm{E}^*\left\{\mathrm{E}_{\mathbf{Z}_p}\{\mathrm{E}_{\mathbf{Z}_{\overline{p}}}\{\mathcal{H}(\boldsymbol{\zeta})|\mathbf{Z}_p\}^2\}\right\} = \\ &= \mathrm{E}^*\left\{\mathrm{E}_{\mathbf{Z}_p}\left\{\int\cdots\int_{\mathcal{D}_{\boldsymbol{\zeta}_{\overline{p}}}} \mathcal{H}(\boldsymbol{\zeta})\mathrm{d}\mathbf{G}_{\overline{p}}(\boldsymbol{\zeta}_{\overline{p}}) \cdot \int\cdots\int_{\mathcal{D}_{\boldsymbol{\zeta}'_{\overline{p}}}} \mathcal{H}(\boldsymbol{\zeta}^\star)\mathrm{d}\mathbf{G}'_{\overline{p}}(\boldsymbol{\zeta}'_{\overline{p}})\right\}\right\} \\ &= \mathrm{E}^*\left\{\mathrm{E}_{\mathbf{Z}_p}\left\{I(\boldsymbol{\zeta}_p)\right\}\right\}, \end{aligned} \tag{8.58}$$

[8)] Dies folgt aus dem Verschiebungssatz, d.h. $\mathrm{Var}\{X|Y\} = \mathrm{E}\{X^2|Y\} - \mathrm{E}\{X|Y\}^2$.

mit

$$I(\boldsymbol{\zeta}_p) := \int\cdots\int_{\mathcal{D}_{\boldsymbol{\zeta}_{\overline{p}}}} \mathcal{H}(\boldsymbol{\zeta})\mathrm{d}\mathbf{G}_{\overline{p}}(\boldsymbol{\zeta}_{\overline{p}}) \cdot \int\cdots\int_{\mathcal{D}_{\boldsymbol{\zeta}'_{\overline{p}}}} \mathcal{H}(\boldsymbol{\zeta}^\star)\mathrm{d}\mathbf{G}'_{\overline{p}}(\boldsymbol{\zeta}'_{\overline{p}}), \tag{8.59}$$

sowie $\boldsymbol{\zeta} := (\boldsymbol{\zeta}_p, \boldsymbol{\zeta}_{\overline{p}})$, $\boldsymbol{\zeta}^\star := (\boldsymbol{\zeta}_p, \boldsymbol{\zeta}'_{\overline{p}})$ und

$$\mathrm{d}\mathbf{G}_{\overline{p}}(\boldsymbol{\zeta}_{\overline{p}}) := \prod_{\substack{l=1 \\ l\neq p_{(i)} \\ i=1,\ldots,\nu_p}}^{d} f_l(\zeta_l)\,\mathrm{d}\boldsymbol{\zeta}_{\overline{p}},$$

$$\mathrm{d}\mathbf{G}'_{\overline{p}}(\boldsymbol{\zeta}'_{\overline{p}}) := \prod_{\substack{l=1 \\ l\neq p_{(i)} \\ i=1,\ldots,\nu_p}}^{d} f_l(\zeta_l)\,\mathrm{d}\boldsymbol{\zeta}'_{\overline{p}}.$$

Weiterhin folgt aus Gl. (8.58), dass

$$\mathrm{E}^*\left\{\mathrm{E}_{\mathbf{Z}_p}\{I(\boldsymbol{\zeta}_p)\}\right\} = \mathrm{E}^*\left\{\int\cdots\int_{\mathcal{D}_{\boldsymbol{\zeta}_p}} I(\boldsymbol{\zeta}_p)\mathrm{d}\mathbf{G}_p(\boldsymbol{\zeta}_p)\right\}, \tag{8.60}$$

mit

$$\mathrm{d}\mathbf{G}_p(\boldsymbol{\zeta}_p) := \prod_{l=1}^{\nu_p} f_{p_{(l)}}(\zeta_{p_{(l)}})\,\mathrm{d}\boldsymbol{\zeta}_p. \tag{8.61}$$

Somit ergibt sich

$$\begin{aligned}&\mathrm{E}^*\left\{\mathrm{E}_{\mathbf{Z}_p}\{\mathrm{E}_{\mathbf{Z}_{\overline{p}}}\{\mathcal{H}(\boldsymbol{\zeta})|\mathbf{Z}_p\}^2\}\right\} = \\ &= \int\cdots\int_{\mathcal{D}_{\boldsymbol{\zeta}_p}}\int\cdots\int_{\mathcal{D}_{\boldsymbol{\zeta}_{\overline{p}}}}\int\cdots\int_{\mathcal{D}_{\boldsymbol{\zeta}'_{\overline{p}}}} \mathrm{E}^*\{\mathcal{H}(\boldsymbol{\zeta})\mathcal{H}(\boldsymbol{\zeta}^\star)\}\mathrm{d}\mathbf{G}_{\overline{p}}(\boldsymbol{\zeta}_{\overline{p}})\mathrm{d}\mathbf{G}'_{\overline{p}}(\boldsymbol{\zeta}'_{\overline{p}})\mathrm{d}\mathbf{G}_p(\boldsymbol{\zeta}_p),\end{aligned}$$

wobei

$$\mathrm{E}^*\{\mathcal{H}(\boldsymbol{\zeta})\mathcal{H}(\boldsymbol{\zeta}^\star)\} = \mathrm{E}^*\{\mathcal{H}(\boldsymbol{\zeta})\}\,\mathrm{E}^*\{\mathcal{H}(\boldsymbol{\zeta}^\star)\} + \mathrm{Cov}^*\{\mathcal{H}(\boldsymbol{\zeta}),\mathcal{H}(\boldsymbol{\zeta}^\star)\}$$

mit

$$\begin{aligned}\mathrm{E}^*\{\mathcal{H}(\boldsymbol{\zeta})\} &= \mu_i(\boldsymbol{\zeta}),\\ \mathrm{E}^*\{\mathcal{H}(\boldsymbol{\zeta}^\star)\} &= \mu_i(\boldsymbol{\zeta}^\star),\\ \mathrm{Cov}^*\{\mathcal{H}(\boldsymbol{\zeta}),\mathcal{H}(\boldsymbol{\zeta}^\star)\} &= \hat{\sigma^2_{\mathcal{Z}}} \cdot r^*(\boldsymbol{\zeta},\boldsymbol{\zeta}^\star),\end{aligned}$$

sowie $\mu_i(\,\cdot\,)$, $\hat{\sigma^2_{\mathcal{Z}}}$ und $r^*(\boldsymbol{\zeta},\boldsymbol{\zeta}^\star)$ aus Theorem B.1 entnommen werden können.

Schließlich wird der *A-posteriori*-Erwartungswert der Varianz des Zufallsfeldes $\mathcal{H}(\boldsymbol{\zeta})$, $\mathrm{E}^*\left\{\mathrm{Var}\{\mathcal{H}(\boldsymbol{\zeta})\}\right\} = \mathrm{E}\left\{\mathrm{Var}\{\mathcal{H}(\boldsymbol{\zeta})\}|\mathbf{H}\right\}$ aus Gl. (8.55) berechnet. Dies wird jedoch nur für den Fall von *nicht-informativen A-priori*-Verteilungen von $\boldsymbol{\beta}|\sigma^2_{\mathcal{Z}}$ und $\sigma^2_{\mathcal{Z}}$ gezeigt. Es gilt

$$\mathrm{E}^*\left\{\mathrm{Var}\{\mathcal{H}(\boldsymbol{\zeta})\}\right\} = \mathrm{E}^*\{\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})^2\}\} - \mathrm{E}^*\left\{\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}^2\right\}, \tag{8.62}$$

wobei der zweite Term in Gl. (8.57) berechnet wird, und für den ersten Term gilt

$$\begin{aligned} K_2 &:= \mathrm{E}^*\{\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})^2\}\} \\ &= \int\cdots\int_{\mathcal{D}_{\boldsymbol{\zeta}}} \mathrm{E}^*\left\{\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})^2\}\right\} \cdot f_1(\zeta_1)\cdot\ldots\cdot f_d(\zeta_d)\mathrm{d}\zeta_1\cdots\mathrm{d}\zeta_d \\ &= \int\cdots\int_{\mathcal{D}_{\boldsymbol{\zeta}}} \mathrm{E}\left\{\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})^2\}|\mathbf{H}\right\} \cdot f_1(\zeta_1)\cdot\ldots\cdot f_d(\zeta_d)\mathrm{d}\zeta_1\cdots\mathrm{d}\zeta_d. \end{aligned} \tag{8.63}$$

Der Integrand aus Gl. (8.63) muss jedoch noch bestimmt werden. Dieser ist der *A-posteriori*-Erwartungswert von $\mathcal{H}(\boldsymbol{\zeta})^2$. Da in dieser Arbeit für jedes $\boldsymbol{\zeta}_0 \in \mathcal{D}_{\boldsymbol{\zeta}}$ der Wert des skalaren Zufallsfeldes $\mathcal{H}(\boldsymbol{\zeta}_0)|\mathbf{H} =: H_0|\mathbf{H}$ eine Zufallsvariable mit einer *a-posteriori nicht-zentralen t*-Verteilung ist, besitzt die Zufallsvariable $H_0^2|\mathbf{H}$ eine *nicht-zentrale F*-Verteilung. Da dies die Angabe einer geschlossenen Form des *A-posteriori*-Erwartungswertes der Varianz des Zufallsfeldes $\mathrm{E}^*\{\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})^2\}\}$ praktisch unmöglich macht, wird eine Schätzung des Integranden durch

$$\hat{K}_2 := \mathrm{E}_{\sigma^2_{\mathcal{Z}}|\mathbf{H}}\left\{\mathrm{E}\left\{\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})^2\}|\mathbf{H},\sigma^2_{\mathcal{Z}}\right\}\right\} \tag{8.64}$$

verwendet. Diese Vorgehensweise wurde in [34] vorgestellt und basiert auf der Tatsache, dass

$$\begin{aligned} &\mathrm{E}\left\{\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})^2\}|\mathbf{H},\sigma^2_{\mathcal{Z}}\right\} \\ &\quad= \mathrm{E}\left\{\int\cdots\int_{\mathcal{D}_{\boldsymbol{\zeta}}} \mathcal{H}(\boldsymbol{\zeta})^2\cdot f_1(\zeta_1)\cdot\ldots\cdot f_d(\zeta_d)\mathrm{d}\zeta_1\cdots\mathrm{d}\zeta_d|\mathbf{H},\sigma^2_{\mathcal{Z}}\right\} \\ &\quad= \int\cdots\int_{\mathcal{D}_{\boldsymbol{\zeta}}} \mathrm{E}\left\{\mathcal{H}(\boldsymbol{\zeta})^2|\mathbf{H},\sigma^2_{\mathcal{Z}}\right\} \cdot f_1(\zeta_1)\cdot\ldots\cdot f_d(\zeta_d)\mathrm{d}\zeta_1\cdots\mathrm{d}\zeta_d, \end{aligned}$$

gilt, wobei für jedes $\boldsymbol{\zeta}_0 \in \mathcal{D}_{\boldsymbol{\zeta}}$ der Erwartungswert $\mathrm{E}\left\{\mathcal{H}(\boldsymbol{\zeta}_0)^2|\mathbf{H},\sigma^2_{\mathcal{Z}}\right\}$ in analytischer Form angebbar ist. Dies liegt daran, dass die Zufallsvariable

$\mathcal{H}(\boldsymbol{\zeta}_0)|\mathbf{H},\sigma_{\mathcal{Z}}^2$ normalverteilt ist, mit

$$\mathcal{H}(\boldsymbol{\zeta}_0)|\mathbf{H},\sigma_{\mathcal{Z}}^2 \sim \mathcal{N}(\mu_4(\boldsymbol{\zeta}_0),\sigma_4^2(\boldsymbol{\zeta}_0))$$

und, dass

$$\begin{aligned}\mathrm{E}\left\{\mathcal{H}(\boldsymbol{\zeta}_0)^2|\mathbf{H},\sigma_{\mathcal{Z}}^2\right\} &= \mu_4(\boldsymbol{\zeta}_0)^2 + \sigma_4^2(\boldsymbol{\zeta}_0)\\ &= \mu_4(\boldsymbol{\zeta}_0)^2 + \sigma_{\mathcal{Z}}^2 r^*(\boldsymbol{\zeta}_0,\boldsymbol{\zeta}_0)\end{aligned}$$

gilt. Dabei können die Parameter $\mu_4(\boldsymbol{\zeta}_0)$ und $r^*(\boldsymbol{\zeta}_0,\boldsymbol{\zeta}_0)$ aus Theorem B.1 entnommen werden. Eine Schätzung des gesuchten *A-posteriori*-Erwartungswertes $\mathrm{E}\left\{\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})^2\}|\mathbf{H}\right\}$ bildet dessen Erwartungswert bezüglich der *A-posteriori*-Wahrscheinlichkeitsdichte von $\sigma_{\mathcal{Z}}^2$, d.h. von $\sigma_{\mathcal{Z}}^2|\mathbf{H}$. Es folgt

$$\begin{aligned}\hat{K}_2 &:= \mathrm{E}_{\sigma_{\mathcal{Z}}^2|\mathbf{H}}\left\{\mathrm{E}\left\{\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})^2\}|\mathbf{H},\sigma_{\mathcal{Z}}^2\right\}\right\}\\
&= \mathrm{E}_{\sigma_{\mathcal{Z}}^2|\mathbf{H}}\left\{\int\cdots\int_{\mathcal{D}_{\boldsymbol{\zeta}}} \mathrm{E}\left\{\mathcal{H}(\boldsymbol{\zeta})^2|\mathbf{H},\sigma_{\mathcal{Z}}^2\right\}\cdot f_1(\zeta_1)\cdot\ldots\cdot f_d(\zeta_d)\mathrm{d}\zeta_1\cdots\mathrm{d}\zeta_d\right\}\\
&= \mathrm{E}_{\sigma_{\mathcal{Z}}^2|\mathbf{H}}\left\{\int\cdots\int_{\mathcal{D}_{\boldsymbol{\zeta}}} \left(\mu_4(\boldsymbol{\zeta})^2 + \sigma_{\mathcal{Z}}^2 r^*(\boldsymbol{\zeta},\boldsymbol{\zeta})\right)\cdot f_1(\zeta_1)\cdot\ldots\cdot f_d(\zeta_d)\mathrm{d}\zeta_1\cdots\mathrm{d}\zeta_d\right\}\\
&= \int\cdots\int_{\mathcal{D}_{\boldsymbol{\zeta}}} \mu_4(\boldsymbol{\zeta})^2\cdot f_1(\zeta_1)\cdot\ldots\cdot f_d(\zeta_d)\mathrm{d}\zeta_1\cdots\mathrm{d}\zeta_d +\\
&\quad + \mathrm{E}_{\sigma_{\mathcal{Z}}^2|\mathbf{H}}\left\{\int\cdots\int_{\mathcal{D}_{\boldsymbol{\zeta}}} \sigma_{\mathcal{Z}}^2 r^*(\boldsymbol{\zeta},\boldsymbol{\zeta})\cdot f_1(\zeta_1)\cdot\ldots\cdot f_d(\zeta_d)\mathrm{d}\zeta_1\cdots\mathrm{d}\zeta_d\right\}\\
&= \int\cdots\int_{\mathcal{D}_{\boldsymbol{\zeta}}} \mu_4(\boldsymbol{\zeta})^2\cdot f_1(\zeta_1)\cdot\ldots\cdot f_d(\zeta_d)\mathrm{d}\zeta_1\cdots\mathrm{d}\zeta_d\\
&\quad + \int\cdots\int_{\mathcal{D}_{\boldsymbol{\zeta}}} \mathrm{E}_{\sigma_{\mathcal{Z}}^2|\mathbf{H}}\left\{\sigma_{\mathcal{Z}}^2\right\} r^*(\boldsymbol{\zeta},\boldsymbol{\zeta})\cdot f_1(\zeta_1)\cdot\ldots\cdot f_d(\zeta_d)\mathrm{d}\zeta_1\cdots\mathrm{d}\zeta_d.\end{aligned} \tag{8.65}$$

Wie in [52] gezeigt, hat die Zufallsvariable $\sigma_{\mathcal{Z}}^2|\mathbf{H}$ eine inverse Chi-quadrat Verteilung, d.h.

$$\sigma_{\mathcal{Z}}^2|\mathbf{H} \sim Q_4^2\cdot\chi_{N-q}^{-2},$$

mit $q := \mathrm{Rang}(\mathbf{F})$ und Erwartungswert

$$\mathring{\sigma}_Z^2 := \mathrm{E}\{\sigma_{\mathcal{Z}}^2|\mathbf{H}\} = Q_4^2/(N-q-2), \tag{8.66}$$

wobei der Skalar $Q_4^2$ aus Theorem B.1 zu entnehmen ist. Für den Schätzer aus Gl. (8.65) ergibt sich schließlich

$$\hat{K}_2 = \int\cdots\int_{\mathcal{D}_\zeta} \mu_4(\boldsymbol{\zeta})^2 \cdot f_1(\zeta_1) \cdot \ldots \cdot f_d(\zeta_d) \mathrm{d}\zeta_1 \cdots \mathrm{d}\zeta_d + \int\cdots\int_{\mathcal{D}_\zeta} \mathring{\sigma}_Z^2 r^*(\boldsymbol{\zeta},\boldsymbol{\zeta}) \cdot f_1(\zeta_1) \cdot \ldots \cdot f_d(\zeta_d) \mathrm{d}\zeta_1 \cdots \mathrm{d}\zeta_d.$$

Somit kann der *A-posteriori*-Erwartungswert der Varianz des skalaren Zufallsfeldes aus Gl. (8.62) vollständig berechnet werden.

### 8.3.3 Beispiel: Sensitivitätsanalyse und Prädiktion einer Funktion mit zwei Variablen

Folgendes Beispiel illustriert den partiellen Bayes'schen Ansatz und die Sensitivitätsanalyse anhand der folgenden Funktion mit zwei Variablen aus [64, Beispiel 4.2]

$$\eta = h(\boldsymbol{\zeta}) = 2\zeta_1^3\zeta_2^2, \quad \boldsymbol{\zeta} := [\zeta_1 \;\; \zeta_2]^\top \in [-1,1] \times [-1,1]. \tag{8.67}$$

Der Ausgang wird als skalares Gauß'sches Zufallsfeld

$$\mathcal{H}(\boldsymbol{\zeta}) = \beta + \mathcal{Z}(\boldsymbol{\zeta}) \tag{8.68}$$

modelliert, wobei der freie Koeffizient $\beta \in \mathbb{R}$ unbekannt und $\mathcal{Z}(\boldsymbol{\zeta})$ ein mittelwertfreies skalares Zufallsfeld mit einer noch unbekannten konstanten Varianz $\sigma_{\mathcal{Z}}^2 > 0$ ist. Für die Interpolation werden $N = 20$ Design-Punkte $\mathcal{X}_D := \{\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_7\}$ durch *Latin-Hypercube-Sampling* (vgl. Abschnitt B.5 (Anhang)) erzeugt, welche die Trainingsdaten $\boldsymbol{\eta} := [\eta_1 \;\; \cdots \;\; \eta_{20}]$ generieren.

Da das Produkt von Korrelationsfunktionen wiederum eine Korrelationsfunktion generiert, wird bei diesem Beispiel die Korrelationsfunktion

$$R(\mathbf{d}|\boldsymbol{\psi}) = R(d_1|\psi_1) \cdot R(d_2|\psi_2) \tag{8.69}$$

verwendet, wobei $R(d_1|\psi_1)$ und $R(d_2|\psi_2)$ Gauß'sche Korrelationsfunktionen mit unbekannten Parametern $\psi_1$ und $\psi_2$ darstellen.

Als *A-priori*-Verteilung $[\boldsymbol{\beta},\sigma_{\mathcal{Z}}^2]$ wird die *nicht-informative* Verteilung (4) aus Tabelle 8.2 verwendet. Der Wert des skalaren Zufallsfeldes an einem Test-Punkt $\boldsymbol{\zeta}_0 \notin \mathcal{X}_D$ wird durch die Zufallsvariable $H_0 := H(\boldsymbol{\zeta}_0)$ modelliert. Unter diesen Annahmen besitzt die bedingte Zufallsvariable $H_0|\mathbf{H}$

eine $t$-Verteilung mit $\nu_4 = 5$ Freiheitsgraden, dessen Nichtzentralitätsparameter $\mu_4(\boldsymbol{\zeta}_0,\boldsymbol{\psi})$ und Skalierungsparameter $\sigma_4^2(\boldsymbol{\zeta}_0,\boldsymbol{\psi})$ aus Theorem B.1 (Anhang) entnommen werden können. Es gilt folglich

$$\frac{H_0|\mathbf{H} - \mu_4(\boldsymbol{\zeta}_0,\boldsymbol{\psi})}{\sigma_4(\boldsymbol{\zeta}_0,\boldsymbol{\psi})} \sim \mathcal{T}_1(\nu_4,0,1), \tag{8.70}$$

wobei $\boldsymbol{\psi} := [\psi_1,\psi_2]$ den noch unbekannten Parametervektor der Korrelationsfunktion $R(\mathbf{h}|\boldsymbol{\psi})$ darstellt. Der Parametervektor $\boldsymbol{\psi}$ wird mit Hilfe der Maximum Likelihood Methode aus Abschnitt 8.2.1 auf $\hat{\boldsymbol{\psi}} = [0.6645, 2.3251]$ geschätzt.

Bild 8.3 (links) zeigt die wahre Funktion aus Gl. (8.67) und (rechts) die Prädiktion $\hat{H}_0 = \mathrm{E}\{H_0|\mathbf{H}\}$ für 625 äquidistante Test-Punkte, sowie, auf beiden Seiten die Menge der Design-Punkte $\mathcal{X}_D \subset [0,1] \times [0,1]$. An den Design-Punkten stimmen, wie erwartet, die Prädiktion und die wahre Funktion überein. Dies kann man auch anhand der in den beiden Bildern dargestellten Niveaulinien sehen, welche die gleichen Niveaus darstellen. Die Untersuchung der Prädiktionsgenauigkeit, welche anhand der Maße aus Abschnitt 8.2.4 quantifiziert wird, ergibt einen empirischen quadratischen Mittelwert des Prädiktionsfehlers von ERMSPE = 0.1828 und einen Anteil der erzielten Deckung von AC = 0.3776.

Für die Sensitivitätsanalyse wird angenommen, dass die Eingangsvariablen $\zeta_1$ und $\zeta_2$ zwei Umgebungsvariablen darstellen und durch zwei **unabhängige** standard-normalverteilte Zufallsvariablen - bezeichnet durch $Z_1$, $Z_2$ - repräsentiert sind, d.h. $Z_1, Z_2 \sim \mathcal{N}_1(0,1)$. Es gilt

$$\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\} = \int_{\mathcal{D}_{\zeta_1}} \int_{\mathcal{D}_{\zeta_2}} \mathcal{H}(\boldsymbol{\zeta}) f_1(\zeta_1) f_2(\zeta_2) \mathrm{d}\zeta_1 \mathrm{d}\zeta_2, \tag{8.71}$$

$$\mathrm{E}_{Z_{\overline{1}}}\{\mathcal{H}(\boldsymbol{\zeta})|Z_1\} = \int_{\mathcal{D}_{\zeta_2}} \mathcal{H}(\boldsymbol{\zeta}) f_2(\zeta_2) \mathrm{d}\zeta_2, \tag{8.72}$$

$$\mathrm{E}_{Z_{\overline{2}}}\{\mathcal{H}(\boldsymbol{\zeta})|Z_2\} = \int_{\mathcal{D}_{\zeta_1}} \mathcal{H}(\boldsymbol{\zeta}) f_1(\zeta_1) \mathrm{d}\zeta_1 \tag{8.73}$$

und die Haupteffekte sind

$$\mathcal{C}_1(\zeta_1) = \mathrm{E}_{Z_{\overline{1}}}\{\mathcal{H}(\boldsymbol{\zeta})|Z_1\} - \mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\},$$
$$\mathcal{C}_2(\zeta_2) = \mathrm{E}_{Z_{\overline{2}}}\{\mathcal{H}(\boldsymbol{\zeta})|Z_2\} - \mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}.$$

Da bei diesem Beispiel die wahre Funktion - aus Gl. (8.67) - bekannt ist, können auch der Erwartungswert $\mathrm{E}\{h(\boldsymbol{\zeta})\}$, sowie die entsprechenden

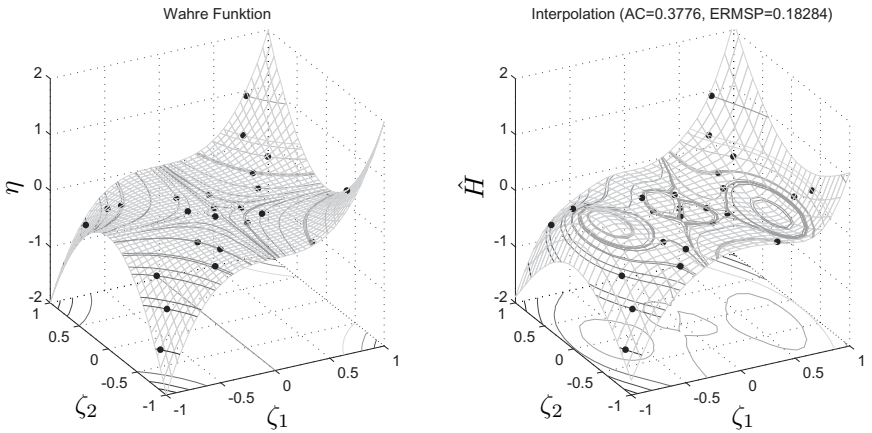


**Bild 8.3:** Wahre Funktion $\eta = h(\boldsymbol{\zeta})$ und Menge der Design-Punkte $\boldsymbol{\eta}$ für das Beispiel einer Funktion mit zwei Variablen (links), sowie Prädiktion $\hat{H}$ und die gleiche Menge der Design-Punkte $\boldsymbol{\eta}$ (rechts).

Haupteffekte $c_1(\zeta_1)$ und $c_2(\zeta_2)$ basierend auf der *wahren* Funktion $h(\boldsymbol{\zeta})$ berechnet werden. Insbesondere ergibt sich für $\mathrm{E}\{h(\boldsymbol{\zeta})\}$

$$\mathrm{E}\{h(\boldsymbol{\zeta})\} = \frac{1}{\pi} \int_{-1}^{1} \zeta_1^3 \exp\left(-\frac{1}{2}\zeta_1^2\right) \mathrm{d}\zeta_1 \int_{-1}^{1} \zeta_2^2 \exp\left(-\frac{1}{2}\zeta_2^2\right) \mathrm{d}\zeta_2.$$

Für das zweite Integral aus der obigen Gleichung gilt

$$\int_{-1}^{1} \zeta_2^2 \exp\left(-\frac{1}{2}\zeta_2^2\right) \mathrm{d}\zeta_2 = \sqrt{2\pi}(2\Phi(1) - 1) \approx 1.71,$$

wobei $\Phi(\,\cdot\,)$ die Verteilungsfunktion der Standardnormalverteilung ist. Für das erste Integral gilt

$$\int_{-1}^{1} \zeta_1^3 \exp\left(-\frac{1}{2}\zeta_1^2\right) \mathrm{d}\zeta_1 = -(\zeta_1^2 + 1) \cdot \exp\left(-\frac{1}{2}\zeta_1^2\right)\Big|_{-1}^{1} = 0.$$

Somit gilt $\mathrm{E}\{h(\boldsymbol{\zeta})\} = 0$. Ebenso ergibt sich $c_2(\zeta_2) = 0$. Schließlich ergibt sich für $c_1(\zeta_1)$

$$\begin{aligned} c_1(\zeta_1) &= 2\zeta_1^3 \int_{-1}^{1} \zeta_2^2 f_2(\zeta_2)\mathrm{d}\zeta_2 = \sqrt{\frac{2}{\pi}}\zeta_1^3 \int_{-1}^{1} \zeta_2^2 \exp\left(-\frac{1}{2}\zeta_2^2\right)\mathrm{d}\zeta_2 \\ &= 2(2\Phi(1)-1)\zeta_1^3 \approx 1.37 \cdot \zeta_1^3. \end{aligned}$$

Die Bayes'sche Inferenz für die Haupteffekte ergibt

$$\begin{aligned} \mathrm{E}^*\{\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}\} &= \int_{-1}^{1}\int_{-1}^{1} \mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})|\mathbf{H}\} f_1(\zeta_1) f_2(\zeta_2)\mathrm{d}\zeta_1\mathrm{d}\zeta_2 \\ &= \int_{-1}^{1}\int_{-1}^{1} \left(\hat{\beta} + \hat{\mathbf{r}}(\boldsymbol{\zeta})^\top \hat{\mathbf{R}}^{-1}(\boldsymbol{\eta} - \mathbf{1}_N\hat{\beta})\right) f_1(\zeta_1) f_2(\zeta_2)\mathrm{d}\zeta_1\mathrm{d}\zeta_2 \\ &= s\hat{\beta} + \mathbf{t}^\top \hat{\mathbf{R}}^{-1}(\boldsymbol{\eta} - \mathbf{1}_N\hat{\beta}), \end{aligned} \tag{8.74}$$

wobei $\mathbf{1}_N \in \mathbb{R}^n$ ein Vektor von Einsen ist, und

$$\begin{aligned} s &= \int_{-1}^{1}\int_{-1}^{1} f_1(\zeta_1) f_2(\zeta_2)\mathrm{d}\zeta_1\mathrm{d}\zeta_2 = (2\Phi(1)-1)^2, \\ \mathbf{t}^\top &= \int_{-1}^{1}\int_{-1}^{1} \hat{\mathbf{r}}^\top(\boldsymbol{\zeta}) f_1(\zeta_1) f_2(\zeta_2)\mathrm{d}\zeta_1\mathrm{d}\zeta_2, \end{aligned}$$

sowie

$$\mathrm{E}^*\{\mathrm{E}_{Z_{\overline{i}}}\{\mathcal{H}(\boldsymbol{\zeta})|Z_i\}\} = g_i\hat{\beta} + \mathbf{t}_i^\top(\zeta_i)\hat{\mathbf{R}}^{-1}(\boldsymbol{\eta} - \mathbf{1}_N\hat{\beta}), \quad i = \{1,2\},$$

mit

$$\begin{aligned} g_i &= \int_{-1}^{1} f_{\overline{i}}(\zeta_{\overline{i}})\mathrm{d}\zeta_{\overline{i}} = 2\Phi(1) - 1, \\ \mathbf{t}_i^\top(\zeta_i) &= \int_{-1}^{1} \hat{\mathbf{r}}^\top(\boldsymbol{\zeta}) f_{\overline{i}}(\zeta_{\overline{i}})\mathrm{d}\zeta_{\overline{i}}. \end{aligned}$$

Der Index $\overline{i}$ bezeichnet den zum Index $i$ komplementären Index, d.h. falls $i = 1$, ist $\overline{i} = 2$. Die konstanten Faktoren $s$ und $s_i$ sind nur von den Verteilungen der Zufallsvariablen $Z_1$ und $Z_2$ abhängig.

Die Haupteffektindizes sind

$$\begin{aligned} \mathrm{Var}_{Z_i}\{\mathrm{E}_{Z_{\overline{i}}}\{\mathcal{H}(\boldsymbol{\zeta})|Z_i\}\} &= \mathrm{E}_{Z_i}\{\mathrm{E}_{Z_{\overline{i}}}\{\mathcal{H}(\boldsymbol{\zeta})|Z_i\}^2\} - \mathrm{E}_{Z_i}\{\mathrm{E}_{Z_{\overline{i}}}\{\mathcal{H}(\boldsymbol{\zeta})|Z_i\}\}^2 \\ &= \mathrm{E}_{Z_i}\{\mathrm{E}_{Z_{\overline{i}}}\{\mathcal{H}(\boldsymbol{\zeta})|Z_i\}^2\} - \mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}^2, \quad i \in \{1,2\}, \end{aligned}$$

wobei der Term $\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}$ in Gl. (8.71) berechnet wird. Der erste Term ist

$$\begin{aligned}&\mathrm{E}_{Z_i}\{\mathrm{E}_{Z_{\overline{i}}}\{\mathcal{H}(\boldsymbol{\zeta})|Z_i\}^2\}\\&\quad=\int_{-1}^{1}\left[\int_{-1}^{1}\int_{-1}^{1}\mathcal{H}(\boldsymbol{\zeta})\mathcal{H}(\boldsymbol{\zeta}^\star)f_{\overline{i}}(\zeta_{\overline{i}})f_{\overline{i}}(\zeta_{\overline{i}}')\mathrm{d}\zeta_{\overline{i}}\mathrm{d}\zeta_{\overline{i}}'\right]f_i(\zeta_i)\mathrm{d}\zeta_i,\end{aligned}\tag{8.75}$$

wobei $\boldsymbol{\zeta}:=(\zeta_i,\zeta_{\overline{i}})$, $\boldsymbol{\zeta}^\star:=(\zeta_i,\zeta_{\overline{i}}')$. Die Bayes'sche Inferenz für die Haupteffektindizes ergibt

$$\mathrm{E}^*\{\mathrm{Var}_{Z_i}\{\mathrm{E}_{Z_{\overline{i}}}\{\mathcal{H}(\boldsymbol{\zeta})|Z_i\}\}\}=\mathrm{E}^*\{\mathrm{E}_{Z_i}\{\mathrm{E}_{Z_{\overline{i}}}\{\mathcal{H}(\boldsymbol{\zeta})|Z_i\}^2\}\}-\mathrm{E}^*\{\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}^2\},$$

für alle $i\in\{1,2\}$. Für den zweiten Term $\mathrm{E}^*\{\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}^2\}$ gilt

$$\mathrm{E}^*\{\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}^2\}=\mathrm{E}\{\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}^2|\mathbf{H}\}=\mathrm{Var}\{\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}|\mathbf{H}\}+\mathrm{E}\{\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}|\mathbf{H}\}^2,$$

wobei

$$\mathrm{Var}\{\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}|\mathbf{H}\}=\mathrm{Var}^*\{\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}\}=\left(\int_{-1}^{1}\int_{-1}^{1}\sigma_i^2(\boldsymbol{\zeta})f_1(\zeta_1)f_2(\zeta_2)\mathrm{d}\zeta_1\mathrm{d}\zeta_2\right)^2$$

und - aus Gl. (8.74) -

$$\mathrm{E}\{\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}|\mathbf{H}\}^2=\mathrm{E}^*\{\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}\}^2=\left[s\hat{\beta}+\mathbf{t}^\top\hat{\mathbf{R}}^{-1}(\boldsymbol{\eta}-\mathbf{I}_N\hat{\beta})\right]^2.$$

Für den ersten Term gilt aus Gl. (8.75)

$$\begin{aligned}&\mathrm{E}^*\left\{\mathrm{E}_{Z_i}\left\{\mathrm{E}_{Z_{\overline{i}}}\{\mathcal{H}(\boldsymbol{\zeta})|Z_i\}^2\right\}\right\}\\&\quad=\mathrm{E}^*\left\{\int_{-1}^{1}\left[\int_{-1}^{1}\int_{-1}^{1}\mathcal{H}(\boldsymbol{\zeta})\mathcal{H}(\boldsymbol{\zeta}^\star)f_{\overline{i}}(\zeta_{\overline{i}})f_{\overline{i}}(\zeta_{\overline{i}}')\mathrm{d}\zeta_{\overline{i}}\mathrm{d}\zeta_{\overline{i}}'\right]f_i(\zeta_i)\mathrm{d}\zeta_i\right\},\\&\quad=\int_{-1}^{1}\left[\int_{-1}^{1}\int_{-1}^{1}\mathrm{E}^*\{\mathcal{H}(\boldsymbol{\zeta})\mathcal{H}(\boldsymbol{\zeta}^\star)\}f_{\overline{i}}(\zeta_{\overline{i}})f_{\overline{i}}(\zeta_{\overline{i}}')\mathrm{d}\zeta_{\overline{i}}\mathrm{d}\zeta_{\overline{i}}'\right]f_i(\zeta_i)\mathrm{d}\zeta_i\\&\quad=\int_{-1}^{1}\left[\int_{-1}^{1}\int_{-1}^{1}\left(\mu_i(\boldsymbol{\zeta})\mu_i(\boldsymbol{\zeta}^\star)+\hat{\sigma}_{\mathcal{Z}}^2r^*(\boldsymbol{\zeta},\boldsymbol{\zeta}^\star)\right)f_{\overline{i}}(\zeta_{\overline{i}})f_{\overline{i}}(\zeta_{\overline{i}}')\mathrm{d}\zeta_{\overline{i}}\mathrm{d}\zeta_{\overline{i}}'\right]\\&\qquad\qquad\cdot f_i(\zeta_i)\mathrm{d}\zeta_i,\end{aligned}\tag{8.76}$$

wobei $\boldsymbol{\zeta}:=(\zeta_i,\zeta_{\overline{i}})$, $\boldsymbol{\zeta}^\star:=(\zeta_i,\zeta_{\overline{i}}')$ und $\mu_i(\,\cdot\,)$, sowie $\sigma_{\mathcal{Z}}^2$ und $r^*(\boldsymbol{\zeta},\boldsymbol{\zeta}^\star)$ aus Theorem B.1 entnommen werden.

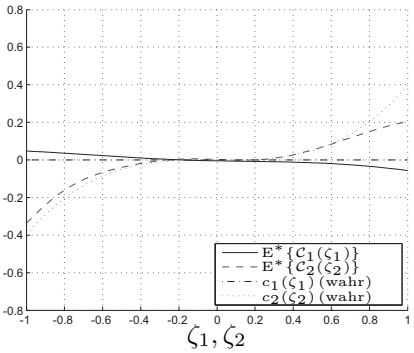


**Bild 8.4:** Wahre Haupteffekte und die jeweiligen Bayes'schen Inferenzen.

Bild 8.4 zeigt die wahren Werte der Sensitivitätsmaße $c_1(\zeta_1)$ und $c_2(\zeta_2)$, die sich ergeben wenn $h(\boldsymbol{\zeta})$ statt $\mathcal{H}(\boldsymbol{\zeta})$ in Gl. (8.71)-(8.73) eingesetzt wird, sowie die entsprechende Bayes'sche Inferenz $\mathrm{E}^*\{\mathcal{C}_1(\zeta_1)\}$ und $\mathrm{E}^*\{\mathcal{C}_2(\zeta_2)\}$. Tabelle 8.3 zeigt die *wahren* Haupteffektindizes $S_i$ und die geschätzten Haupteffektindizes $\hat{S}_i$. Es ist ersichtlich, dass der Beitrag der Zufallsvariable $Z_1$ - unabhängig von $Z_2$ - etwa 35% der Gesamtvarianz $\mathrm{Var}\{\mathcal{H}(\boldsymbol{\zeta})\}$ erklärt, wohingegen die Zufallsvariable $Z_2$ keinen Einfluss unabhängig von $Z_1$ auf die Varianz hat. Die inferierten Werte überschätzen dabei die *wahren* Haupteffektindizes. Diese Differenz ist abhängig von der Wahl der Design-Punkte $\mathcal{X}_\mathcal{D}$.

**Tabelle 8.3:** Die *wahren* und die geschätzten Haupteffektindizes.

| | $100S_i$ | $100\hat{S}_i$ |
|---|---|---|
| $Z_1$ | 35.1735 | 39.3996 |
| $Z_2$ | 0 | 1.1861 |

## 8.4 Anwendungsbeispiel: Sensitivitätsanalyse und Performanceprädiktion in einem Streckenensemble

Der Einsatz Bayes'scher Methoden wird im Folgenden anhand eines *Computerexperiments* dargestellt, welches die Performance einer Regelmethode für ein Streckenensemble analysiert. In Abschnitt 8.4.1 wird die Performance einer Regelmethode innerhalb eines Streckenensembles an mehreren Strecken (Design-Punkten) berechnet und im Übrigen interpoliert. Das Beispiel zeigt, wie man die Performance eines nicht-simulierten Regelkreises analysieren kann. In Abschnitt 8.4.2 wird eine Sensitivitätsanalyse für die gebietsabhängige Konvergenzrate innerhalb eines Streckenensembles durchgeführt. Das Beispiel zeigt welchen Einfluß der Reglerparameter $v \in [\varepsilon,1]$ und der Ensembleparameter $\theta \in [-1,1]$ auf die prädizierte Konvergenzrate sowie auf deren Varianz haben. Schließlich wird in Abschnitt 8.4.3 ein empirischer Vergleich zwischen mehreren Prädiktoren anhand von 93 *Computerexperimenten* gezeigt, welche jeweils die Prädiktion für eine nicht-simulierte Strecke als Ziel haben. Dabei wird mit Hilfe von Boxplots gezeigt, wie die Prädiktionsgenauigkeit durch verschiedene prädiktive Verteilungen, sowie Korrelationsfunktionen und empirische Schätzmethoden der Korrelationsparameter variiert.

### 8.4.1 Prädiktion für eine nicht-simulierte Regelstrecke

Eine Prädiktion der Performance für eine nicht-simulierte Regelstrecke aus einem Regelstreckenensemble kann mit folgenden Schritten durchgeführt werden:

**Schritt 1: Wahl eines Streckenensembles** Wähle für eine bestimmte Ordnung $n \geq 1$ die Matrizen $\mathbf{A}_i \in \mathbb{R}^{n\times n}$, $i = 0,\ldots,a$ und die Vektoren $\mathbf{b}_i \in \mathbb{R}^n$, $i = 0,\ldots,b$, welche ein Streckenensemble der Form

$$\dot{\mathbf{x}} = \sum_{i=0}^{a} \zeta^i \mathbf{A}_i \mathbf{x} + \sum_{i=0}^{b} \zeta^i \mathbf{b}_i \cdot u, \quad \mathbf{x} \in \mathbb{R}^n, \zeta \in [-1,1] \tag{8.77}$$

definieren.

**Schritt 2: Überprüfung der Stabilisierbarkeit des gesamten Ensembles** Überprüfe wie im Abschnitt 6 dargestellt, ob für jede mögliche Strecke aus diesem Streckenensemble ein stabilisierendes Regelgesetz

existiert. Falls dies nicht der Fall ist, ändere das Intervall $\theta \in [-1,1]$ oder gehe zu Schritt 1 und wähle ein anderes Streckenensemble.

**Schritt 3: Generierung von Trainingsdaten** Wähle eine Menge von $N$ Design-Punkten $\mathcal{X}_\mathcal{D} := \{\zeta_1,\ldots,\zeta_N\} \subset \mathcal{C} = [-1,1]$ und entwerfe für jeden Design-Punkt $\zeta_k$, mit $k = 1,\ldots,N$, den jeweiligen zeitoptimalen und den konvergenzoptimalen Regler. Berechne anschließend für den geschlossenen Regelkreis die Performance-Maße $J_{\mathrm{t_a}}(\zeta_k)$ (relative *Einschwingzeit*) und $J^{\mathrm{n}}_{\mathrm{H}}(\zeta_k)$ (Fehlklassifikationsanteil durch die konvergenzoptimale Schaltfunktion) aus Gl. (7.14) bzw. Gl. (7.13). Diese bilden die Trainingsdaten für die jeweiligen Prädiktionen, d.h. $\boldsymbol{\eta} := \{J_*(\zeta_1),\ldots,J_*(\zeta_N)\}$, wobei das Symbol $*$ stellvertretend für $\mathrm{t_a}$ oder H verwendet wird.

**Schritt 4: Bildung der prädiktiven *A-posteriori*-Verteilung** Bilde eine oder mehrere prädiktive Verteilungen mit Hilfe des Theorems B.1 (Anhang) unter Verwendung der im Schritt 3 ausgewählten Design-Punkte $\mathcal{X}_\mathcal{D}$.

**Schritt 5: Berechnung der Prädiktionsgenauigkeit** Berechne für $M$ äquidistante Testpunkte aus dem Intervall $\zeta \in [-1,1]$ die wahren Performance-Maße $J$ und $J_\Sigma$, sowie die Prädiktionen $\hat{J}$ und $\hat{J}_\Sigma$. Darauf basierend, berechne die Prädiktionsgenauigkeit, d.h. den empirischen quadratischen Mittelwert des Prädiktionsfehlers (ERMSPE) und die erzielte Deckung des wahren Wertes durch das $1-\alpha$ Konfidenzintervall des Prädiktors (AC), vgl. Abschnitt 8.2.4.

*Bemerkung* 8.8 (Zu Schritt 1). In diesem Beispiel werden die analysierten Streckenensembles zufällig generiert. Die Elemente der Matrizen $\mathbf{A}_i$, mit $i = 1,\ldots,a$ und der Vektoren $\mathbf{b}_i$, mit $i = 1,\ldots,b$, sind dabei standard normalverteilte Zufallszahlen. Die Koeffizienten der hier analysierten Streckenensembles sind im Anhang C.2 zu finden. △

Die Berechnung der Performance-Indizes $J_{\mathrm{t_a}}(\zeta)$ und $J^{\mathrm{n}}_{\mathrm{H}}(\zeta)$ erfolgt numerisch. Die relative *Einschwingzeit* wird in Gl. (7.14) definiert, wobei die jeweiligen *Einschwingzeiten* nach einer Simulation in der Toolumgebung Matlab/Simulink gemessen werden. Dabei werden jeweils 10 Anfangsauslenkungen simuliert, welche äquidistant auf dem oberen Rand[9)] des maximalen Einzugsgebiets (in diesem Fall eine Ellipse) verteilt sind. Die Simulation wurde jeweils mit einer konstanten Schrittweite von 0.001 ausgeführt.

[9)] Der obere Rand ist durch den Rand oberhalb einer der Halbachsen definiert.

Aufgrund der Reglerstruktur entsteht bei den beiden Reglern ein Rattern um die Ruhelage, welches sich auf die Bemessung der *Einschwingzeit* auswirkt. Darüber hinaus stellt die *Einschwingzeit* den Zeitpunkt dar, wann die jeweilige Zustandsnorm 5% der Norm der Anfangsauslenkung (zum letzten Mal) erreicht. So ist es möglich, dass - auch aufgrund numerischer Ungenauigkeiten - der konvergenzoptimale Regler schneller eine kürzere *Einschwingzeit* als der (angenäherte) zeitoptimale Regler hat. Bild 8.5 zeigt ein solches Beispiel. Dabei wird jeweils die zeitliche Änderung der Zustandsnorm auf einer logarithmischen Skala, sowie der Grenzwert $0.05\|\mathbf{x}(0)\|$ gezeigt. Es ist ersichtlich, dass in diesem Fall die *Einschwingzeit* des konvergenzoptimalen Reglers kürzer als die des zeitoptimalen Reglers ist.

Bild 8.6 zeigt ein Beispiel des Fehlklassifikationsanteils durch die konvergenzoptimale Schaltfunktion für eine andere Strecke aus dem obigen Streckenensemble mit $\theta = 1$. Dieser entspricht der Fläche zwischen den beiden Schaltfunktionen, wo die Regelgesetze unterschiedliche Vorzeichen aufweisen. Die Berechnung der zeitoptimalen Schaltfunktion für lineare Strecken mit komplex konjugierten Eigenwerten kann u.a. in [7] gefunden werden.

Tabelle 8.4 zeigt die jeweiligen Parameter und die erzielte Prädiktionsgenauigkeit anhand des empirischen quadratischen Mittelwertes des Prädiktionsfehlers und der erzielten Deckung. Beide Maße wurden anhand von jeweils $M = 25$ Test-Punkten berechnet.

**Tabelle 8.4:** Prädiktionsparameter für das Streckenensemble im Fall der Gauß'schen Korrelationsfunktion und der nicht-informativen prädiktiven Verteilung aus Satz B.1.

| | $\hat{\psi}$ | $\nu_4$ | $\alpha$ | $t_{\nu_4}^{\alpha/2}$ | ERMSPE | AC |
|---|---|---|---|---|---|---|
| $J_{\mathrm{t_a}}(\zeta)$ | 0.0836 | 9 | 0.05 | 2.2622 | 0.0199 | 0.9200 |
| $J_{\mathrm{H}}^{\mathrm{n}}(\zeta)$ | 0.2360 | 9 | 0.05 | 2.2622 | 0.0296 | 0.8000 |

Bild 8.7 zeigt die Prädiktion von $J_{\mathrm{t_a}}(\zeta)$ und $J_{\mathrm{H}}^{\mathrm{n}}(\zeta)$ für das gesamte Streckenensemble. Die Prädiktion an den Design-Punkten entspricht, wie erwartet, der tatsächlichen Performance der Regelmethode. Für die Prädiktion wurde die Gauß'sche Korrelationsfunktion und die nicht-informative *A-posteriori*-Verteilung verwendet. Es ist ersichtlich, dass die Prädiktion von $J_{\mathrm{t_a}}(\zeta)$ eine höhere Genauigkeit aufweist. Dies ist jedoch nicht immer der Fall. Eine empirische Analyse verschiedener Prädiktoren wird im

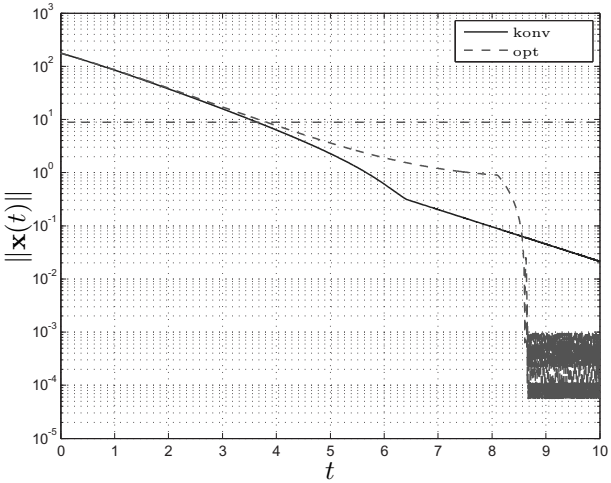


**Bild 8.5:** Vergleich der *Einschwingzeiten* im Fall einer Strecke zweiter Ordnung mit negativ reellen Eigenwerten. Gezeigt werden die zeitlichen Änderungen der jeweiligen Zustandsnormen für den konvergenzoptimalen Regler (bezeichnet mit *konv*, -) und den zeitoptimalen Regler (bezeichnet mit *opt*, - -), sowie die Grenze $0.05\|\mathbf{x}(0)\|$ (-.).

letzten Abschnitt anhand von mehreren Streckenensembles für die beiden Performance-Maße durchgeführt.

Mit Hilfe der Prädiktion kann auch die erwartete Performance einer nicht-simulierten Regelstrecke untersucht werden. Diese entspricht dem Erwartungswert der *A-posteriori*-Wahrscheinlichkeitsdichte des Performance-Maßes. Das Konfidenzintervall der Prädiktion kann ebenfalls angegeben werden. Für $\zeta^* = 0.5833$ ergibt sich beispielsweise die Strecke

$$\dot{\mathbf{x}} = \begin{bmatrix} -1.0477 & -0.1971 \\ 2.5108 & -1.0402 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0.5515 \\ -1.1724 \end{bmatrix} u, \tag{8.78}$$

deren Systemmatrix $\mathbf{A}$ die Eigenwerte $\lambda(\mathbf{A}) = -1.0439 \pm 0.7035j$ aufweist. Die Prädiktion beider Performance-Indizes kann aus Tabelle 8.5 entnommen werden. Gemäß des angenommenen Modells beträgt die Wahrscheinlichkeit, dass der wahre Wert des Performance-Maßes außerhalb des in Tabelle 8.5 angegebenen Intervalls $\mu_4(\zeta^*) \pm \sigma_4(\zeta^*, \hat{\psi}) t_{\nu_4}^{\alpha/2}$ liegt, $\alpha = 0.05$.

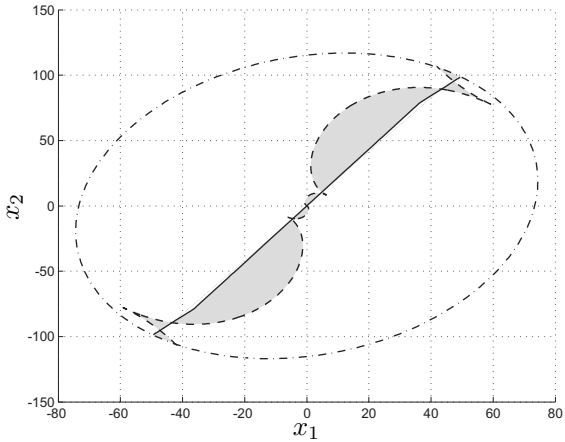


**Bild 8.6:** Fehlklassifikationsanteil durch die konvergenzoptimale Schaltfunktion im Fall einer Strecke zweiter Ordnung mit konjugiert komplexen Eigenwerten. Gezeigt werden der Rand des Einzugsgebiets der Ruhelage (-.), die zeitoptimale Schaltfunktion (- -) und die konvergenzoptimale Schaltfunktion (-). Die Fläche zwischen den beiden Funktionen stellt den Fehlklassifikationsanteil dar. Der normierte Wert ist $J_{\mathrm{H}}^{\mathrm{n}}(\zeta_L) = 0.1018$.

**Tabelle 8.5:** Wahre und prädizierte Werte der Performance-Maße für die Strecke aus Gl. (8.78).

| $\zeta^* = 0.5833$ | $J_{\mathrm{t_a}}$ | $J_{\mathrm{H}}^{\mathrm{n}}$ |
|---|---|---|
| $\eta$ (wahrer Wert) | 1.0028 | 0.2112 |
| $\hat{\eta} = \mu_4(\zeta^*)$ (Prädiktion) | 0.9905 | 0.2912 |
| $\sigma_4(\zeta^*,\hat{\psi})$ | 0.0256 | 0.0476 |
| $\mu_4(\zeta^*) \pm \sigma_4(\zeta^*,\hat{\psi}) t_{\nu_4}^{\alpha/2}$ | [0.9327,1.0483] | [0.1835, 0.3988] |

Dieses Intervall hängt sowohl von den gewählten Design-Punkten, als auch von der Wahl der Korrelationsfunktion und ihrer Parameterschätzung ab.

## 8.4.2 Sensitivitätsanalyse

Als Anwendungsbeispiel für die Sensitivitätsanalyse wird die Prädiktion der Konvergenzrate für ein Streckenensemble gewählt. Diese hängt von

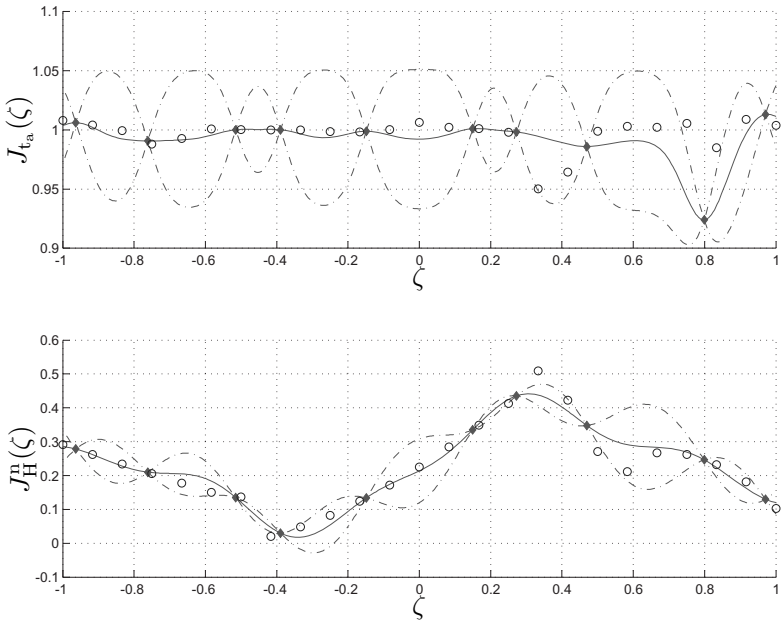


**Bild 8.7:** Prädiktion von $J_{t_a}(\zeta)$ (-, oben) und $J_H^n(\zeta)$ (-, unten), sowie jeweils das Konfidenzintervall (-.), die wahren Funktionswerte (∘) und die Trainingsdaten (⧫).

zwei Parametern ab, dem Parameter des Streckenensembles $\zeta_1 := \theta \in [-1{,}1]$ und dem Selektionsparameter $\zeta_2 := v \in [\varepsilon{,}1]$.

Bild 8.8 zeigt erstens die Prädiktion der gebietsabhängigen Konvergenzrate für das gleiche Streckenensemble aus dem vorherigen Abschnitt. Dafür wurden $N = 10$ Design-Punkte verwendet, welche ebenfalls im Bild dargestellt sind. Die Untersuchung der Prädiktionsgenauigkeit, welche anhand der Maße aus Abschnitt 8.2.4 quantifiziert wird, ergibt einen empirischen quadratischen Mittelwert des Prädiktionsfehlers von ERMSPE $\approx 0.28$ und einen Anteil der erzielten Deckung AC $\approx 0.23$ bei einer Anzahl von $M = 25 \times 25$ Test-Punkten. Es ist ersichtlich, dass die Prädiktion an den Design-Punkten der tatsächlichen Konvergenzrate entspricht. Dies ist anhand der im Bild dargestellten Niveaulinien zu sehen, die unterschiedlichen Flächen (der wahren bzw. der interpolierten Fläche) unterliegen, jedoch die gleichen Niveaus zeigen.

Bild 8.9 zeigt die Haupteffekte sowie die jeweilige Bayes'sche Inferenz, d.h. die *A-posteriori*-Erwartungswerte $\mathrm{E}^*\{\mathcal{C}_1(\zeta_1)\}$ und $\mathrm{E}^*\{\mathcal{C}_2(\zeta_2)\}$ und Bild 8.10 zeigt die Interaktionen zwischen den beiden Eingangsvariablen. Dabei entspricht $\zeta_1$ dem Parameter $\theta \in [-1,1]$ des Streckenensembles und $\zeta_2$ dem Reglerparameter $v \in [\varepsilon,1]$. Es ist ersichtlich, dass die Bayes'sche Inferenz sehr nah an dem wahren Verlauf ist. Diese Sensitivitätsmaße zeigen den Einfluss des Eingangs $\zeta_i$ auf die Variation der gebietsabhängigen Konvergenzrate. Da im Rahmen der Sensitivitätsanalyse die Eingangswerte als Realisierungen unabhängiger Zufallsvariablen betrachtet werden, hängen die Sensitivitätsmaße auch von der Wahl der Verteilung der jeweiligen Eingangsgröße ab. Für den ersten Eingang (der Parameter $\theta \in [-1,1]$) wurde die Standardnormalverteilung, d.h. $Z_1(\zeta_1) \sim \mathcal{N}_1(0,1)$, und für den zweiten (der Reglerparameter $v \in [\varepsilon,1]$) die Exponentialverteilung gewählt, d.h. $Z_2 \sim \mathrm{Exp}(1)$.

Aus Bild 8.9 (links) ist beispielsweise ersichtlich, dass mit steigendem $\theta =: \zeta_1$ der Einfluss dieses Parameters auf die Konvergenzrate unabhängig von $v =: \zeta_2$ sinkt. Der Einfluss der Interaktion zwischen den beiden Parametern, $\mathrm{E}^*\{\boldsymbol{\mathcal{C}}_{12}(\boldsymbol{\zeta})\}$, sinkt mit steigendem $\theta$ jedoch nicht für alle $v$. Bild 8.4 zeigt z.B., dass der Einfluss der Interaktion zwischen $\theta$ und $v$ auf die Konvergenzrate mit steigendem $\theta$ für $v = 1$ sinkt, jedoch für $v = \varepsilon$

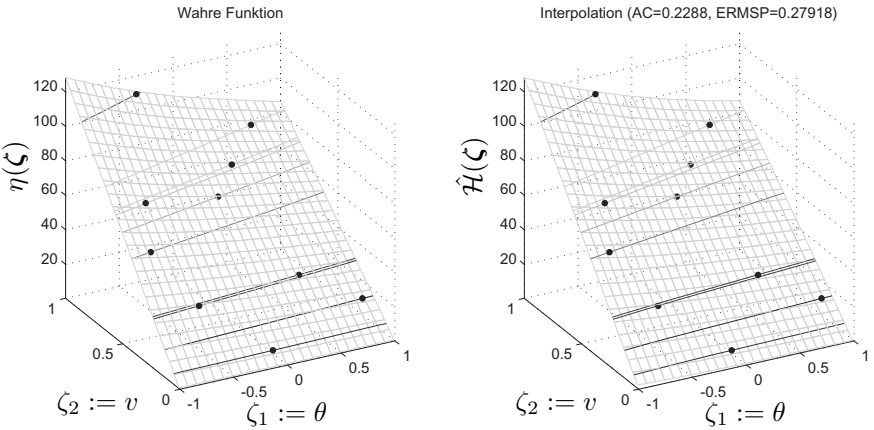


**Bild 8.8:** Prädiktion $\hat{H}$ und Menge der Design-Punkte $\boldsymbol{\eta}$ für die gebietsabhängige Konvergenzrate innerhalb eines Streckenensembles.

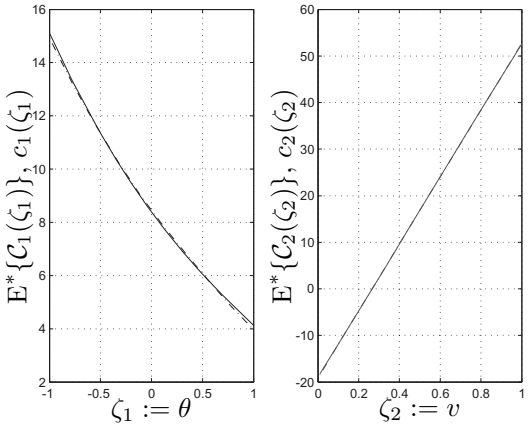


**Bild 8.9:** Bayes'sche Inferenz für die Sensitivitätsmaße.

steigt. Tabelle 8.6 verdeutlicht dies. Diese Bilder stellen also ein nützliches

**Tabelle 8.6:** Ausgangszerlegung in Haupteffekte und Interaktionen für die Ecken des untersuchten Parameterbereichs.

| $\theta$ | $v$ | $\mathrm{E}^*\{\mathrm{E}\{\mathcal{H}(\boldsymbol{\zeta})\}\}$ | $\mathrm{E}^*\{\mathcal{C}_1(\zeta_1)\}$ | $\mathrm{E}^*\{\mathcal{C}_2(\zeta_2)\}$ | $\mathrm{E}^*\{\boldsymbol{\mathcal{C}}_{12}(\boldsymbol{\zeta})\}$ |
|---|---|---|---|---|---|
| $-1$ | $\varepsilon$ | 18.8280 | 15.1027 | -18.2647 | -13.4647 |
| 1 | | | 4.1281 | | -4.6448 |
| $-1$ | 1 | | 15.1027 | 52.6291 | 41.6111 |
| 1 | | | 4.1281 | | 11.3320 |

Instrument dar, um die verschiedenen Komponenten des Modellausgangs, d.h. die Haupteffekte und Interaktionen, zu visualisieren, und damit den Einfluß jedes Eingangs zu analysieren.

Schließlich zeigt Tabelle 8.7 das Ausmaß der Varianzreduktion des Ausgangs durch die Fixierung des jeweiligen Eingangs $Z_i$. Es ist ersichtlich, dass die einzelnen Eingänge nicht die gesamte Varianz des Modells erklären können. Die restliche Varianz von etwa 26.1% wird durch Interaktionen zwischen den beiden Eingängen erklärt.

Diese Beispiele verdeutlichen die Anwendbarkeit der *Computerexperimente* unter Verwendung Bayes'scher Interpolationsmethoden für die Performance-Analyse von Regelungsmethoden. Dabei ist es leicht ersichtlich,

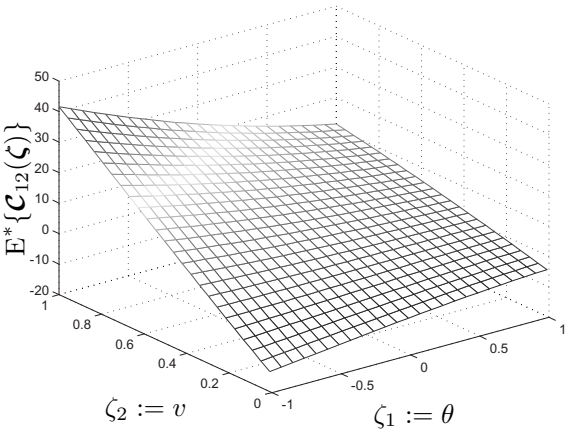


**Bild 8.10:** *A-posteriori*-Erwartungswert der Interaktion zwischen den beiden Eingängen, $\mathrm{E}^*\{\boldsymbol{\mathcal{C}}_{12}(\boldsymbol{\zeta})\}$.

welche Vorteile diese Analyse-Methode mit sich bringt. Die Performance einer Regelungsmethode für ein gesamtes Streckenensemble kann an einzelnen Strecken überprüft und im Übrigen interpoliert werden. Mittels Bayes'scher Interpolationsmethoden ist es auch möglich, jeweils Konfidenzintervalle der Prädiktionen anzugeben. Wie das Beispiel der Sensitivitätsanalyse gezeigt hat, ist es darüber hinaus möglich, den Einfluss eines oder mehrerer Parameter der Strecke und/oder des Reglers auf die Performance der Regelmethode zu veranschaulichen und/oder zu quantifizieren.

### 8.4.3 Empirischer Vergleich von Prädiktoren

Um den Unterschied zwischen verschiedenen prädiktiven Verteilungen, sowie Korrelationsfunktionen und empirischen Schätzmethoden der Korrelationsparameter zu zeigen, werden mehrere Prädiktionen anhand von 93

**Tabelle 8.7:** Die *wahren* und die geschätzten Haupteffektindizes.

| | $100S_i$ | $100\hat{S}_i$ |
|---|---|---|
| $Z_1$ | 19.8235 | 19.8559 |
| $Z_2$ | 54.1266 | 54.1273 |

Streckenensembles durchgeführt. Bei jedem Ensemble wird die Prädiktion nicht nur für eine Strecke durchgeführt, wie im Abschnitt 8.4.1 gezeigt wurde, sondern für 25 Strecken (Test-Strecken), welche darüber hinaus simuliert werden, um die Genauigkeit der Prädiktion quantifizieren und vergleichen zu können.

Jedes Ensemble ist wie in Gl. (8.77) durch ein parametrisches LTI-System beschrieben. Als Elemente der jeweiligen Systemmatrizen $\mathbf{A}_i$ und $\mathbf{b}_i$ (aus Gl. (8.77)) werden standard normalverteilte Zufallszahlen gewählt. Für jedes Ensemble wird die Prädiktion der Regelmethode, wie im Abschnitt 8.4.1 gezeigt, anhand von 25 Test-Strecken durchgeführt. Für die Bestimmung der prädiktiven Verteilungen werden jeweils 10 Design-Strecken verwendet. Diese und die 25 Test-Strecken werden darüber hinaus simuliert. Für jedes Ensemble ergeben sich 35 Werte für die *relative Einschwingzeit* und den Fehlklassifikationsanteil. Das Minimum, das Maximum, der Mittelwert, der Median, sowie die Standardabweichung der 35 generierten Beobachtungen werden gebildet und als erste allgemeine Statistik im Folgenden gezeigt. Da dies für 93 Ensembles geschieht, werden Boxplots zur Darstellung der Resultate verwendet.

Bild 8.11 zeigt mehrere Boxplots[10)] über die empirischen Verteilungen der untersuchten statistischen Maße. Der Mittelwert der *relativen Einschwingzeit* über alle Ensembles hinweg liegt beispielsweise bei etwa 1.72, der Median bei 1.66. Beide Kennzahlen sind im Bild 8.11 (links) zu sehen. Diese sind durch die Kreise innerhalb der jeweiligen Boxplots gekennzeichnet. Da insgesamt 3255 Strecken untersucht wurden, stellt eine mittlere *relative Einschwingzeit* von 1.72 ein sehr gutes Ergebnis dar. Der Fehlklassifikationsanteil zeigt zwar eine gute Klassifikation durch die konvergenzoptimale Schaltfunktion, jedoch keine große Korrelation mit der *relativen Einschwingzeit.*

Für beide Performance-Maße, die relative *Einschwingzeit* $J_{\mathrm{t_a}}$ und den Fehlklassifikationsanteil durch die konvergenzoptimale Schaltfunktion $J_{\mathrm{H}}^{\mathrm{n}}$,

[10)] Ein Boxplot ist die grafische Darstellung einer Verteilung. Dieser zeigt in welchem Bereich die Daten liegen (die senkrechten Linien stellen das Minimum und Maximum aus der jeweiligen Datenmenge dar), in welchem Bereich sich die mittleren 50% der Daten befinden (der Block zeigt wo sich die Daten zwischen dem unteren Quartil, d.h. 25% der Datenwerte, und dem oberen Quartil, d.h. 75% der Datenwerte, befinden) und den Median (der Kreis zeigt den Wert, der größer oder gleich 50% aller Datenwerte ist). An der Lage des Kreises innerhalb des Blocks erkennt man beispielsweise, ob die Verteilung symmetrisch oder schief ist. Die roten Plus-Zeichen in der Grafik zeigen Datenpunkte, welche als Ausreißer angenommen wurden und in der Berechnung der Quartile und der Extremwerte nicht berücksichtigt wurden.

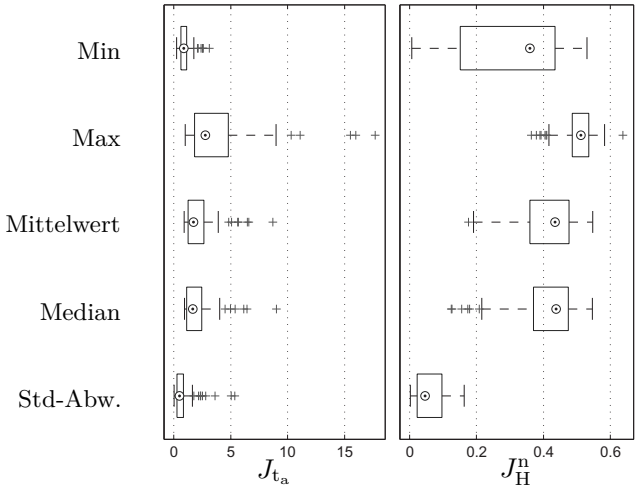


**Bild 8.11:** Boxplot-Statistiken für die untersuchten Streckenensembles.

werden jeweils in einem Trellisplot[11)] alle vier prädiktiven Verteilungen ((1) − (4)) aus Satz B.1 (Anhang), vier Korrelationsfunktionen (die lineare, die exponentiale, sowie die Potenz-Exponentiale mit $p = 1$ und $p = 2$ (Gauß'sche Korrelationsfunktion)), sowie die empirischen Schätzer (die (begrenzte) Maximum Likelihood Methode und die Kreuzvalidierungsmethode) der Parameter der jeweiligen Korrelationsfunktionen verglichen. Die Bilder 8.12 und 8.13, sowie 8.14 und 8.15 zeigen jeweils die erreichte Deckung des wahren Funktionswertes durch das $1-\alpha$ Konfidenzintervall (mit $\alpha = 0.05$) des Prädiktors und den empirischen quadratischen Mittelwert des Prädiktionsfehlers im Fall von 14 Ensembles, die erfolgreich analysiert werden konnten.[12)] Es ist ersichtlich, dass die Gaußsche Korrelationsfunktion generell eine schlechtere Prädiktionsgenauigkeit auf-

[11)] Der Trellisplot zeigt Boxplots über die Verteilung von Beobachtungen (hier die erzielte Deckung und der mittlere quadratische Prädiktionsfehler bei einem Ensemble) im Falle von verschiedenen Faktoren. Die Faktoren sind hier die vier analysierten Korrelationsfunktionen, die vier *A-posteriori*-Wahrscheinlichkeitsdichten, sowie drei empirische Schätzmethoden für die Parameter der Korrelationsfunktionen.

[12)] Die empirische Schätzung der Parameter der Korrelationsfunktionen wurde numerisch durchgeführt. Bei vielen Ensembles war die numerische Optimierung nicht erfolgreich. Für diese 14 aus 93 untersuchten Ensembles wurden alle Korrelationsfunktionen für beide Performance-Maße erfolgreich geschätzt.

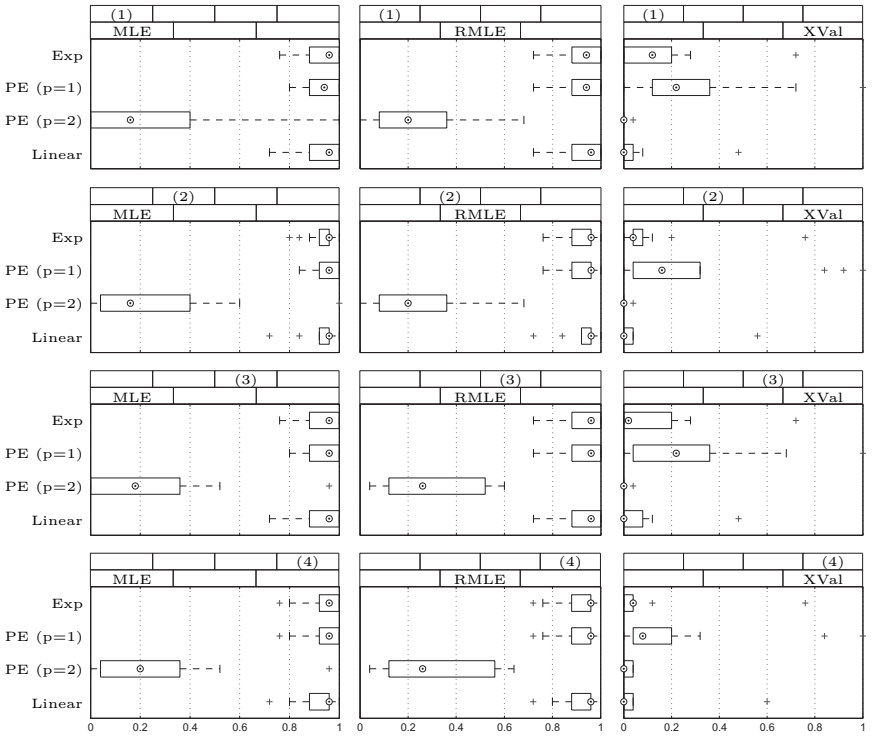


**Bild 8.12:** Vergleich der erzielten Deckung (AC) im Fall der Prädiktion der *relativen Einschwingzeit*, vgl. Abschnitt 8.2.4. Dabei wurden vier Korrelationsfunktionen (lineare, exponentiale, sowie Potenz-Exponential-Familie der Korrelationsfunktionen mit $p = 1$ und $p = 2$), die vier *A-posteriori*-Wahrscheinlichkeitsdichten aus Satz B.1, sowie drei empirische Methoden (MLE, RMLE, XVal) für die Schätzung der Parameter der Korrelationsfunktionen verglichen.

weist. Darüber hinaus erzielt die Kreuzvalidierungsmethode die niedrigsten Deckungsraten und zwischen den *A-posteriori*-Wahrscheinlichkeitsdichten gibt es kaum Unterschiede.

Zusammenfassend lässt sich feststellen, dass eine Performance-Analyse in nichtlinearen Regelkreisen unter Verwendung von *Computerexperimenten* sehr vorteilhaft ist. Dadurch ist es möglich, die Performance einer Re-

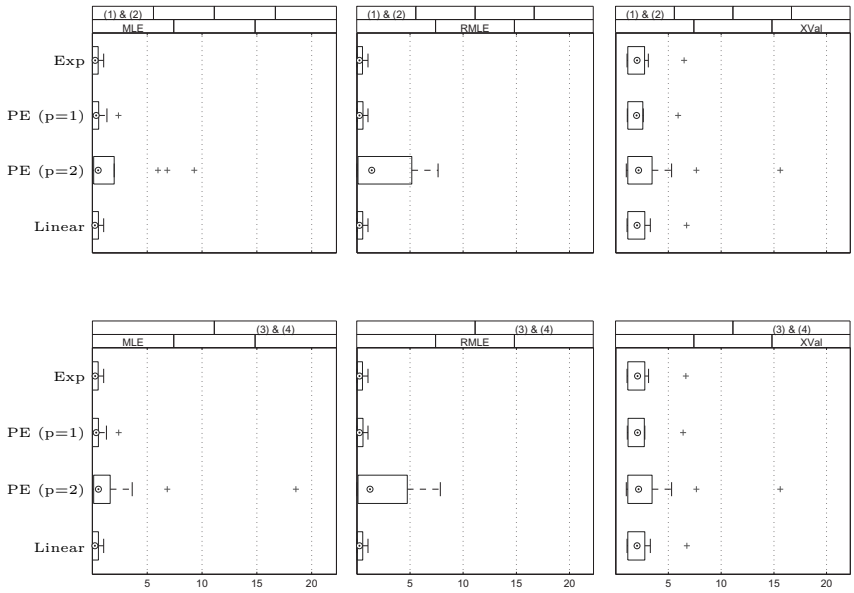


**Bild 8.13:** Vergleich des mittleren quadratischen Prädiktionsfehlers (ERMSPE) im Fall der Prädiktion der *relativen Einschwingzeit*, vgl. Abschnitt 8.2.4. Dabei wurden vier Korrelationsfunktionen (lineare, exponentiale, sowie Potenz-Exponential-Familie der Korrelationsfunktionen mit $p = 1$ und $p = 2$), die vier *A-posteriori*-Wahrscheinlichkeitsdichten aus Satz B.1, sowie drei empirische Methoden (MLE, RMLE, XVal) für die Schätzung der Parameter der Korrelationsfunktionen verglichen.

gelungsmethode für ein gesamtes Regelstreckenensemble zu analysieren. Die erwartete Performance für das Streckenensemble lässt sich dabei in funktionaler Form angeben. Somit ist es möglich, eine Prädiktion der Performance für eine Regelstrecke aus einem vorgegebenem Ensemble ohne Reglerentwurf und Simulation zu machen. Darüber hinaus ist es möglich, die Sensitivität der Performance einer Regelungsmethode infolge der Variation eines Streckenparameters zu quantifizieren. Ein empirischer Vergleich von Prädiktoren im Falle von 93 Streckenensembles zeigt schließlich, dass diese Methode zur Performance-Analyse eine gute Prädiktionsgenauigkeit aufweist und sich als vielversprechend für weitere Untersuchungen zeigt.

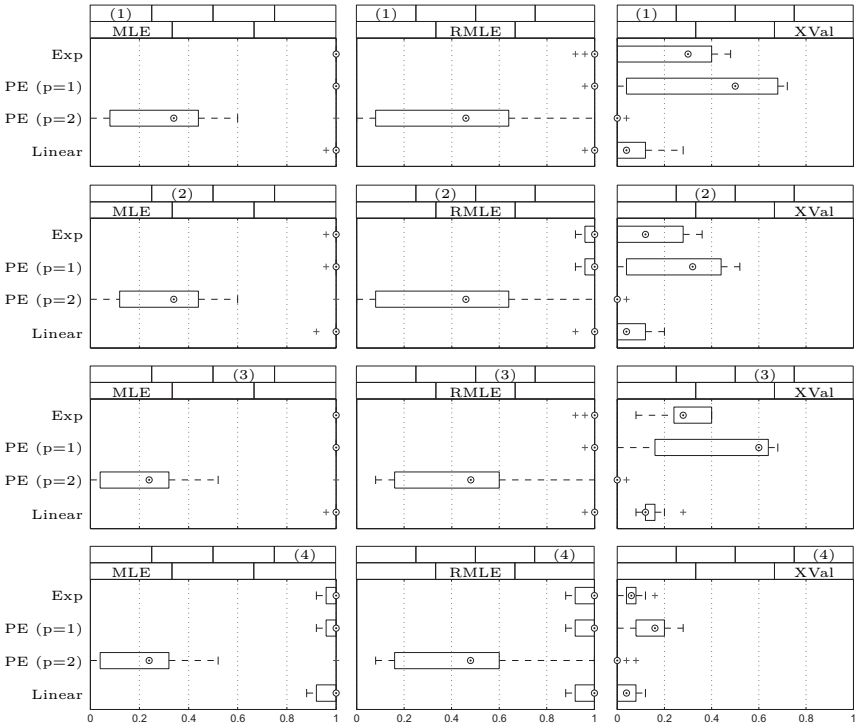


**Bild 8.14:** Vergleich der erzielten Deckung (AC) im Fall der Prädiktion des Fehlklassifikationsanteils durch die konvergenzoptimale Schaltfunktion, vgl. Abschnitt 8.2.4. Dabei wurden vier Korrelationsfunktionen (lineare, exponentiale, sowie Potenz-Exponential-Familie der Korrelationsfunktionen mit $p = 1$ und $p = 2$), die vier *A-posteriori*-Wahrscheinlichkeitsdichten aus Satz B.1, sowie drei empirische Methoden (MLE, RMLE, XVal) für die Schätzung der Parameter der Korrelationsfunktionen verglichen.

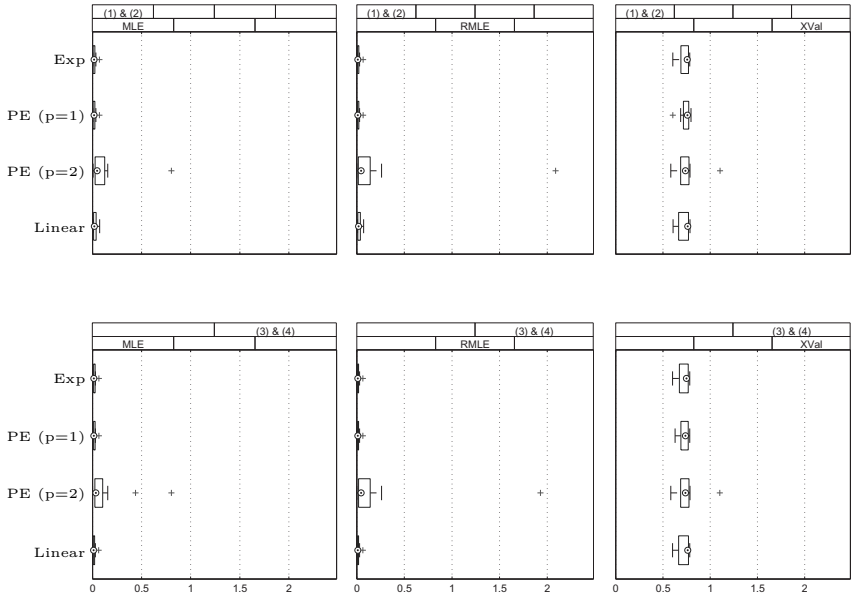


**Bild 8.15:** Vergleich des mittleren quadratischen Prädiktionsfehlers (ERMSPE) im Fall des Fehlklassifikationsanteils durch die konvergenzoptimale Schaltfunktion, vgl. Abschnitt 8.2.4. Dabei wurden vier Korrelationsfunktionen (lineare, exponentiale, sowie Potenz-Exponential-Familie der Korrelationsfunktionen mit $p = 1$ und $p = 2$), die vier *A-posteriori*-Wahrscheinlichkeitsdichten aus Satz B.1, sowie drei empirische Methoden (MLE, RMLE, XVal) für die Schätzung der Parameter der Korrelationsfunktionen verglichen.

# 9 Zusammenfassung und Ausblick

Diese Arbeit stellt mehrere Weiterentwicklungen weicher strukturvariabler Regelungen (WSVR) mittels impliziter Ljapunov-Funktionen (iLF) vor, deren Hauptaugenmerk die Nicht-Konservativität der Regelgesetze bildet. Dabei werden Entwurfsbedingungen vorgestellt, welche nicht nur hinreichend für die Stabilisierung einer linearen Strecke mit Stellgrößenbeschränkung, sondern auch notwendig sind. Aus der Notwendigkeit der Bedingungen folgt, dass im Fall deren Nichterfüllung überhaupt kein Regler dieser Klasse die jeweilige Strecke stabilisieren kann. Dieser Vorteil der Entwurfsbedingungen ist darüber hinaus nützlich, wenn ein beliebiger Startregler für eine Optimierung gesucht wird. So wird beispielsweise gezeigt, dass eine daran anschließende Optimierung der Konvergenzrate fast zeitoptimale Regelgesetze erzielt. Da der Entwurf dieser Regelgesetze mittels äquivalenter linearer Matrixungleichungen (LMIs) erfolgt, welche numerisch gelöst werden, ist der Entwurf für höherdimensionale Systeme unproblematisch. Anders ist es z.B. im Fall zeitoptimaler Regler, bei denen für höherdimensionale Systeme keine exakte Lösung mehr angegeben werden kann.

Die Weiterentwicklungen der WSVR, welche in dieser Arbeit vorgestellt werden, sind die **nicht-konservative** *klassische* WSVR mittels iLF und die **nicht-konservative *invers-polynomiale* WSVR**. Darüber hinaus wird bei jedem dieser Regler gezeigt, dass die Optimierung der Konvergenzrate Zweipunktregler (auch *Bang-Bang*- Regler genannt) mit einer parameterabhängigen Schaltfunktion erzielt. Diese unterscheiden sich daher von dem zeitoptimalen Regler allein durch die Schaltfunktion. Da in der Praxis die Diskontinuität des Regelgesetzes technische Probleme verursachen kann, wird in dieser Arbeit eine **stetige Approximation des konvergenzoptimalen Regelgesetzes** vorgestellt, die auf Kosten einer leichten Verschlechterung der Konvergenzrate einen stetigen Stellgrößenverlauf erzielt. Zwei Beispiele, ein Fusionsreaktor und ein U-Boot, veranschaulichen die Vorteile der vorgestellten Regelungsmethoden.

Der zweite Teil der Arbeit widmet sich der **Performance-Analyse** in nichtlinearen Regelkreisen. Dabei wird erstmal eine Klassifizierung von

Performance-Maßen für nichtlineare Regelkreise vorgestellt. Eine besondere Berücksichtigung erfährt die **Konvergenzrate** eines nichtlinearen Systems, für welche ein theoretischer Rahmen vorgestellt wird und im Fall der vorher entwickelten Regelungsmethoden analysiert wird. Es wird dabei gezeigt, dass im Unterschied zu linearen Regelkreisen, die nichtlinearen Regelkreise eine variable Konvergenzrate aufweisen, die von dem Abstand zur Ruhelage abhängig ist. Daher wird erstens der Begriff der *gebietsabhängigen* Konvergenzrate eingeführt und zweitens gezeigt, wie diese zur Angabe einer oberen Grenze der Zustandsnorm genutzt werden kann.

Der letzte Beitrag der Arbeit stellt zum ersten Mal die Anwendung der in der Praxis weit verbreiteten Theorie über das **Design von *Computerexperimenten*** auf die Performance-Analyse in nichtlinearen Regelkreisen vor. Die *Computerexperimente* bilden neben physikalischen Experimenten eine Methode zur Generierung von Beobachtungen über die Eigenschaften eines Versuchsobjekts, in diesem Fall eines nichtlinearen Regelkreises, infolge der Variation eines oder mehrerer Parameter. Der Zusammenhang zwischen den Eingangs- und Ausgangsvariablen wird in Form eines Rechnercodes basierend auf einem mathematischen Modell beschrieben, dessen Komplexität im Allgemeinen sehr hoch ist. Die Parameter stellen in dieser Arbeit hauptsächlich Streckenparameter dar, welche zwar aus einem vorgegebenen Intervall beliebig gewählt werden können, aber während eines Ausregelvorgangs konstant bleiben. Die Idee dieser freien Parameterauswahl ist, dass sie **Streckenensembles** erzeugt, welche somit eine unendliche Menge an Strecken beinhalten. Ein *Computerexperiment* enthält dabei den Entwurf und die Performance-Analyse an einer endlichen Anzahl von *Design*-Strecken aus diesem Streckenensemble. Mit Hilfe der gesammelten *Trainingsdaten* wird die Performance des gesamten Streckenensembles mittels Bayes'scher Interpolationsmethoden prädiziert. Wesentlich bei diesem Ansatz der **Prädiktion der Performance** einer Regelungsmethode ist die Angabe von Konfidenzintervallen des besten linearen erwartungstreuen Prädiktors (BLUP), welche einerseits von der Wahl der prädiktiven *A-posteriori*-Verteilung und andererseits von der Wahl der Design-Strecken abhängig ist.

Für ein gesamtes Streckenensemble, welches eine unendliche Anzahl an Strecken enthält, werden erstmals (ebenfalls nicht-konservative) Bedingungen vorgestellt, welche in **äquivalente** LMIs transformiert werden können und die Stabilisierbarkeit des gesamten Ensembles mittels *klassischer* oder *invers-polynomialer* WSVRs (je nach Untersuchung) sicherstellen. Anschließend wird mit Hilfe Bayes'scher Interpolationsmethoden die Per-

formance einer Regelmethode für das gesamte Streckenensemble prädiziert. So wird beispielsweise gezeigt, wie man für eine Regelstrecke aus diesem Ensemble die Performance einer Regelmethode vorhersagen kann, ohne dabei einen Regler entwerfen zu müssen.

Die **Prädiktionsgenauigkeit** wird darüber hinaus anhand von 93 Streckenensembles empirisch getestet. Die Ergebnisse dieser empirischen Untersuchung zeigen die Variation der Prädiktionsgenauigkeit bei einer unterschiedlichen Wahl prädiktiver Distributionen, Korrelationsfunktionen, sowie Schätzmethoden der Parameter der Korrelationsfunktionen. Es wird gezeigt, dass die Prädiktionsgenauigkeit, welche durch den mittleren quadratischen Prädiktionsfehler gemessen wird, bis auf wenige Ausreißer hoch ist. Auch die erzielte Deckung des wahren Funktionswertes durch das $1-\alpha$ Konfidenzintervall des Prädiktors wird untersucht. Die empirischen Ergebnisse zeigen auch hier eine hohe Genauigkeit. Darüber hinaus wird anhand der gesammelten Daten aus den 93 Streckenensembles über die Performance der Regelmethode gezeigt, dass die *invers-polynomiale* WSVR eine durchschnittliche *relative Einschwingzeit* von etwa 1.72 aufweist, und dies bei etwa 3255 direkt untersuchten Regelstrecken. Die *relative Einschwingzeit* mißt den Unterschied zwischen den Einschwingzeiten des zeitoptimalen Reglers und der *invers-polynomialer* WSVR. Eine *relative Einschwingzeit* von eins entspricht einer gleich langen *Einschwingzeit* der beiden Regler.

Als eine zweite Anwendung der Theorie über das Design von *Computerexperimenten* wird die **Sensitivitätsanalyse** untersucht. Anhand verschiedener Sensitivitätsmaße wird die (unbekannte) Sensitivität der Performance einer Regelungsmethode infolge der Variation eines Streckenparameters quantifiziert. Die Bayes'sche Inferenz stellt anschließend eine Methode dar, diese unbekannte Funktion zu prädizieren. Die in dieser Arbeit untersuchten Sensitivitätsmaße sind die *Haupteffekte* und *Interaktionen* generiert durch eine diesbezügliche Ausgangszerlegung, sowie die *Haupteffektindizes* und *Interaktionenindizes* generiert durch eine diesbezügliche Zerlegung der Ausgangsvarianzreduktion. Anhand der bereits gesammelten Trainingsdaten erfolgt beispielsweise die Angabe einer prozentualen Ausgangsvarianzreduktion infolge der Fixierung eines oder mehrerer Parameters des Streckenensembles.

Ausblickend kann man über die neu-entwickelte *invers-polynomiale* WSVR feststellen, dass die Methode vielversprechende Ergebnisse relativ zur zeitoptimalen Regelung aufweist. Da diese nicht nur Strecken in Steuerungsnormalform stabilisieren kann, ist es möglich, nicht-konserva-

tive Entwurfsbedingungen für lineare Strecken mit mehreren Eingängen anzugeben.

Bezüglich der Performance-Analyse mittels *Computerexperimenten* wäre auch von Interesse, eine empirische Untersuchung für Strecken höherer Ordnung durchzuführen. Allerdings erweist sich ein automatischer Entwurf zeitoptimaler Regler für solche Systeme zwecks Vergleich mit der jeweiligen Regelmethode als sehr aufwendig. Darüber hinaus kann hinzugefügt werden, dass die Performance-Analyse mittels *Computerexperimenten* keinesfalls auf die in dieser Arbeit untersuchten Performance-Maße oder Regelungsmethoden beschränkt, sondern beliebig anwendbar ist. Allerdings ist die Sicherstellung der Existenz von stabilisierenden Regelgesetzen im gesamten Streckenensemble unerlässlich für die Anwendbarkeit solcher Performance-Prädiktionen. Für andere Regelungsmethoden müssen daher vorerst solche Existenzbedingungen entworfen bzw. überprüft werden. Dabei wäre auch von Interesse, welche Vorteile ein *vollständiger* Bayes'scher Ansatz gegenüber dem in dieser Arbeit untersuchten *partiellen* Bayes'schen Ansatz mit sich bringt.

Auch eine gleichzeitige Untersuchung mehrerer *Ausgänge*, wie z.B. der *relativen Einschwingzeit* und des Volumens des jeweiligen Einzugsgebiets der Ruhelage, kann im Rahmen von *Computerexperimenten* durchgeführt werden. Da solche Ausgänge aber nicht unabhängig sind, muss die prädiktive Verteilung auf den höherdimensionalen Fall überführt werden. Solche höherdimensionalen Verteilungen können beispielsweise mit Hilfe von (Pair)-Copula Verteilungen aufgebaut werden.

Schließlich ist bezüglich der Sensitivitätsanalyse zu erwähnen, dass diese Methode sehr gut geeignet ist, um gezielt nach Streckenparametern zu suchen, welche die Performance einer Regelmethode stärker beeinflussen, oder, anders ausgedrückt, die Robustheit eines Reglers determinieren.

# A Ausgewählte Definitionen und Hilfssätze für die Reglersynthese

## A.1 Ausgewählte Definitionen

### A.1.1 Mengen

**Definition 5** [Offene Kugel um $\mathbf{x} \in \mathbb{R}^n$ mit Radius $\varepsilon > 0$, vgl. [8], S. 681] *Die Menge $\mathcal{B}_\varepsilon(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^n | \|\mathbf{x} - \mathbf{y}\| < \varepsilon\}$, wobei $\|\cdot\|$ eine beliebige Norm auf $\mathbb{R}^n(\mathbb{C}^n)$ ist, heißt offene Kugel um $\mathbf{x} \in \mathbb{R}^n$ mit Radius $\varepsilon > 0$.*

**Definition 6** [Innerer Punkt einer Menge, vgl. [8], Def. 10.1.1] *Ein Vektor $\mathbf{x} \in \mathcal{M} \subseteq \mathbb{R}^n(\mathbb{C}^n)$ heißt innerer Punkt der Menge $\mathcal{M}$ wenn es eine Zahl $\varepsilon > 0$ existiert, sodass $\mathcal{B}_\varepsilon(\mathbf{x}) \subseteq \mathcal{M}$.*

**Definition 7** [Randpunkt einer Menge] *Ein Vektor $\mathbf{x} \in \mathbb{R}^n$ heißt Randpunkt einer Menge $\mathcal{M}$, wenn er nicht innerer Punkt der Menge ist und für alle $\varepsilon > 0$, die Menge $\mathcal{M} \cap \mathcal{B}_\varepsilon(\mathbf{x})$ nichtleer ist.*

**Definition 8** [Rand einer Menge] *Die Menge aller Randpunkte einer Menge $\mathcal{M} \subset \mathbb{R}^n$ heißt der Rand von $\mathcal{M}$ und wird mit $\partial\mathcal{M}$ bezeichnet.*

**Definition 9** [Abgeschlossene Menge] *Eine Menge $\mathcal{M}$ heißt abgeschlossen wenn sie ihren Rand enthält.*

**Definition 10** [Beschränkte Menge, vgl. [8], S. 682] *Eine Menge $\mathcal{M} \subset \mathbb{R}^n(\mathbb{C}^n)$ heißt beschränkt wenn es eine Zahl $\delta > 0$ existiert, sodass $\|\mathbf{x} - \mathbf{y}\| < \delta$ für alle $\mathbf{x}, \mathbf{y} \in \mathcal{M}$.*

**Definition 11** [Kompakte Menge, vgl. [8], S. 682] *Eine Menge $\mathcal{M} \subseteq \mathbb{R}^n(\mathbb{C}^n)$ heißt kompakt wenn diese abgeschlossen und beschränkt ist.*

**Definition 12** [Konvexe Hülle, vgl. [8], S. 98] *Für eine gegebene Menge $\mathcal{M} \subseteq \mathbb{R}^n(\mathbb{C}^n)$, die konvexe Hülle $\operatorname{co}\mathcal{M}$ beschreibt die kleinste konvexe Menge, welche die Menge $\mathcal{M}$ beinhaltet. Falls die Menge $\mathcal{M}$ eine endliche Anzahl von Elementen aufweist, bildet $\operatorname{co}\mathcal{M}$ ein konvexes Polytop.*

**Definition 13** [Kontraktiv invariantes Gebiet] *Ein abgeschlossenes Gebiet*

$$G(V,c) := \{\mathbf{x} \in \mathbb{R}^n | \, V(\mathbf{x}) \leq c\}$$

*heißt kontraktiv invariant für ein System* $\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t))$ *mit der Ruhelage* $\mathbf{x}_\mathrm{R} = \mathbf{0}$*, wenn für eine gegebene Funktion* $V(\mathbf{x})$*, mit* $V(\mathbf{x}) > 0$*,* $\forall \mathbf{x} \in G(V,c)\backslash\{\mathbf{0}\}$*, und* $V(\mathbf{0}) = 0$*,* $\dot{V}(\mathbf{x}) < 0$*,* $\forall \mathbf{x} \in G(V,c)\backslash\{\mathbf{0}\}$ *gilt. Somit ist* $V(\mathbf{x})$ *eine Ljapunov-Funktion des Systems und das Gebiet* $G(V,c)$ *ist Teil des Einzugsgebietes der Ruhelage* $\mathbf{x}_\mathrm{R} = \mathbf{0}$*.*

Bei quadratischen Ljapunov-Funktionen ist dieses Gebiet ellipsoidal, d.h.

$$\mathcal{E}(\mathbf{P},c) := G(V,c) = \{\mathbf{x} \in \mathbb{R}^n | \, \mathbf{x}^\top \mathbf{P} \mathbf{x} \leq c\}.$$

**Definition 14** [Kontraktiv invariantes Ellipsoid] *Für eine gegebene Matrix* $\mathbf{P} \succ \mathbf{0}$ *heißt ein Ellipsoid* $\mathcal{G}(\mathbf{P}) := \{\mathbf{x} \in \mathbb{R}^n | \, \mathbf{x}^\top \mathbf{P} \mathbf{x} < 1\}$ kontraktiv invariant *für ein System* $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$*, wenn für* $V(\mathbf{x}) := \mathbf{x}^\top \mathbf{P} \mathbf{x}$*,* $\dot{V}(\mathbf{x}) = 2\mathbf{x}^\top \mathbf{P}\, \mathbf{f}(\mathbf{x}) < 0$*,* $\forall \mathbf{x} \in \mathcal{G}(\mathbf{P})\backslash\{\mathbf{0}\}$ *gilt. In diesem Fall konvergieren alle Trajektorien, die in* $\mathcal{G}(\mathbf{P})$ *starten, asymptotisch gegen die Ruhelage* $\mathbf{x}_\mathrm{R} = \mathbf{0}$*. Dies ist ein Spezialfall eines kontraktiv invarianten Gebietes aus Def. 13.*

**Definition 15** [Verschachtelte Ellipsoide] *Zwei Ellipsoide* $\mathcal{G}(v_1)$ *und* $\mathcal{G}(v_2)$*, mit* $0 < v_2 < v_1 \leq \overline{v}$*, heißen* verschachtelt *im Intervall* $v \in (0,\overline{v}]$*, wenn deren Ränder keine gemeinsamen Punkte haben, d.h. wenn* $\partial\mathcal{G}(v_1) \cap \partial\mathcal{G}(v_2) = \emptyset$*, und wenn* $\mathcal{G}(v_2) \subset \mathcal{G}(v_1)$ *gilt.*

## A.1.2 Funktionen

**Definition 16** [Limes superior einer Funktion, vgl. [62, S. vi]]

$$\limsup_{x \to a} f(x) := \inf_{\epsilon > 0} \{\sup\{f(x) : x \in \mathcal{B}(a,\epsilon), x \neq a\}\} \in \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$$

**Definition 17** [Rechte Obere Dini-Derivierte einer stetigen Funktion]

$$\begin{aligned} D^+ f(t) &:= \limsup_{h \to 0^+} \frac{f(t+h) - f(t)}{h} \\ &= \inf_{\epsilon > 0} \left\{ \sup \left\{ \frac{f(t+h) - f(t)}{h} : h \in (0,\epsilon) \right\} \right\} \end{aligned}$$

### A.1.3 Matrixdefinitionen und -funktionen

**Definition 18** [Ähnliche Matrizen] *Die Matrizen* $\mathbf{A}, \mathbf{B} \in \mathbb{F}^{n\times n}$ *sind definitionsgemäß* ähnlich *wenn* $\exists\, \mathbf{S} \in \mathbb{F}^{n\times n}$ *(nichtsingulär) existiert, sodass* $\mathbf{A} = \mathbf{S}\mathbf{B}\mathbf{S}^{-1}$.

**Definition 19** [Kongruente Matrizen] *Die Matrizen* $\mathbf{A}, \mathbf{B} \in \mathbb{F}^{n\times n}$ *heißen* kongruent*, wenn* $\exists\, \mathbf{S} \in \mathbb{F}^{n\times n}$ *(nichtsingulär) existiert, sodass* $\mathbf{A} = \mathbf{S}\mathbf{B}\mathbf{S}^{*}$.

*Bemerkung* A.1 (Vgl [8, Fakt 3.4.5, *xi*)]). Sind die Matrizen $\mathbf{A}$ und $\mathbf{B}$ kongruent, dann ist Matrix $\mathbf{A}$ positiv (semi)definit dann und nur dann wenn $\mathbf{B}$ positiv (semi)definit ist. △

**Definition 20** [Voll-Rang-Faktorisierung einer Matrix] *Gegeben sei eine Matrix* $\mathbf{A} \in \mathbb{R}^{m\times n}$*, mit* $\mathrm{Rang}(\mathbf{A}) = r$*. Das Matrizentupel* $(\mathbf{A}_l, \mathbf{A}_r)$*, mit* $\mathbf{A}_l \in \mathbb{R}^{m\times r}$*,* $\mathrm{Rang}(\mathbf{A}_l) = r$ *und* $\mathbf{A}_r \in \mathbb{R}^{r\times n}$*,* $\mathrm{Rang}(\mathbf{A}_r) = r$ *heißt eine Voll-Rang-Faktorisierung der matrix* $\mathbf{A}$ *wenn* $\mathbf{A} = \mathbf{A}_l\mathbf{A}_r$.

**Definition 21** [Kronecker Summe] *Gegeben seien die Matrizen* $\mathbf{A} \in \mathbb{R}^{n\times n}$ *und* $\mathbf{B} \in \mathbb{R}^{m\times m}$*. Die Kronecker Summe ist die* $nm \times nm$ *Matrix*

$$\mathbf{A} \oplus \mathbf{B} := \mathbf{A} \otimes \mathbf{I}_n + \mathbf{I}_m \otimes \mathbf{B}. \tag{A.1}$$

*Bemerkung* A.2. Eine wichtige Eigenschaft der Matrix $\mathbf{A} \oplus \mathbf{A}$ ist, dass ihre Eigenwerte die $n^2$ Zahlen $\lambda_i + \lambda_j$, mit $i,j = 1,\ldots,n$ sind, wobei $\lambda_i$, $\lambda_j$ Eigenwerte der Matrix $\mathbf{A}$ darstellen. △

**Definition 22** [Spalten-Vektorisierung einer Matrix] *Die Spalten-Vektorisierung einer Matrix* $\mathbf{A} \in \mathbb{R}^{m\times n}$ *ist definiert als*

$$\mathrm{vec}(\mathbf{A}) := \left[a_{(1,1)}, \cdots, a_{(m,1)}, a_{(1,2)}, \cdots, a_{(m,2)}, \cdots a_{(1,n)}, \cdots, a_{(m,n)}\right]^{\top}.$$

*Bemerkung* A.3. Folgende Beziehung gilt[1)]

$$\mathrm{vec}(\mathbf{A}\mathbf{X} + \mathbf{X}\mathbf{A}^{\top}) = (\mathbf{A} \oplus \mathbf{A})\,\mathrm{vec}(\mathbf{X}), \tag{A.2}$$

△

**Definition 23** [Norm] *Die Funktion* $\|\cdot\| : \mathbb{R}^n \to [0,\infty)$ *heißt* Norm *falls sie folgende Bedingungen erfüllt:*

*i)* $\|\mathbf{x}\| \geq 0$, $\forall\, \mathbf{x} \in \mathbb{F}^n$.

*ii)* $\|\mathbf{x}\| = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$.

[1)] Vgl. [49, Abschnitt 2.1 und 2.3].

*iii)* $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$, $\forall\,\alpha \in \mathbb{F}$ *und* $\mathbf{x} \in \mathbb{F}^n$.

*iv)* $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$, $\forall\,\mathbf{x},\mathbf{y} \in \mathbb{F}^n$.

**Definition 24** [Induzierte Matrixnorm (für quadratische Matrizen]) *Die Funktion* $\|\cdot\|_i : \mathbb{F}^{n\times n} \to \mathbb{F}$, *mit*

$$\|\mathbf{A}\|_i := \max_{\mathbf{x}\in\mathbb{F}_*^n} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|}$$

*heißt (von einer Vektornorm)* induzierte *Matrixnorm. Für nicht-quadratische Matrizen siehe [8, Definition 9.4.1]. Darüber hinaus ist die induzierte Matrixnorm ebenfalls eine Norm*[a)]*, vgl. [8, Theorem 9.4.2].*

[a)] Für die Definition der Norm siehe Def. 23.

## A.1.4 Parameterabhängige Matrizen und Funktionen

**Definition 25** [Matrixwertige Funktion] *Eine Funktion* $\mathcal{A}(\mathbf{z}) : \mathbb{R}^d \to \mathbb{R}^{n\times n}$, *deren Funktionswert analytisch für alle* $\mathbf{z} \in \mathbb{R}^d$ *ist, wird in dieser Arbeit unter dem Begriff* ein- ($d = 1$) oder mehrdimensionale ($d > 1$) matrixwertige Funktion *verwendet.*

**Definition 26** [Charakteristisches Polynom eines matrixwertigen Funktionwertes] *Für einen matrixwertigen Funktionswert* $\mathcal{A}(\mathbf{z})$ *ist das charakteristische Polynom durch*

$$p(\mathbf{z},\lambda) := \det(\lambda\mathbf{I}_n - \mathcal{A}(\mathbf{z})) = a_n(\mathbf{z})\lambda^n + \cdots + a_1(\mathbf{z})\lambda + a_0(\mathbf{z}), \qquad \text{(A.3)}$$

*definiert, wobei die Koeffizienten* $a_0(\mathbf{z}),\ldots,a_n(\mathbf{z})$ *analytische Funktionen in* $\mathbf{z}$ *sind.*

**Definition 27** [Polynomiell parameterabhängige quadratische Funktion (PPDQ-Funktion], [14]) *Eine polynomiell parameterabhängige quadratische Funktion, im Weiteren PPDQ-Funktion genannt, ist jede quadratische Funktion* $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}$, $f(\mathbf{z},\mathbf{x}) = \mathbf{x}^\top\mathbf{M}(\mathbf{z})\mathbf{x}$, *sodass*

$$\mathbf{M}(\mathbf{z}) := \sum_{i_1,i_2,\ldots,i_d=0}^{k-1} z_1^{i_1}\cdots z_d^{i_d}\mathbf{M}_{i_1,\ldots,i_d}, \quad \mathbf{M}_{i_1,\ldots,i_d} \in \mathrm{Sym}^n. \qquad \text{(A.4)}$$

*Dabei wird die Zahl $k-1 \in \mathbb{Z}$ als Grad der PPDQ-Funktion bezeichnet. Die Funktion kann umgeformt werden zu*

$$\begin{aligned}\mathbf{M}(\mathbf{z}) &:= (\mathbf{z}_d^{[k]} \otimes \cdots \otimes \mathbf{z}_1^{[k]} \otimes \mathbf{I}_n)^\top \mathbf{M}_k \, (\mathbf{z}_d^{[k]} \otimes \cdots \otimes \mathbf{z}_1^{[k]} \otimes \mathbf{I}_n), \qquad \text{(A.5)}\\ \mathbf{z}_i^{[k]} &:= \begin{bmatrix} 1 & z_{(i)} & \cdots & z_{(i)}^{k-1} \end{bmatrix}^\top, \forall i = 1,\ldots,d,\\ \mathbf{M}_k &\in \mathrm{Sym}^{k^d n}\,.\end{aligned}$$

*Die Umformung ist nicht eindeutig.*

## A.1.5 Andere Funktionen

**Lemma A.1 [Adjunkte einer polynomialen Matrix, vgl. [39]].** *Gegeben sei die polynomiale Matrix*

$$\mathbf{A}_v := \sum_{i=\mathrm{N_l}}^{\mathrm{N_u}} v^i \mathbf{A}_i, \quad \mathrm{N_l} < 0,\ \mathrm{N_l} \leq \mathrm{N_u}\,,\ \mathbf{A}_i \in \mathbb{R}^{n\times n}. \qquad \text{(A.6)}$$

*Die Adjunkte von $\mathbf{A}_v$ ist*

$$\begin{aligned}\mathbf{A}_v^A &= v^{\mathrm{N_l}(n-1)} \left( \sum_{i=0}^{\mathrm{N_u}-\mathrm{N_l}} v^i \mathbf{A}_{i+\mathrm{N_l}} \right)^A \qquad \text{(A.7)}\\ &= v^{\mathrm{N_l}(n-1)} \sum_{i=0}^{\mu} v^i \mathbf{N}_i, \quad \mathbf{N}_i \in \mathbb{R}^{n\times n}, \qquad \text{(A.8)}\end{aligned}$$

*wobei*

$$\mu \leq (\mathrm{N_u} - \mathrm{N_l}) \min\{n-1, n-q\}, \quad q := \dim \left[ \bigcap_{i=1}^{\mathrm{N_u}-\mathrm{N_l}} \mathcal{N}(\mathbf{A}_{i+\mathrm{N_l}}) \right] \qquad \text{(A.9)}$$

*und das $(i,j)$-Element der Matrix $\mathbf{A}_v^A$ ist gegeben durch*

$$(\mathbf{A}_v^A)_{(i,j)} = (-1)^{i+j} v^{\mathrm{N_l}(n-1)} \det(\tilde{\mathbf{A}}_{v_{[j,i]}}), \qquad \text{(A.10)}$$

*wobei $\tilde{\mathbf{A}}_{v_{[j,i]}}$ der Kofaktor zum Element $\mathbf{A}_{v_{(j,i)}}$ der Matrix $\mathbf{A}_v$ ist, d.h. die Matrix, die entsteht, wenn bei $\mathbf{A}_v$ die $j$-te Zeile und $i$-te Spalte gestrichen werden.*

**Lemma A.2 [Adjunkte einer polynomialen Matrix mit Grad 1, vgl. [77, Korollar 2.2]].** *Seien die Matrizen* $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$ *und der Parameter* $\rho \in \mathbb{R}$*. Dann gilt*

$$(\mathbf{A}+\rho\mathbf{B})^A = \mathbf{A}^A + \rho^{n-1}\mathbf{B}^A + \sum_{i=1}^{n-2} \rho^i \Gamma^i_{n-1}(\mathbf{A}/\mathbf{B}^i)^A,$$

*wobei das* $(k,j)$*-te Element der Matrix* $\Gamma^i_{n-1}(\mathbf{A}/\mathbf{B}^i)^A$ *ist*

$$\left(\Gamma^i_{n-1}(\mathbf{A}/\mathbf{B}^i)^A\right)_{kj} = (-1)^{k+j}\Gamma^i_{n-1}\det(\mathbf{A}_{jk}/\mathbf{B}^i_{jk}),$$

$\mathbf{A}_{jk}$ *und* $\mathbf{B}_{jk}$ *Matrizen nach Elimination der* $j$*-ten Reihe und* $k$*-ten Spalte der Matrizen* $\mathbf{A}$ *bzw.* $\mathbf{B}$ *sind, und* $\Gamma^i_{n-1}\det(\mathbf{A}_{jk}/\mathbf{B}^i_{jk})$ *die Summe der Determinanten, wobei die* $i$*-ten Reihen der Matrix* $\mathbf{A}_{jk}$ *durch diejenigen der Matrix* $\mathbf{B}_{jk}$ *substituiert werden.*

**Lemma A.3 [Newton-Regel].** *Gegeben sei das reelle Polynom* $m$*-ter Ordnung*

$$p(\mathbf{x},\theta) = a_m(\mathbf{x})\theta^m + a_{m-1}(\mathbf{x})\theta^{m-1} + \ldots + a_0(\mathbf{x})$$

*mit* $m > 0$, $a_i(\mathbf{x}) : \mathcal{X} \to \mathbb{R}$, $\mathcal{X} \subseteq \mathbb{R}^n$, $\forall\, i = 0,\ldots,m$, $a_m(\mathbf{x}) \neq 0$, $\forall\, \mathbf{x} \in \mathcal{X}$ *und eine Zahl* $L \in \mathbb{R}$*. Wenn an der Stelle* $\theta = L$ *das Polynom und seine sämtlichen partiellen Ableitungen* $\partial^i_\theta p(\mathbf{x}, L) \leq 0$, $\forall i \in \{0,1,\ldots,m\}$ *und* $\forall \mathbf{x} \in \mathcal{X}$ *sind, dann ist* $p(\mathbf{x},\theta) < 0$, $\forall \theta > L$, $\forall \mathbf{x} \in \mathcal{X}$*.*

*Beweis.* Das Lemma ist sehr ähnlich zur Lemma 4.1 aus [25].

Für $\theta = L + \epsilon$, mit $\epsilon \geq 0$ ist die Entwicklung des Polynoms $p(\mathbf{x},\theta)$ nach $\theta$ an der Stelle $\theta = L$ gegeben durch

$$p(\mathbf{x},\theta) = p(\mathbf{x},L) + \epsilon\partial_\theta p(\mathbf{x},L) + \frac{\epsilon^2}{2}\partial^2_\theta p(\mathbf{x},L) + \cdots + \frac{\epsilon^n}{n}\partial^m_\theta p(\mathbf{x},L).$$

Da $\partial^i_\theta p(\mathbf{x}, L) \leq 0$, $\forall i \in \{0,1,\ldots,m\}$ gilt und die $m$-te partielle Ableitung $\partial^m_\theta p(\mathbf{x},L) = m! \cdot a_m(\mathbf{x}) \neq 0$, $\forall\, \mathbf{x} \in \mathcal{X}$, folgt, dass $p(\mathbf{x},\theta) < 0$, $\forall\, \theta > L$ und $\mathbf{x} \in \mathcal{X}$. □

## A.2 Hilfssätze

**Lemma A.4.** *Folgende Aussagen sind equivalent:*

***i)*** *Die Ruhelage* $\mathbf{x} = \mathbf{0}$ *des Systems* $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$, $\mathbf{x} \in \mathbb{R}^n$ *ist asymptotisch stabil.*

***ii)*** $\exists\, \mathbf{P} \in \mathbb{P}^n$, *sodass* $\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^\top \prec \mathbf{0}$.

***iii)*** $\forall\, \mathbf{Q} \in \mathbb{P}^n$, $\exists!\, \mathbf{P} \in \mathbb{P}^n$, *welche die Gleichung* $\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^\top = -\mathbf{Q}$ *erfüllt.*

*Beweis.* iii) $\Rightarrow$ ii) ist offensichtlich. ii) $\Rightarrow$ i) folgt aus der Anwendung der direkten Methode von Ljapunov,[2)] mit der Ljapunov Funktion $V(\mathbf{x}) = \mathbf{x}^\top\mathbf{P}\mathbf{x}$, und unter Berücksichtigung der Tatsache, dass die Matrizen $\mathbf{A}$ und $\mathbf{A}^\top$ die gleichen Eigenwerte besitzen. i) $\Rightarrow$ iii) Ist die Matrix $\mathbf{A}$ asymptotisch stabil, so ist auch $\mathbf{A} \oplus \mathbf{A} = \mathbf{A} \otimes \mathbf{I}_n + \mathbf{I}_n \otimes \mathbf{A}$.[3)] Folglich sind alle Eigenwerte der Matrix $\mathbf{A} \oplus \mathbf{A}$ negativ und somit die Matrix $\mathbf{A} \oplus \mathbf{A}$ nichtsingulär. Gleichung $\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^\top = -\mathbf{Q}$ ist im Weiteren äquivalent zu[4)]

$$\operatorname{vec}(\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^\top) + \operatorname{vec}(\mathbf{Q}) = (\mathbf{A} \oplus \mathbf{A}) \operatorname{vec}(\mathbf{P}) + \operatorname{vec}(\mathbf{Q}) = \mathbf{0}.$$

Da $\mathbf{A} \oplus \mathbf{A}$ nichtsingulär ist, ist die Lösung $\mathbf{P} = -\operatorname{vec}^{-1}((\mathbf{A} \oplus \mathbf{A})^{-1} \operatorname{vec}(\mathbf{Q}))$ eindeutig. Darüber hinaus gilt

$$\begin{aligned}\int_0^t \mathbf{A} e^{\tau\mathbf{A}} \mathrm{d}\tau &= \int_0^t \mathbf{A} \sum_{k=0}^{\infty} \frac{1}{k!} (\tau\mathbf{A})^k \mathrm{d}\tau \\ &= \int_0^t \sum_{k=0}^{\infty} \frac{1}{k!} \tau^k \mathbf{A}^{k+1} \mathrm{d}\tau \\ &= \sum_{k=0}^{\infty} \frac{t^{k+1}}{(k+1)!} \mathbf{A}^{k+1} = e^{t\mathbf{A}} - \mathbf{I}_n\end{aligned}$$

und folglich, weil $\mathbf{A}$ Hurwitz und somit auch nichtsingulär ist

$$\lim_{t\to\infty} \int_0^t e^{\tau\mathbf{A}} \mathrm{d}\tau = \lim_{t\to\infty} \mathbf{A}^{-1}(e^{\mathbf{A}t} - \mathbf{I}_n) = -\mathbf{A}^{-1}.$$

[2)] Vgl. z.B. [8, Satz 11.7.2].
[3)] Vgl. [8, Satz 11.18.32].
[4)] Vgl. dazu Gl. (A.2).

Darüber hinaus gilt für die Lösung der Ljapunov-Gleichung $\mathbf{AP} + \mathbf{PA}^\top = -\mathbf{Q}$

$$\begin{aligned}\mathbf{P} &= \operatorname{vec}^{-1}\left(-(\mathbf{A} \oplus \mathbf{A})^{-1} \operatorname{vec}(\mathbf{Q})\right) = \operatorname{vec}^{-1}\left(\int_0^\infty e^{t(\mathbf{A}\oplus\mathbf{A})} \mathrm{d}t \operatorname{vec}(\mathbf{Q})\right) \\ &= \int_0^\infty \operatorname{vec}^{-1}\left(e^{t(\mathbf{A}\oplus\mathbf{A})} \operatorname{vec}(\mathbf{Q})\right) \mathrm{d}t = \int_0^\infty \operatorname{vec}^{-1}\left[\left(e^{t\mathbf{A}} \otimes e^{t\mathbf{A}}\right) \operatorname{vec}(\mathbf{Q})\right] \mathrm{d}t \\ &= \int_0^\infty \operatorname{vec}^{-1}\left[\operatorname{vec}\left(e^{t\mathbf{A}}\mathbf{Q}e^{t\mathbf{A}^\top}\right)\right] \mathrm{d}t = \int_0^\infty e^{t\mathbf{A}}\mathbf{Q}e^{t\mathbf{A}^\top} \mathrm{d}t.\end{aligned}$$

Da $\mathbf{Q} \succ 0$ ist, existiert immer eine nichtsinguläre Matrix $\mathbf{C}$, sodass $\mathbf{Q} = \mathbf{CC}^\top$. Somit gilt für einen beliebigen Vektor $\mathbf{x} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$,

$$\mathbf{x}^\top \mathbf{P} \mathbf{x} = \int_0^\infty \mathbf{x}^\top e^{t\mathbf{A}} \mathbf{C}\mathbf{C}^\top e^{t\mathbf{A}^\top} \mathbf{x} \, \mathrm{d}t = \int_0^\infty \left\| \mathbf{C}^\top e^{t\mathbf{A}^\top} \mathbf{x} \right\|^2 \mathrm{d}t > 0,$$

d.h. die Matrix $\mathbf{P}$ ist positiv definit. Schließlich, weil die Matrix $\mathbf{Q}$ symmetrisch angenommen wurde und somit,

$$(\mathbf{AP} + \mathbf{PA}^\top) - (\mathbf{AP} + \mathbf{PA}^\top)^\top = \mathbf{A}(\mathbf{P} - \mathbf{P}^\top) + (\mathbf{P} - \mathbf{P}^\top)\mathbf{A}^\top = \mathbf{0}$$

gilt, folgt, dass $\mathbf{P} - \mathbf{P}^\top = \mathbf{0}$, d.h. die Matrix $\mathbf{P}$ ist auch symmetrisch. □

*Bemerkung* A.4 (Analytische Ljapunov-Funktion). Für ein gegebenes konstantes $\theta \in \Theta$ ist die Lösung $\mathbf{P}(\theta) \in \mathbb{R}^{n \times n}$ einer parameterabhängigen Ljapunov-Gleichung (PDLG)

$$\mathbf{A}(\theta)\mathbf{P}(\theta) + \mathbf{P}(\theta)\mathbf{A}^\top(\theta) + \mathbf{Q}(\theta) = \mathbf{0}, \tag{A.11}$$

gegeben durch

$$\mathbf{P}(\theta) = \int_0^\infty e^{\mathbf{A}(\theta)t} \mathbf{Q}(\theta) e^{\mathbf{A}^\top(\theta)t} \mathrm{d}t,$$

analytisch in $\theta$ wenn $\mathbf{A}(\theta)$ und $\mathbf{Q}(\theta)$ analytisch in $\theta$ sind. In diesem Fall kann die Matrix $\mathbf{P}(\theta)$ durch

$$\mathbf{P}(\theta) = \mathbf{P}_0 + \theta\mathbf{P}_1 + \theta^2\mathbf{P}_2 + \ldots = \sum_{i=0}^{\infty} \theta^i \mathbf{P}_i \tag{A.12}$$

approximiert werden. Ist darüber hinaus die Menge $\Theta$ kompakt, so kann man die Potenzreihe aus Gl. (A.12) ab einem hinreichend großen aber endlichen Grad $m < \infty$ abbrechen, sodass die Ljapunov-Ungleichung aus Lemma A.4 ii) erfüllt ist, vgl. [78]. △

**Satz A.5** [Satz von Finsler, vgl. [68, Satz 2.3.10]] *Es seien die Matrizen* $\mathbf{B} \in \mathbb{C}^{n\times m}$, *mit* $\mathrm{Rang}\,(\mathbf{B}) = r < n$ *und* $\mathbf{B}^{\perp} \in \mathbb{C}^{(n-r)\times n}$, *sodass* $\mathbf{B}^{\perp}\mathbf{B} = \mathbf{0}$ *und* $\mathbf{B}^{\perp}\mathbf{B}^{\perp^*} \prec \mathbf{0}$. *und* $\mathbf{Q} \in \mathbb{H}^n$. *Angenommen sei im Weiteren, dass* $(\mathbf{B}_r,\mathbf{B}_l)$ *eine Voll-Rang-Faktorisierung*[a)] *der Matrix* $\mathbf{B}$ *ist und* $\mathbf{D} := (\mathbf{B}_r\mathbf{B}_r^*)^{-1/2}\mathbf{B}_l^+$. *Dann sind folgende Aussagen äquivalent:*

***i)*** $\exists\,\mu \in \mathbb{R}$, *sodass*

$$\mu\mathbf{B}\mathbf{B}^* - \mathbf{Q} \succ \mathbf{0}. \tag{A.13}$$

***ii)*** *Es gilt*

$$\mathbf{R} := \mathbf{B}^{\perp}\mathbf{Q}\mathbf{B}^{\perp^*} \prec \mathbf{0}. \tag{A.14}$$

*Wenn Gl.* (A.13) *und* (A.14) *gelten, dann gilt auch*

$$\mu > \mu_{\min} := \lambda_{\max}\left[\mathbf{D}(\mathbf{Q} - \mathbf{Q}\mathbf{B}^{\perp^*}\mathbf{R}^{-1}\mathbf{B}^{\perp}\mathbf{Q})\mathbf{D}^*\right]. \tag{A.15}$$

[a)] Vgl. Def. 20 (Anahng).

*Beweis.* Die Notwendigkeit der Bedingung (A.14) kann durch eine Kongruenztransformation[5)] der Matrix $\mu\mathbf{B}\mathbf{B}^* - \mathbf{Q} \succ \mathbf{0}$ aus Gl. (A.13), mit $\mathbf{S} = [\mathbf{D}\ \ \mathbf{B}^{\perp}]^{\top}$ nachgewiesen werden. Es folgt[6)]

$$\begin{aligned}
\mathbf{S}(\mu\mathbf{B}\mathbf{B}^* - \mathbf{Q})\mathbf{S}^* &= \begin{bmatrix}\mathbf{D}\\ \mathbf{B}^{\perp}\end{bmatrix}(\mu\mathbf{B}\mathbf{B}^* - \mathbf{Q})\begin{bmatrix}\mathbf{D} & \mathbf{B}^{\perp}\end{bmatrix}\\
&= \begin{bmatrix}\mathbf{D}(\mu\mathbf{B}\mathbf{B}^* - \mathbf{Q})\mathbf{D}^* & \mathbf{D}(\mu\mathbf{B}\mathbf{B}^* - \mathbf{Q})\mathbf{B}^{\perp^*}\\ \mathbf{B}^{\perp}(\mu\mathbf{B}\mathbf{B}^* - \mathbf{Q})\mathbf{D}^* & \mathbf{B}^{\perp}(\mu\mathbf{B}\mathbf{B}^* - \mathbf{Q})\mathbf{B}^{\perp^*}\end{bmatrix}\\
&= \begin{bmatrix}\mu\,\mathbf{D}\mathbf{B}\mathbf{B}^*\mathbf{D}^* - \mathbf{D}\mathbf{Q}\mathbf{D}^* & -\mathbf{D}\mathbf{Q}\mathbf{B}^{\perp^*}\\ -\mathbf{B}^{\perp}\mathbf{Q}\mathbf{D}^* & -\mathbf{B}^{\perp}\mathbf{Q}\mathbf{B}^{\perp^*}\end{bmatrix} \succ \mathbf{0}.
\end{aligned}$$

Dabei gilt

$$\mathbf{D}\mathbf{B}\mathbf{B}^*\mathbf{D}^* = (\mathbf{B}_r\mathbf{B}_r^*)^{-1/2}\underbrace{\mathbf{B}_l^+\mathbf{B}_l}_{\mathbf{I}_r}\mathbf{B}_r\mathbf{B}_r^*\underbrace{\mathbf{B}_l^*\mathbf{B}_l^{+^*}}_{\mathbf{I}_r}[(\mathbf{B}_r\mathbf{B}_r^*)^{-1/2}]^* = \mathbf{I}_r.$$

[5)] Vgl. Def. 19.

[6)] Vgl. auch die Bemerkung A.1 bezüglich der positiven Definitheit von zwei *kongruenten* Matrizen.

Somit ist $\mathbf{S}(\mu\mathbf{B}\mathbf{B}^* - \mathbf{Q})\mathbf{S}^* \succ \mathbf{0}$ äquivalent zu[7)]

$$\begin{aligned}
&\mathbf{R} := \mathbf{B}^{\perp}\mathbf{Q}\mathbf{B}^{\perp^*} \prec \mathbf{0},\\
&(\mu\,\mathbf{I}_r - \mathbf{D}\mathbf{Q}\mathbf{D}^*)\,|(\mathbf{S}(\mu\mathbf{B}\mathbf{B}^* - \mathbf{Q})\mathbf{S}^*)\\
&\qquad = \mu\,\mathbf{I}_r - \mathbf{D}\mathbf{Q}\mathbf{D}^* + \mathbf{D}\mathbf{Q}\mathbf{B}^{\perp^*}\mathbf{R}^{-1}\mathbf{B}^{\perp}\mathbf{Q}\mathbf{D}^* \succ \mathbf{0}, \qquad \text{(A.16)}
\end{aligned}$$

es folgt also Gl. (A.14).

Bedingung (A.14) ist auch hinreichend, da für eine beliebige negativ definite Matrix $\mathbf{R} \prec \mathbf{0}$ immer eine Zahl $\mu \in \mathbb{R}$ existiert, sodass Gl. (A.16) erfüllt ist, und zwar ist diese Zahl gegeben durch Gl. (A.15). Es folgt daraus Gl. (A.13). □

*Bemerkung* A.5. Es gilt noch, vgl. [68], $\mu_{\min} \leq 0$ dann und nur dann wenn $\mathbf{Q} \preceq \mathbf{0}$. Dies folgt aus

$$\begin{aligned}
\left\{ \begin{array}{l} \mu_{\min} \leq 0 \Leftrightarrow \mathbf{D}(\mathbf{Q} - \mathbf{Q}\mathbf{B}^{\perp^*}\mathbf{R}^{-1}\mathbf{B}^{\perp}\mathbf{Q})\mathbf{D}^* \preceq \mathbf{0} \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \mathbf{R} \prec \mathbf{0} \end{array} \right.\\
\Leftrightarrow \begin{bmatrix} \mathbf{D}\mathbf{Q}\mathbf{D}^* & \mathbf{D}\mathbf{Q}\mathbf{B}^{\perp^*} \\ \mathbf{B}^{\perp}\mathbf{Q}\mathbf{D}^* & \mathbf{R} \end{bmatrix} \preceq \mathbf{0}.
\end{aligned}$$

Dies ist äquivalent zu $\mathbf{S}\mathbf{Q}\mathbf{S}^* \preceq \mathbf{0}$ und, somit, zu $\mathbf{Q} \preceq \mathbf{0}$. △

## A.3 Umwandlung der unendlich- in endlich-dimensionale LMIs

**Lemma A.6 [*S*-Prozedur].** *Gegeben seien die Matrizen $\mathbf{Q} \in \mathbb{H}^n$ und $\mathbf{S}_i \in \mathbb{H}^n$, mit $i = 1,\ldots,m$. Die Ungleichung*

$$\mathbf{x}^{\top}\mathbf{Q}\mathbf{x} < 0, \quad \forall\, \mathbf{x} \in \mathcal{X} := \{\mathbf{x} \in \mathbb{R}^n_* | \mathbf{x}^{\top}\mathbf{S}_i\mathbf{x} \leq 0, \forall\, i = 1,\ldots,m\} \qquad \text{(A.17)}$$

*ist genau dann erfüllt, wenn die Skalare $\tau_i > 0$ existieren, sodass*

$$\mathbf{Q} \prec \sum_{i=1}^{m} \tau_i \mathbf{S}_i \qquad \text{(A.18)}$$

[7)] Vgl. [8, Fakt 8.2.4] für die positive Definitheit des Schur-Komplements einer positiv definiten Blockmatrix.

*gilt.*

Die Bestimmung der Skalare $\tau_i$ stellt ein endlich-dimensionales konvexes Validierungsproblem dar. In mehreren Spezialfällen ist Gl. (A.18) nicht nur hinreichend - wie in Lemma A.6 dargestellt - sondern auch notwendig für die Erfüllung der Gl. (A.17).

Die Umwandlung einer parameterabhängigen LMI der Form (A.5) in eine parameterunabhängige LMI basiert auf die in [38] vorgestellte Verallgemeinerung der $S$-Prozedur. Für den eindimensionalen Fall, d.h. für $d = 1$, wurde die Umwandlung in [77] gezeigt. Die Lemmas A.7 bis A.9 fassen dieses Ergebnis zusammen. Die entsprechenden Beweise können in [77] gefunden werden.

**Lemma A.7 [[78], Erweiterung der verallgemeinerten *S*-Prozedur].** *Gegeben seien die Matrizen* $\mathbf{\Sigma} \in \mathrm{Sym}^n$ *und* $\mathbf{J}, \mathbf{C} \in \mathbb{R}^{k\times n}$. *Die folgenden Aussagen sind äquivalent:*

*i)* $\zeta^\top \mathbf{\Sigma} \zeta < 0, \quad \forall \zeta \in \mathcal{Z} := \{\zeta \in \mathbb{R}^n \backslash \{\mathbf{0}\} \,|(\mathbf{J} - \delta \mathbf{C})\zeta = \mathbf{0}, \delta \in \mathbb{R}, |\delta| \leq 1\}.$

*ii)* $\exists\, \mathbf{D} \in \mathbb{P}^k, \mathbf{G} \in \mathrm{Skew}^k$, *sodass*

$$\mathbf{\Sigma} \prec \begin{bmatrix} \mathbf{C} \\ \mathbf{J} \end{bmatrix}^\top \begin{bmatrix} -\mathbf{D} & \mathbf{G} \\ \mathbf{G} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{C} \\ \mathbf{J} \end{bmatrix}.$$

Lemma A.7 gibt eine notwendige und hinreichende Bedingung für $i$) in Form einer linearen Matrixungleichung (LMI) aus $ii$) an. Dass diese Bedingung somit nicht konservativ ist, folgt aus der verallgemeinerten $S$-Prozedur, vgl. [38, Theorem 1]. Die Menge $\mathcal{G}$ der in dem Teil $i$) des Lemmas A.7 betrachteten Vektoren $\zeta \in \mathbb{R}^n \backslash \{\mathbf{0}\}$ ist für bestimmte Matrizen $\mathbf{J}$ und $\mathbf{C}$ gleich zu der gesuchten Menge aus Gl. (A.19), wie es in Lemma A.8 dargestellt ist.

### A.3.1 Einparametriger Fall ($d = 1$)

Gesucht ist eine parameterunabhängige Bedingung für die Erfüllung der Ungleichung

$$(\mathbf{z}^{[k]} \otimes \mathbf{I}_n)^\top \mathbf{M}_k \, (\mathbf{z}^{[k]} \otimes \mathbf{I}_n) < \mathbf{0}, \quad \mathbf{z}^{[k]} = \begin{bmatrix} 1 & z & z^2 & \cdots & z^{k-1} \end{bmatrix}^\top, \forall\, z \in [-1{,}1], \tag{A.19}$$

wobei die Matrix $\mathbf{M}_k \in \mathrm{Sym}^{kn}$ vorgegeben ist. Dafür wird folgendes Lemma aus [78] verwendet:

**Lemma A.8 [[78]].** *Folgende Mengen sind gleich:*

$$\begin{aligned}
\mathcal{C}_1 := \Big\{ \zeta \in \mathbb{R}^{nk} \,\Big|\, & (\mathbf{J} - \delta\mathbf{C})\zeta = \mathbf{0}, \mathbf{J} = \check{\mathbf{J}}_{k-1} \otimes \mathbf{I}_n, \mathbf{C} = \hat{\mathbf{J}}_{k-1} \otimes \mathbf{I}_n, \\
& \check{\mathbf{J}}_{k-1} = [\mathbf{0}_{k-1,1} \;\; \mathbf{I}_{k-1}], \hat{\mathbf{J}}_{k-1} = [\mathbf{I}_{k-1} \;\; \mathbf{0}_{k-1,1}], \delta \in [-1,1] \Big\}, \\
\mathcal{C}_2 := \Big\{ \zeta \in \mathbb{R}^{nk} \,\Big|\, & \zeta = (\mathbf{z}^{[k]} \otimes \mathbf{I}_n)\mathbf{v}, \mathbf{z}^{[k]} = \begin{bmatrix} 1 & z & z^2 & \cdots & z^{k-1} \end{bmatrix}^\top, \\
& z \in [-1,1], \mathbf{v} \in \mathbb{R}^n \Big\}.
\end{aligned}$$

Schließlich, unter Verwendung der Lemmas A.7 und A.8, folgt

**Lemma A.9 [[77], Lemma 4.12].** *Die Ungleichung*

$$(\mathbf{z}^{[k]} \otimes \mathbf{I}_n)^\top \mathbf{\Sigma} \, (\mathbf{z}^{[k]} \otimes \mathbf{I}_n) < \mathbf{0}, \quad \mathbf{z}^{[k]} = \begin{bmatrix} 1 & z & z^2 & \cdots & z^{k-1} \end{bmatrix}^\top \tag{A.20}$$

*ist für jedes $z \in [-1,1]$ erfüllt dann und nur dann wenn es existieren die Matrizen $\mathbf{D} \in \mathbb{P}^{n(k-1)}$ und $\mathbf{G} \in \mathrm{Skew}^{n(k-1)}$ sodass*

$$\mathbf{\Sigma} \prec \begin{bmatrix} \mathbf{C} \\ \mathbf{J} \end{bmatrix}^\top \begin{bmatrix} -\mathbf{D} & \mathbf{G} \\ \mathbf{G} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{C} \\ \mathbf{J} \end{bmatrix}. \tag{A.21}$$

# B Ausgewählte stochastische Grundlagen

## B.1 Die multivariate Normalverteilung

Sei der Zufallsvektor $\mathbf{X} := (X_1, \ldots, X_r)$ gebildet aus $r$ standardnormalverteilten Zufallsvariablen $X_i \sim \mathcal{N}_1(0,1)$, $i = 1, \ldots, r$. Der Zufallsvektor

$$\mathbf{W} := \mathbf{L}\mathbf{X} + \boldsymbol{\mu}, \tag{B.1}$$

mit $\mathbf{L} \in \mathbb{R}^{m \times r}$ und $\boldsymbol{\mu} \in \mathbb{R}^m$, heißt *multivariat normalveteilt*. Dies wird durch $\mathbf{W} \sim \mathcal{N}_m(\boldsymbol{\mu}, \Sigma)$ bezeichnet. Die Wahrscheinlichkeitsdichtefunktion des Zufallsvektors lautet

$$f_W(\mathbf{w}) = \frac{1}{(2\pi)^{m/2}\sqrt{\det(\Sigma)}} \exp\left\{-\frac{1}{2}(\mathbf{w} - \boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{w} - \boldsymbol{\mu})\right\}, \tag{B.2}$$

wobei $\mathbf{w} \in \mathbb{R}^m$ und $\mathrm{Rang}(\Sigma) = m$.

## B.2 Die Chi-Quadrat Verteilung

Seien $X_1, \ldots, X_n$ unabhängige und identisch verteilte Zufallsvariablen, mit $X_i \sim \mathcal{N}_1(0,1)$. Die Verteilung der Zufallsvariable

$$W := X_1^2 + \ldots + X_n^2 \tag{B.3}$$

heißt *Chi-Quadrat* mit $n$ Freiheitsgraden. Dies wird durch $W \sim \chi_n^2$ bezeichnet. Es gilt dabei $\mathrm{E}\{W\} = n$ und $\mathrm{Var}\{W\} = 2n$, vgl. auch [26]. Die Verteilung der Zufallsvariable

$$Y := \frac{1}{W} \tag{B.4}$$

heißt *inverse-Chi-Quadrat* mit $n$ Freiheitsgraden und wird mit $Y \sim \chi_n^{-2}$ bezeichnet. Es gilt dabei

$$\mathrm{E}\{Y\} = \frac{1}{n-2}, \quad n > 2, \tag{B.5}$$

$$\mathrm{Var}\{Y\} = \frac{2}{(n-2)^2(n-4)}, \quad n > 4. \tag{B.6}$$

## B.3 Die nicht-zentrale $t$-Verteilung

Seien die unabhängigen Zufallsvariablen $X \sim \mathcal{N}_1(0,1)$ und $Z \sim \chi_n^2$. Die Verteilung der Zufallsvariable

$$\tilde{W} := \frac{X}{\sqrt{Z/n}} \tag{B.7}$$

heißt *$t$-Verteilung* mit $n$ Freiheitsgraden. Diese wird mit $\tilde{W} \sim \mathcal{T}_1(\nu_W,0,1)$ bezeichnet, vgl. [26].

Eine Zufallsvariable $W \sim \mathcal{T}_1(\nu_W,\mu_W,\sigma_W^2)$ heißt *nicht-zentral* $t$-verteilt. Ihre Wahrscheinlichkeitsdichtefunktion lautet

$$f_W(w) = \frac{\Gamma((\nu_W+1)/2)}{\sqrt{\sigma_W^2 \pi \nu_W}\,\Gamma(\nu_W/2)} \left(1 + \frac{1}{\nu_W}\frac{(w-\mu_W)^2}{\sigma_W^2}\right)^{-(\nu_W+1)/2}. \tag{B.8}$$

Es gilt dabei

$$W \sim \mathcal{T}_1(\nu_W,\mu_W,\sigma_W^2) \Leftrightarrow \frac{W-\mu_W}{\sigma_W^2} \sim \mathcal{T}_1(\nu_W,0,1). \tag{B.9}$$

Weitere Eigenschaften sind

$$\mathrm{E}\{W\} = \mu_W, \text{ falls } \nu_W > 1, \tag{B.10}$$

$$\mathrm{Var}\{W\} = \sigma_W^2 \frac{\nu_W}{\nu_W - 2}, \text{ falls } \nu_W > 2. \tag{B.11}$$

## B.4 Prädiktive Distributionen

**Satz B.1** [Vgl. [64, Theorem 4.1.2]] *Es sei angenommen, dass der Zufallsvektor $(H_0,\mathbf{H})$ eine multivariate Normalverteilung besitzt, d.h.*

$$(H_0,\mathbf{H})|\boldsymbol{\beta},\sigma_Z^2 \sim \mathcal{N}_{N+1}\left(\begin{bmatrix}\mathbf{f}_0^\top \\ \mathbf{F}\end{bmatrix}\boldsymbol{\beta},\sigma_Z^2\begin{bmatrix}1 & \mathbf{r}_0^\top \\ \mathbf{r}_0 & \mathbf{R}\end{bmatrix}\right), \tag{B.12}$$

*wobei* $\beta \in \mathbb{R}^p$ *und* $\sigma_Z^2 > 0$ *unbekannt sind, aber der Zufallsvektor* $(\boldsymbol{\beta},\sigma_Z^2)$ *eine der Verteilungen* $(1)-(4)$ *aus Tabelle 8.2 besitzt, wobei die Parameter* $\boldsymbol{\beta}_0 \in \mathbb{R}^p$, $\boldsymbol{\Sigma}_0 \succeq 0$, $c_0 > 0$ *und* $\nu_0 > 2$ *bekannt sind. Darüber hinaus sind der Vektor* $\mathbf{r}_0 := [R(\boldsymbol{\zeta}_0 - \boldsymbol{\zeta}_1) \quad \cdots \quad R(\boldsymbol{\zeta}_0 - \boldsymbol{\zeta}_N)]^\top$ *und die Matrix* $\mathbf{R}$*, mit den Elementen* $\mathbf{R}_{(i,j)} := R(\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_j)$*, mit* $i,j = 1,\ldots,N$*, abhängig von einer parametrischen Korrelationsfunktion* $R(\boldsymbol{\zeta} - \boldsymbol{\zeta}') = r(\boldsymbol{\zeta},\boldsymbol{\zeta}';\boldsymbol{\psi})$ *mit einem unbekannten Parametervektor* $\boldsymbol{\psi}$*.*

*Die Zufallsvariable* $H_0|\mathbf{H}$ *besitzt dann eine eindimensionale nicht-zentrale t-Verteilung gegeben durch*

$$[H_0|\mathbf{H}] \sim \mathcal{T}_1(\nu_i,\mu_i,\sigma_i^2), \tag{B.13}$$

*mit* $\nu_i$ *Freiheitsgraden, Nichtzentralitätsparameter* $\mu_i$ *und Skalierungsparameter* $\sigma_i^2$*. Der Parameter* $\nu_i$ *ergibt sich aus*

$$\nu_i = \begin{cases} N + \nu_0, & i = (1), \\ N, & i = (2), \\ N - p + \nu_0, & i = (3), \\ N - p, & i = (4). \end{cases} \tag{B.14}$$

*Der Nichtzentralitätsparameter* $\mu_i$ *ergibt sich aus*

$$\mu_i(\boldsymbol{\zeta}_0) = \mathrm{E}\{H_0|\mathbf{H}\} = \mathbf{f}_0^\top\hat{\boldsymbol{\beta}} + \mathbf{r}_0^\top\mathbf{R}^{-1}(\boldsymbol{\eta} - \mathbf{F}\hat{\boldsymbol{\beta}}), \tag{B.15}$$

*wobei - in diesem Fall - der Parametervektor* $\hat{\boldsymbol{\beta}}$ *aus Gl.* (8.12) *die analytische Form*

$$\hat{\boldsymbol{\beta}} = \begin{cases} \left(\mathbf{F}^\top\mathbf{R}^{-1}\mathbf{F} + \boldsymbol{\Sigma}_0^{-1}\right)^{-1}\left(\mathbf{F}^\top\mathbf{R}^{-1}\boldsymbol{\eta} + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\beta}_0\right), & i = (1) \text{ oder } (2) \\ \left(\mathbf{F}^\top\mathbf{R}^{-1}\mathbf{F}\right)^{-1}\left(\mathbf{F}^\top\mathbf{R}^{-1}\boldsymbol{\eta}\right), & i = (3) \text{ oder } (4) \end{cases} \tag{B.16}$$

*hat. Schließlich ist der Parameter* $\sigma_i^2$

$$\sigma_i^2 := \sigma_i^2(\boldsymbol{\zeta}_0) = \hat{\sigma}_Z^2 \cdot r^*(\boldsymbol{\zeta}_0,\boldsymbol{\zeta}_0) = \hat{\sigma}_Z^2 \cdot \left\{1 - \begin{bmatrix}\mathbf{f}_0 \\ \mathbf{r}_0\end{bmatrix}^\top \begin{bmatrix}\boldsymbol{\Sigma}_i & \mathbf{F}^\top \\ \mathbf{F} & \mathbf{R}\end{bmatrix}^{-1} \begin{bmatrix}\mathbf{f}_0 \\ \mathbf{r}_0\end{bmatrix}\right\}, \tag{B.17}$$

*mit*

$$\hat{\sigma}_Z^2 = \frac{Q_i^2}{\nu_i}, \tag{B.18}$$

$$r^*(\boldsymbol{\zeta},\boldsymbol{\zeta}') := r(\boldsymbol{\zeta},\boldsymbol{\zeta}'|\mathbf{H}) = r(\boldsymbol{\zeta},\boldsymbol{\zeta}') - \begin{bmatrix}\mathbf{f}_{\boldsymbol{\zeta}}\\ \mathbf{r}_{\boldsymbol{\zeta}}\end{bmatrix}^\top \begin{bmatrix}\boldsymbol{\Sigma}_i & \mathbf{F}^\top\\ \mathbf{F} & \mathbf{R}\end{bmatrix}^{-1} \begin{bmatrix}\mathbf{f}_{\boldsymbol{\zeta}'}\\ \mathbf{r}_{\boldsymbol{\zeta}'}\end{bmatrix}, \tag{B.19}$$

$$r(\boldsymbol{\zeta},\boldsymbol{\zeta}') = R(\boldsymbol{\zeta}-\boldsymbol{\zeta}'), \tag{B.20}$$

$$\boldsymbol{\Sigma}_i = \begin{cases} -\boldsymbol{\Sigma}_0^{-1}, & i = (1)\ \textit{oder}\ (2)\\ \mathbf{0}, & i = (3)\ \textit{oder}\ (4)\end{cases} \tag{B.21}$$

*und*

$$Q_i^2 := \begin{cases} c_0 + Q_2^2, & i = (1),\\ Q_4^2 + (\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}})^\top \left[\boldsymbol{\Sigma}_0 + \left(\mathbf{F}^\top\mathbf{R}^{-1}\mathbf{F}\right)^{-1}\right]^{-1} (\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}), & i = (2),\\ c_0 + Q_4^2, & i = (3),\\ \boldsymbol{\eta}^\top \left[\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{F}\left(\mathbf{F}^\top\mathbf{R}^{-1}\mathbf{F}\right)^{-1}\mathbf{F}^\top\mathbf{R}^{-1}\right]\boldsymbol{\eta}, & i = (4).\end{cases} \tag{B.22}$$

Der Erwartungswert und Varianz der Zufallsvariable $H_0|\mathbf{H}$ können somit aus Gl. (B.10) und (B.11) berechnet werden.

### B.4.1 Berechnung der bedingten prädiktiven Verteilung $[H_0|\mathbf{H}]$ aus Schritt 5a

Die bedingte Verteilung $[H_0|\mathbf{H}]$ aus Schritt 5*a* kann aus

$$[H_0|\mathbf{H}] = \int_{\mathcal{D}_{\sigma_Z^2}} \left[H_0,\sigma_Z^2|\mathbf{H}\right] \mathrm{d}\sigma_Z^2,$$

berechnet werden, wobei für den Integrand

$$[H_0,\sigma_Z^2|\mathbf{H}] = [H_0|\mathbf{H},\sigma_Z^2]\cdot[\sigma_Z^2|\mathbf{H}] \tag{B.23}$$

gilt. Die Verteilung $[H_0|\mathbf{H},\sigma_Z^2]$ kann dabei wie folgt berechnet werden:

$$[H_0|\mathbf{H},\sigma_Z^2] = \int\cdots\int_{\mathcal{D}_{\boldsymbol{\beta}}} [H_0,\boldsymbol{\beta}|\mathbf{H},\sigma_Z^2]\mathrm{d}\boldsymbol{\beta},$$

wobei

$$[H_0,\boldsymbol{\beta}|\mathbf{H},\sigma_Z^2] = [H_0|\mathbf{H},\boldsymbol{\beta},\sigma_Z^2]\cdot[\boldsymbol{\beta}|\mathbf{H},\sigma_Z^2]. \tag{B.24}$$

Die erste Verteilung auf der rechten Seite der Gl. (B.24) ist bekannt. Die zweite kann mit Hilfe des Bayes'schen Satzes aus

$$[\boldsymbol{\beta}|\mathbf{H},\sigma_Z^2] = [\mathbf{H},\sigma_Z^2|\boldsymbol{\beta}] \cdot [\boldsymbol{\beta}], \tag{B.25}$$

berechnet werden, wobei die Verteilung $[\mathbf{H},\sigma_Z^2|\boldsymbol{\beta}]$ aus der Äquivalenz der Zerlegungen

$$[\mathbf{H},\sigma_Z^2,\boldsymbol{\beta}] = [\boldsymbol{\beta}] \cdot [\sigma_Z^2|\boldsymbol{\beta}] \cdot [\mathbf{H}|\sigma_Z^2,\boldsymbol{\beta}] = [\boldsymbol{\beta}] \cdot [\boldsymbol{\beta}|\sigma_Z^2] \cdot [\sigma_Z^2] \cdot [\mathbf{H}|\sigma_Z^2,\boldsymbol{\beta}],$$
$$[\mathbf{H},\sigma_Z^2,\boldsymbol{\beta}] = [\boldsymbol{\beta}] \cdot [\mathbf{H},\sigma_Z^2|\boldsymbol{\beta}]$$

berechnet werden kann, d.h.

$$[\mathbf{H},\sigma_Z^2|\boldsymbol{\beta}] = [\boldsymbol{\beta}|\sigma_Z^2] \cdot [\sigma_Z^2] \cdot [\mathbf{H}|\sigma_Z^2,\boldsymbol{\beta}].$$

Dabei sind die Verteilungen $[\mathbf{H}|\sigma_Z^2,\boldsymbol{\beta}]$, $[\boldsymbol{\beta}|\sigma_Z^2]$ und $[\sigma_Z^2]$ bekannt. Schließlich kann die Verteilung $[\sigma_Z^2|\mathbf{H}]$ aus Gl. (B.23) aus der Äquivalenz der Zerlegungen

$$[\mathbf{H},\sigma_Z^2,\boldsymbol{\beta}] = [\boldsymbol{\beta}|\mathbf{H},\sigma_Z^2] \cdot [\mathbf{H},\sigma_Z^2] = [\boldsymbol{\beta}|\mathbf{H},\sigma_Z^2] \cdot [\sigma_Z^2|\mathbf{H}] \cdot [\mathbf{H}],$$
$$[\mathbf{H},\sigma_Z^2,\boldsymbol{\beta}] = [\mathbf{H}|\boldsymbol{\beta},\sigma_Z^2] \cdot [\boldsymbol{\beta}|\sigma_Z^2] \cdot [\sigma_Z^2],$$

berechnet werden, d.h.

$$[\sigma_Z^2|\mathbf{H}] = \frac{[\mathbf{H}|\boldsymbol{\beta},\sigma_Z^2] \cdot [\boldsymbol{\beta}|\sigma_Z^2] \cdot [\sigma_Z^2]}{[\boldsymbol{\beta}|\mathbf{H},\sigma_Z^2] \cdot [\mathbf{H}]},$$

wobei die Verteilung $[\boldsymbol{\beta}|\mathbf{H},\sigma_Z^2]$ in Gl. (B.25) berechnet wird, und die Verteilungen $[\mathbf{H}|\boldsymbol{\beta},\sigma_Z^2]$ und $[\boldsymbol{\beta}|\sigma_Z^2]$ bekannt sind. Der Erwartungswert

$$\mu_{H_0|\mathbf{H}}(\boldsymbol{\zeta}_0) = \mu_{H_0|\mathbf{H}}(\boldsymbol{\zeta}_0|\boldsymbol{\psi}) \tag{B.26}$$

und die Varianz

$$\sigma^2_{H_0|\mathbf{H}}(\boldsymbol{\zeta}) = \sigma^2_{H_0|\mathbf{H}}(\boldsymbol{\zeta}_0|\boldsymbol{\psi}) \tag{B.27}$$

der *A-posteriori*-Wahrscheinlichkeitsdichte $[H_0|\mathbf{H}]$ hängen vom Parametervektor $\boldsymbol{\psi}$ der Korrelationsfunktion $R(\,\cdot\,|\boldsymbol{\psi})$ ab.

## B.5 Latin-Hypercube-Sampling (LHS)

Die Methode wird für den zweidimensionalen Fall erläutert. Ist beispielsweise der Parameterbereich $[0,1]^2$, so kann man darin $N$ Design-Punkte durch *Latin-Hypercube-Sampling* wie folgt generieren. In einem ersten

Schritt wird über den untersuchten Bereich ein Gitter gelegt, mit jeweils $N$ gleich großen Intervallen pro Dimension. Es ergeben sich somit $N^2$ Zellen. Jede Zelle wird mit einer Zahl aus der Menge $\{1,2,\ldots,N\}$ versehen, sodass ein *lateinisches Quadrat* entsteht. Dieses ist dadurch gekennzeichnet, dass jede Zahl nur einmal in jeder Zeile und jeder Spalte erscheint. Anschließend wählt man zufällig eine Zahl aus der Menge $\{1,2,\ldots,N\}$, und in jeder damit versehenen Zelle wird ein Punkt zufällig gewählt.

# C Parameter der Beispiele

Die Folgenden LMI-basierten Validierungs- und Optimierungsprobleme wurden unter Anwendung des Parsers YALMIP, vgl. [46], und des Solvers SDPT3, vgl. [71], in der Programmiersprache MATLAB formuliert und gelöst.

## C.1 Parameter für das Beispiel 5.5.2

### C.1.1 die *klassische* WSVR mittels iLF

Die Lösung der Optimierungsproblems (3.14)-(3.16) mit $\varepsilon = 0.01$ und Optimierungsziel

$$\min_{\mathbf{P}\in \mathrm{Sym}^2} \mathrm{Spur}(\mathbf{P})$$

ergibt die (in Originalkoordinaten transformierte) Matrix

$$\mathbf{P} = \begin{bmatrix} 2.3584 & 9.7426 \\ 9.7426 & 46.1678 \end{bmatrix}.$$

Das maximale Einzugsgebiet ist durch das kleinste $d > \mathbf{b}^\top \mathbf{P}^{-1}\mathbf{b}/4$, d.h. durch ein $d_{\min} > 0.5412$, bestimmt. Der Parameter $d$ kann aber auch so gewählt werden, dass eine gewünschte Anfangsauslenkung innerhalb oder auf dem Rand des Einzugsgebiets liegt, jedoch solange Gl. (3.18) und (3.19) erfüllt sind. Für $d = 0.5682$ und $\nu = 0.99$ ergibt sich $c = 0.5072$ und $r = 0.2868$. Der Parameter $d$ wurde so gewählt, dass die gewünschte Anfangsauslenkung $\mathbf{x}_0 = [0.7,2.8]^\top$ auf dem Rand des Einzugsgebiets liegt. Im Bild 5.3 wird die sich ergebende Ellipse, $\mathcal{E}(d\mathbf{P}^{-1})$ (-.) gezeigt.

### C.1.2 Die *invers-polynomiale* WSVR

Die Lösung des Validierungsproblems (4.10)-(4.12) mit $\mathrm{M_l} = -1$ und $\mathrm{M_u} = 0$, sowie $\varepsilon = 0.01$ ergibt die Matrizen

$$\mathbf{P}_{c_{-1}} = \begin{bmatrix} -0.1000 & -0.4897 \\ -0.4897 & -2.4573 \end{bmatrix},$$
$$\mathbf{P}_{c_0} = \begin{bmatrix} 15.5918 & 76.5334 \\ 76.5334 & 383.5764 \end{bmatrix}.$$

Für $v = 1$ und $v = \varepsilon$ ergeben sich die Matrizen

$$\mathbf{P}_1^{-1} = \begin{bmatrix} 2.0076 & 0.0052 \\ 0.0052 & 0.0130 \end{bmatrix}, \quad \mathbf{P}_\varepsilon^{-1} = \begin{bmatrix} 8.0040 & 0.0078 \\ 0.0078 & 0.0358 \end{bmatrix}.$$

Die Koeffizienten des Matrixpolynoms $\mathbf{R}_v^A := \mathbf{P}_v^A$ aus Korollar 5.3 sind

$$\mathbf{N}_0 = \mathbf{P}_{\mathrm{c}_{-1}}^A = \begin{bmatrix} -0.2246 & -0.4419 \\ -0.4419 & -0.8842 \end{bmatrix}, \quad \mathbf{N}_1 = \mathbf{P}_{\mathrm{c}_0}^A = \begin{bmatrix} 34.9526 & 68.9648 \\ 68.9648 & 138.0588 \end{bmatrix}.$$

Die Überprüfung der Bedingungen aus Korollar 5.3 mit dem Skalierungsfaktor $d = 1.1058$ ergibt

$$\max_{\mathbf{x}^\top \mathbf{P}_1^{-1} \mathbf{x} = d} G(\mathbf{x},1) = -52.2830,$$
$$\max_{\mathbf{x}^\top \mathbf{P}_1^{-1} \mathbf{x} = d} \partial_{\tilde{v}} G(\mathbf{x},1) = -52.6196.$$

## C.2 Parameter für das Beispiel 8.4.1

Das Streckenensemble aus Gl. (8.77) ist gegeben durch

$$\mathbf{A}_0 = \begin{bmatrix} -1.4095 & 0.3255 \\ 1.7701 & -1.1190 \end{bmatrix}, \quad \mathbf{A}_1 = \begin{bmatrix} 0.6204 & -0.8960 \\ 1.2698 & 0.1352 \end{bmatrix}$$

und

$$\mathbf{b}_0 = \begin{bmatrix} -0.1390 \\ -1.1634 \end{bmatrix}, \quad \mathbf{b}_1 = \begin{bmatrix} 1.1837 \\ -0.0154 \end{bmatrix}.$$

Die Überprüfung der Stabilisierbarkeitsbedingungen aus Gl. (6.34)-(6.36) hat für $r = 1$, sowie $\mathrm{m_l} = -1$ und $\mathrm{m_u} = 0$ folgende Matrizen $\mathbf{P}_{c_{ij}}$ ergeben:

$$\mathbf{P}_{c_{-1,0}} = \begin{bmatrix} -0.3238 & -0.2778 \\ -0.2778 & -0.8295 \end{bmatrix}, \quad \mathbf{P}_{c_{-1,1}} = \begin{bmatrix} -0.0357 & -0.0178 \\ -0.0178 & -0.0711 \end{bmatrix}, \quad \text{(C.1)}$$

$$\mathbf{P}_{c_{0,0}} = \begin{bmatrix} 103.6145 & 32.0932 \\ 32.0932 & 255.2761 \end{bmatrix}, \quad \mathbf{P}_{c_{1,1}} = \begin{bmatrix} 12.6196 & -1.9560 \\ -1.9560 & 31.8894 \end{bmatrix}. \quad \text{(C.2)}$$

# Index

# Literaturverzeichnis


[1] Adamy, J. : *Strukturvariable Regelungen mittels impliziter Lyapunov-Funktionen*. Düsseldorf : VDI Verlag, 1991 (Fortschr.-Ber. VDI, Reihe 8, Nr. 271). – Diss. 7, 10, 14

[2] Adamy, J. : Implicit Lyapunov functions and isochrones of linear systems. In: *IEEE Transactions on Automatic Control* 50 (2005), Nr. 6, S. 874–879 3, 7, 10, 11, 12, 13, 29, 36, 100, 102, 103, 194, 195

[3] Adamy, J. : *Nichtlineare Systeme und Regelungen*. 2. Auflage. Springer Vieweg, 2014 31, 83

[4] Adamy, J. ; Flemming, A. : Soft variable-structure controls: a survey. In: *Automatica* 40 (2004), S. 1821–1844 7, 13

[5] Adamy, J. ; Lens, H. : Stabilitätsnachweis für weiche strukturvariable Regelungen mit Zustandsbeobachter. In: *at-Automatisierungstechnik* 55 (2007), Nr. 3, S. 107–118 100

[6] Adler, R. : *The Geometry of Random Fields*. New York : John Wiley & Sons, 1981 117, 120, 122

[7] Athans, M. ; Falb, P. : *Optimal Control. An Introduction to the Theory and Its Applications*. McGraw-Hill, 1966 155

[8] Bernstein, D. : *Matrix mathematics: theory, facts, and formulas with application to linear systems theory*. New Jersey : Princeton University Press, 2005 X, 8, 21, 22, 32, 39, 40, 42, 53, 54, 55, 92, 96, 97, 104, 108, 172, 174, 175, 178, 181, 193, 194

[9] Bernstein, D. : *Matrix Mathematics. Errata and Addenda for the Second Edition*. `http://www.engin.umich.edu/aero/people/files/matrix-math-errata`. Version: July 2014, Abruf: 23.01.2015 104

[10] Bernstein, D. ; Michel, A. : A chronological bibliography on saturating actuators. In: *International Journal of Robust and Nonlinear Control* 5 (1995), S. 375–381 1

[11] Bittel, R. ; Siljak, D. : An application of the Krylov-Bogoliubov method to linear time-varying systems. In: *International Journal of Control* 11 (1970), Nr. 3, S. 423–429 78

[12] Blake, A. ; Mumtaz, H. : *Technical Handbook - No. 4 Applied Bayesian econometrics for central bankers*. Threadneedle Street, London, EC2R 9AH: Centre for Central Banking Studies, Bank of England, September 2012 132

[13] Blanchini, F. ; Miani, S. : Constrained stabilization of continuous-time linear systems. In: *Systems & Control Letters* 28 (1996), Nr. 2, S. 95–102 1

[14] Bliman, P.-A. : A convex approach to robust stability for linear systems with uncertain scalar parameters. In: *SIAM Journal on Control and Optimization* 42 (2004), Nr. 6, S. 2016–2042 175, 194

[15] Boiko, I. : On frequency-domain criterion of finite-time convergence of second-order sliding mode algorithms. In: *Automatica* 47 (2011), S. 1969–1973 78

[16] Boyd, S. ; Ghaoui, L. ; Feron, E. ; Balakrishnan, V. : *Linear matrix inequalities in system and control theory*. 1994 (SIAM studies in applied mathematics, Vol 15) 83, 84

[17] Boyd, S. ; Vandenberghe, L. : *Convex Optimization*. New York : Cambridge University Press, 2004 18, 25

[18] Bronštejn, I. ; Grosche, G. ; Zeidler, E. : *Springer-Taschenbuch der Mathematik. Begründet von I.N. Bronštejn und K.A. Semendjaew Weitergeführt von G. Grosche, V. Ziegler und D. Ziegler. Herausgegeben von E. Zeidler*. 3. Springer Vieweg, 2013 11

[19] Buhl, M. ; Joos, P. ; Lohmann, B. : Sättigende weiche strukturvariable Regelung. In: *at-Automatisierungstechnik* 56 (2008), Nr. 6, S. 316–323 7

[20] Chin, S. ; Pengilley, C. : The application of harmonic linearization to optimal control problems. In: *International Journal of Control* 12 (1970), Nr. 6, S. 999–1007 78

[21] Currin, C. ; Mitchell, T. ; Morris, M. ; Ylvisaker, D. : A Bayesian Approach to the Design and Analysis of Computer Experiments / Oak Ridge National Laboratory. 1988. – Forschungsbericht 117

[22] Currin, C. ; Mitchell, T. ; Morris, M. ; Ylvisaker, D. : Bayesian Prediction of Deterministic Functions With Applications to the Design and Analysis of Computer Experiments. In: *Journal of the American Statistical Association* 86 (1991), Nr. 416, S. 953–963 117

[23] Dahlquist, G. ; Björck, Å. : *Numerical Methods*. Englewood Cliffs, New Jersey : Prentice-Hall, 1974 (Series in Automatic Computation) 90

[24] Desoer, C. ; Haneda, H. : The measure of a matrix as a tool to analyse computer algorithms for circuit analysis. In: *IEEE Transactions on Circuit Theory* 19 (1972), Nr. 5, S. 480–486 90, 194

[25] Domont-Yankulova, D. : *Entwurf strukturvariabler Regelungen mittels linearer Matrixungleichungen*. Düsseldorf : VDI Verlag, 2010 (Fortschr.-Ber. VDI, Reihe 8, Nr. 1175). – Diss. 7, 177

[26] Fahrmeir, L. ; Künstler, R. ; Pigeot, I. ; Tutz, G. : *Statistik. Der Weg zur Datenanalyse*. 3. Auflage. Springer-Verlag, 2001 184, 185

[27] Fang, K.-T. ; Li, R. ; Sudjianto, A. : *Design and Modeling for Computer Experiments*. Boca Raton : Taylor & Francis Group, LLC, 2006 113

[28] Favez, J.-Y. : *Enhancing the control of tokamaks via a continuous nonlinear control law*, EPFL Lausanne, Diss., 2004 61, 62, 63

[29] Föllinger, O. : *Nichtlineare Regelungen I. Grundlagen und Harmonische Balance*. München : Oldenbourg Verlag, 1969 (Methoden der Regelungstechnik) 80

[30] Föllinger, O. : *Optimale Regelung und Steuerung*. München : Oldenbourg Verlag, 1994 (Methoden der Regelungs- und Automatisierungstechnik) 83, 84, 85, 86

[31] Gelb, A. ; Vander Velde, W. : *Multiple-Input Describing Functions and Nonlinear System Design*. McGraw-Hill, 1968 78

[32] Gussner, T. : *Weiche strukturvariable Regelung mittels impliziter Ljapunov-Funktionen für Mehrgrößensystemen*, Technische Universität Darmstadt, Diplomarbeit, 2007 26

[33] Handcock, M. ; Stein, M. : A Bayesian Analysis of Kriging. In: *Technometrics* 35 (1993), Nr. 4, S. 403–410 131

[34] Haylock, R. : *Bayesian inference about outputs of computationally expensive algorithms with uncertainty on the inputs*, University of Nottingham, Diss., 1997 145

[35] Hu, T. ; Lin, Z. : *Control systems with Actuator saturation. Analysis and design.* Boston : Birkhäuser, 2001 1, 2, 10, 35, 81, 83

[36] Hu, T. ; Lin, Z. : On improving the performance with bounded continuous feedback laws. In: *IEEE Transactions on Automatic Control* 47 (2002), Nr. 9, S. 1570–1575 2, 4, 7, 26, 61

[37] Hu, T. ; Lin, Z. ; Shamash, Y. : On Maximizing the convergence rate for linear systems with input saturation. In: *IEEE Transactions on Automatic Control* 48 (2003), Nr. 7, S. 1249–1253 36, 52, 111, 195

[38] Iwasaki, T. ; Meinsma, G. ; Fu, M. : Generalized S-procedure and finite frequency KYP lemma. In: *Mathematical Problems in Engineering* 6 (2000), Nr. 2-3, S. 305–320 22, 29, 47, 68, 71, 182

[39] Iwasaki, T. ; Tsiotras, P. ; Zhang, X. : State-feedback controller synthesis for parameter-dependent LTI systems. In: *American Control Conference, 2005. Proceedings of the 2005*, 2005, S. 593–597 vol. 1 7, 10, 43, 56, 64, 176, 195

[40] Jasniewicz, B. : *Über weiche strukturvariable Regelung mittels impliziter Lyapunov-Funktionen - von der impliziten zur expliziten Regelung.* Düsseldorf : VDI Verlag, 2010 (Fortschr.-Ber. VDI, Reihe 8, Nr. 1169). – Diss. 7, 26, 29, 65

[41] Jasniewicz, B. ; Adamy, J. ; D. Domont-Yankulova: Vereinfachte schnelle Regelung von linearen Systemen mit Stellgrößenbegrenzungen. In: *at-Automatisierungstechnik* 59 (2011), Nr. 2, S. 84–93 15, 16

[42] Kiendl, H. ; Schneider, G. : Synthese nichtlinearer Regler für die Regelstrecke $const/s^2$ aufgrund ineinandergeschachtelter abgeschlossener Gebiete beschränkter Stellgröße. In: *Regelungstechnik und Prozeß-Datenverarbeitung* 20 (1972), S. 289–296 13

[43] Krylov, N. ; Bogoliubov, N. : *Introduction to nonlinear mechanics.* Princeton : Princeton University Press, 1947 78

[44] Lens, H. ; Adamy, J. : Schnelle Regelung von linearen Systemen mit Stellgrößenbeschränkungen. In: *at-Automatisierungstechnik* 57 (2009), S. 70–79 7

[45] Lens, H. ; Adamy, J. ; Domont-Yankulova, D. : A fast nonlinear control method for linear systems with input saturation. In: *Automatica* 47 (2011), S. 857–860 40

[46] Löfberg, J. : YALMIP : A Toolbox for Modeling and Optimization in MATLAB. In: *Proceedings of the CACSD Conference.* Taipei, Taiwan, 2004 190

[47] Lunze, J. : *Regelungstechnik 1.* 7. Auflage. Springer-Lehrbuch, 2008 80

[48] Murty, I. : A Unified Krylov-Bogoliubov Method for Solving Second-Order Non-Linear Systems. In: *International Journal of Non-Linear Mechanics* 6 (1971), Nr. 1, S. 45–53 78

[49] Mustafa, D. : Block Lyapunov sum with applications to integral controllability and maximal stability of singularly perturbed systems. In: *International Journal of Control* 61 (1995), Nr. 1, S. 47–63 174

[50] Oakley, J. : Eliciting Gaussian Process Priors For Complex Computer Codes. In: *Journal of the Royal Statistical Society. Series D, The Statistician* 51 (2002), Nr. 1, S. 81–97 119

[51] Oakley, J. ; O'Hagan, A. : Probabilistic sensitivity analysis of complex models: a Bayesian approach. In: *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 66 (2004), Nr. 3, S. 751–769 117, 138, 143

[52] O'Hagan, A. : Bayes-Hermite quadrature. In: *Journal of Statistical Planning and Inference* 29 (1991), S. 245–260 146

[53] Ortseifen, A. ; Adamy, J. : Eine L2-optimale Beobachtertechnik zur Vermeidung von Regler-Windup. In: *at-Automatisierungstechnik* 59 (2011), S. 114–123 84

[54] Pettersson, S. ; Lennartson, B. : Exponential Stability Analysis of Nonlinear Systems using LMIs. In: *Proceedings of the 36th Conference on Decision and Control*, 1997, S. 199–204 100

[55] Polyakov, A. ; Efimov, D. ; Perruquetti, W. : Finite-time stabilization using implicit Lyapunov function technique. In: *IFAC Nolcos 2013, Toulouse, Frankreich* (2013), S. 140–145 16

[56] Popow, P. ; Paltow, J. : *Näherungsmethoden zur Untersuchung nichtlinearer Regelungssysteme*. Leipzig : Akademische Verlagsgesellschaft, 1963 78

[57] Röthig, A. : Extension of the Krylov-Bogoliubov Method and Its Application to the Decay Rate Analysis of Nonlinear Control Algorithms. In: *12th IEEE European Control Conference (ECC), 17.-19. July, Zürich, Switzerland. Proceedings of the 12th IEEE European Control*, 2013, S. 1693 – 1698 78

[58] Röthig, A. ; Adamy, J. : Entwurf konvergenzoptimaler weich strukturvariabler Regelungen für lineare Systeme mit Stellgrößenbeschränkung. In: *at-Automatisierungstechnik* 62 (2014), Nr. 12, S. 851–864 36

[59] Röthig, A. ; Adamy, J. : Nicht-konservative weich strukturvariable Regelungen für Streckenensembles. In: *Workshop des VDI/VDE-GMA-Fachausschusses 1.40*. Anif/Salzburg, September 2015 69

[60] Röthig, A. ; Adamy, J. : Nicht-konservative weich strukturvariable Regelungen mit invers-polynomialen Selektionsstrategien. In: *at-Automatisierungstechnik* 64 (2016), Nr. 1, S. 3–18 28

[61] Röthig, A. ; Adamy, J. : On stabilizing linear systems with input saturation via soft variable structure control laws. In: *Systems & Control Letters* 89 (2016), Nr. 3, S. 47–54 16

[62] Rouche, N. ; Habets, P. ; Laloy, M. : *Stability Theory by Liapunov's direct method*. New York : Springer-Verlag, 1977 (Applied mathematical sciences, Nr. 22) 103, 173, 194

[63] Saltelli, A. ; Ratto, M. ; Andres, T. ; Campolongo, F. ; Cariboni, J. ; Gatelli, D. ; Saisana, M. ; Tarantola, S. : *Global Sensitivity Analysis. The Primer.* John Wiley & Sons, Chichester, 2008 139

[64] Santner, T. ; Williams, B. ; Notz, W. : *The Design and Analysis of Computer Experiments.* New York : Springer-Verlag, 2003 113, 114, 115, 117, 119, 120, 121, 122, 123, 125, 128, 130, 133, 135, 147, 185, 194, 195

[65] Schwarz, H. : *Einführung in die Systemtheorie nichtlinearer Regelungen.* Aachen : Shaker Verlag, 1999 (Berichte aus der Steuerungs- und Regelungstechnik) 9, 78, 81

[66] Seber, G. A. F.: *A matrix handbook for statisticians.* Hoboken, New Jersey : John Wiley & Sons, 2007 108

[67] Shamsul, A. : A modified and compact form of Krylov-Bogoliubov-Mitropolskii unified method for solving an nth order non-linear differential equation. In: *International Journal of Non-Linear Mechanics* 39 (2004), Nr. 8, S. 1343–1357 78

[68] Skelton, R. E. ; Iwasaki, T. ; Grigoriadis, K. M.: *A unified algebraic approach to linear control design.* New York : Taylor & Francis, 1997 21, 32, 38, 42, 47, 180, 181, 195

[69] Slotine, J.-J. ; Li, W. : *Applied nonlinear control.* Englewood Cliffs : Prentice Hall, 1991 82

[70] Tarbouriech, S. ; Turner, M. : Anti-windup design: an overview of some recent advances and open problems. In: *IET Control Theory and Applications* 3 (2007), Nr. 1, S. 1–19 1

[71] Toh, K. ; Todd, M. ; Tutuncu, R. : SDPT3 - A Matlab software package for semidefinite programming. In: *Optimization Methods and Software* 11 (1999), S. 545–581 190

[72] Vidyasagar, M. : On matrix measures and convex Lyapunov functions. In: *Journal of mathematical analysis and applications* 62 (1978), S. 90–103 90, 93, 94, 193, 194

[73] Vidyasagar, M. : *Nonlinear systems analysis.* 2. SIAM, 2002 (Classics in applied mathematics, Nr. 42) 90, 91, 194, 195

[74] Voigt, C. ; Adamy, J. : *Formmelsammlung der Matrizenrechung.* Oldenbourg, 2007 XI

[75] Vukić, Z. ; Kuljača, L. ; Đonlagić, D. ; Tešnjak, S. ; Munro, N. (Hrsg.): *Nonlinear control systems.* New York : Marcel Dekker, Inc., 2003 (Control Engineering) 80

[76] Zemskow, V. ; Zemskova, N. : Asymptotic method of investigation of dynamic performance of nonlinear systems with fast variation of damping coefficient and frequency. In: *Automation and Remote control* 39 (1978), Nr. 12, S. 1874–1881 78

[77] Zhang, X. : *Parameter-dependent Lyapunov functions and stability analysis of linear parameter-dependent dynamical systems*, School of Aerospace Engineering, Georgia Institute of Technology, Diss., 2003 29, 47, 50, 71, 177, 182, 183, 195

[78] Zhang, X. ; Tsiotras, P. ; Iwasaki, T. : Parameter-dependent Lyapunov function for exact stability analysis of single-parameter dependent LTI systems. In: *Decision and Control, 2003. Proceedings. 42nd IEEE Conference on* Bd. 5, 2003. – ISSN 0191–2216, S. 5168–5173 Vol.5 22, 24, 68, 179, 182, 183, 195

# Lebenslauf

| | |
|---|---|
| Name: | Andreea Violeta Röthig |
| Adresse: | Balduinstr. 83 |
| | 60599 Frankfurt |
| Geburtsdatum: | 21. Dezember 1980 |
| Geburtsort: | Bukarest, Rumänien |

| | |
|---|---|
| 01/2010-12/2015 | Wissenschaftliche Mitarbeiterin am Institut für Automatisierungstechnik und Mechatronik, Fachgebiet Regelungsmethoden und Robotik an der Technischen Universität Darmstadt |
| 08/2015-12/2015 | Data Science Spezialisierung an der Johns Hopkins University (Coursera) |
| 05/2003-12/2009 | Studium des Wirtschaftsingenieurwesens an der Technischen Universität Darmstadt, Fachrichtung Elektrotechnik |
| 09/2000-04/2003 | Studium des Wirtschaftsingenieurwesens an der Universität Politehnica Bukarest, Fakultät für Ingenieurwissenschaften in Fremdsprachen, Deutsche Abteilung (Vordiplom) |
| 09/1999-08/2000 | Vorbereitungsjahr - Deutsch als Fremdsprache an der Universität Politehnica Bukarest |
| 08/1999 | Erwerb der Hochschulreife |

# Online-Buchshop für Ingenieure

**VDI nachrichten** **BUCHSHOP**

Online-Shops

**Fachliteratur und mehr - jetzt bequem online recherchieren & bestellen unter: www.vdi-nachrichten.com/Der-Shop-im-Ueberblick**


**Täglich aktualisiert: Neuerscheinungen VDI-Schriftenreihen**


Im Buchshop von vdi-nachrichten.com finden Ingenieure und Techniker ein speziell auf sie zugeschnittenes, umfassendes Literaturangebot.

Mit der komfortablen Schnellsuche werden Sie in den VDI-Schriftenreihen und im Verzeichnis lieferbarer Bücher unter 1.000.000 Titeln garantiert fündig.

Im Buchshop stehen für Sie bereit:

**VDI-Berichte** und die Reihe **Kunststofftechnik:**

Berichte nationaler und internationaler technischer Fachtagungen der VDI-Fachgliederungen

**Fortschritt-Berichte VDI:**

Dissertationen, Habilitationen und Forschungsberichte aus sämtlichen ingenieurwissenschaftlichen Fachrichtungen

**Newsletter „Neuerscheinungen":**

Kostenfreie Infos zu aktuellen Titeln der VDI-Schriftenreihen bequem per E-Mail

**Autoren-Service:**

Umfassende Betreuung bei der Veröffentlichung Ihrer Arbeit in der Reihe Fortschritt-Berichte VDI

**Buch- und Medien-Service:**

Beschaffung aller am Markt verfügbaren Zeitschriften, Zeitungen, Fortsetzungsreihen, Handbücher, Technische Regelwerke, elektronische Medien und vieles mehr – einzeln oder im Abo und mit weltweitem Lieferservice

## Die Reihen der Fortschritt-Berichte VDI:

1 Konstruktionstechnik/Maschinenelemente
2 Fertigungstechnik
3 Verfahrenstechnik
4 Bauingenieurwesen
5 Grund- und Werkstoffe/Kunststoffe
6 Energietechnik
7 Strömungstechnik
8 Mess-, Steuerungs- und Regelungstechnik
9 Elektronik/Mikro- und Nanotechnik
10 Informatik/Kommunikation
11 Schwingungstechnik
12 Verkehrstechnik/Fahrzeugtechnik
13 Fördertechnik/Logistik
14 Landtechnik/Lebensmitteltechnik
15 Umwelttechnik
16 Technik und Wirtschaft
17 Biotechnik/Medizintechnik
18 Mechanik/Bruchmechanik
19 Wärmetechnik/Kältetechnik
20 Rechnerunterstützte Verfahren (CAD, CAM, CAE CAQ, CIM ...)
21 Elektrotechnik
22 Mensch-Maschine-Systeme
23 Technische Gebäudeausrüstung

ISBN 978-3-18-525208-2