
Eine Methodik zur stochastischen adaptiven Qualitätsregelung

Eric Lenz

Institut für Automatisierungstechnik und Mechatronik
Technische Universität Darmstadt



TECHNISCHE
UNIVERSITÄT
DARMSTADT

REGELUNGSTECHNIK *rtm*
UND MECHATRONIK

Eine Methodik zur stochastischen adaptiven Qualitätsregelung



Eine Methodik zur stochastischen adaptiven Qualitätsregelung

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

von

Dipl.-Ing. Eric Lenz

Geboren am 02.12.1980 in Rüsselsheim

Referent: Prof. Dr.-Ing. Ulrich Konigorski

Korreferent: Prof. Dr.-Ing. Edgar Dörsam

Tag der Einreichung: 13. Mai 2013

Tag der mündlichen Prüfung: 18. Juli 2013

D17
Darmstadt 2013

Impressum

Copyright: © 2013 Eric Lenz

Elektronische Version:

Bereitgestellt von tuprints, E-Publishing-Service der TU Darmstadt.

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

tuprints@ulb.tu-darmstadt.de

URL: <http://tuprints.ulb.tu-darmstadt.de/id/eprint/3609>

URN: <urn:nbn:de:tuda-tuprints-36091>

Gebundene Version:

epubli GmbH, Berlin, www.epubli.de

ISBN 978-3-8442-6706-8

Vorwort

Die vorliegende Arbeit hat ihren Ursprung in einem gemeinsamen Projekt des Fachgebietes Regelungstechnik und Mechatronik der Technischen Universität Darmstadt und der Heidelberger Druckmaschinen AG.

Mein besonderer Dank gilt Herrn Prof. Dr.-Ing. Ulrich Konigorski für sein konstruktives Interesse bei der Betreuung dieser Arbeit sowie seinen vorbildlichen Führungsstil, der maßgeblich zu dem motivierenden Arbeitsklima am Fachgebiet beigetragen hat.

Herrn Prof. Dr.-Ing. Edgar Dörsam danke ich für die Herstellung des Kontaktes zur Heidelberger Druckmaschinen AG wie auch insbesondere für die bereitwillige Übernahme des Korreferates.

Auf Seiten der Heidelberger Druckmaschinen AG möchte ich besonders Herrn Nikolaus Pfeiffer für die sehr gute Zusammenarbeit und auch für die vielen kritischen Nachfragen, die in diese Arbeit eingeflossen sind, meinen Dank aussprechen.

Meinen Kollegen des Fachgebietes Regelungstechnik und Mechatronik sowie der Forschungsgruppe Regelungstechnik und Prozessautomatisierung, insbesondere denen der AG tpm, danke ich für viele fachliche Diskussionen und die sehr gute Arbeitsatmosphäre. Keinesfalls vergessen werden dürfen dabei auch die Mitarbeiterinnen des Sekretariats sowie die Mitarbeiter der Elektro- und Metallwerkstatt des Instituts, die durch ihre engagierte Arbeit meine eigene Tätigkeit erheblich vereinfacht haben.

Ganz besonders danke ich meinen Eltern, die durch ihre stete Unterstützung und Förderung die Basis für diese Arbeit gelegt haben.

Eric Lenz

Darmstadt, im September 2013



Kurzfassung

Ausgangspunkt dieser Arbeit ist das Bestreben, Einrichtvorgänge bei Druckmaschinen mit minimaler Makulatur, d. h. minimalen Ausschuss durchzuführen. Als Qualitätsmerkmal wird dabei die Färbung betrachtet. Dies bedeutet, dass eine gegebene Sollfärbung mit einer möglichst geringen Anzahl an Regelschritten erreicht werden soll. Aufgrund der in Anbetracht des unvermeidbaren Prozessrauschens relativ hohen Anforderungen an die Regelgüte ist dabei eine stochastische Betrachtung der Regelung notwendig.

Diese hier kurz skizzierte Problemstellung wird als allgemeineres zeitdiskretes Qualitätsregelungsproblem statischer Prozesse beschrieben und in zwei Schritten bearbeitet. Zunächst erfolgt die Definition und Analyse von Größen zur Bewertung der Einrichtvorgänge. Aus diesen ergibt sich, dass der mittlere quadratische Fehler als globales Gütekriterium für den Reglerentwurf geeignet ist. Darauf basierend werden in dieser Arbeit verschiedene Reglerkonfigurationen untersucht und bewertet.

Anschließend wird die Identifikation der Prozessparameter während der Regelvorgänge betrachtet, womit eine Adaption der Regler möglich wird. Basis des Identifikationsverfahrens ist die bekannte Methode der kleinsten Quadrate. Aufgrund der Eigenschaften des Anwendungsfalles ist die Leistung des Standardverfahrens jedoch nicht zufriedenstellend. Hervorzuheben ist dabei die Multikollinearität, d. h. die starke Korrelation der verschiedenen Prozesseingänge, sowie die Möglichkeit sprungförmiger Störungen zu Beginn neuer Aufträge. Diese Probleme können in dieser Arbeit gelöst werden, so dass ein leistungsfähiger Schätzalgorithmus angegeben werden kann, der die wenigen Messdaten sehr effektiv ausnutzt.

Darüber hinaus ist auch der Bias, der aufgrund der Regelung im geschlossenen Regelkreis auftritt, zu beachten. Dieser wird untersucht und eine kennfeldbasierte Kompensationsmethode vorgeschlagen.

Um auch bei großen Anfangsabweichungen der angenommenen Prozessparameter sowie Ausreißern bei den Messungen noch ein gutes Regel- und Adaptionsverhalten zu erreichen, wird das vorgeschlagene Schätzverfahren zuletzt noch um geeignete Maßnahmen erweitert.

Abstract

This work is motivated by the demand to perform the makeready procedure of printing machines with a minimum of paper waste. The considered quality criterion is the coloration. This means that a given coloration should be attained by as few as possible control steps. Because of – considering the inevitable process noise – high demands on the controller performance, a stochastic controller design is mandatory.

This shortly outlined problem is described as a general discrete-time quality control problem of static processes and is solved in two steps. First, quantities for assessing the performance of makeready procedures are defined and analysed. These lead to the mean squared error as a global optimization criterion. Based on this, different controller schemes are analysed and evaluated.

Second, the closed-loop identification of the process parameters is considered to allow adaptive control. This identification is based on the well-known least squares method. Because of the demanding requirements and certain properties of the considered application, a direct application of this method does not result in a satisfying performance. Particular challenges are the multicollinearity, i. e. the strong correlation of different process inputs, and the possible occurrence of stepwise disturbances at the beginning of new jobs. These problems are solved within this work, resulting in a capable estimation method which uses the few available measurement data highly effectively.

Furthermore, the bias caused by the closed-loop identification has to be considered. This bias is examined and a compensation method based on a look-up table is proposed.

To achieve a good performance even under large initial deviations of the assumed process parameters and measurement outliers the proposed method is extended by suitable procedures.

Inhaltsverzeichnis

Abkürzungen und Formelzeichen	XIII
1 Einführung	1
1.1 Motivation und Thema	1
1.2 Problemstellung	1
1.3 Vorgehen und Gliederung der Arbeit	2
2 Festlegung des Untersuchungsgegenstandes	4
2.1 Verwendete Begriffe	4
2.2 Problemdefinition	4
2.2.1 Prozessmodell	4
2.2.2 Prozessunsicherheiten und Rauschen	7
2.2.3 Weitere Randbedingungen	9
2.2.4 Ablauf eines Einrichtungsvorgangs	10
2.2.5 Regelziel	11
2.3 Anwendungsbeispiel	11
2.4 Qualitätsregelung	15
3 Regelung des ungestörten Systems	17
3.1 Regelziele	17
3.2 Streckenmodell	17
3.3 Voreinstellung	19
3.4 Regelung	19
3.5 Kompensation der sekundären Eingänge	24
3.6 Zusammenfassung	25
4 Modellierung der Störungen	27
4.1 Stochastik	27
4.1.1 Zufallsvariablen	27
4.1.2 Zufallsprozesse	32
4.2 Beschreibung der Störungen	33
4.3 Normierte Darstellung	34
4.4 Verwendete Gleichungen	36
4.5 Diskussion des Ausgangspunktes für die Analyse	37
4.5.1 Bedingte Verteilung	38
4.5.2 Beispiel Bias	40
4.6 Zusammenfassung	42
5 Regelung des gestörten Systems	43
5.1 Regelziele	43
5.1.1 Erreichen des OK-Bereichs	43
5.1.2 Verbleiben im OK-Bereich	45
5.2 Klassische Ansätze zur Regelung stochastischer Systeme	49
5.2.1 Verfahren	50

5.2.2	Anwendung	53
5.2.3	Fazit	55
5.3	Entwurf anhand der Regelziele bezüglich des Erreichens des OK-Bereichs	56
5.4	Entwurf durch Minimierung des Mean-Squared-Errors	61
5.4.1	I-Regler mit konstanter Verstärkung k_I – Analyse der Verteilungseigenschaften	61
5.4.2	Wahl des MSE als Gütekriterium	63
5.4.3	I-Regler mit konstanter Verstärkung k_I – Bestimmung der optimalen Verstärkung	64
5.4.4	I-Regler mit zeitvariabler Verstärkung $k_{I,k}$	67
5.4.5	I-Regler mit Filter	70
5.4.6	I-Regler mit abweichungsabhängiger Verstärkung $k_I(\tilde{\varepsilon})$	78
5.4.7	Fazit	83
5.5	Verbleiben im OK-Bereich	84
5.5.1	Testfall	84
5.5.2	Simulationen	85
5.5.3	Fazit	88
5.6	Zusammenfassung	88
6	Online-Identifikation und Adaption	90
6.1	Einführung	90
6.1.1	Identifikationsaufgabe	91
6.1.2	Benennung der Größen	91
6.2	Literatur	92
6.3	Schätzung über die Methode der kleinsten Quadrate	93
6.3.1	Schätzgleichung	93
6.3.2	Eigenschaften	94
6.3.3	Varianz	97
6.3.4	Bewertung des Schätzers für das vorliegende Schätzproblem	97
6.4	Identifikation während eines Auftrags	98
6.4.1	Überprüfen der Varianz	99
6.4.2	Implementierung	101
6.4.3	Einflussfaktoren auf die Varianz	105
6.4.4	Least-Squares-Schätzung bei teilweise festgehaltenen Parametern	108
6.4.5	Schätzen der sekundären Parameter bei festem primären Parameter	112
6.4.6	Schätzen der Varianz des Prozess- und Messrauschens	116
6.4.7	Weitere Schätzalgorithmen	117
6.5	Identifikation über mehrere Aufträge	118
6.5.1	Fortführen eines Schätzalgorithmus über mehrere Aufträge	118
6.5.2	Mittelung der für jeden Auftrag einzeln geschätzten Parameter	123
6.5.3	Simulation	135
6.5.4	Auswahl eines Verfahrens	144
6.6	Untersuchung des Bias	145
6.6.1	Prinzipieller Einfluss des geschlossenen Regelkreises	146
6.6.2	Bias für $k = 2$	147
6.6.3	Bias für $k > 2$	150
6.6.4	Berücksichtigung der Varianz-Bedingung	153
6.6.5	Entwicklung des Bias über mehrere Aufträge	154
6.6.6	Kompensation	161
6.7	Erweiterungen des Schätzverfahrens	163
6.7.1	Test des Schätzwertes $\hat{\theta}_{1,r}$ auf Plausibilität	165
6.7.2	Berücksichtigen der geschätzten Varianz des Messrauschens/Ausreißer	170

6.7.3	Diskussion der Erweiterungen	174
6.8	Zusammenfassung des Algorithmus	175
6.9	Adaptive Regelung	178
6.10	Messungen	178
6.11	Zusammenfassung	179
7	Zusammenfassung	182
A	Stochastik	184
A.1	Verteilungen	184
A.2	Rechnungen zur bedingten Verteilung	187
A.3	Optimale gewichtete Mittelung von Zufallszahlen	189
B	Das Least-Squares-Verfahren	190
B.1	Einführung	190
B.1.1	Schätzung der Varianz des Messrauschens	193
B.2	Erweiterungen	194
B.2.1	Weighted Least Squares	194
B.2.2	Exponentielles Vergessen	195
B.3	Iterative Berechnung	195
B.4	Rekursives Least-Squares-Verfahren	196
B.5	Varianz bei teilweise festgehaltenen Parametern	196
B.6	Weiterführen einer Schätzung bei bekannten Parameteränderungen	198
B.7	Extended-Kalmanfilter	198
C	Herleitungen	200
C.1	Beweis zur Mittelung sekundärer Parameter	200
C.2	Empirische Varianz und Kovarianz	201
C.3	Varianz bei Drei-Parameter-Schätzung	202
C.4	Äquivalenz von Gl. (6.35) und (6.36)	203
C.5	Bias im Fall $k > 2$	204
D	Schätzung von q_i	207
D.1	Möglichkeiten zur Schätzung von q_i	207
D.2	Quotienten von Zufallsvariablen	208
D.3	Vergleich der Schätzer für q_2	210
D.3.1	Bias von \hat{q}_2 bei der Zwei-Parameter-Schätzung	213
D.3.2	Fazit	216
	Literaturverzeichnis	217



Abkürzungen und Formelzeichen

Abkürzungen

LS	Least-Squares
MSE	Mean-Squared-Error
OK	Fertigteil erfüllt Anforderungen
$\overline{\text{OK}}$	Nicht OK, Fertigteil erfüllt Anforderungen nicht

Allgemeine Schreibweisen

*	Platzhalter für einen nicht weiter interessierenden Wert	
$\sum x_{k'}$	$= \sum_{k'=1}^k x_{k'}$	
$\sum x_{l'}$	$= \sum_{l'=1}^l x_{l'}$	
$\leq^!$	Darstellung der Bedingung \leq im Fließtext, andere Bedingungen entsprechend	
$(\mathbf{A})_{ij}$	Element der i -ten Zeile und j -ten Spalte der Matrix \mathbf{A}	
$\mathbf{A}_{[ab],[cd]}$	Matrix bestehend aus den Elementen $(\mathbf{A})_{ac}$, $(\mathbf{A})_{ad}$, $(\mathbf{A})_{bc}$ und $(\mathbf{A})_{bd}$ der Matrix \mathbf{A} . (Bei Verwechslungsgefahr wird auch hierbei eine Klammer verwendet: $(\mathbf{A})_{[ab],[cd]}$)	
$\{x\}_r$	Zufallsvariable	4.1.1
$\{x_k\}_r$	Zufallsprozess	4.1.2
$\{x\}_r^\omega$	Realisierung einer Zufallsvariablen	4.1.1
$\{x y_0\}_r$	Bedingte Zufallsvariable	4.1.1, S. 31

Allgemeine Akzente und Indizes

$(\tilde{\cdot})$	gemessene Größe
$(\hat{\cdot})$	geschätzte Größe
$(\cdot)_m$	Mittelwert
$(\cdot)_r$	Zur Regelung verwendeter Wert
$(\cdot)_+$	Mittelung aus Mittelwert $(\cdot)_m$ zurückliegender Aufträge und dem aktuellen Wert aus laufendem Auftrag
$(\cdot)_{i,\{\emptyset,m,r,+ \},\{k,l\}}$	Aufbau allgemeine Indizierung: Eingang bzw. Parameter, Art, Unabhängige (Messung oder Auftrag)

Skalare Größen

d_k	Normierte Störung	4.3
$d_{y,k}$	Störung	2.2.1
e_k	Regelabweichung bezüglich tatsächlicher Ausgangsgröße	3.1

\tilde{e}_k	Regelabweichung bezüglich gemessener Ausgangsgröße	3.1
\hat{e}_k	Schätzfehler, $\hat{e}_k = \tilde{y}_k - \hat{y}_k$	6.3.1
e_{OK}	Betrag der maximal zulässigen Regelabweichung für OK	3.1
$f_{\sigma_y, \max}$	Relative Grenze für geschätzte Standardabweichung $\hat{\sigma}_y < f_{\sigma_y, \max} \cdot \sigma_y$	6.7.2
g	Allgemein Gewicht für Mittelung der Form $x = (x_1 g + x_2)/(g + 1)$	6.4.5, S. 113
$g_{i,m,l}$	Gewicht des gemittelten Schätzwertes der zurückliegenden Aufträge	6.5.2, S. 127
i	Indexvariable für Prozesseingang	
k, k', k''	Unabhängige („Zeitvariable“) bei Betrachtung von Anläufen	
k_I	Reglerverstärkung	3.4
k_{Job}	Anzahl der Messungen je Auftrag in Simulation	6.6.5, S. 156
k_{OK}	Regelschritt zu dem Produktionsphase beginnt	5.1.2, S. 47
k_{OK}^*	Regelschritt zu dem Regler-OK erreicht ist	5.1.2, S. 47
l, l', l''	Unabhängige („Zeitvariable“) bei Betrachtung von Aufträgen	
M	(maximale) Filterordnung	5.4.5, S. 72
m_k	Filterordnung im k -ten Schritt	5.4.5, S. 73
n_k	Normiertes Mess- und Prozessrauschen	4.3
$n_{y,k}$	Mess- und Prozessrauschen, $n_{y,k} = n_{p,k} + n_{M,k}$	2.2.2
$n_{A,k}$	Anlaufschwankungen (Prozessrauschen während Einrichtphase)	2.2.2
$n_{M,k}$	Messrauschen	2.2.2
$n_{p,k}$	Prozessrauschen	2.2.2
$n_{RS, \max}$	Maximal zulässige Anzahl an Regelschritten bis OK erreicht ist	3.1
$n_{S,k}$	Produktionsschwankungen (Prozessrauschen während Produktionsphase)	2.2.2
\bar{n}_{RS}	Mittlere Anzahl an Regelschritten bis zum Erreichen von OK	5.1.2
p	Anzahl Eingangsgrößen	2.2.1
$p_{i,m,l}$	Varianz des aus den ersten l Aufträgen gemittelten Schätzwertes	6.5.2, S. 132
$p_{i, \min}$	Untergrenze für zu berücksichtigende Varianz bei kombinierter Mittelung	6.5.2, S. 133
$p_{H_0, \min}$	Grenze zum Ablehnen der Hypothese H_0 beim Plausibilitätstest	6.7.1
p_{\min}	Minimale Erfüllungsquote der Regelanforderungen	5.1.1
$p_{\overline{OK}}$	Wahrscheinlichkeit dafür, dass ein vermessenes Fertigteil \overline{OK} ist	5.1.2
$p_{\overline{OK}, \max}$	Maximal zulässige Ausschussquote $p_{\overline{OK}}$	5.1.2, S. 47
$p_{\overline{OK}, > x\%}$	Anteil der Versuche, bei denen nach dem letzten Regelschritt die Wahrscheinlichkeit für \overline{OK} größer als $x\%$ ist.	5.1.2
q_i	Verhältnis $\theta_i/\theta_1, i = 2, \dots, p$	6.1.1
$q_{\overline{OK}, x\%}$	Die Wahrscheinlichkeit für \overline{OK} , die von $x\%$ der Versuche erreicht oder unterschritten wurde.	5.1.2
$q_{RS, x\%}$	Anzahl Regelschritte, nach der mehr als $x\%$ der Anfahrvorgänge abgeschlossen sind	5.1.2

r_y	relative Veränderung der Regelabweichung	3.1
$r_{OS,max}$	Maximales Überschwingen, als relative Veränderung ausgedrückt	3.1
$r_{US,max}$	Minimale relative Verbesserung, als relative Veränderung ausgedrückt	3.1
$u_{1,v}$	Stellgröße aus Voreinstellung	3.3
$u_{i,k}$	Eingangsgröße (Eingangsgröße, die letztlich auf den Prozess wirkt)	2.2.1
$u_{i,r,k}$	Stellgröße (Eingangsgröße, die Regler vorgibt)	3.2
$u_{i,N}$	Nennwert (typischer Wert) für Eingangsgröße	2.2.1, 3.2
$u_{i,min}, u_{i,max}$	Eingangsgrößenbeschränkung	2.2.1, S. 6
v_k	Abkürzung für $u_{1,k} - u_{1,N}$	2.2.1
w_k	Abkürzung für $u_{2,k} - u_{2,N}$	C.3
Y_k	Summe der quadrierten gemessenen Ausgangsgröße, $Y_k = \sum \tilde{y}_{k'}^2$	6.4.6
y_k	Ausgangsgröße ohne Prozessrauschen	2.2.2
y'_k	Ausgangsgröße mit Prozess- aber ohne Messrauschen	2.2.2
\tilde{y}_k	gemessene Ausgangsgröße (Ausgangsgröße mit Prozess- und Messrauschen)	2.2.2
$\tilde{y}_{f,k}$	Modifizierte Messgröße bei teilweiser Schätzung	6.4.4
y_{soll}	Sollgröße	3.1
z	Unabhängige im Bildbereich der z -Transformation	3.4
z_k	Eingangsgröße eines Filters	5.4.5
α_k	Auf ideale Reglerverstärkung bezogene Reglerverstärkung	4.3
$\tilde{\alpha}_k$	Anhand $\hat{\theta}_{1,r}$ gewählte bezogene Reglerverstärkung	5.3, S. 60
$\alpha_{min}, \alpha_{max}$	Grenzen der normierten Reglerverstärkung	5.3, S. 58
γ	Allgemein Gewicht für Mittelung der Form $x = (1 - \gamma) \cdot x_1 + \gamma x_2$	6.4.5, S. 115
Δk_{OK}	Differenz $k_{OK} - k_{OK}^*$	5.1.2, S. 47
$\Delta \hat{\theta}_{i,r}$	Fehler der zur Regelung verwendeten Schätzwerte, $\Delta \hat{\theta}_i = \hat{\theta}_{i,r} - \theta_i$	4.2
$\Delta \hat{\vartheta}_{i,k}$	Schätzfehler, $\Delta \hat{\vartheta}_{i,k} = \hat{\vartheta}_{i,k} - \vartheta_i$	4.5.2
$\Delta \hat{\vartheta}_{i,mean,k}$	Bias des Schätzers für den normierten i -ten Parameter nach der k -ten Messung	4.5.2
$\varepsilon_{1,min}$	Minimale anfängliche Regelabweichung für Schätzung	6.6
ε_k	Normierte Regelabweichung	4.3
$ \varepsilon_k _{max}$	Maximale zulässiger Betrag der normierten Regelabweichung, um Toleranz ε_{OK} mit gegebener Quote $p_{\overline{OK,max}}$ einhalten zu können	5.1.2, S. 48
ε_{OK}	Maximal zulässige normierte Regelabweichung für OK-Fertigteil	4.3
$\varepsilon_{max,k}$	Maximal zulässige normierte Regelabweichung nach dem k -ten Schritt	5.1.1, S. 45
ε_{∞}	$\rightarrow \text{Var}\{\varepsilon_{\infty}\}_r$	
η_k	Normierte Ausgangsgröße	4.3

θ_0	Prozessparameter, Wert der Ausgangsgröße am Nennpunkt ($u_i = u_{i,N}, i = 1, \dots, p$)	2.2.1
θ_{00}	Prozessparameter, Wert der Ausgangsgröße bei „Nulleingang“ ($u_i = 0, i = 1, \dots, p$)	2.2.1
θ_{0S}	Prozessparameter, Wert der Ausgangsgröße bei $u_1 = u_{1,N}$ und aktuellen sekundären Eingangsgrößen	2.2.1
θ_i	Prozessparameter zur i -ten Eingangsgröße, $i = 1, \dots, p$	2.2.1
$\theta_{i,\min}, \theta_{i,\max}$	Grenzen für Prozessparameter	2.2.3
$\hat{\theta}_{i,k}$	Parameterschätzwert nach k -ter Messung aus aktuellem Auftrag	6.4.1
$\hat{\theta}_{i,l}$	Parameterschätzwert aus l -tem Auftrag	6.5.2, S. 124
$\hat{\theta}_{i,+}$	Mittelung aus Schätzwert zurückliegender Aufträge und dem aktuellen Schätzwert aus laufendem Auftrag	6.5.2, S. 124
$\hat{\theta}_{i,m,l}$	Aus mehreren Aufträgen gemittelter Parameterschätzwert	6.5.2, S. 124
$\hat{\theta}_{i,r,k}$	Zur Regelung verwendeter Parameterschätzwert	6.4.1
ϑ_i	Normierter Streckenparameter, $\vartheta_i = \theta_i / \sigma_y$	4.3
κ_I	Reglerverstärkung für normierte Darstellung	4.3
κ_θ	Relativer Schätzfehler von θ_1 , $\hat{\theta}_1 = \theta_1 \cdot \kappa_\theta$	5.3, S. 60
λ_i	Vergessensfaktor für Mittelung der Schätzwerte des i -ten Parameters	6.5.2, S. 127
μ_x	Abkürzende Schreibweise für $E\{x\}_r$	4.1.1, S. 29
ρ_{xy}	Korrelationskoeffizient	4.1.1, S. 30
$\sigma_{\hat{\theta}_{i,r},\max}^2$	Maximale Varianz, mit der ein Schätzwert für θ_i zur Regelung verwendet werden darf	6.4.1
$\sigma_{\hat{\theta}_{i,m},\max}^2$	Maximale Varianz, mit der ein Schätzwert für θ_i zur Mittelung verwendet werden darf	6.5.2, S. 126
$\sigma_{\hat{\theta}_{1,m},\max}^2 \Big _{\max,l}$		6.5.2, S. 130
σ_S^2	Varianz von n_S (Produktionsschwankungen)	5.5.1
σ_x^2	Abkürzende Schreibweise für $\text{Var}\{x\}_r$	4.1.1, S. 29
σ_y^2	Varianz des Mess- und Prozessrauschen	4.2, S. 34
$\hat{\sigma}_{y,k}^2$	Schätzwert für Varianz des Mess- und Prozessrauschens	6.4.6, S. 116
ω	Elementarereignis	4.1.1

Vektor- und matrixwertige Größen

A_k	Zustandsmatrix einer Zustandsraumdarstellung (LZV, diskret)	5.2.1
B_k	Eingangsmatrix einer Zustandsraumdarstellung (LZV, diskret)	5.2.1
C_k	Ausgangsmatrix einer Zustandsraumdarstellung (LZV, diskret)	5.2.1
$C_{e f}$	Korrekturmatrix für Schätzwerte $\hat{\theta}_e$ aus teilweiser Schätzung zur Anpassung bei Änderungen von $\hat{\theta}_f$	6.4.5
$c_{i f}^T$	Korrekturvektor für Schätzwerte des i -ten Parameters aus teilweiser Schätzung zur Anpassung bei Änderungen von $\hat{\theta}_f$	6.5.2, S. 135
I	Einheitsmatrix	

\mathbf{K}_k	Rückführmatrix einer statischen Zustandsregelung	5.2.1
\mathbf{L}_k	Informationsmatrix $\Psi_k^T \Psi_k$	6.3.1
$\mathbf{L}_{e,k}$	Informationsmatrix bei teilweiser Schätzung	6.4.4
\mathbf{P}_k	Kovarianzmatrix der Parameterschätzwerte	6.3.2, S. 95
\mathbf{Q}_k	Gewichtsmatrix der Zustände für Optimalreglerentwurf	5.2.1
\mathbf{R}_k	Gewichtsmatrix der Eingangsgrößen für Optimalreglerentwurf	5.2.1
\mathbf{r}_k	Rechte Seite der Normalgleichung, $\mathbf{r}_k = \Psi_k^T \tilde{\mathbf{y}}_k$	6.3.1
$\mathbf{r}_{e f,k}$	Rechte Seite der Normalgleichung bei teilweiser Schätzung	6.4.4
\mathbf{u}_k	Eingangsvektor einer Zustandsraumdarstellung	5.2.1
$\mathbf{u}_{S,k}$	Vektor der sekundären Eingangsgrößen $[u_{2,k} \cdots u_{p,k}]^T$	2.2.1
$\mathbf{u}_{S,\text{soll}}$	Sollwerte bzw. vorgegebene Werte für sekundäre Eingangsgrößen	3.2
\mathbf{v}_k	Vektor der auf die Systemzustände wirkenden Störungen	5.2.1
\mathbf{w}_k	Vektor der auf die Systemausgänge wirkenden Störungen	5.2.1
\mathbf{x}_k	Zustandsvektor einer Zustandsraumdarstellung	5.2.1
$\boldsymbol{\theta}$	Vektor der Parameter $[\theta_0 \cdots \theta_p]^T$	2.2.1
$\boldsymbol{\theta}_e$	Vektor der Parameter, die bei teilweiser Schätzung geschätzt werden	6.4.4
$\boldsymbol{\theta}_f$	Vektor der Parameter, die bei teilweiser Schätzung festgehalten werden	6.4.4
$\hat{\boldsymbol{\theta}}_{e f}$	Vektor der bei teilweiser Schätzung geschätzten Parameter	6.4.4
$\boldsymbol{\theta}_S$	Vektor der Parameter zu den sekundären Eingängen $[\theta_2 \cdots \theta_p]^T$	2.2.1
$\boldsymbol{\psi}_k^T$	Eingangsvektor zur Parameterschätzung	6.3.1
Ψ_k	Eingangsmatrix zur Parameterschätzung	6.3.1
Funktionen		
$\text{Cov}(\{x\}_r, \{y\}_r)$	Kovarianz der Zufallsvariablen $\{x\}_r$ und $\{y\}_r$	4.1.1, S. 29
C_{xy}	Kovarianzmatrix der vektoriellen Zufallsgrößen $\{\mathbf{x}\}_r$ und $\{\mathbf{y}\}_r$	4.1.1, S. 31
$C_{xy}(k, l)$	Kovarianzmatrix der vektoriellen Zufallsprozesse $\{\mathbf{x}_k\}_r$ und $\{\mathbf{y}_l\}_r$	4.1.2
$C_{xy}(k)$	Abkürzung für $C_{xy}(k, k)$	4.1.2
$E\{x\}_r$	Erwartungswert der Zufallsvariablen $\{x\}_r$	4.1.1, S. 29
$\bar{E}\{x\}_r$	Erwartungswert, der unter Berücksichtigung bestimmter Bedingungen bestimmt wird	4.5.2
$F_{\{x\}_r}(a)$	Verteilungsfunktion der Zufallsvariablen $\{x\}_r$	4.1.1, S. 28
$f_{\{x\}_r}(a)$	Verteilungsdichtefunktion der Zufallsvariablen $\{x\}_r$	4.1.1, S. 28
$\text{MSE}\{e\}_r$	Mean-Squared-Error, $\{e\}_r$ stellt dabei den betrachteten Fehler dar, $\text{MSE}\{e\}_r = (E\{e\}_r)^2 + \text{Var}\{e\}_r$	5.4.2
$P(\{x\}_r^\omega \leq a)$	Wahrscheinlichkeit dafür, dass $\{x\}_r^\omega$ einen Wert kleiner oder gleich a annimmt	4.1.1
$\text{sat}(v, l, u)$	Sättigungsfunktion, $\text{sat}(v, l, u) = \min\{\max\{v, l\}, u\}$	
$\text{Var}\{x\}_r$	Varianz der Zufallsvariablen $\{x\}_r$	4.1.1
$\text{Var}\{\varepsilon_\infty\}_r$	$\lim_{k \rightarrow \infty} \text{Var}\{\varepsilon_k\}_r$	

$\alpha_a(\tilde{\varepsilon}_k)$	Kennlinie „a“ der regelabweichungsabhängigen, normierten Reglerverstärkung	5.4.6
$\alpha_a^*(\tilde{\varepsilon}_k)$	Kennlinie „a“ der regelabweichungsabhängigen, normierten Reglerverstärkung mit zusätzlicher Bedingung	5.4.6
$\Phi(a; \mu, \sigma^2)$	Verteilungsfunktion der Normalverteilung $N(\mu, \sigma^2)$	4.1.1, S. 30
$\Phi(a)$	Verteilungsfunktion der Standardnormalverteilung $N(0, 1)$	4.1.1, S. 30
$\varphi(a; \mu, \sigma^2)$	Verteilungsdichtefunktion der Normalverteilung $N(\mu, \sigma^2)$	4.1.1, S. 30
$\varphi(a)$	Verteilungsdichtefunktion der Standardnormalverteilung $N(0, 1)$	4.1.1, S. 30
Verteilungen		
$G(\mu, a)$	Gleichverteilung auf Intervall $[\mu - a, \mu + a]$	4.1.1, S. 31
$N(\mu, \sigma^2)$	Normalverteilung mit Erwartungswert μ und Varianz σ^2	4.1.1, S. 30

1 Einführung

In diesem Kapitel wird zunächst in Abschnitt 1.1 die Motivation und das Thema der Arbeit vorgestellt. Anschließend wird in Abschnitt 1.2 die abstrahierte Problemstellung beschrieben, auf deren Basis im folgenden Kapitel der Untersuchungsgegenstand definiert wird.

Das Kapitel schließt mit einer Übersicht über den Aufbau dieser Arbeit.

1.1 Motivation und Thema

Die in dieser Arbeit behandelte Thematik ergab sich aus einem Projekt mit der Heidelberger Druckmaschinen AG, das die Einrichtvorgänge von Druckmaschinen mit Anilox-Farbwerken zum Inhalt hatte. Ziel war es dabei, die Anzahl der notwendigen Anläufe und damit die Makulatur möglichst gering zu halten, wozu ein entsprechender Algorithmus für die Steuerung zu entwickeln war, der die Nachführungen optimal einstellt.

Dabei wurden aus dem Stillstand der Maschine immer einige Bogen gedruckt, bis sich ein stationärer Zustand eingestellt hatte. Einer dieser „stabilen“ Bogen wurde vermessen, und anhand der festgestellten Farbdichte dieses Bogens sowie der vermessenen Bogen aus den vorangegangenen Anläufen sollte die Stellgröße, in diesem Fall die Temperatur der Rasterwalze des Farbwerks, möglichst optimal gewählt werden.

Da zu einer guten Regelung die Parameter, d. h. hier der Einfluss der Temperatur der Rasterwalze sowie der Druckgeschwindigkeit auf die Farbdichte, möglichst gut bekannt sein sollten, wurden auch Verfahren zur Identifikation dieser Parameter entwickelt.

Diese Thematik wird in dieser Arbeit losgelöst von dem speziellen Anwendungsfall behandelt und ist im folgenden Abschnitt abstrahiert dargestellt. Wie später noch genauer dargelegt wird, ergibt sich als Thema die Regelung und Identifikation eines statischen Prozesses der affin in den Eingangsgrößen ist. Dies ist zunächst ein Standardproblem, das in der Literatur zur Regelungstechnik und Identifikation auf der einen, und in der zur Regressionsrechnung auf der anderen Seite behandelt wird. Die Besonderheiten im vorliegenden Fall, auf die diese Arbeit explizit eingeht, sind die äußerst geringe Anzahl an Messungen sowie relativ starke Störungen und Rauschen. Wie diesen begegnet werden kann, bzw. inwieweit neue Lösungen erforderlich sind, wird an den entsprechenden Stellen begründet.

Zur Veranschaulichung wird die Anilox-Druckmaschine bzw. das Anilox-Farbwerk im Weiteren als Beispiel herangezogen. Auch wurden an dieser die Algorithmen getestet und deren Tauglichkeit bestätigt.

1.2 Problemstellung

In dieser Arbeit werden Einrichtvorgänge von Produktionsmaschinen betrachtet. Dabei ist mit „Einrichten“ der Wiederstart der Produktion nach einer Pause oder nach Umrüsten auf andere zu produzierende Fertigteile gemeint, nicht aber das Inbetriebnehmen einer Produktionsmaschine nach Aufbau oder Wartung.

Ziel ist es, beim Einrichten der Produktion möglichst schnell eine Mindestqualität der Fertigteile zu erreichen, die sich durch bestimmte Eigenschaften auszeichnet. Wie in Abbildung 1.1 angedeutet, werden die Eigenschaften der Fertigteile durch die eingesetzten Roh- und Hilfsstoffe, die Prozesssteuerparameter

und Umgebungsbedingungen beeinflusst.¹ Darüber hinaus führt auch eine nicht ideale Prozessbeherrschung bei komplexeren Prozessen zu Abweichungen der Eigenschaften.

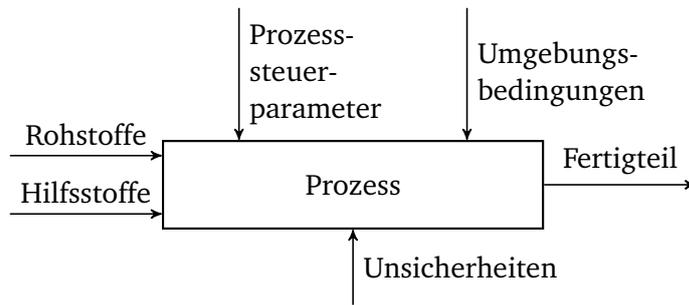


Abbildung 1.1: Allgemeiner Prozess

Die Prozesssteuerparameter können an der Maschine eingestellt werden. Prozesssteuerparameter² stellen beispielsweise Bearbeitungsgeschwindigkeiten und -drücke oder Temperaturen von Werkzeug oder Werkstück dar.

Wesentlich für die Anwendung einer automatisierten Regelung ist es, dass die erreichte Qualität eines Fertigteils möglichst objektiv, idealerweise automatisiert, quantifiziert werden kann. (Vgl. auch [KIPPHAN, 2000, S. 103].)

Um das betrachtete Problem sinnvoll einzugrenzen, wird davon ausgegangen, dass die Nachführung der Prozesssteuerparameter typischerweise im Stillstand geschieht. Ein Schritt sieht damit so aus, dass die (neuen) Prozesssteuerparameter gewählt werden, diese eingestellt werden und dann der Produktionsprozess gestartet wird. Nach Erreichen eines stabilen Zustands wird ein Fertigteil entnommen und dessen Eigenschaften überprüft. Damit wird der Prozess als statisch betrachtet.

Eine weitere Eigenschaft der hier behandelten Prozesse ist es, dass die Qualität der Fertigteile durch eine einzige Eigenschaft beschrieben wird. Aus regelungstechnischer Sicht besitzen die Prozesse damit nur einen Ausgang. Sinnvollerweise ist dann einer der Prozesssteuerparameter der primäre Parameter bzw. Eingang zur Regelung, während die übrigen Prozesssteuerparameter als sekundäre Eingänge bezeichnet werden. Der primäre Eingang wird dabei in jedem Schritt angepasst werden, während die sekundären Eingänge in den meisten Schritten konstant gehalten werden, und nur dann angepasst werden, wenn bspw. der Stellbereich des primären Eingangs nicht mehr ausreicht, oder die Bearbeitungsgeschwindigkeit oder Taktrate nach dem Einrichten erhöht werden soll.

1.3 Vorgehen und Gliederung der Arbeit

Zur Lösung der beschriebenen Problemstellung wird in dieser Arbeit eine Methodik vorgestellt, die die Auslegung eines stochastischen adaptiven Reglers beschreibt.

Da ein relativ hohes Rauschen, bezogen auf den zulässigen Toleranzbereich, zugelassen werden soll, ist eine explizite Berücksichtigung des Rauschens im Reglerentwurf erforderlich. Somit ist eine stochastische Regelung zu entwerfen.

Die Regelung soll auch bei unbekanntem oder sich verändernden Prozessparametern ein gutes Verhalten zeigen. Daher wird die Regelung adaptiv gestaltet. Genauer gesagt wird eine indirekte adaptive Regelung entworfen. Dies bedeutet, dass der Regler anhand als bekannt vorausgesetzter Streckenparameter entworfen wird und im Betrieb dann eine Anpassung an aktuelle Schätzwerte der Parameter erfolgt. [TAO, 2003, S. 295]

¹ Die Darstellung nach Abbildung 1.1 ist von Darstellungen wie bspw. in [KAMISKE und BRAUER, 2011, S. 339] motiviert.

² Hier wird der Begriff „Prozesssteuerparameter“ verwendet, da der Begriff „Prozessparameter“ in dieser Arbeit anderweitig belegt ist. (Siehe auch Seite 5)

Der wesentliche Schritt der Adaption ist daher die Identifikation der Parameter, der in dieser Arbeit entsprechend Platz eingeräumt wird, und es begründet sich der unten skizzierte Aufbau der Arbeit, nach dem der Reglerentwurf und die Adaption bzw. Identifikation hintereinander abgehandelt werden.

Bei der Darstellung des Reglerentwurfes und dem Entwurf des Parameterschätzers wird besonderer Wert auf die Darlegung des Weges zur Lösungsfindung, d. h. die Methodik gelegt. In Abhängigkeit der Prozesseigenschaften und den Anforderungen sind teilweise verschiedene Lösungsvarianten vorzuziehen, und mit der Darstellung der Methodik wird die Übernahme der Ergebnisse auf andere Anwendungen vereinfacht.

Gliederung der Arbeit

Im Kapitel 2 wird der Untersuchungsgegenstand dieser Arbeit, der sich an der beschriebenen Problemstellung orientiert, definiert. In diesem Kapitel wird auch der Beispielprozess der Druckmaschine genauer vorgestellt.

Im Anschluss wird zunächst die Regelung und dann die Identifikation und Adaption behandelt. Dabei wird in Kapitel 3 mit der Regelung des ungestörten Systems begonnen. Bevor in Kapitel 5 mit der Regelung unter Berücksichtigung der Störungen fortgefahren wird, wird in Kapitel 4 die Modellierung des Rauschens und der Störungen besprochen sowie die verwendeten Grundlagen aus der Stochastik eingeführt. Die Identifikation und Adaption sind Thema des sechsten Kapitels.

Die Arbeit schließt mit einer Zusammenfassung in Kapitel 7.

2 Festlegung des Untersuchungsgegenstandes

Ausgehend von der im vorherigen Kapitel beschriebenen Problemstellung wird in diesem Kapitel der in dieser Arbeit behandelte Untersuchungsgegenstand definiert.

Dazu werden in Abschnitt 2.1 einige wesentliche Begriffe festgelegt, die in der weiteren Arbeit verwendet werden. Daran schließt sich in Abschnitt 2.2 die eigentliche Problemdefinition, d. h. die Festlegung eines mathematischen Modells, an.

In Abschnitt 2.3 wird das Beispiel der Druckmaschine wieder aufgenommen und kurz beschrieben.

Zuletzt wird in Abschnitt 2.4 der Begriff „Qualitätsregelung“ erläutert und diese Arbeit in den Kontext anderer Arbeiten gestellt, die sich mit diesem Thema befassen.

2.1 Verwendete Begriffe

Im Rahmen dieser Arbeit werden folgende gebräuchlichen Begriffe in der angegebenen Bedeutung verwendet.

Einrichten: Anfahren der Produktion nach einer längeren Pause oder nach Umrüsten der Maschine. Ein Einrichtvorgang setzt sich aus einem oder mehreren Anläufen zusammen, bis dieser in die fortlaufende Produktion übergeht.

Anlauf: Ein Anlauf ist das einmalige Starten der Maschine aus dem Stillstand bzw. Standby-Modus, bis zum Erreichen eines stabilen Zustandes der Fertigteile.

Ausschuss: Fertigteil bzw. Gesamtheit der Fertigteile, welche die geforderten Eigenschaften nicht erfüllt.

Produktionsphase: Fortlaufende, stabile Fertigung nach Beenden des letzten Anlaufs des Einrichtens.

Auftrag: Ein Auftrag umfasst das Einrichten und die Produktionsphase.

2.2 Problemdefinition

Aus der oben gegebenen Beschreibung der Problemstellung sowie den Eigenschaften des Beispielsystems ergibt sich folgende Formulierung und Spezifizierung der Problemdefinition, die in dieser Arbeit betrachtet wird.

2.2.1 Prozessmodell

Es wird der statische, diskrete Prozess

$$y_k = \theta_{00} + \theta_1 u_{1,k} + \sum_{i=2}^p \theta_i u_{i,k} + d_{y,k} \quad (2.1)$$

betrachtet. Die unabhängige Variable k , die in dieser Arbeit immer im Index geschrieben wird, entspricht dabei keinen festen Abtastpunkten, sondern den Messungen. Dabei wird in der Regel bei jedem Anlauf nach Erreichen eines eingeschwungenen Zustands der Fertigteile eine Messung vorgenommen. In der Produktionsphase erfolgen die Messungen in gewissen zeitlichen Abständen.

Die einzige Ausgangsgröße des Prozesses ist y . Der Prozess besitzt p Eingangsgrößen u_i , wobei die Eingangsgröße u_1 die primäre Eingangsgröße ist, was auch in Abbildung 2.1 dargestellt ist. Die Größe d_y beinhaltet Teile der Prozessunsicherheiten. Diese hängen u. a. davon ab, in welcher Phase sich der Prozess befindet, also ob beispielsweise zwischen verschiedenen Anläufen gemessen wird, oder ob die Eigenschaft y verschiedener Fertigteile in der Produktionsphase betrachtet wird. Dies wird in Abschnitt 2.2.2 genauer diskutiert.

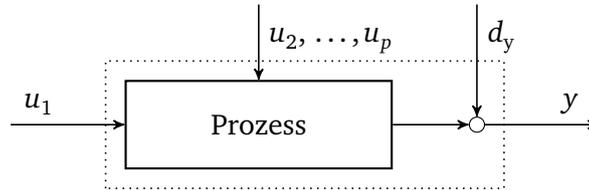


Abbildung 2.1: Prozess aus regelungstechnischer Sicht

Es handelt sich bei dem Prozess (2.1) – auch unter Vernachlässigung von d_y – nicht um einen linearen, sondern einen affinen Zusammenhang bezüglich der Eingänge u_i . Bezüglich der Parameter¹ θ_i ist der Prozess jedoch linear, wenn d_y vernachlässigt wird.

Allgemeiner können auch nichtlineare Abhängigkeiten von Eingangsgrößen u'_i berücksichtigt werden, sofern y weiterhin linear in den Parametern θ_i ist, d. h. sich als Summe von Termen der Art $\theta_i \cdot \varphi_i(u'_1, \dots, u'_p)$ darstellen lässt. Mit $u_i = \varphi_i(u'_1, \dots, u'_p)$ führt dies wieder auf die Form (2.1).²

Es kann ein Nennpunkt $\mathbf{u}_N = [u_{1,N} \ u_{2,N} \ \dots \ u_{p,N}]^T$ definiert werden, der typische oder mittlere Eingangswerte beschreibt. Damit kann Gl. (2.1) auch als

$$y_k = \theta_0 + \theta_1 \cdot (u_{1,k} - u_{1,N}) + \sum_{i=2}^p \theta_i \cdot (u_{i,k} - u_{i,N}) + d_{y,k} \quad (2.2)$$

geschrieben werden. Ohne Störungen $d_{y,k}$ würde für $u_{i,k} = u_{i,N}$ damit $y_k = \theta_0$ gelten.

Alternative, abkürzende Schreibweisen der Prozessgleichung

Die sekundären Eingänge und die dazugehörigen Parameter können auch in den Vektoren

$$\mathbf{u}_S = [u_2 \ \dots \ u_p]^T \quad \text{und} \quad \boldsymbol{\theta}_S = [\theta_2 \ \dots \ \theta_p]^T$$

zusammengefasst werden. Damit kann die Gleichung auch als

$$y_k = \theta_0 + \theta_1 \cdot (u_{1,k} - u_{1,N}) + \boldsymbol{\theta}_S^T \cdot (\mathbf{u}_{S,k} - \mathbf{u}_{S,N}) + d_{y,k} \quad (2.3)$$

geschrieben werden.

In den Untersuchungen wird häufig der Fall betrachtet, dass der Prozess über den primären Eingang $u_{1,k}$ geregelt werden soll, wobei die sekundären Eingänge $\mathbf{u}_{S,k}$ einen festen, sich nicht ändernden Wert

¹ Es ist zwischen den „Prozesssteuerparametern“ auf der einen und den „Prozessparametern“ oder kurz „Parametern“ auf der anderen Seite zu unterscheiden. Erstere bezeichnen die Einstell- und Konfigurationsmöglichkeiten des Fertigungsprozesses, letztere die Koeffizienten der Prozessgleichung (2.1).

² Dies ist dadurch motiviert, dass für die lineare Schätzung als wesentlicher Schritt des Verfahrens nur die Linearität in den Parametern gefordert ist [WOOLDRIDGE, 2003, S. 42], [ISERMANN und MÜNCHHOF, 2011, S. 210ff]. Das prinzipielle Vorgehen bleibt damit gleich, wobei sich jedoch nicht alle Ergebnisse (direkt) übertragen lassen. Der Einfachheit halber wird im Weiteren immer von den Eingangsgrößen u_i (und nicht u'_i) ausgegangen.

$\mathbf{u}_{S,k} \equiv \mathbf{u}_{S,c}$ annehmen. Diese können damit fest eingesetzt werden und man erhält die etwas kompaktere Gleichung

$$y_k = \theta_{0S} + \theta_1 \cdot (u_{1,k} - u_{1,N}) + d_{y,k}, \quad (2.4)$$

wobei $\theta_{0S} = \theta_0 + \boldsymbol{\theta}_S^T \cdot (\mathbf{u}_{S,c} - \mathbf{u}_{S,N})$ ist, bzw.

$$y_k = \theta_{0S} + \theta_1 \cdot v_k + d_{y,k}, \quad (2.5)$$

wenn noch

$$v_k = u_{1,k} - u_{1,N}$$

definiert wird.

Eine weitere Schreibweise der Prozessgleichung (2.2) wird in Anlehnung an eine übliche Notation zur Prozessidentifikation (z. B. in [ISERMANN und MÜNCHHOF, 2011] verwendet) angegeben. Dabei werden alle Prozessparameter in

$$\boldsymbol{\theta} = [\theta_0 \ \cdots \ \theta_p]^T \quad (2.6)$$

und alle Eingänge zusammen mit der Konstanten Eins zum „Eingangsvektor“

$$\boldsymbol{\psi}_k^T = [1 \ u_{1,k} - u_{1,N} \ u_{2,k} - u_{2,N} \ \cdots \ u_{p,k} - u_{p,N}] \quad (2.7)$$

zusammengefasst. Somit lässt sich Gl. (2.1) auch kompakter als

$$y_k = \boldsymbol{\psi}_k^T \boldsymbol{\theta} + d_{y,k} \quad (2.8)$$

schreiben.

Zeitvarianz

Es ist davon auszugehen, dass sich der Parameter θ_0 von Anfahrvorgang zu Anfahrvorgang stark ändern kann. Während der Anläufe eines Anfahrvorgangs ist er in der Regel jedoch konstant. Die in dieser Arbeit verwendeten Verfahren müssen aber robust gegenüber dem Fall sein, dass sich θ_0 auch während eines Anfahrvorgangs ändert. Diese Zeitvarianz des Parameters wird der Prozessunsicherheit d_y zugeschlagen und gleich in Abschnitt 2.2.2 noch genauer besprochen.

Die weiteren Parameter θ_1 bis θ_p sind ebenfalls als zeitvariant anzusehen, allerdings mit deutlich langsameren Änderungsraten.

Beschränkungen

Die Wertebereiche der Eingänge u_i sind mit

$$u_i \in [u_{i,\min}, u_{i,\max}], \quad i = 1, \dots, p,$$

limitiert. Dabei ist hauptsächlich die Sättigung des primären Eingangs u_1 zu berücksichtigen, da dieser durch den Regler verwendet wird. Die sekundären Eingänge werden durch den Benutzer oder einer übergeordneten Steuerung vorgegeben.

2.2.2 Prozessunsicherheiten und Rauschen

Unter den Prozessunsicherheiten werden alle die Effekte zusammengefasst, die dazu führen, dass auch bei gleichen Eingangswerten u_i Schwankungen der Eigenschaft y der Fertigteile auftreten. Solche Schwankungen können durch wechselnde Umgebungseinflüsse, Variation der Qualität der Roh- und Hilfsstoffe, nicht-ideale Stellglieder und unterlagerte Regelungen und allgemein nicht vollständig bekanntes bzw. beherrschbares Verhalten komplexer Fertigungsprozesse entstehen.

Die Prozessunsicherheiten können nach ihrer Ausprägung unterschieden werden. So gibt es Unsicherheiten, die eine über einen längeren Zeitraum andauernde Änderung der (mittleren) Ausgangsgröße erzeugen. Zum anderen gibt es Unsicherheiten bzw. Störungen, die ähnlich einem Rauschen sind, d. h. die bei aufeinanderfolgenden Messungen Schwankungen der Ausgangsgröße um einen Mittelwert erzeugen. Die erstgenannte Art der Unsicherheiten werden im Weiteren als Störungen bezeichnet und der schon eingeführten Größe d_y zugeordnet. Letztere Art wird mit n_p bezeichnet und Prozessrauschen genannt.

Wird

$$y'_k = y_k + n_{p,k}$$

definiert, dann kann y in diesem Sinne als Mittelwert angesehen werden, der sich über einen gewissen, längeren Zeitraum einstellt, wenn die Eingangsgrößen konstant gehalten werden. Die Größe y' bezeichnet den tatsächlichen Ausgang.

Die nicht-ideale Messung führt dazu, dass sich im Allgemeinen der Messwert \tilde{y} vom wahren Ausgangswert y' unterscheidet,

$$\tilde{y}_k = y'_k + n_{M,k}, \quad (2.9)$$

wobei $n_{M,k}$ das Messrauschen bezeichnet. In Abbildung 2.3 sind die Wirkungen der Prozessunsicherheiten und des Prozessrauschens sowie das Messrauschen dargestellt.

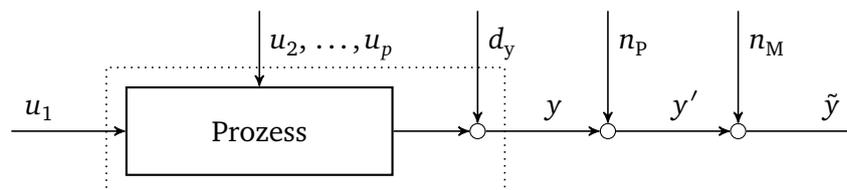


Abbildung 2.2: Prozess mit Unsicherheiten und Rauschen

Für die Betrachtungen ist es meist unerheblich, ob die Schwankungen aufeinanderfolgender Messungen vom Messrauschen oder dem Prozessrauschen hervorgerufen werden. Daher werden diese beiden Effekte zum Prozess- und Messrauschen

$$n_{y,k} = n_{p,k} + n_{M,k}$$

zusammengefasst (siehe auch Abbildung 2.3). Danach gilt

$$\tilde{y}_k = y_k + n_{y,k}. \quad (2.10)$$

Im Allgemeinen hängt die Art und Größe der Schwankungen auch davon ab, welcher Zeitraum betrachtet wird. Hier wird unterschieden zwischen³

³ Diese Aufteilung sowie die folgenden Beschreibungen sind durch die Erfahrungen motiviert, die während der Messungen an der Beispielmachine gemacht wurden.

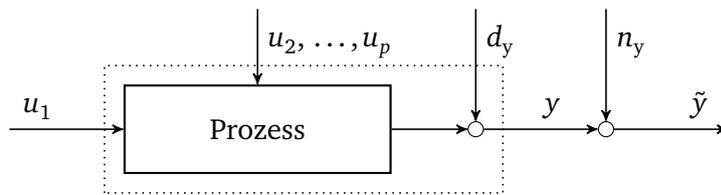


Abbildung 2.3: Prozess mit Unsicherheiten und Rauschen

- den Schwankungen aufeinanderfolgender Fertigteile („Prozessrauschen im engeren Sinne“),
- den Schwankungen innerhalb einer Produktionsphase („Produktionsschwankungen“),
- den Schwankungen zwischen verschiedenen Anläufen eines Einrichtungsvorgangs („Anlaufschwankungen“) sowie
- den Schwankungen zwischen verschiedenen Aufträgen („Auftragsschwankungen“).

Es wird angenommen, dass sich die Wirkung dieser Schwankungen jeweils aus einem mittelwertfreien Rauschen und einem Drift oder auch sprungförmigen Änderungen des Parameters θ_0 zusammensetzt.

Prozessrauschen im engeren Sinne

Das Prozessrauschen im engeren Sinne beschreibt die Änderungen der Ausgangsgröße (bei konstanten Eingangsgrößen), die zwischen direkt hintereinander produzierten Fertigteilen auftreten. Je nach Wirkungsweise der Maschine können aber auch zyklische Störungen mit einer Periodenlänge von wenigen Fertigteilen im Prozessrauschen enthalten sein.

Dieses Rauschen spielt eine untergeordnete Rolle für diese Untersuchung, da primär der Fall betrachtet wird, dass nicht alle Fertigteile vermessen werden, sondern Probestücke nur in längeren Intervallen entnommen werden.

Produktionsschwankungen

Die Produktionsschwankungen aus Sicht der Regelung beschreiben die Änderungen der Ausgangsgröße, die zwischen Fertigteilen zu verschiedenen Zeitpunkten während einer Produktionsphase auftreten. Diese Zeitpunkte sind durch die Intervalle gegeben, in denen Fertigteile gezogen und vermessen werden.

Die Schwankungen können durch wechselnde Umgebungsbedingungen, Qualitätsänderungen von Betriebs- und Hilfsstoffen oder der Änderung von Prozessparametern, beispielsweise aufgrund der Erwärmung der Maschine, entstehen. Diese Effekte erzeugen gewisse Drifts in der Ausgangsgröße, die durch die „Abtastung“ in Form der Messung in bestimmten Intervallen wie Sprünge erscheinen.

Somit bestehen diese Schwankungen aus einem rauschartigen Anteil n_s , der etwas höher ist als das reine Prozessrauschen, und, bei längeren Produktionsphasen, einem Störanteil d_s , der den Drift der Parameter beschreibt.

Diese Schwankungen, die mit n_s bezeichnet werden, sind für die Regelung in der Produktionsphase zu Grunde zu legen. D. h. in dieser Phase gilt $n_p = n_s$.

Anlaufschwankungen

Führt man direkt hintereinander mehrere Anläufe mit konstanten Eingangsgrößen aus, so ist die Variation der Ausgangsgröße bei dem Beispielprozess höher als das reine Prozessrauschen. Ändert man bei mehreren direkt hintereinander durchgeführten Anläufen die Eingangsgrößen zyklisch und betrachtet die Ausgangsgröße immer für Gruppen gleicher Eingangsgrößen, so steigt die Streuung noch etwas an.

Diese Schwankungen werden als Anlaufschwankungen n_A bezeichnet und sind das Prozessrauschen n_p , das der Regel- und Identifikationsalgorithmus in der Anfahrphase „sieht“.

Aufgrund des kurzen betrachteten Zeitraums ist kein Drift der Parameter berücksichtigt. Dennoch besteht immer die, wenn auch seltene, Möglichkeit einer deutlichen sprungförmigen Störung, z. B. aufgrund einer Maschinenstörung oder eines Fehlers bei der Durchführung der Messung. Die Auslegung des Reglers und der Identifikation müssen sich nicht an diesen unwahrscheinlicheren Störungen orientieren, jedoch muss sichergestellt werden, dass das Auftreten einer solchen Störung nicht zu einer dauerhaften Verschlechterung der Regelgüte und der Parameterschätzung führt.

Im Weiteren wird dazu so vorgegangen, dass diese Störungen bei der Auslegung in der Regel ignoriert werden, und deren Wirkung dann im Anschluss betrachtet wird und Gegenmaßnahmen ergriffen werden.

Auftragsschwankungen

Zwischen zwei Aufträgen zu Produktionen der gleichen Fertigteile können sich viele Parameter ändern. So braucht nicht sichergestellt sein, dass die Roh- oder Hilfsstoffe aus der gleichen Charge und vom selben Hersteller kommen. An der Maschine oder am Werkzeug können Wartungsarbeiten durchgeführt worden sein, die das Verhalten beeinflussen. Auch können sich die klimatischen Bedingungen stärker ändern. Dies führt dazu, dass mit größeren Änderungen von Anfahrvorgang zu Anfahrvorgang zu rechnen ist.

Diese Störungen werden hier als sprungförmige Prozessstörungen d_y betrachtet und sind zu berücksichtigen, wenn die Schätzung von Parametern über einzelne Aufträge hinaus betrachtet wird.

Messrauschen

Das Messrauschen n_M wirkt aus zeitlicher Sicht wie das Prozessrauschen im eigentlichen Sinne und wird im Weiteren meist gegenüber den anderen Einflüssen vernachlässigt. D. h. je nach Phase gilt $n_y \approx n_A$ oder $n_y \approx n_S$.

2.2.3 Weitere Randbedingungen

Korrelation zwischen u_1 und u_S

Die Eigenschaft y der Fertigteile soll in der Regel konstant gehalten werden, auch wenn sich sekundäre Eingänge ändern. D. h. die sekundären Eingänge werden zusammen mit dem primären Eingang so angesteuert, dass sich der Ausgang (möglichst) nicht ändert. Daher sind diese untereinander und mit u_1 relativ stark korreliert.

Parameterbegrenzungen

Aus den Erfahrungen des Anwenders mit der Maschine können sich ggf. auch Grenzen für die Prozessparameter angeben lassen,

$$\theta_i \in [\theta_{i,\min}, \theta_{i,\max}], \quad i = 0, \dots, p.$$

Implementierung

Da die neuen Stellgrößen nicht innerhalb weniger Millisekunden nach der Messung vorliegen müssen, ist die Echtzeitfähigkeit zunächst weniger entscheidend. Dafür ist eine effiziente Speicherung der Parametersätze sowie der für das Lernen der Parameter notwendigen Daten wesentlich, d. h. diese soll möglichst kompakt erfolgen. Dabei ist darüber hinaus vorteilhaft, wenn die Anzahl der für einen Auftrag sowie der über einen Auftrag hinaus zu speichernden Daten fest ist, und nicht beliebig anwachsen kann.

Handeingriffe des Bedieners

Der Bediener soll zu jedem Schritt die Möglichkeit haben, den Regler zu überstimmen und beliebige, ihm in der vorliegenden Situation sinnvoll erscheinenden Eingangsgrößen vorgeben zu können. Die Regelung und Online-Identifikation muss damit entsprechend umgehen können.

2.2.4 Ablauf eines Einrichtvorgangs

In Abbildung 2.4 ist der prinzipielle Ablauf eines Einrichtvorgangs aus Sicht der Regelung dargestellt.

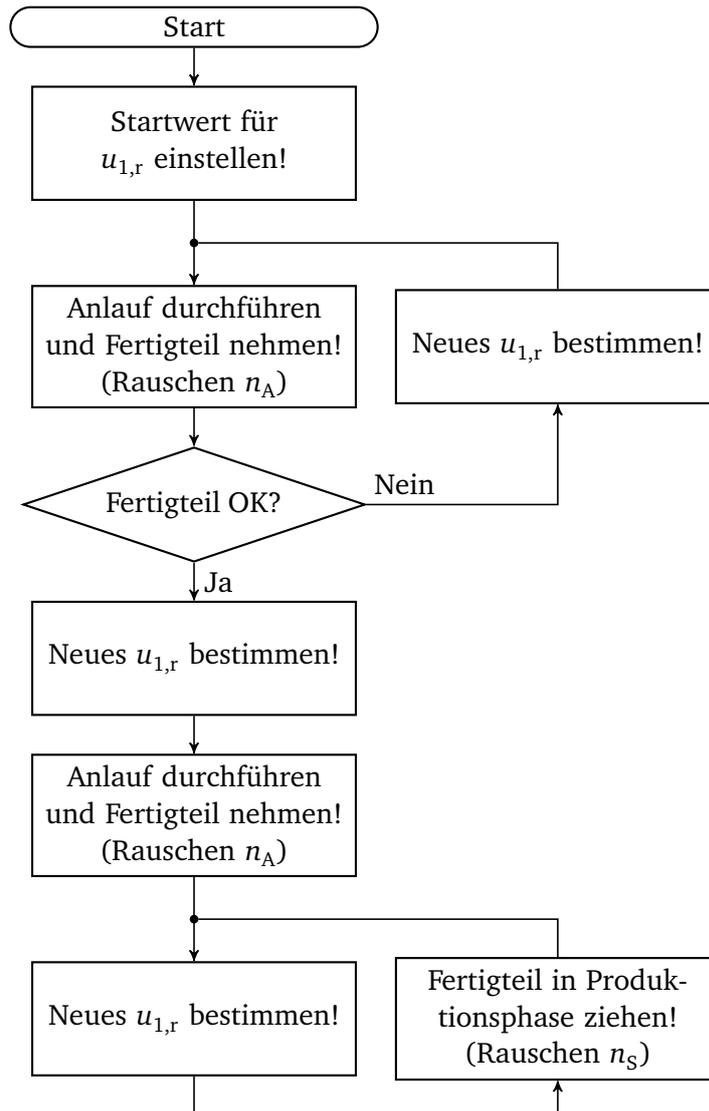


Abbildung 2.4: Ablauf eines Einrichtvorgangs

Als erster Schritt wird der Startwert der primären Eingangsgröße bestimmt. Bei (ungefähr) bekannten Prozessparametern wird diese sinnvollerweise aus der Sollgröße und den Werten der sekundären Eingangsgrößen berechnet. Alternativ kann auch der Bediener der Maschine den Startwert vorgeben.

Es wird dann ein Anlauf durchgeführt und ein Fertigteil vermessen. Dabei ist das reglerrelevante Rauschen $n_y = n_A$ anzunehmen. Es wird geprüft, ob die durch y beschriebene Qualität des Fertigteils innerhalb eines zulässigen Toleranzbereiches liegt. Ist dies der Fall, so wird es als „OK“ bezeichnet. Liegt y außerhalb des Toleranzbereiches, so ist es „nicht OK“, bzw. „ $\overline{\text{OK}}$ “.

Ist das Fertigteil \overline{OK} , so wird mit dem Regler ein neuer Wert für die primäre Eingangsgröße bestimmt und ein weiterer Anlauf durchgeführt. (Ggf. auch unter Berücksichtigung geänderter sekundärer Eingänge.)

Ist das Fertigteil OK, dann wird der Regler nochmals eine Korrektur der Eingangsgröße berechnen und es wird ein weiterer Anlauf durchgeführt (mit reglerrelevantem Rauschen $n_y = n_A$), wobei die Maschine dann aber natürlich nicht mehr gestoppt wird, sondern es beginnt die Produktionsphase, so dass für die folgenden Messungen das Rauschen $n_y = n_S$ anzunehmen ist.

2.2.5 Regelziel

Primäres Regelziel ist es, den OK-Zustand mit möglichst wenigen Regelschritten zu erreichen. Daneben können aber auch Anforderungen an den Verlauf der Regelabweichung bis zum Erreichen dieses Zustandes gegeben sein. So sind mögliche Anforderungen an die Regelung beispielsweise⁴

- „OK-Fertigteil nach maximal zwei Regelschritten!“,
- „Kein Überschwingen der Ausgangsgröße!“,
- „Ein Regelschritt verringert die Abweichung um mindestens 50%!“ und
- „Kompensation von Änderungen der sekundären Eingänge!“ .

Dabei werden die Anforderungen, insbesondere bezüglich der maximalen Anzahl an Regelschritten, strenger sein, wenn von bekannten Prozessparametern ausgegangen wird. D. h. das Lernen und die Adaption der Prozessparameter sollen selbstverständlich auch zu einer Verbesserung der Regelgüte führen.

Betrachtet man die Störungen und das Rauschen, so kommt noch die Forderung dazu, dass dieses am Ausgang möglichst wenig zu erkennen sein soll.

Bezüglich der Störungen und dem Rauschen ist auch zu beachten, dass sich, je nach Annahme der Unsicherheiten d_y und n_y , nicht garantieren lässt, dass nach dem ersten OK-Fertigteil auch alle folgenden Teile der Produktion die OK-Bedingung einhalten werden.

Die mathematische Formulierung der Regelziele wird noch zurückgestellt und zu Beginn der Kapitel nachgeholt, die sich mit der Regelung befassen.

2.3 Anwendungsbeispiel

Zur Veranschaulichung der zuvor bewusst allgemein gehaltenen Betrachtung wird in dieser Arbeit eine Offset-Bogendruckmaschine mit Anilox-Farbwerken betrachtet, an welcher die entwickelten Algorithmen auch getestet wurden.

Druckverfahren

Beim Offsetdruck wird eine ebene Druckform verwendet, bei der die mit Farbe abzubildenden Bereiche eine hydrophobe (wasserabstoßende) Oberfläche haben und die anderen Bereiche hydrophil (wasserliebend) sind. Vor dem Auftragen der Farbe wird Feuchtmittel (Wasser mit Zusätzen) auf die Druckform aufgebracht, das sich auf den hydrophilen Bereichen als dünne Schicht ausbreitet und so verhindert, dass dort Farbe aufgetragen wird. [KIPPHAN, 2000, S. 214ff], [STIEBNER, 1992, S. 228ff]

Die Farbe, die über eine Walze an die hydrophoben Bereiche der Druckform abgegeben wurde, wird dann über eine Walze, die mit einem Gummituch bespannt ist, auf den Bedruckstoff übertragen, weshalb der Offsetdruck auch zu den indirekten Druckverfahren gezählt wird. [KIPPHAN, 2000, S. 214ff], [STIEBNER, 1992, S. 228ff]

⁴ Die hier aufgeführten Regelziele ergaben sich aus den Anforderungen des Projektpartners.

Im klassischen Offsetdruck werden hochviskose Druckfarben verwendet, die durch ein Farbwerk mit einer Vielzahl von Walzen laufen müssen, um sie richtig dosiert an die Druckform abgeben zu können. Auch muss die Farbzufuhr über die Breite des Druckwerkes in vielen einzelnen Zonen gesteuert bzw. geregelt werden, abhängig von der Flächendeckung der Farbe in der jeweiligen Zone, um eine gleichmäßige Schichtdicke zu erhalten. [KIPPHAN, 2000, S. 221ff], [STIEBNER, 1992, S. 228ff]

Beim Anilox-Offsetdruck werden dagegen Druckfarben mit niedriger Viskosität verwendet, die Dosierung erfolgt hierbei über eine Rasterwalze. Diese besitzt auf der Oberfläche kleine Nöpfchen, die eine definierte Menge an Farbe aufnehmen können. Durch diese Dosierung soll dieses Druckverfahren unempfindlicher gegenüber unterschiedlicher Flächendeckung der Druckform sein, so dass keine zonenweise Einstellung nötig ist [AULL, 2008, S. 157]. Das Farbwerk besteht bei Anilox-Druckwerken darüber hinaus nur aus wenigen Walzen, so dass auch von „zonenlosen Kurzfarbwerken“ gesprochen wird. Ein Vorteil des Anilox-Drucks ist damit, dass die Anlaufmakulatur geringer ist als beim klassischen Offsetdruck, da sich die Farbe zunächst über weniger Walzen verteilen muss. Nachteilig ist die Notwendigkeit der Verwendung von niedrigviskosen Farben, deren Eigenschaften teils hinter denen hochviskoser Farben zurückstehen. [KIPPHAN, 2000, S. 224f]

Der Anilox-Offsetdruck hat erst in jüngerer Zeit eine größere Verbreitung erreicht. So ist in [STIEBNER, 1992] der Anilox-Offsetdruck noch gar nicht erwähnt, in [KIPPHAN, 2000] nur im Zusammenhang mit dem Zeitungsdruck. [KIPPHAN, 2000, S. 309] deutete aber schon an, dass dieser auch im Akzidenzdruck Verbreitung finden wird.

Druckwerk

Eine schematische Darstellung eines Druckwerks mit Anilox-Farbwerk ist in Abbildung 2.5 dargestellt. Die folgende Beschreibung folgt den oben angegebenen Quellen.

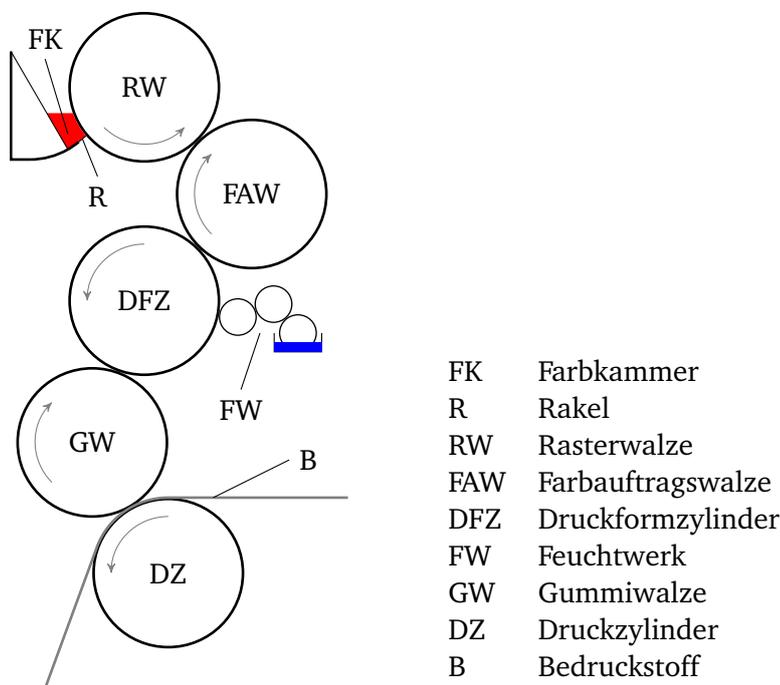


Abbildung 2.5: Schematische Darstellung eines Druckwerks mit Anilox-Farbwerk (an Anicolor-Speedmaster von Heidelberg orientiert)

Die Druckfarbe befindet sich im Farbkasten (FK) und wird von der drehenden Rasterwalze (RW) mit kleinen Nöpfchen, die sich auf dieser befinden, aufgenommen. Das Rakel (R), bzw. Kammerrakel, da es die Farbkammer abschließt, streift überschüssige Farbe sauber ab, so dass nur noch die Nöpfchen der

Rasterwalze definiert gefüllt find. Von der Rasterwalze kommt dann die Farbe auf die Farbauftragswalze (FAW).

Über die Farbauftragswalze wird die Farbe an die Druckform übertragen, die sich auf dem Druckformzylinder (DFZ) befindet. Die Druckform wird dabei zunächst am Feuchtwerk (FW) vorbeigeführt, mit dem die Stellen benetzt werden, die keine Farbe übertragen sollen. Über eine mit einem Gummituch bespannte Walze (GW) wird die Farbe letztlich auf den Bedruckstoff übertragen, der durch den Druckzylinder an die Gummiwalze gepresst wird.

Die Menge der Farbe, die von der Rasterwalze auf die Farbauftragswalze übertragen wird, hängt neben der Größe der Nöpfchen auch von der Temperatur der Rasterwalze ab. Je höher die Temperatur der Rasterwalze ist, desto mehr Farbe wird übertragen, und die Schichtdicke der Farbe auf dem Bedruckstoff wird entsprechend dicker. [AULL, 2008, S. 157]

Farbmessung

Als Ergebnis des Prozesses liegt der bedruckte Bogen vor. Wesentliches Qualitätsmerkmal ist der Farbeindruck des Druckbildes. Dieser wird, neben der Art des Papiers und der Farbe, von der Gleichmäßigkeit und Dicke der aufgetragenen Farbschicht beeinflusst, so dass dies die zu bewerteten Qualitätsmerkmale sind.

Um die Farbschicht objektiv zu beurteilen muss diese vermessen werden. Dazu werden hier zwei Verfahren genannt, die im Rahmen des Projektes auch verwendet wurden.

So kann die Messung mittels der Densitometrie erfolgen, bei der die optische Dichte bestimmt wird [KIPPHAN, 2000, S. 103f] [STIEBNER, 1992, S. 114]. Die optische Dichte ist als

$$D = \log \frac{1}{\beta} = \log \frac{I_0}{I}$$

definiert, wobei β die Remission bzw. I die Lichtintensität des bedruckten Bogens und I_0 die Lichtintensität bei einem weißem (unbedrucktem) Bogen darstellt. Die Logarithmusbildung wird angewandt, um das Maß dem Empfinden des menschlichen Auges anzupassen. Sollen bunte Farben damit vermessen werden, werden entsprechende Filter verwendet. [KIPPHAN, 2000, S. 103f]

Diese Art der Farbmessung hat jedoch den Nachteil, dass der Farbeindruck auch bei gleicher optischer Dichte z. B. von dem Papier abhängen kann. Dies kann mit der aufwendigeren spektralen Farbmessung vermieden werden. Der Vorteil liegt hierbei darin, dass diese Art der Farbmessung der Wahrnehmung des menschlichen Sehempfindens entspricht. [KIPPHAN, 2000, S. 109]

Auch ermöglicht die spektrale Messung die Regelung der Färbung nicht anhand einer vorgegebenen Farbdichte, sondern anhand eines vorgegebenen Farbortes, d. h. einem in einem geeigneten Koordinatensystem beschriebenen Sollspektrums. [KIPPHAN, 2000, S. 319], [PFEIFFER, 1995]

Beschreibung der Prozessgrößen

Der Prozess ist bei der Beispielanwendung das Bedrucken der Papierbogen im Druckwerk. Roh- und Hilfsstoffe im Sinne der Abbildung 1.1 beim betrachteten Druckprozess sind Farbe, Papier und das Feuchtemittel. Der bedruckte Bogen stellt das Fertigteil dar.

Wichtige Prozesssteuerparameter sind die Temperatur der Rasterwalze sowie die Druckgeschwindigkeit. Weitere Prozesssteuerparameter sind beispielsweise die Druckform bzw. das Druckbild, die Einstellung des Feuchtwerks sowie die Temperatur der Farbauftragswalze, die ebenfalls temperiert ist. (Wobei es nicht vorgesehen ist, diese Temperatur zu ändern.)

Umgebungsbedingungen sind im Wesentlichen die Temperatur und Luftfeuchtigkeit, die hier aus Sicht der Prozesssteuerung zu den Unsicherheiten zählen, da diese nicht gemessen werden.

Es ergeben sich immer kleinere Abweichungen der Färbung von Bogen zu Bogen. Zwischen zwei Anläufen treten ebenfalls Abweichungen auf, die etwas größer sind als die erstgenannten. Ein Beispiel für eine Ursache der Abweichungen über einen mittleren Zeitraum ist der Wassergehalt in der Farbe [DIECKHOFF, 1999, S. 99]. Auch traten bei Messungen Abweichungen der Färbung auf, wenn die Maschine umgebaut wurde oder beispielsweise das Gummituch gereinigt wurde.

Anwendung der Problemdefinition auf die Beispielanwendung

Ausgangsgröße

Ausgangsgröße ist entweder die Farbdichte, die densitometrisch bestimmt wird, oder eine relative notwendige Änderung der Farbschichtdicke, die mit spektralen Methoden bestimmt wird [PFEIFFER, 1995]. Für das Verfahren ist der Unterschied zwischen beiden Möglichkeiten nicht wesentlich, da sich in Versuchen mit beiden eine (näherungsweise) affine Abhängigkeit der jeweiligen Ausgangsgröße zur Temperatur der Rasterwalze, sowie der Druckgeschwindigkeit ergeben hat. Daher wird im Weiteren meist von „Färbung“ gesprochen, wenn eine der beiden genannten Größen gemeint ist.

Eingangsgrößen

Aus den Prozesssteuerparametern werden die Eingangsgrößen gewählt. Die Temperatur der Rasterwalze wird als primäre Eingangsgröße u_1 verwendet, da diese in gegebenen Grenzen beliebig gewählt werden kann. Die Druckgeschwindigkeit ist die (einzige) sekundäre Eingangsgröße u_2 . Im Gegensatz zur Temperatur der Rasterwalze wird diese nur vom Bediener der Druckmaschine manuell geändert.

Messungen zeigen einen (näherungsweise) affinen Zusammenhang zwischen diesen beiden Prozesssteuerparametern und der Färbung (siehe auch [BERTI et al., 2010], [MICHELS et al., 2011]).

Parameter

Mit der Wahl der Eingangsgrößen sind auch die Parameter θ_0 , θ_1 und θ_2 festgelegt, wobei θ_1 die Änderung der Färbung bei Änderung der Rasterwalzentemperatur und θ_2 die Abhängigkeit der Färbung von der Druckgeschwindigkeit beschreiben.

Jedoch haben auch andere Prozesssteuerparameter wie beispielsweise die Papiersorte und Farbserie einen Einfluss auf die Färbung. Meist haben diese aber nur einen (wesentlichen) Einfluss auf das Grundniveau θ_0 der Färbung, nicht aber auf die Sensitivitäten, d. h. θ_1 und θ_2 .

Die Papiersorte und die Farbserie haben jedoch auch einen wesentlichen Einfluss auf θ_1 und θ_2 .

Man könnte daher für jeden Prozesssteuerparameter eine Eingangsgröße definieren und somit dessen Einfluss berücksichtigen. Jedoch würde dann die Anzahl der Eingangsgrößen stark steigen, so dass eine gute Schätzung deutlich erschwert würde.

Auch kann beispielsweise die Farbserie nicht einer kontinuierlichen Größe zugeordnet werden, d. h. es müsste für jede Farbserie eine eigene Eingangsgröße definiert werden, die entweder null oder eins ist. (Siehe [WOOLDRIDGE, 2003, S. 218ff] zu „binären“ Größen.) Dadurch würde die Anzahl der Eingangsgrößen noch weiter steigen.

Daher wird hier so vorgegangen, dass für jede Kombination aus Papierart und Farbserie ein Parametersatz bestehend aus θ_0 , θ_1 und θ_2 gelernt wird. Die Sensitivitäten θ_1 und θ_2 sind dabei als konstant anzusehen, da mit der Papierart und Farbserie die wesentlichen Einflussfaktoren auf diese Parameter zur Festlegung der Parametersätze verwendet werden.

Dagegen kann sich θ_0 für einen Parametersatz von Auftrag zu Auftrag ändern. Die entwickelten Algorithmen sind daher so ausgelegt, dass sie Sprünge in θ_0 von Auftrag zu Auftrag erwarten.

Störungen

Neben den systematischen Störungen aufgrund der beschriebenen Definition der Parametersätze, die eine Änderung von θ_0 von vorneherein in Kauf nehmen, wirken sich noch die oben angesprochenen Unsicherheiten als Störungen oder Rauschen auf die Ausgangsgröße aus.

Farbregelungen bei Druckmaschinen

Das hohe Interesse bezüglich des Themas der Farbregelung bei Druckmaschinen zeigt sich durch die hohe Anzahl von Patenten bzw. Patentanmeldungen zu diesem Themenbereich. Im Wesentlichen werden darin Messverfahren und Methoden zur Bestimmung der notwendigen Änderung der Schichtdicke in Abhängigkeit der gemessenen Dichte oder des gemessenen Spektrums vorgestellt.

Beispielhaft, da auch nahe an dem hier bearbeiteten Thema, werden mit [MAYER und PFEIFFER, 2003] und [ELTER und PFEIFFER, 2008] zwei Patentanmeldungen genannt, in denen relativ allgemein lernende Verfahren beschrieben werden, die Messungen nutzen, um die Parameter anzupassen und so zu besseren Ergebnissen zu kommen. Die neue Methodik dieser Arbeit findet sich in [PFEIFFER und LENZ, 2013] wieder und wurde darüber hinaus in [LENZ, 2013] vorgestellt.

Speziell mit der Regelung von Anilox-Farbwerken (bzw. Kurzfarbwerken) befassen sich die Patentanmeldungen [BERTI et al., 2010] und [MICHELS et al., 2011]. Diese beschreiben im Wesentlichen Verfahren, die Freiheitsgrade, die dadurch vorhanden sind, dass mehrere Stellgrößen auf die Färbung wirken, zu nutzen, um ein schnelleres Ausregeln zu erreichen oder einen größeren Bereich der regelbaren Färbung zu ermöglichen.

2.4 Qualitätsregelung

Da das Regelziel hier darin besteht, gewisse Qualitätsmerkmale in der Produktion einzuhalten, wurde der Titel „Qualitätsregelung“ gewählt. Unter diesem Schlagwort finden sich auch Arbeiten mit ähnlichen Zielen, jedoch anderen methodischen Ansätzen.

Ein weitbeachtetes Thema in diesem Kontext ist die statistische Prozesslenkung (SPC, Statistical Process Control) [DIETRICH und SCHULZE, 2003, S. 1]. [LOVE, 2007, S. 837] beschreibt das Ziel der Statistischen Prozesslenkung prägnant als das Feststellen von Veränderungen des Prozesses, wobei diese von den statistischen Effekten befreit werden. Dies ermögliche das Einleiten von Gegenmaßnahmen.

Wesentlicher Bestandteil der statistischen Prozesslenkung sind Qualitätsregelkarten, in die statistische Größen wie z. B. der Mittelwert einer messbaren Eigenschaft von in regelmäßigen Abständen gezogenen Stichproben eingetragen werden. Diese erlauben zum einen die Prozessüberwachung und zum anderen auch ein Ergreifen von Gegenmaßnahmen. Letzteres führt damit zum Schließen eines Regelkreises. [DIETRICH und SCHULZE, 2003], [WESTKÄMPER, 1991]

In [KOCZY, 1998] wird die statistische Prozessregelung in Bezug auf die Kleinserienfertigung behandelt. Die dort eingangs beschriebenen besonderen Probleme gegenüber der Serien- und Massenfertigung treffen im Prinzip auch auf die Herstellung von Druckerzeugnissen zu.

Prinzipiell eignet sich die statistische Prozessregelung jedoch nicht zur Lösung der beschriebenen Problemstellung. Zum einen wäre der zulässige Toleranzbereich in dieser Anwendung zu gering und zum anderen wirkt die statistische Prozessregelung erst zu spät auf das System zurück, weshalb [SAWODNY, 1996, S. 13] einen modellbasierten Ansatz vorzieht.

Der Begriff Qualitätsregelung wird in einer sehr weiten Bedeutung verwendet, und kann beispielsweise auch die „Rückführung“ der Kundenmeinung beinhalten. [WESTKÄMPER, 1991, S. 12] Der hier betrachtete Aspekt wird in der genannten Quelle als „prozessnahe Qualitätsregelung“ bezeichnet.

Der prinzipielle Ansatz dieser Arbeit ist ähnlich dem von [PFEIFFER, 1988] und [SAWODNY, 1996], die sich beide mit der prozessnahen Qualitätsregelung bei der spanenden Fertigung befassen und dabei eine direkte Rückführung gemessener Ausgangsgrößen verwenden.

Insbesondere [SAWODNY, 1996] ist zu erwähnen, da auch hier eine Parameterschätzung stattfindet. Das Prozessmodell ist in dieser Arbeit nichtlinear, womit bei der Verwendung eines linear ausgelegten Reglers eine Adaption erforderlich ist, um eine zufriedenstellende Robustheit zu erreichen [SAWODNY, 1996, S. 83, 84ff]. Auch lassen sich mit einem adaptiven Regleransatz die Anzahl fertigbarer Geometrien erhöhen [SAWODNY, 1996, S. 90ff]. Ein Unterschied zwischen dieser und der vorliegenden Arbeit ist jedoch, dass die Regelung dort während der Fertigung eines Teiles aktiv ist und diese unterstützt, während hier nicht der Prozess der eigentlichen Fertigung, sondern nur das Ergebnis relevant ist. D. h. die hier entworfene Regelung hat nur einen Einfluss auf noch zu fertigende Teile.

Im Vergleich zu letztgenannter Arbeit ist das angenommene Prozessmodell dieser Arbeit deutlich einfacher. Dafür muss hier ein relativ starkes Rauschen, gemessenen an der einzuhaltenden Toleranz, berücksichtigt werden, womit eine stochastische Betrachtung des Reglers notwendig ist. Auch ist dadurch der Einsatz von Parameterschätzverfahren erschwert, weshalb dies einen Schwerpunkt dieser Arbeit darstellt.

3 Regelung des ungestörten Systems

In diesem Kapitel wird kurz auf den Reglerentwurf des ungestörten Systems eingegangen. Dazu werden zunächst die in Abschnitt 2.2.5 formulierten Anforderungen in eine mathematische Form gebracht, die für den Reglerentwurf verwendbar ist. Basierend auf dem in Abschnitt 2.2.1 gegebenen Prozessmodell wird das Streckenmodell aufgestellt, welches dem Regler zugrunde liegt.

Die Komponenten „Vorsteuerung“, „Regler“ und „Kompensation der sekundären Eingänge“ werden kurz besprochen, wobei die einfache Dynamik der Strecke eine sehr kurze Behandlung zulässt. Ebenfalls wird auf ein mögliches Windup sowie die Benutzerinteraktion eingegangen.

3.1 Regelziele

Der Sollwert y_{soll} soll zügig und ohne zu starkes Schwingen erreicht werden. Um diese Anforderungen mathematisch zu beschreiben wird die relative Veränderung

$$r_{y,k} = \frac{e_{k+1}}{e_k} = \frac{y_{\text{soll}} - y_{k+1}}{y_{\text{soll}} - y_k} \quad (3.1)$$

der Regelabweichung e_k eingeführt.¹

Die Anforderung „zügig“ bedeutet genauer, dass in jedem Schritt eine „Mindestkorrektur“ eingehalten werden soll, d. h. in jedem Schritt soll sich die vorliegende Regelabweichung um einen bestimmten Prozentwert verringern. Zudem soll der gesamte Regelvorgang nach einer maximalen Anzahl von Schritten abgeschlossen sein. Die Anforderung „ohne zu starkes Schwingen“ wird hier als maximales prozentuales Überschwingen bezüglich der zu Beginn jeden Schrittes vorliegenden Regelabweichung aufgefasst.

Damit sind folgende Anforderungen zu erfüllen:

- i) maximales Überschwingen von $r_{\text{OS,max}}$ ($r_{\text{OS,max}} = 0$ bedeutet kein Überschwingen, $r_{\text{OS,max}} = 0,1$ bedeutet maximal 10 % Überschwingen)
- ii) ein Regelschritt verringert die Regelabweichung um mindestens $1 - r_{\text{US,max}}$ ($r_{\text{US,max}} = 0,4$ bedeutet, dass mindestens 60 % der Regelabweichung im folgenden Schritt korrigiert werden müssen)
- iii) OK-Zustand (Regelabweichung $|e| \leq e_{\text{OK}}$) nach $n_{\text{RS,max}}$ Regelschritten

Mit der relativen Veränderung der Regelabweichung nach Gl. (3.1) können die Bedingungen i) und ii) in der Form

$$r_{y,k} \overset{!}{\in} [-r_{\text{OS,max}}, r_{\text{US,max}}] \quad (3.2)$$

ausgedrückt werden.

3.2 Streckenmodell

In Abschnitt 2.2 wurde das statische Prozessmodell

$$y_k = \theta_0 + \theta_1 \cdot (u_{1,k} - u_{1,N}) + \boldsymbol{\theta}_S^T \cdot (\mathbf{u}_{S,k} - \mathbf{u}_{S,N}) + d_{y,k} \quad (3.3)$$

¹ Mit e_k wird hier immer die Regelabweichung des tatsächlichen Ausgangs y_k bezeichnet, $e_k = y_{\text{soll}} - y_k$. Die anhand des gemessenen Ausgangs \tilde{y}_k bestimmte Regelabweichung wird als $\tilde{e}_k = y_{\text{soll}} - \tilde{y}_k$ bezeichnet. Im Fall des ungestörten Systems sind diese beiden Größen identisch.

definiert. Für die weitere Betrachtung ist es darüber hinaus nötig, die zeitliche Abfolge zwischen der Stellgröße, die der Regler am Ausgang fordert und der Stellgröße, die die Maschine eingestellt hat, zu beachten. Außerdem sind Stellgrößenbeschränkungen zu berücksichtigen. Der geschlossene Regelkreis hat den in Abbildung 3.1 gezeigten Aufbau.

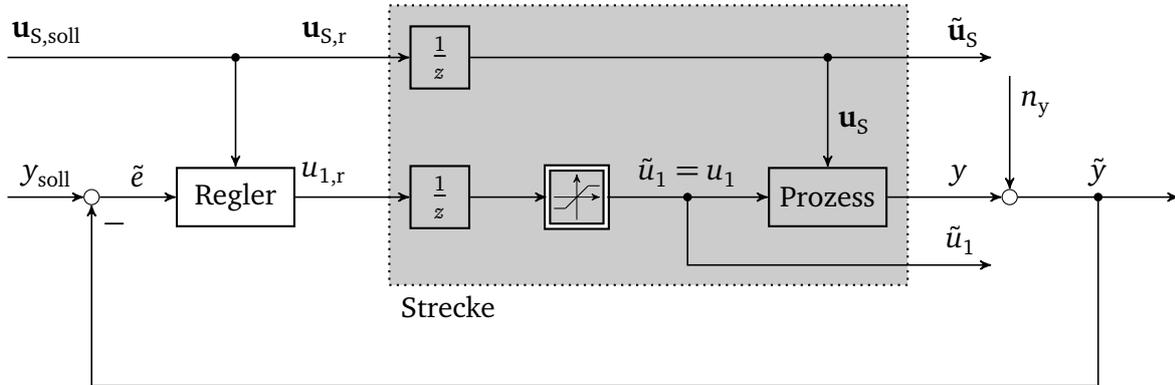


Abbildung 3.1: Betrachteter Regelkreis

Es handelt sich systemtheoretisch um ein Abtastsystem, bei dem die Abtastzeit allerdings keine feste Zeitdauer ist, sondern gerade einem Anlauf entspricht.

Die Verzögerungen um einen Abtastschritt, die Sättigung und der Prozess werden, wie in Abbildung 3.1 dargestellt, zur betrachteten „Strecke“ zusammengefasst.

Nach dem Durchführen eines Anlaufs und dem Vermessen eines Fertigteils, d. h. zu einem Abtastpunkt, werden die mit einer Tilde gekennzeichneten (Mess-)Größen \tilde{y} , \tilde{u}_1 und \tilde{u}_S an die Maschinensteuerung übertragen. Diese bestimmt daraus den neuen Wert der primären Eingangsgröße. In Abbildung 3.1 wird dazu nur die Größe \tilde{y} verwendet, auf die Verwendung der Größen \tilde{u}_1 und \tilde{u}_S wird später eingegangen. Die „Ausgangsgröße“ $u_{1,r}$ des Reglers, die dieser an die Maschinensteuerung übermittelt, wird mit „r“ indiziert. Entsprechend wird auch der neue, gewählte Vektor $\mathbf{u}_{S,r}$ der sekundären Eingangsgrößen mit „r“ indiziert. Da letzterer hier vom Bediener oder einer übergeordneten Steuerung vorgegeben wird, gilt damit auch $\mathbf{u}_{S,r} = \mathbf{u}_{S,soll}$. Aus Abbildung 3.1 können die Zusammenhänge

$$\tilde{u}_{1,k+1} = \text{sat}(u_{1,r,k}, u_{1,\min}, u_{1,\max}) \quad \text{und} \\ \tilde{\mathbf{u}}_{S,k+1} = \mathbf{u}_{S,r,k}$$

zwischen den Tilde-Größen und den „r“-Größen abgelesen werden. Im Folgenden wird die Sättigung meist nicht direkt berücksichtigt, so dass $\tilde{u}_{1,k+1} = u_{1,r,k}$ gesetzt werden kann. Auf die Auswirkung der Sättigung wird in Abschnitt 3.4 auf Seite 22 eingegangen.

Der in Abbildung 3.1 durch die gepunktete Box markierte Bereich wird im Weiteren als „Strecke“ bezeichnet, wobei die „Ausgänge“ \tilde{u}_1 und \tilde{u}_S in der Regel nicht eingezeichnet werden. Für die Strecke gilt damit (ohne Berücksichtigung der Sättigung)

$$y_k = \theta_0 + \theta_1 \cdot (u_{1,r,k-1} - u_{1,N}) + \boldsymbol{\theta}_S^T \cdot (\mathbf{u}_{S,r,k-1} - \mathbf{u}_{S,N}) + d_{y,k}.$$

Für die Auslegung des Reglers sind Annahmen über die Parameter des Druckwerks vorzusetzen. Die tatsächlichen Parameter sind im Allgemeinen nicht bekannt, so dass anstelle von θ_0 , θ_1 und $\boldsymbol{\theta}_S$ mit den Schätzwerten $\hat{\theta}_{0,r}$, $\hat{\theta}_{1,r}$ und $\hat{\boldsymbol{\theta}}_{S,r}$ gearbeitet wird. Es wird aber davon ausgegangen, dass Intervalle angegeben werden können, in denen die Parameter liegen:

$$\theta_i \in [\theta_{i,\min}, \theta_{i,\max}], \quad i = 0, \dots, p. \quad (3.4)$$

Zunächst wird das in Abbildung 3.2 dargestellte System ohne Berücksichtigung des Rauschens n_y und der Störungen d_y betrachtet. Die Gleichung der Strecke lautet in diesem Fall

$$y_k = \theta_0 + \theta_1 \cdot (u_{1,r,k-1} - u_{1,N}) + \boldsymbol{\theta}_S^T \cdot (\mathbf{u}_{S,r,k-1} - \mathbf{u}_{S,N}).$$

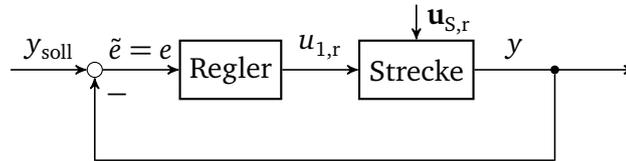


Abbildung 3.2: Regelkreis mit ungestörtem System

3.3 Voreinstellung

Zur Voreinstellung, d. h. der Bestimmung des ersten Wertes $u_{1,V} = u_{1,r,0} = u_{1,1}$ der primären Stellgröße, wird die Streckenfunktion einfach nach $u_{1,r}$ aufgelöst, und man erhält

$$u_{1,V} = \frac{y_{\text{soll}} - \hat{\theta}_{0,r} - \hat{\boldsymbol{\theta}}_{S,r}^T \cdot (\mathbf{u}_{S,\text{soll}} - \mathbf{u}_{S,N})}{\hat{\theta}_{1,r}} + u_{1,N}. \quad (3.5)$$

3.4 Regelung

Hierbei wird das System bei konstanten sekundären Eingängen betrachtet, d. h. es wird von

$$y_k = \theta_{0S} + \theta_1 \cdot (u_{1,r,k-1} - u_{1,N}) \quad (3.6)$$

ausgegangen, wobei $\theta_{0S} = \theta_0 + \boldsymbol{\theta}_S^T \cdot (\mathbf{u}_{S,\text{soll}} - \mathbf{u}_{S,N})$ mit den konstanten sekundären Eingängen $\mathbf{u}_{S,\text{soll}}$ ist.

I-Regler

Das Regelgesetz des einfachen I-Reglers lautet

$$u_{1,r,k} = u_{1,r,k-1} + k_I \cdot (y_{\text{soll}} - y_k). \quad (3.7)$$

Bildet man die Differenz der für die Zeitpunkte $k+1$ und k aufgeschriebenen Gleichung (3.6) ergibt sich

$$\begin{aligned} y_{k+1} - y_k &= \theta_1 \cdot (u_{1,r,k} - u_{1,r,k-1}) \\ &= \theta_1 k_I \cdot (y_{\text{soll}} - y_k), \end{aligned}$$

wobei im zweiten Schritt das Regelgesetz (3.7) verwendet wurde. Umsortieren ergibt die Differenzgleichung

$$y_{k+1} + (\theta_1 k_I - 1)y_k = \theta_1 k_I y_{\text{soll}} \quad (3.8)$$

des geregelten Systems. Diese ist von erster Ordnung mit dem Pol

$$z_0 = 1 - \theta_1 k_I. \quad (3.9)$$

In Abhängigkeit der Wahl von (positiven) k_I ergeben sich folgende Pollagen und damit mögliche Systemverhalten [ISERMANN, 1989, S. 52]:

- $k_I = 0 \Rightarrow z_0 = 1$
Der Regler ist ohne Wirkung ($y_{k+1} = y_k$).
- $\frac{1}{\theta_1} > k_I > 0 \Rightarrow 0 < z_0 < 1$
Der Systemausgang konvergiert monoton gegen den Sollwert.
- $k_I = \frac{1}{\theta_1} \Rightarrow z_0 = 0$
Dead-Beat-Verhalten: Gl. (3.8) wird zu $y_{k+1} = y_{\text{soll}}$, d. h. für jeden beliebigen Wert y_k wird einen Zeitschritt später der Sollwert erreicht.
- $\frac{1}{\theta_1} < k_I < \frac{2}{\theta_1} \Rightarrow 0 > z_0 > -1$
Der Systemausgang konvergiert gegen den Endwert, allerdings mit alternierendem Vorzeichen. Es würde also ein nicht gewünschtes Schwingen auftreten.
- $\frac{2}{\theta_1} \leq k_I \Rightarrow z_0 \geq -1$
Das System wäre instabil.

In Abbildung 3.3a sind beispielhaft Verläufe für verschiedene k_I gezeigt. Alternativ zu dieser Vorgehensweise kann man den Regler und die (um ein beliebiges y linearisierte) Strecke in den z -Bereich transformieren [FÖLLINGER, 1993, S. 43ff, 114ff]. Mit dem Regler

$$G_I(z) = k_I \frac{z}{z-1}$$

und der Strecke

$$G_S(z) = \theta_1 \frac{1}{z}$$

ergibt sich für den offenen Regelkreis in Abbildung 3.2 die Übertragungsfunktion

$$G_o(z) = G_I(z) \cdot G_S(z) = k_I \theta_1 \frac{1}{z-1},$$

woraus sich die in Abbildung 3.3b dargestellte Wurzelortskurve zeichnen lässt, [LUNZE, 2007, S. 450ff]. Aus dieser sind die gleichen Ergebnisse abzulesen.

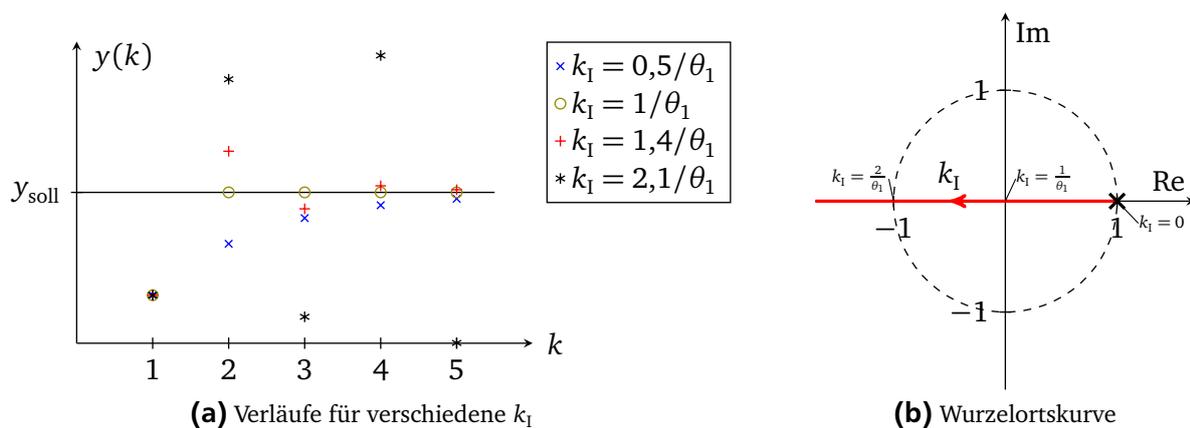


Abbildung 3.3: I-Regler

Idealerweise würde $k_I = 1/\theta_1$ gewählt werden, was allerdings natürlich nur dann möglich ist, wenn θ_1 genau bekannt ist. Um die Bedingungen i) (Vermeidung von Überschwingern) und ii) (Mindestverbesserung) zu erfüllen, muss zunächst die durch Gl. (3.2) definierte „relative Veränderung der Regelabweichung“ $r_{y,k}$ durch Strecken- und Reglerparameter ausgedrückt werden.

Dazu ersetzt man in $y_{\text{soll}} - y_{k+1}$ den Wert y_{k+1} durch die nach y_{k+1} umgestellte Gl. (3.8). Damit folgt

$$\begin{aligned} y_{\text{soll}} - y_{k+1} &= y_{\text{soll}} - \theta_1 k_I y_{\text{soll}} + \theta_1 k_I y_k - y_k \\ &= (y_{\text{soll}} - y_k) \cdot (1 - \theta_1 k_I) \end{aligned}$$

und somit

$$r_{y,k} = \frac{y_{\text{soll}} - y_{k+1}}{y_{\text{soll}} - y_k} = 1 - \theta_1 k_I .$$

Das geregelte System ist ein PT_1 -Glied. Die relative Veränderung ist daher unabhängig vom Startwert und dem aktuellen Zeitpunkt bzw. Abtastschritt k .

Damit lässt sich die Bedingung i) als

$$-r_{\text{OS,max}} \stackrel{!}{\leq} 1 - \theta_1 k_I$$

schreiben, was auf

$$k_I \leq \frac{1 + r_{\text{OS,max}}}{\theta_1}$$

führt. Damit ergibt sich

$$\begin{aligned} k_I &\leq \min \left\{ \frac{1 + r_{\text{OS,max}}}{\theta_1} \mid \theta_1 \in [\theta_{1,\text{min}}, \theta_{1,\text{max}}] \right\} \\ &\leq \frac{1 + r_{\text{OS,max}}}{\theta_{1,\text{max}}} . \end{aligned} \tag{3.10}$$

Analog ergibt sich für die Bedingung ii),

$$1 - \theta_1 k_I \stackrel{!}{\geq} r_{\text{US,max}} ,$$

die untere Grenze

$$\begin{aligned} k_I &\geq \max \left\{ \frac{1 - r_{\text{US,max}}}{\theta_1} \mid \theta_1 \in [\theta_{1,\text{min}}, \theta_{1,\text{max}}] \right\} \\ &\geq \frac{1 - r_{\text{US,max}}}{\theta_{1,\text{min}}} \end{aligned} \tag{3.11}$$

für den Reglerparameter k_I .

Um möglichst schnell zu regeln, ist es sinnvoll, den größten zulässigen Wert zu nehmen, also

$$k_I = \frac{1 + r_{\text{OS,max}}}{\theta_{1,\text{max}}} . \tag{3.12}$$

Aus (3.10) und (3.11) folgt auch, dass

$$\frac{\theta_{1,\text{max}}}{\theta_{1,\text{min}}} \leq \frac{1 + r_{\text{OS,max}}}{1 - r_{\text{US,max}}}$$

gelten muss. Für größere zulässige Intervalle für mögliche Werte von θ_1 können die Bedingungen i) und ii) nicht gleichzeitig eingehalten werden.

Für $r_{OS,max} = 0$ und $r_{US,max} = 0,5$ ergibt sich beispielsweise $k_1 = \frac{1}{\theta_{1,max}}$ und $\frac{\theta_{1,max}}{\theta_{1,min}} \leq 2$.

Die Bedingungen i) und ii) sind mit dieser Wahl immer erfüllt, so dass nur noch die Bedingung iii) betrachtet werden muss. Der einzige Reglerparameter k_1 ist aber schon gewählt, so dass nur noch überprüft werden kann, ob die Bedingung erfüllt wird.

Für die Regelabweichung $e_{n_{RS}+1}$ nach dem n_{RS} -ten Regelschritt folgt

$$e_{n_{RS}+1} = z_0^{n_{RS}} e_1,$$

wobei z_0 der oben bestimmte Pol des geschlossenen Regelkreises und e_1 der Ausgangswert nach der Voreinstellung ist. Es wird nun nach dem n_{RS} gefragt, für das $e_{n_{RS}+1} \leq e_{OK}$ erfüllt ist. Dieses kann über

$$n_{RS} \geq \log_{z_0} \frac{e_{OK}}{e_1}$$

berechnet werden. Dabei hängt z_0 von der Wahl für k_1 bezogen auf die tatsächliche Streckenkonstante θ_1 ab. Der Wert e_1 wird durch die Güte der Voreinstellung beeinflusst.

Windup

Das Strukturbild der mit dem I-Regler geregelten linearen Strecke ist in Abbildung 3.4a gezeigt.

Da am realen Streckeneingang jedoch eine Sättigung vorliegt, müssen Maßnahmen ergriffen werden, den Windup-Effekt zu vermeiden. D. h. es ist zu verhindern, dass sich der Integrator des Reglers über den Sättigungswert auflädt. Dies hätte zur Folge, dass sich der Integrator erst langsam wieder abbauen muss, auch wenn die Regelabweichung wieder das Vorzeichen wechselt. [HIPPE, 2006, S. 3ff]

Dem wird hier einfach dadurch begegnet, dass bei der Berechnung des Stellgröße $u_{1,r,k}$ nicht die Stellgröße des letzten Schrittes $u_{1,r,k-1}$ verwendet wird, sondern die aktuell gemessene bzw. rückgemeldete Temperatur $\tilde{u}_{1,k}$,

$$u_{1,r,k} = \tilde{u}_{1,k} + k_1 \cdot (y_{soll} - y_k).$$

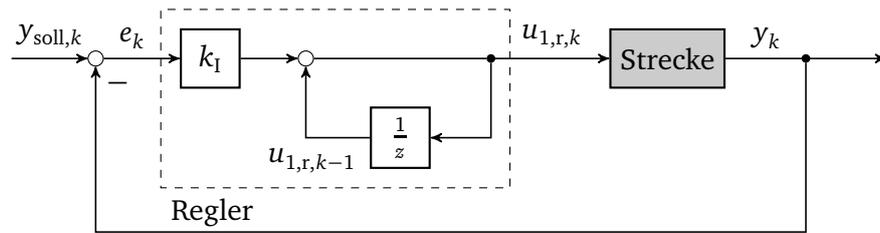
Das sich damit ergebende Blockschaltbild ist in Abbildung 3.4b dargestellt. Aufgrund des einfachen Regelgesetzes sind keine weitergehenden Maßnahmen nötig, um bei dem Verlassen des Sättigungsbereiches sofort wieder das lineare Verhalten zu erhalten.

Der in Abbildung 3.4a als „Strecke“ bezeichneter Block entspricht bis auf die Sättigung der in Abbildung 3.4b hervorgehobenen Strecke. Damit ist in dem Block „Strecke“ in Abbildung 3.4a am Eingang ebenfalls eine Verzögerung um einen Abtastschritt enthalten. Das Gesamtsystem in Abbildung 3.4a besitzt daher zwei Verzögerungen, während das Gesamtsystem in Abbildung 3.4b nur eine Verzögerung enthält. Sofern sich der Streckeneingang jedoch nicht in der Sättigung befindet sind dennoch beide Strukturen äquivalent. Dies liegt daran, dass die Eingänge beider Verzögerungen im Gesamtsystem aus Abbildung 3.4a immer den gleichen Wert besitzen, womit nach dem ersten Abtastschritt auch die Ausgänge beider Verzögerungen, die als Zustände des Gesamtsystems interpretiert werden können, denselben Wert besitzen. Somit ist die Differenz dieser Zustände immer Null und nicht steuerbar.

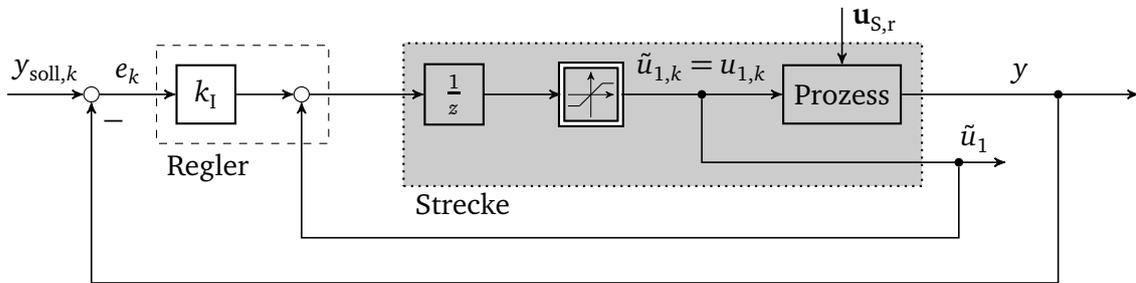
Erweiterungen des I-Reglers

IT-Regler

Das Hinzufügen weiterer Pole, also die Erweiterung zu einem IT-Regler, bringt zum jetzigen Zeitpunkt beim ungestörten System keine Vorteile, sondern würde das System nur langsamer machen, wie auch aus Abbildung 3.5 zu erkennen ist.



(a) Struktur bei idealer Strecke



(b) Struktur bei Sättigung

Abbildung 3.4: Maßnahme zur Vermeidung des Windup-Effekts

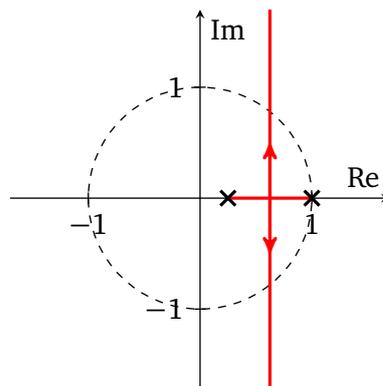


Abbildung 3.5: Wurzelortskurve bei Regelung mit IT-Regler

PI-Regler

Der Zusammenhang zwischen Regelabweichung e_k und der Stellgröße $u_{1,r,k}$ lautet beim PI-Regler

$$u_{1,r,k} = k_I x_{I,k} + k_P e_k,$$

wobei

$$x_{I,k} = x_{I,k-1} + e_k$$

das diskretisierte Integral über der Regelabweichung ist. Aus diesen beiden Gleichungen ergibt sich

$$u_{1,r,k} = u_{1,r,k-1} + k_I \cdot e_k + k_P \cdot (e_k - e_{k-1}),$$

und daraus die Übertragungsfunktion

$$G_{PI}(z) = \frac{z \cdot (k_I - k_P) + k_P}{(z - 1)}$$

im z -Bereich, womit sich

$$G_o(z) = \theta_1 \cdot \frac{z \cdot (k_I - k_p) + k_p}{(z - 1) \cdot z}$$

für den offenen Regelkreis ergibt. Bei diesem liegt die Nullstelle nicht mehr bei null, wie beim reinen I-Regler, sondern bei

$$n_z = \frac{k_p}{k_p - k_I}.$$

Je nach Verhältnis von k_I zu k_p ergibt sich einer der beiden in Abbildung 3.6 gezeigten prinzipiellen Verläufe der Wurzelortkurve, wobei $k_I > 0$ und $k_p > 0$ vorausgesetzt wird. Auch der P-Anteil bringt demnach keinen Vorteil, sondern würde das Systemverhalten nur verschlechtern.

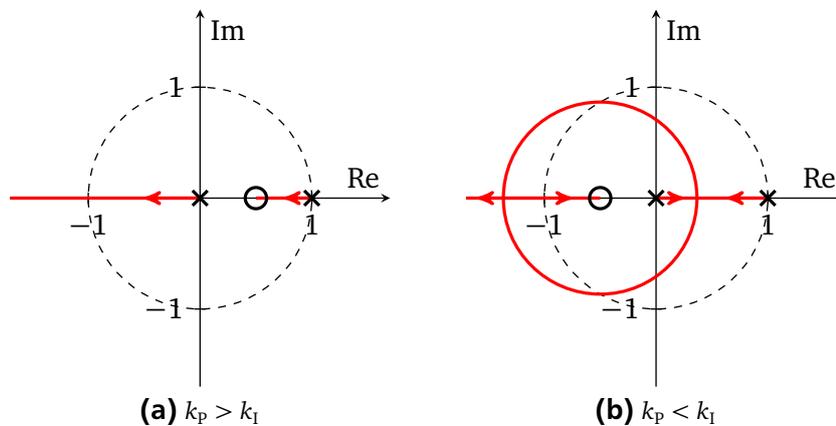


Abbildung 3.6: Wurzelortkurven bei Regelung mit PI-Regler

3.5 Kompensation der sekundären Eingänge

Ändert sich mindestens ein sekundärer Eingang im k -ten Schritt, so ergibt sich dadurch bei gleichem primären Eingang $u_{1,k} = u_{1,k-1}$ eine Änderung des Ausgangs von

$$\Delta y_S = \boldsymbol{\theta}_S^T \cdot (\mathbf{u}_{S,r,k} - \mathbf{u}_{S,r,k-1}) = \boldsymbol{\theta}_S^T \cdot \Delta \mathbf{u}_{S,k}.$$

Diese kann kompensiert werden, wenn $u_{1,r,k}$ um ein $\Delta u_{1,S}$ korrigiert wird, dass über die Bedingung

$$\theta_1 \Delta u_{1,S} = -\Delta y_S$$

bestimmt wird. Somit ergibt sich für die ideale Kompensation

$$\Delta u_{1,S} = -\frac{\boldsymbol{\theta}_S^T \cdot \Delta \mathbf{u}_{S,k}}{\theta_1}.$$

Für die reale Kompensation kann natürlich nur auf die angenommenen Parameterwerte $\hat{\theta}_{1,r}$ und $\hat{\boldsymbol{\theta}}_{S,r}^T$ zurückgegriffen werden, so dass man

$$\Delta u_{1,S} = -\frac{\hat{\boldsymbol{\theta}}_{S,r}^T \cdot \Delta \mathbf{u}_{S,k}}{\hat{\theta}_{1,r}}$$

erhält.

3.6 Zusammenfassung

Fasst man die Reglergleichung mit der Kompensation zusammen, so erhält man

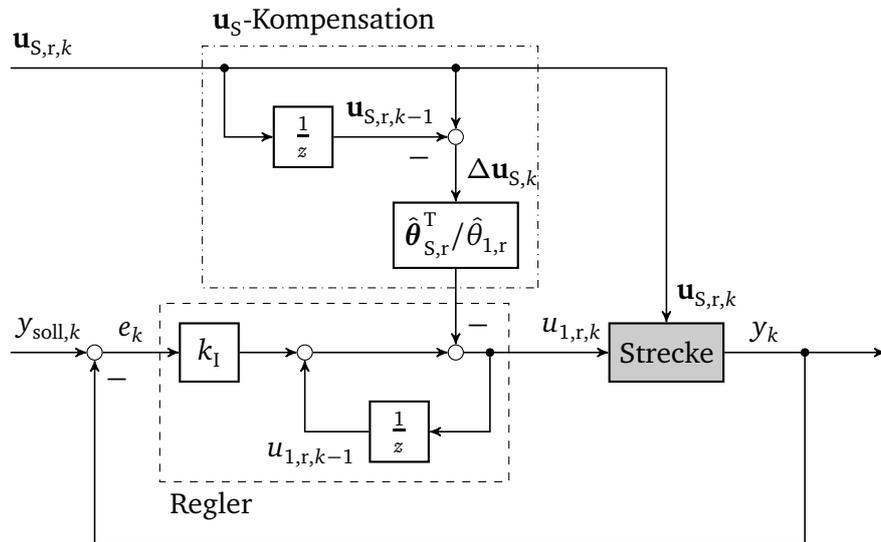
$$u_{1,r,k} = \underbrace{u_{1,r,k-1} + \frac{1}{\hat{\theta}_{1,r}} \cdot (y_{\text{soll}} - \tilde{y}_k)}_{\text{Regler}} - \underbrace{\frac{1}{\hat{\theta}_{1,r}} \hat{\boldsymbol{\theta}}_{S,r}^T \cdot (\mathbf{u}_{S,r,k} - \mathbf{u}_{S,r,k-1})}_{\mathbf{u}_S\text{-Kompensation}} \quad (3.13)$$

bzw.

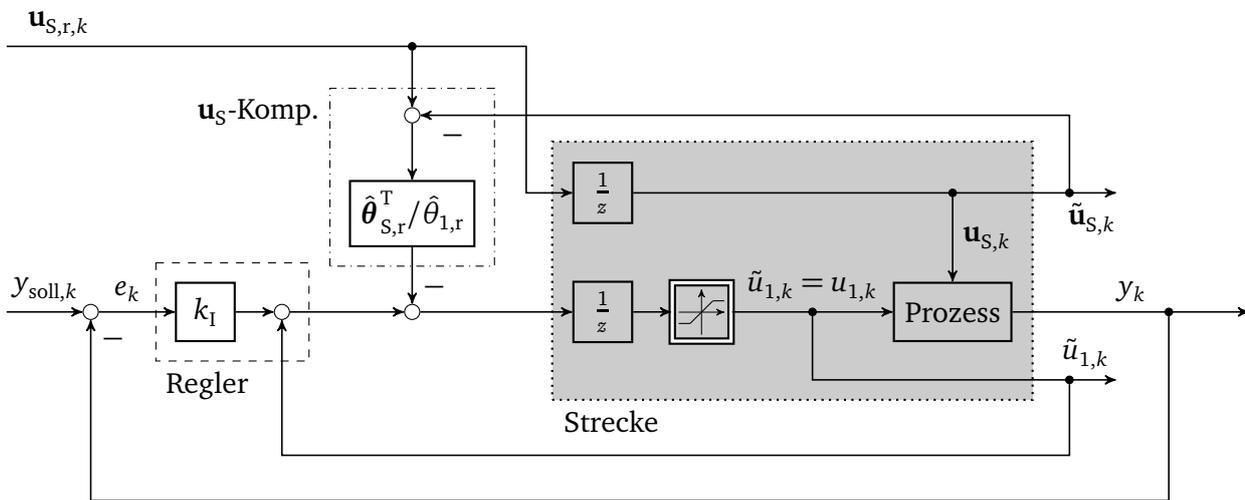
$$u_{1,r,k} = \underbrace{\tilde{u}_{1,r,k} + \frac{1}{\hat{\theta}_{1,r}} \cdot (y_{\text{soll}} - \tilde{y}_k)}_{\text{Regler}} - \underbrace{\frac{1}{\hat{\theta}_{1,r}} \hat{\boldsymbol{\theta}}_{S,r}^T \cdot (\mathbf{u}_{S,r,k} - \tilde{\mathbf{u}}_{S,r,k})}_{\mathbf{u}_S\text{-Kompensation}}, \quad (3.14)$$

wobei bei der zweiten Variante in (3.14) nicht $u_{1,r,k-1}$ und $\mathbf{u}_{S,r,k-1}$, sondern die gemessenen bzw. rückgemeldeten Größen $\tilde{u}_{1,r,k}$ und $\tilde{\mathbf{u}}_{S,r,k}$ verwendet werden. Für die primäre Eingangsgröße ist dies nötig, um ein Windup des Integrators der Regelung zu vermeiden. Daneben hat dies den Vorteil, dass der Bediener den Regler problemlos „überstimmen“ kann, ohne dass es in folgenden Schritten zu einem schlechteren Verhalten führt.

Beide Varianten sind in Abbildung 3.7 dargestellt.



(a) Struktur bei idealer Strecke nach Gl. (3.13)



(b) Struktur mit Maßnahmen gegen Windup und Handeingriffe nach Gl. (3.14)

Abbildung 3.7: Struktur des Reglerkreises mit Kompensation der sekundären Eingänge

4 Modellierung der Störungen

Im vorhergehenden Kapitel wurde die Regelung des ungestörten Systems betrachtet. Bevor im nächsten Kapitel die Regelung des gestörten Systems behandelt wird, ist dieses Kapitel der Modellierung der Störungen gewidmet. Diese Modellierung ist Grundlage der Regelung sowie der in Kapitel 6 behandelten Identifikation.

Neben der Darstellung der gewählten Modelle der Störungen wird in diesem Kapitel auch eine normierte Version der Systemgleichungen eingeführt, die sich in diesem Kontext anbietet. Ein weiterer wichtiger Punkt ist die Betrachtung der bedingten stochastischen Verteilungen des Rauschens.

4.1 Stochastik

In diesem Abschnitt werden kurz die in dieser Arbeit verwendeten Notationen eingeführt und die wesentlichen, verwendeten Eigenschaften stochastischer Variablen und Prozesse zusammengestellt. Dabei wird auf eine mathematische Einführung in die Wahrscheinlichkeitsrechnung verzichtet. Eine solche ist z. B. in [HÄNSLER, 2001] oder, mathematisch etwas strenger aber gut lesbar, in [HOFFMANN et al., 2006, S. 631ff, 683ff] zu finden.

4.1.1 Zufallsvariablen

Eine Zufallsvariable, die in dieser Arbeit in der Form

$$\{x\}_r$$

geschrieben wird, hat die Eigenschaft, dass sie bei jeder Wiederholung eines Experiments einen anderen Wert x annehmen kann, der nicht vorhersagbar ist. Der Wert x , den eine Zufallsvariable $\{x\}_r$ bei einer bestimmten Ausführung des Experiments annimmt, wird als Realisierung der Zufallsvariablen bezeichnet¹ und auch als

$$x = \{x\}_r^\omega$$

geschrieben², wenn in einem Ausdruck die Realisierungen von Zufallsvariablen deutlich gemacht werden sollen.³

Die Wahrscheinlichkeit, dass eine Realisierung x bzw. $\{x\}_r^\omega$ einer Zufallsvariablen $\{x\}_r$ einen Wert kleiner oder gleich einer Konstanten a annimmt, wird hier als

$$P(\{x\}_r^\omega \leq a)$$

geschrieben.⁴ Falls, was hier meist der Fall ist, mit reellen (und nicht diskreten) Zufallsvariablen mit stetigen Verteilungsfunktionen gearbeitet wird, ist $P(\{x\}_r^\omega = a) = 0$ und damit gilt auch $P(\{x\}_r^\omega < a) = P(\{x\}_r^\omega \leq a)$.

¹ Genauer ist die Realisierung der Wert, den die Zufallsvariable bei Eintreten eines bestimmten Elementarereignisses ω annimmt. Und bei verschiedenen Ausführungen eines Experiments kann das gleiche Elementarereignis auftreten. [HÄNSLER, 2001, S. 15, 19]

² Zwischen dieser und der Notation von [HÄNSLER, 2001] gilt damit: $\{x\}_r \Leftrightarrow \mathbf{x}$ und $\{x\}_r^\omega \Leftrightarrow \mathbf{x}(\eta)$.

³ Teilweise wird im Fließtext auch „ x “ geschrieben, wenn eigentlich die Zufallsvariable $\{x\}_r$ gemeint ist und dies aus dem Kontext hervorgeht.

⁴ Dies wird hier als etwas kompaktere Schreibweise für $P(\{\omega \mid \{x\}_r^\omega \leq a\})$ verwendet.

Wahrscheinlichkeitsverteilung

Die Verteilungsfunktion oder kurz Verteilung einer Zufallsvariablen $\{x\}_r$ wird mit $F_{\{x\}_r}(x)$ bezeichnet, wobei das Argument hier nicht eine Realisierung von $\{x\}_r$, sondern einen bestimmten, vorgegebenen Wert meint. Dieser kann daher natürlich auch anders benannt werden, z. B. $F_{\{x\}_r}(a)$. Die Verteilung $F_{\{x\}_r}(a)$ gibt an, mit welcher Wahrscheinlichkeit eine Realisierung x der Zufallsvariablen $\{x\}_r$ einen Wert annimmt, der kleiner oder gleich a ist,

$$F_{\{x\}_r}(a) = P(\{x\}_r^\omega \leq a)$$

[HÄNSLER, 2001, S. 21]. In Abbildung 4.1a sind die Verteilungsfunktionen für die Standardnormalverteilung und für eine Gleichverteilung auf dem Intervall $[-1, 3]$ dargestellt.

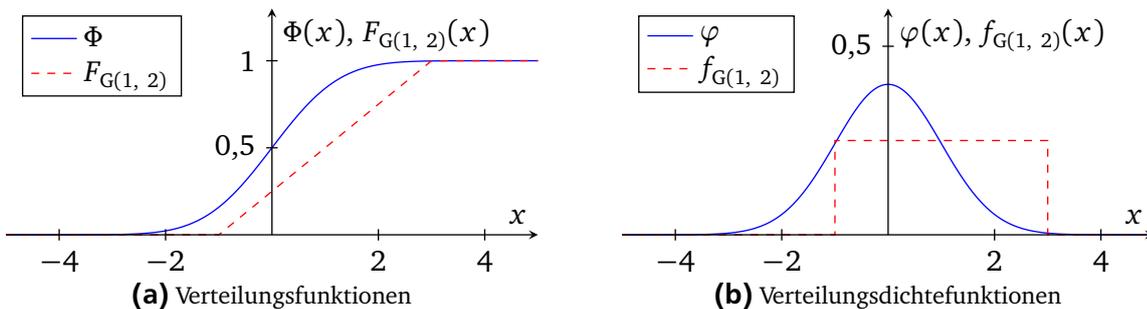


Abbildung 4.1: Standardnormalverteilung $\Phi \sim N(0, 1)$ und Gleichverteilung $G(1, 2)$

Die Verteilungsdichtefunktion ist die Ableitung der Verteilungsfunktion,

$$f_{\{x\}_r}(x) = \frac{dF_{\{x\}_r}(x)}{dx}$$

[HÄNSLER, 2001, S. 22]. In Abbildung 4.1b sind die Verteilungsdichtefunktionen zu den Verteilungen aus Abbildung 4.1a gezeigt.

Besitzen zwei Zufallsvariablen $\{x\}_r$ und $\{y\}_r$ die gleiche Verteilung, so wird dies hier über

$$\{x\}_r \sim \{y\}_r$$

ausgedrückt. Dagegen soll die Schreibweise

$$\{x\}_r = \{y\}_r$$

bedeuten, dass für jedes Experiment bzw. jedes Elementarereignis die Realisierungen $\{x\}_r^\omega$ und $\{y\}_r^\omega$ von $\{x\}_r$ bzw. $\{y\}_r$ den gleichen Wert annehmen, d. h. $\{x\}_r^\omega = \{y\}_r^\omega \forall \omega$. Damit impliziert $\{x\}_r = \{y\}_r$ auch $\{x\}_r \sim \{y\}_r$.

Gemeinsame Verteilungen

Betrachtet man zwei Zufallsvariablen $\{x\}_r$ und $\{y\}_r$, so kann man eine gemeinsame Verteilung

$$F_{\{x\}_r, \{y\}_r}(x, y) = P(\{x\}_r^\omega < x \wedge \{y\}_r^\omega < y)$$

angeben. Für die gemeinsame Verteilungsdichte gilt

$$f_{\{x\}_r, \{y\}_r}(x, y) = \frac{\partial^2 F_{\{x\}_r, \{y\}_r}(x, y)}{\partial x \partial y}.$$

Integriert man über eine der Variablen, so erhält man die Randdichte der anderen Variablen,

$$f_{\{x\}_r}(x) = \int_{-\infty}^{\infty} f_{\{x\}_r, \{y\}_r}(x, y) \cdot dy.$$

[HÄNSLER, 2001, S. 25ff]

Erwartungswert

Der Erwartungswert ist der Mittelwert aller Realisierungen und kann über

$$E\{x\}_r = \int_{-\infty}^{\infty} x \cdot f_{\{x\}_r}(x) \cdot dx$$

berechnet werden [HÄNSLER, 2001, S. 34]. In dieser Arbeit wird der Erwartungswert teils auch mit

$$\mu_x = E\{x\}_r$$

abgekürzt.

Eine wesentliche Eigenschaft der Erwartungswertbildung ist, dass diese linear ist, d. h. mit zwei beliebigen Zufallsvariablen $\{x\}_r$ und $\{y\}_r$ sowie der Konstanten a gilt

$$E\{ax\}_r = a \cdot E\{x\}_r$$

und

$$E\{x + y\}_r = E\{x\}_r + E\{y\}_r$$

[HÄNSLER, 2001, S. 35], [HOFFMANN et al., 2006, S. 706ff]. Hierbei ist hervorzuheben, dass diese Beziehungen immer gelten, unabhängig davon, welche Verteilungen $\{x\}_r$ und $\{y\}_r$ besitzen, oder in welcher Beziehung diese zueinander stehen.

Varianz

Die Varianz einer Zufallsvariablen $\{x\}_r$ ist über

$$\text{Var}\{x\}_r = E\{(x - E\{x\}_r)^2\}_r$$

definiert. Es wird in dieser Arbeit für die Varianz auch häufig die Abkürzung

$$\text{Var}\{x\}_r = \sigma_x^2$$

verwendet. [RINNE, 2008, S. 191]

Betrachtet man zwei Zufallsvariablen $\{x\}_r$ und $\{y\}_r$, so kann auch die Kovarianz

$$\text{Cov}(\{x\}_r, \{y\}_r) = E\{(x - E\{x\}_r) \cdot (y - E\{y\}_r)\}_r$$

definiert werden.

Für die Varianz der Zufallsvariablen $\{ax\}_r$ und $\{x + y\}_r$ ergibt sich damit

$$\text{Var}\{ax\}_r = a^2 \cdot \text{Var}\{x\}_r$$

und

$$\text{Var}\{x + y\}_r = \text{Var}\{x\}_r + \text{Var}\{y\}_r + 2 \cdot \text{Cov}(\{x\}_r, \{y\}_r).$$

[RINNE, 2008, S. 191]⁵

⁵ $\text{Var}\{ax\}_r = E\{(ax - a\mu_x)^2\}_r = a^2 \cdot E\{(x - \mu_x)^2\}_r = a^2 \cdot \text{Var}\{x\}_r$
 $\text{Var}\{x + y\}_r = E\{(x + y - \mu_x - \mu_y)^2\}_r = E\{(x - \mu_x)^2 + (y - \mu_y)^2 + 2 \cdot (x - \mu_x) \cdot (y - \mu_y)\}_r = \text{Var}\{x\}_r + \text{Var}\{y\}_r + 2 \cdot \text{Cov}(\{x\}_r, \{y\}_r)$

Unkorrelierte und unabhängige Variablen

Kann man die Erwartungswertbildung und die Multiplikation zweier Zufallsvariablen $\{x\}_r$ und $\{y\}_r$ vertauschen,

$$E\{x \cdot y\}_r = E\{x\}_r \cdot E\{y\}_r,$$

so nennt man $\{x\}_r$ und $\{y\}_r$ unkorreliert. [HÄNSLER, 2001, S. 39]

Aus dieser Festlegung folgt, dass für unkorrelierte Zufallsvariablen $\{x\}_r$ und $\{y\}_r$ die Kovarianz verschwindet, $\text{Cov}(\{x\}_r, \{y\}_r) = 0$. Damit folgt weiter

$$\text{Var}\{x + y\}_r = \text{Var}\{x\}_r + \text{Var}\{y\}_r.$$

Der Korrelationskoeffizient ist über

$$\rho_{xy} = \frac{\text{Cov}(\{x\}_r, \{y\}_r)}{\sqrt{\text{Var}\{x\}_r \cdot \text{Var}\{y\}_r}}$$

gegeben und ist Null, wenn $\{x\}_r$ und $\{y\}_r$ unkorreliert sind. Sind beide Zufallsvariablen ideal (negativ) korreliert, so ist der Korrelationskoeffizient 1 (bzw. -1). [HÄNSLER, 2001, S. 41f]

Zwei Zufallsvariablen werden statistisch unabhängig genannt, wenn

$$f_{\{x\}_r, \{y\}_r}(x, y) = f_{\{x\}_r}(x) \cdot f_{\{y\}_r}(y)$$

gilt. Unabhängigkeit ist eine stärkere Eigenschaft als Unkorreliertheit, d. h. zwei unabhängige Zufallsvariablen sind immer unkorreliert, die Umkehrung gilt aber nicht. [HÄNSLER, 2001, S. 28, 39]

Normalverteilung

Die Normalverteilung $N(\mu, \sigma^2)$ wird über deren Erwartungswert μ und Varianz σ^2 vollständig beschrieben. Deren Verteilungs- und Dichtefunktion sind in Anhang A.1 angegeben.

Für die Verteilungs- und Dichtefunktion einer Normalverteilung werden auch die Schreibweisen

$$\Phi(x; \mu, \sigma^2) = F_{N(\mu, \sigma^2)}(x) \quad \text{und} \quad \varphi(x; \mu, \sigma^2) = f_{N(\mu, \sigma^2)}(x)$$

verwendet. Eine besondere Stellung nimmt die Standardnormalverteilung $N(0, 1)$ ein, bei der auch kurz

$$\Phi(x) = F_{N(0, 1)}(x) \quad \text{und} \quad \varphi(x) = f_{N(0, 1)}(x)$$

geschrieben wird.

Eine wesentliche Eigenschaft der Normalverteilung ist, dass das Produkt einer normalverteilten Zufallsvariablen mit einer Konstanten sowie die Summe von zwei normalverteilten Zufallsvariablen jeweils wieder eine normalverteilte Zufallsvariable ergibt [HÄNSLER, 2001, S. 98]. Mit den oben angegebenen allgemeinen Beziehungen für den Erwartungswert und die Varianz ergibt sich Folgendes:

- Wenn $\{x\}_r \sim N(\mu_x, \sigma_x^2)$, dann gilt

$$\{a \cdot x\}_r \sim N(a \cdot \mu_x, a^2 \cdot \sigma_x^2). \quad (4.1)$$

- Wenn $\{x\}_r \sim N(\mu_x, \sigma_x^2)$ und $\{y\}_r \sim N(\mu_y, \sigma_y^2)$, dann gilt

$$\{x + y\}_r \sim N(\mu_x + \mu_y, \sigma_x^2 + 2\sigma_x\sigma_y\rho_{xy} + \sigma_y^2),$$

mit dem Korrelationskoeffizienten ρ_{xy} . Für unkorrelierte Zufallsvariablen gilt $\rho_{xy} = 0$ und damit

$$\{x + y\}_r \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2). \quad (4.2)$$

Gleichverteilung

Bei einer Gleich- oder Rechtecksverteilung $\{x\}_r \sim G(\mu, \Delta)$ kann x nur Werte im Intervall $[\mu - \Delta, \mu + \Delta]$ annehmen, wobei die Wahrscheinlichkeitsdichte innerhalb dieses Intervalls konstant ist. [RINNE, 2008, S. 237]

In dem in Abbildung 4.1 dargestellten Beispiel ist $\{x\}_r \sim G(1, 2)$, und damit hat die Wahrscheinlichkeitsdichte im Intervall $[-1, 3]$ den konstanten Wert $1/4$, und sie ist Null außerhalb dieses Intervalls.

Bedingte Verteilungen

Es seien die Zufallsvariablen $\{x\}_r$ und $\{y\}_r$ gegeben, die im Allgemeinen korreliert sind. Wenn die Verteilung von $\{x\}_r$ betrachtet wird, die unter der Bedingung vorliegt, dass die Realisierung der Zufallsvariable $\{y\}_r$ den (festen) Wert y_0 annimmt, so wird diese Verteilung ausführlich als

$$\{x|\{y\}_r^\omega = y_0\}_r$$

bezeichnet. Um eine etwas kürzere Notation zu erhalten und da Verwechslungen hier ausgeschlossen sind, wird im Weiteren kurz

$$\{x|y_0\}_r$$

geschrieben. Dies ist besonders zu beachten, wenn für die Zufallsvariable und Wert der speziellen Realisierung derselbe Name verwendet wird, d. h. $\{x|\{y\}_r^\omega = y\}_r = \{x|y\}_r$.

Für die abkürzenden Schreibweisen für Mittelwerte und Varianzen gilt dieselbe Notation entsprechend, d. h. $E\{x|y\}_r = \mu_{x|y}$ und $\text{Var}\{x|y\}_r = \sigma_{x|y}^2$.

Die bedingte Dichte der Verteilung $\{x|y_0\}_r$ ist über

$$f_{\{x|y_0\}_r} = \frac{f_{\{x\}_r, \{y\}_r}(x, y_0)}{f_{\{y\}_r}(y_0)}$$

gegeben. Sind $\{x\}_r$ und $\{y\}_r$ unabhängig, dann gilt

$$\{x|y\}_r = \{x\}_r.$$

[RINNE, 2008, S. 200]

Vektorielle Zufallsvariablen

Eine vektorielle Zufallsvariable ist zunächst eine einfache Zusammenfassung mehrerer Zufallsvariablen in einem Vektor.

Für eine n -dimensionale vektorielle Zufallsvariable $\{\mathbf{x}\}_r$ ist die Kovarianzmatrix

$$\text{Cov}(\{\mathbf{x}\}_r, \{\mathbf{x}\}_r) = E\{(\mathbf{x} - E\{\mathbf{x}\}_r) \cdot (\mathbf{x} - E\{\mathbf{x}\}_r)^T\}_r = \begin{bmatrix} \text{Var}\{x_1\}_r & \cdots & \text{Cov}(\{x_1\}_r, \{x_n\}_r) \\ \vdots & \ddots & \vdots \\ \text{Cov}(\{x_n\}_r, \{x_1\}_r) & \cdots & \text{Var}\{x_n\}_r \end{bmatrix}$$

definiert [BRAMMER und SIFFLING, 1975, S. 142]. Für $\text{Cov}(\{\mathbf{x}\}_r, \{\mathbf{x}\}_r)$ kann auch kürzer $\text{Cov}\{\mathbf{x}\}_r$ oder $C_{\mathbf{xx}}$ geschrieben werden. Sind die einzelnen in \mathbf{x} zusammengefassten Zufallsvariablen unkorreliert, so ist $\text{Cov}\{\mathbf{x}\}_r$ eine Diagonalmatrix.

Allgemein kann die Kovarianzmatrix auch für zwei vektorielle Zufallsvariablen angegeben werden,

$$\text{Cov}(\{\mathbf{x}\}_r, \{\mathbf{y}\}_r) = E\{(\mathbf{x} - E\{\mathbf{x}\}_r) \cdot (\mathbf{y} - E\{\mathbf{y}\}_r)^T\}_r = \begin{bmatrix} \text{Cov}(\{x_1\}_r, \{y_1\}_r) & \cdots & \text{Cov}(\{x_1\}_r, \{y_m\}_r) \\ \vdots & \ddots & \vdots \\ \text{Cov}(\{x_n\}_r, \{y_1\}_r) & \cdots & \text{Cov}(\{x_n\}_r, \{y_m\}_r) \end{bmatrix}.$$

Hierbei ist $\{\mathbf{y}\}_r$ eine m -dimensionale vektorielle Zufallsvariable. [BRAMMER und SIFFLING, 1975, S. 145] Eine alternative Schreibweise ist $\text{Cov}(\{\mathbf{x}\}_r, \{\mathbf{y}\}_r) = C_{\mathbf{xy}}$.

4.1.2 Zufallsprozesse

Bei diskreten Zufallsprozessen $\{x_k\}_r$ gehört zu jeder Realisierung nicht ein einzelner Wert, sondern eine Folge von Variablen x_k . [HÄNSLER, 2001, S. 52f]

Für ein festes k ergibt sich damit eine Zufallsvariable, für die auch die schon vorgestellten Größen wie z. B. der Erwartungswert und die Varianz definiert sind. [HÄNSLER, 2001, S. 58ff]

Korreliertheit

Zwei Zufallsprozesse sind unkorreliert, wenn

$$E\{x_k \cdot y_l\}_r = E\{x_k\}_r \cdot E\{y_l\}_r \quad \forall k, l$$

gilt. [HÄNSLER, 2001, S. 63]

Desweiteren wird hier auch von *einem* unkorrelierten Zufallsprozess gesprochen, wenn die zu zwei unterschiedlichen Zeitpunkten gehörenden Zufallsvariablen $\{x_k\}_r$ und $\{x_l\}_r$ eines Zufallsprozesses unkorreliert sind, d. h. wenn

$$E\{x_k \cdot x_l\}_r = E\{x_k\}_r \cdot E\{x_l\}_r \quad \text{für } k \neq l$$

gilt.

Weißes Rauschen

Als „Weißes Rauschen“ wird ein Prozess bezeichnet, dessen einzelne Verteilungen $\{x_k\}_r$ zu den Zeitpunkten k unabhängig und identisch verteilt sind. [RINNE, 2008, S. 392] Aus der Unabhängigkeit folgt

$$\text{Cov}(\{x_k\}_r, \{x_l\}_r) = \begin{cases} \text{Var}\{x_k\}_r & \text{für } k = l \\ 0 & \text{für } k \neq l, \end{cases}$$

wobei aufgrund der identischen Verteilung $\text{Var}\{x_k\}_r$ für alle k denselben Wert besitzt.

Sind darüber hinaus die einzelnen Variablen $\{x_k\}_r$ normalverteilt, so spricht man von einem „Gauß’schen Weißen Rauschen“. [BRAMMER und SIFFLING, 1975, S. 196]

Vektorielle Zufallsprozesse

Wie bei Zufallsvariablen kann auch ein Zufallsprozess aus vektoriellen Variablen bestehen. In diesem Kontext werden die Bezeichnungen

$$\text{Cov}(\{\mathbf{x}_k\}_r, \{\mathbf{x}_l\}_r) = \mathbf{C}_{\mathbf{xx}}(k, l) = E\{(\mathbf{x}_k - E\{\mathbf{x}_k\}_r) \cdot (\mathbf{x}_l - E\{\mathbf{x}_l\}_r)^T\}_r$$

und

$$\text{Cov}(\{\mathbf{x}_k\}_r, \{\mathbf{y}_l\}_r) = \mathbf{C}_{\mathbf{xy}}(k, l) = E\{(\mathbf{x}_k - E\{\mathbf{x}_k\}_r) \cdot (\mathbf{y}_l - E\{\mathbf{y}_l\}_r)^T\}_r$$

verwendet. [BRAMMER und SIFFLING, 1975, S. 180f]

Wird die Kovarianzmatrix für beide Prozesse zum gleichen Zeitpunkt $k = l$ ausgewertet, so wird auch kürzer $\mathbf{C}_{\mathbf{xx}}(k) = \mathbf{C}_{\mathbf{xx}}(k, k)$ bzw. $\mathbf{C}_{\mathbf{xy}}(k) = \mathbf{C}_{\mathbf{xy}}(k, k)$ geschrieben.

4.2 Beschreibung der Störungen

In Abschnitt 2.2.2 wurde auf die verschiedenen Arten von Störungen eingegangen. Diese wurden dabei in „rauschartige“ Störungen n_y und sprungförmige Störungen d_y eingeteilt, die in Gl. (2.5),

$$y_k = \theta_{0S} + \theta_1 \cdot (u_{1,k} - u_{1,N}) + d_y,$$

und (2.10),

$$\tilde{y}_k = y_k + n_{y,k},$$

wirken.

Es wird hier zwar formal zwischen der Ausgangsgröße y_k des Prozesses und der Messgröße \tilde{y}_k unterschieden, jedoch ist zu bedenken, dass das Rauschen $n_{y,k}$ nicht nur Messrauschen, sondern auch tatsächliche Variationen der Ausgangsgröße umfasst. In wieweit dies bei der Definition der Regelziele zu berücksichtigen ist, wird im Kapitel 5 besprochen.

Im Normalfall wird man versuchen, $u_{1,1} = u_{1,V}$ mit Gl. (3.5) so vorzugeben, dass schon die erste Regelabweichung $e_1 = 0$ ist, d. h. $u_{1,1} = u_{1,1}(y_{\text{soll}}, \hat{\theta}_{0S,r}, \hat{\theta}_{1,r})$. Da die Streckenparameter θ_{0S} und θ_1 jedoch nicht exakt bekannt sind, d. h. $\hat{\theta}_{0S,r} = \theta_{0S} + \Delta\hat{\theta}_{0S,r}$ bzw. $\hat{\theta}_{1,r} = \theta_1 + \Delta\hat{\theta}_{1,r}$, und weitere Störungen d_y^* wirken, wird im Allgemeinen $e_1 = 0$ nicht gelten. Die Parameterunsicherheiten und Störungen können dann zur Störung d_y zusammengefasst werden, d. h.

$$\begin{aligned} e_1 &= y_{\text{soll}} - (\theta_{0S} + \Delta\hat{\theta}_{0S,r}) - (\theta_1 + \Delta\hat{\theta}_{1,r}) \cdot (u_{1,1} - u_{1,N}) - d_y^* \\ &= \underbrace{y_{\text{soll}} - \theta_{0S} - \theta_1 \cdot (u_{1,1} - u_{1,N})}_{e_1^*} - \underbrace{\Delta\hat{\theta}_{0S,r} - \Delta\hat{\theta}_{1,r} \cdot (u_{1,1} - u_{1,N})}_{-d_y} = e_1^* - d_y. \end{aligned}$$

Im Weiteren wird $e_1^* = 0$ gesetzt, da e_1^* die „Zielabweichung“ beschreibt, die mit der Voreinstellung $u_{1,1}$ erreicht werden soll, womit

$$e_1 = -d_y \tag{4.3}$$

und

$$\tilde{e}_1 = e_1 - n_{y,1} = -d_y - n_{y,1} \tag{4.4}$$

gilt.

Wenn nicht versucht wird, $u_{1,1}$ so zu wählen, dass $e_1 = 0$ ist, beispielsweise indem man prinzipiell mit einem festen $u_{1,V}$ beginnt, so wird der dadurch entstehende systematische Fehler der Störung d_y zugeschlagen.

Störung d_y

Im Folgenden wird in der Regel davon ausgegangen, dass die Störung d_y über einen Auftrag konstant ist. Auf die Behandlung des Falles, dass auch während eines Auftrags sprungförmige Störungen auftreten, wird an den entsprechenden Stellen explizit eingegangen. Von Auftrag zu Auftrag ändert sich d_y jedoch, d. h. $d_y = d_{y,l'}$, wobei l' für den Auftrag steht.

Die Verteilung von $d_{y,l'}$ über mehrere Aufträge ist schwer zu bestimmen. Beispielsweise hängen die Schwankungen von $d_{y,l'}$ auch davon ab, ob die Aufträge direkt hintereinander bearbeitet werden, ob

gleichartige Aufträge durchgeführt werden⁶ und wie sich die Güte der Parameterschätzwerte über die Aufträge ändert.

Später in diesem Kapitel wird diskutiert, was aus einem stark schwankenden bzw. einem (über mehrere Aufträge) konstanten $d_{y,l'}$ folgt. Ansonsten werden über $d_{y,l'}$ keine Annahmen getroffen. Was dies für die Analyse bedeutet, wird ebenfalls am Ende dieses Kapitels diskutiert.

Rauschen n_y

Das Rauschen wird im Folgenden meist als Gauß'sches Weißes Rauschen mit der Varianz

$$\text{Var}\{n_{y,k}\}_r = \sigma_y^2 \quad (4.5)$$

modelliert.

Ob diese Annahme berechtigt ist, muss an dem jeweiligen betrachteten Prozess geprüft werden, z. B. wie in [HARTUNG, 2009, S. 690f] beschrieben. Ggf. ist ein Modell des Rauschprozesses zu identifizieren, siehe z. B. [HARTUNG, 2009, S. 678ff].

Das weitere Vorgehen in dieser Arbeit besteht darin, zunächst das geregelte System unter Störeinfluss analytisch zu untersuchen. Damit dies überhaupt bzw. mit vertretbarem Aufwand möglich ist, ist die Annahme eines unkorrelierten Rauschens notwendig. Damit lassen sich Erwartungswert und Varianz entweder geschlossen oder zumindest iterativ über die Regelschritte k bestimmen.

Da auch nicht-lineare Konzepte untersucht werden, wird auch mit Simulationen gearbeitet. Dabei können dann prinzipiell beliebige Prozesse für das Rauschen angesetzt werden. Auch muss das Rauschen nicht mehr weiß sein, sondern kann auch korreliert sein.

Für die Annahme einer Normalverteilung spricht prinzipiell der Zentrale Grenzwertsatz. Eine Formulierung des Zentralen Grenzwertsatzes sagt aus, dass die Summe einer großen Anzahl unabhängiger Zufallsvariablen gegen eine Normalverteilung geht, sofern keine der Zufallsvariablen die übrigen „dominiert“. [HÄNSLER, 2001, S. 100], [STORM, 2001, S. 69]

Auch wenn die Normalverteilung nur eine grobe Näherung für das tatsächliche Rauschen darstellen sollte, so geben die Ergebnisse, die unter der Annahme einer Normalverteilung entstanden sind, dennoch einen guten Überblick über die Eigenschaften des Systems unter Rauschen.

4.3 Normierte Darstellung

Eine wesentliche Kenngröße des Rauschens n_y ist dessen Varianz σ_y^2 bzw. Standardabweichung σ_y . Daher bietet es sich an, die Systemgleichung auf die Standardabweichung σ_y des Rauschens zu normieren.

Dabei müssen auch die Regelziele entsprechend normiert werden, was zu sehr anschaulichen Werten führt. So wird die zulässige Toleranz e_{OK} der Regelabweichung nicht mehr absolut angegeben, sondern als dimensionsloser Wert $\varepsilon_{OK} = e_{OK}/\sigma_y$, mit dem der Toleranzbereich als Vielfaches der Standardabweichung ausgedrückt wird.

Die Normierung, die in diesem Abschnitt durchgeführt wird, erleichtert das Verallgemeinern auf andere Prozesse.

⁶ Beim Beispielprozess der Druckmaschine wird eine Kombination von Papier und Farbe als ein Parametersatz aufgefasst, d. h. angenommen, dass die Steigungen θ_1 und θ_2 konstant sind, wobei der Parameter θ_0 aber durchaus schwanken kann. Dieser kann beispielsweise von der Flächendeckung der Druckform oder dem Wassergehalt der Farbe abhängen. Werden mehrere gleichartige Aufträge hintereinander abgearbeitet, dann ist zu erwarten, dass die Vorsteuerung bessere Ergebnisse liefert, als wenn sich die Prozessparameter stark ändern.

Strecke

Teilt man die Gleichungen (2.5) und (2.10) durch die Standardabweichung σ_y^2 des Rauschens, so erhält man mit den Abkürzungen⁷

$$\begin{aligned} n_k &= n_{y,k}/\sigma_y, & d &= d_y/\sigma_y, \\ \eta_k &= y_k/\sigma_y, & \tilde{\eta}_k &= \tilde{y}_k/\sigma_y \end{aligned}$$

sowie

$$\vartheta_p = \theta_p/\sigma_y$$

die normierten Gleichungen

$$\eta_k = \vartheta_{0S} + \vartheta_1 v_k + d. \quad (4.6)$$

und

$$\tilde{\eta}_k = \eta_k + n_k \quad (4.7)$$

bzw. zusammengenommen

$$\tilde{\eta}_k = \vartheta_{0S} + \vartheta_1 v_k + d + n_k. \quad (4.8)$$

Die Regelabweichungen $e_k = y_{\text{soll}} - y_k$ bzw. $\tilde{e}_k = y_{\text{soll}} - \tilde{y}_k$ können ebenfalls normiert werden,

$$\varepsilon_k = e_k/\sigma_y \quad \text{und} \quad \tilde{\varepsilon}_k = \tilde{e}_k/\sigma_y.$$

Regler

Die Eingänge u_i bleiben unverändert, und damit muss für die normierten Gleichungen die Reglerverstärkung über

$$\kappa_I = k_I \cdot \sigma_y$$

angepasst werden, womit sich für das Regelgesetz des I-Reglers mit zeitvarianter Verstärkung in normierter Darstellung

$$u_{1,k} = u_{1,k-1} + \kappa_{I,k-1} \cdot (\eta_{\text{soll}} - \tilde{\eta}_{k-1}) \quad (4.9)$$

ergibt. Wichtiger im Folgenden ist allerdings der Faktor

$$\alpha = k_I \cdot \theta_1 = \kappa_I \cdot \vartheta_1,$$

der sehr häufig auftreten wird. Die, zumindest aus der deterministischen Betrachtung, ideale Reglerverstärkung $k_I = 1/\theta_1$ entspricht $\alpha = 1$.

⁷ Es ist zu beachten, dass die unterschiedlichen Schreibweisen des Buchstaben Theta hier unterschiedliche Bedeutungen haben, $\vartheta_i \neq \theta_i$.

4.4 Verwendete Gleichungen

In diesem Abschnitt werden ausgehend von der Systemgleichung

$$y_k = \theta_{0S} + \theta_1 \cdot (u_{1,k} - u_{1,N}) + d_y \quad (4.10)$$

und dem Regelgesetz (bei dem hier von einem zeitvariablen Verstärkungsfaktor ausgegangen wird)

$$u_{1,k} = u_{1,k-1} + k_{1,k-1} \cdot (y_{\text{soll}} - \tilde{y}_{k-1}) \quad (4.11)$$

verschiedene Formen der Differenzgleichung für den Streckenausgang bzw. die Regelabweichung angeben. Die verschiedenen Formen unterscheiden sich darin, ob der gemessene oder tatsächliche Streckenausgang betrachtet wird.

Aus der Systemgleichung (4.10) folgt

$$y_k - y_{k-1} = \theta_1 \cdot (u_{1,k} - u_{1,k-1}).$$

Die Differenz des Eingangssignals kann dabei durch

$$\begin{aligned} u_{1,k} - u_{1,k-1} &= k_{1,k-1} \cdot (y_{\text{soll}} - \tilde{y}_{k-1}) \\ &= k_{1,k-1} \cdot (y_{\text{soll}} - y_{k-1}) - k_{1,k-1} n_{y,k-1} \end{aligned}$$

ausgedrückt werden, was auf

$$\begin{aligned} y_k - y_{k-1} &= \theta_1 k_{1,k-1} \cdot (y_{\text{soll}} - y_{k-1}) - \theta_1 k_{1,k-1} n_{y,k-1} \\ &= \alpha_{k-1} \cdot (y_{\text{soll}} - y_{k-1}) - \alpha_{k-1} \cdot n_{y,k-1} \end{aligned}$$

führt. Dabei wurde in der zweiten Zeile die Abkürzung $\alpha = k_1 \theta_1$ verwendet. Umstellen ergibt

$$y_k = (1 - \alpha_{k-1}) \cdot y_{k-1} + \alpha_{k-1} \cdot (y_{\text{soll}} - n_{y,k-1}). \quad (4.12)$$

Multipliziert man diese Gleichung mit -1 durch und addiert y_{soll} auf beiden Seiten, so erhält man

$$y_{\text{soll}} - y_k = (1 - \alpha_{k-1}) \cdot (y_{\text{soll}} - y_{k-1}) + \alpha_{k-1} \cdot n_{y,k-1}.$$

Damit ist eine Differenzgleichung für die Regelabweichungen erhalten,

$$e_k = (1 - \alpha_{k-1}) \cdot e_{k-1} + \alpha_{k-1} \cdot n_{y,k-1}. \quad (4.13)$$

Ersetzt man in Gl. (4.13) e_k und/oder e_{k-1} durch den jeweiligen Wert mit Rauschen (\tilde{e}_k bzw. \tilde{e}_{k-1}), so erhält man die Gleichungen

$$\tilde{e}_k = (1 - \alpha_{k-1}) \cdot \tilde{e}_{k-1} + n_{y,k-1} - n_{y,k}, \quad (4.14)$$

und

$$\tilde{e}_k = (1 - \alpha_{k-1}) \cdot e_{k-1} + \alpha_{k-1} \cdot n_{y,k-1} - n_{y,k}. \quad (4.15)$$

Normierte Darstellung

In der normierten Darstellung ergeben sich für (4.13), (4.14) und (4.15) die Gleichungen

$$\varepsilon_k = (1 - \alpha_{k-1}) \cdot \varepsilon_{k-1} + \alpha_{k-1} \cdot n_{k-1}, \quad (4.16)$$

$$\tilde{\varepsilon}_k = (1 - \alpha_{k-1}) \cdot \tilde{\varepsilon}_{k-1} + n_{k-1} - n_k, \quad (4.17)$$

und

$$\tilde{\varepsilon}_k = (1 - \alpha_{k-1}) \cdot \varepsilon_{k-1} + \alpha_{k-1} \cdot n_{k-1} - n_k. \quad (4.18)$$

4.5 Diskussion des Ausgangspunktes für die Analyse

Ausgangspunkt für die Untersuchung des geregelten Prozesses ist meist eine der in Abschnitt 4.4 aufgeführten Gleichungen für ε_k oder $\tilde{\varepsilon}_k$ für den Zeitpunkt $k = 2$. Damit muss ε_1 oder $\tilde{\varepsilon}_1$ als Anfangswert vorgegeben werden und es stellt sich die Frage, welcher der beiden Werte gewählt werden soll.

Aus Sicht der Anwendung wäre es vorteilhaft, mit $\tilde{\varepsilon}_1$ anstelle von ε_1 arbeiten zu können, da letztere Größe nicht gemessen werden kann. Mit der Normierung von Gl. (4.3) und (4.4) gilt

$$\varepsilon_1 = -d \quad (4.19)$$

bzw.

$$\tilde{\varepsilon}_1 = \varepsilon_1 - n_1 = -d - n_1. \quad (4.20)$$

Die Größen ε_1 und $\tilde{\varepsilon}_1$ sind beide Zufallsvariablen, da d und n_1 Zufallsvariablen sind.

Das Ersetzen von ε_1 durch $-d$ ist nicht sinnvoll, da über die Verteilung von d keine sicheren Aussagen gemacht werden können. Daher muss bei dieser Variante ε_1 als bekannt vorausgesetzt werden und die Ergebnisse werden immer von ε_1 abhängen. Da d in den betrachteten Gleichungen nicht auftritt, muss nicht berücksichtigt werden, dass $\{d\}_r \approx \{d|\varepsilon_1\}_r$ ist. Und da das Rauschen n_1 unabhängig von ε_1 ist, gilt $\{n_1\}_r \sim \{n_1|\varepsilon_1\}_r$. Es ist damit unproblematisch, ε_1 als Startwert vorzugeben.

Nachteilig ist aber, dass ε_1 nicht gemessen wird, so dass man nur qualitative Aussagen darüber erhalten kann, wie sich das System bei niedrigen oder hohen Anfangsabweichungen verhält. Es ist nicht möglich, die Analysen zunächst für ein ε_1 durchzuführen, und diese Ergebnisse mit der Argumentation $E\{\tilde{\varepsilon}_1\}_r = \varepsilon_1$ auf die gemessene Größe $\tilde{\varepsilon}_1$ zu übertragen, da das Rauschen n_1 sowohl in $\tilde{\varepsilon}_1 = \varepsilon_1 - n_1$ als auch in den der Analyse zugrundeliegenden Gleichungen vorkommt und so eine Korrelation besteht.

Wählt man dagegen die Variante mit $\tilde{\varepsilon}_1$ als Startwert, dann hätte dies den Vorteil, dass man in Abhängigkeit der gerade gemessenen Regelabweichung die optimale Strategie bestimmen und anwenden könnte. Wählt man für jedes $\tilde{\varepsilon}_1$ die optimale Strategie, so hat man auch insgesamt die beste Strategie.

Allerdings müsste bei gegebenen $\tilde{\varepsilon}_1$ beachtet werden, dass $\tilde{\varepsilon}_1$ von n_1 abhängt, und damit im Allgemeinen $\{n_1\}_r \approx \{n_1|\tilde{\varepsilon}_1\}_r$ gilt. Diese bedingte Verteilung hängt wiederum von d ab, was im Folgenden gezeigt wird. Da über d keine sicheren Aussagen getroffen werden können, kann im Allgemeinen $\{n_1|\tilde{\varepsilon}_1\}_r$ nicht bestimmt werden.

Ausgangspunkt für die Analyse wird daher immer eine gegebene (ungestörte) Anfangsabweichung ε_1 sein.

4.5.1 Bedingte Verteilung

Im Folgenden wird zunächst für den Fall, dass $\{d\}_r$ gleichverteilt ist, die Dichte der bedingten Verteilung $\{n_1|\tilde{\varepsilon}_1\}_r$ für zwei Werte von $\tilde{\varepsilon}_1$ beispielhaft dargestellt. Die zur Bestimmung dieser Verläufe vollzogenen Rechnungen finden sich im Anhang A.2. Im Anschluss wird der Fall betrachtet, dass $\{d\}_r$ normalverteilt ist. Dies hat den Vorteil, dass die Rechnungen einfacher sind, allerdings sind bei der ersten Variante die zu besprechenden Effekte in den Graphen deutlicher zu sehen.

$\{\varepsilon_1\}_r$ gleichverteilt

Das Rauschen $\{n_1\}_r$ wird als standardnormalverteilt angenommen. Für die Störung $\{d\}_r$ werden für dieses Beispiel zwei verschiedene Gleichverteilungen angenommen, die jeweils den Mittelwert von 2,5 besitzen und sich in der Breite der Streuung unterscheiden. Diese werden als „hohe Streuung“ bzw. „geringe Streuung“ bezeichnet:

„Hohe Streuung“: $\{-d\}_r = \{\varepsilon_1\}_r \sim G(2,5, 5)$

„Geringe Streuung“: $\{-d\}_r = \{\varepsilon_1\}_r \sim G(2,5, 0,1)$

In Abbildung 4.2 sind für beide Fälle die Verteilungsdichten von $\{n_1\}_r$ und $\{\varepsilon_1\}_r$, sowie die Dichte der Verteilung $\{\tilde{\varepsilon}_1\}_r = \{\varepsilon_1 - n_1\}_r$ geplottet.

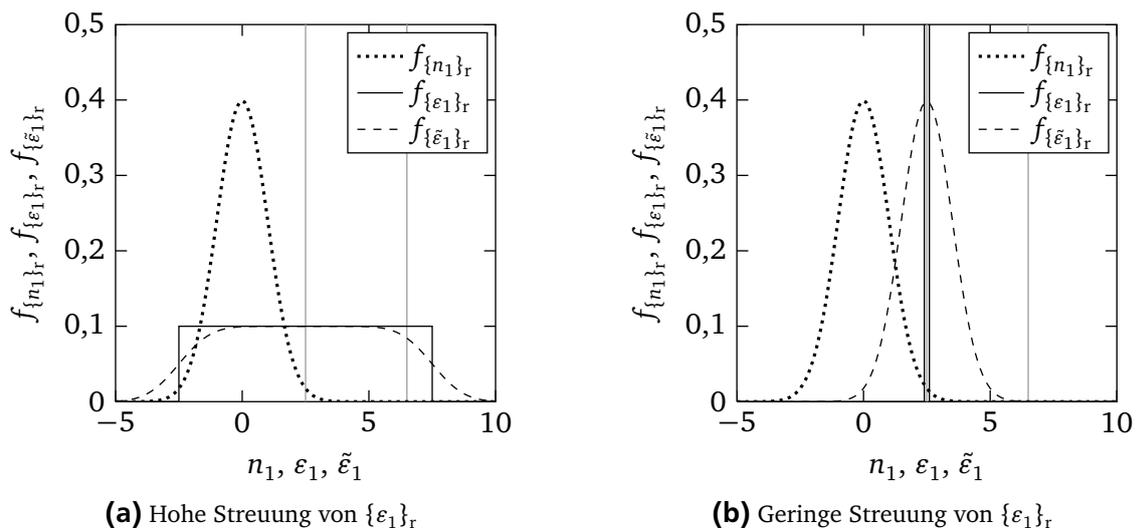


Abbildung 4.2: Verteilungsdichten für $\{n_1\}_r$, $\{\varepsilon_1\}_r$ und $\{\tilde{\varepsilon}_1\}_r$

(Die beiden senkrechten grauen Linien markieren die Punkte $\tilde{\varepsilon}_1 = 2,5$ und $\tilde{\varepsilon}_1 = 6,5$.)

Im Falle der hohen Streuung von $\{\varepsilon_1\}_r$ ändert sich der Verlauf der Dichte von $\{\tilde{\varepsilon}_1\}_r$ deutlich, wohingegen bei der geringen Streuung von $\{\varepsilon_1\}_r$ der Dichteverlauf von $\{\tilde{\varepsilon}_1\}_r$ dem von $\{n_1\}_r$ bis auf eine Verschiebung sehr ähnlich sieht.

In Abbildung 4.3 sind die mit $\tilde{\varepsilon}_1 = 2,5$ und $\tilde{\varepsilon}_1 = 6,5$ bedingten Verteilungen $\{n_1|\tilde{\varepsilon}_1\}_r$ im Vergleich zur unbedingten Verteilung $\{n_1\}_r$ gezeigt.⁸

Bei der hohen Streuung und dem Wert von $\tilde{\varepsilon}_1 = 2,5$, der genau dem Mittelwert von $\{\varepsilon_1\}_r$ entspricht, unterscheiden sich die unbedingte und bedingte Verteilung nicht sichtbar. Geht man aber von $\tilde{\varepsilon}_1 = 6,5$ aus, so zeigt sich, dass der Verlauf der bedingten Verteilung quasi „abgeschnitten“ ist, und sich somit von der unbedingten Verteilung unterscheidet.

⁸ Für die Berechnung siehe Anhang A.2.

Bei der geringen Streuung von $\{\varepsilon_1\}_r$ zeigt sich, dass die bedingte Verteilung $\{n_1|\tilde{\varepsilon}_1\}_r$ für beide Werte für $\tilde{\varepsilon}_1$ ebenfalls kaum streut, wobei der Mittelwert von $\tilde{\varepsilon}_1$ abhängt. Damit unterscheidet sich der Verlauf der bedingten Verteilung deutlich von der unbedingten Verteilung $\{n_1\}_r$.

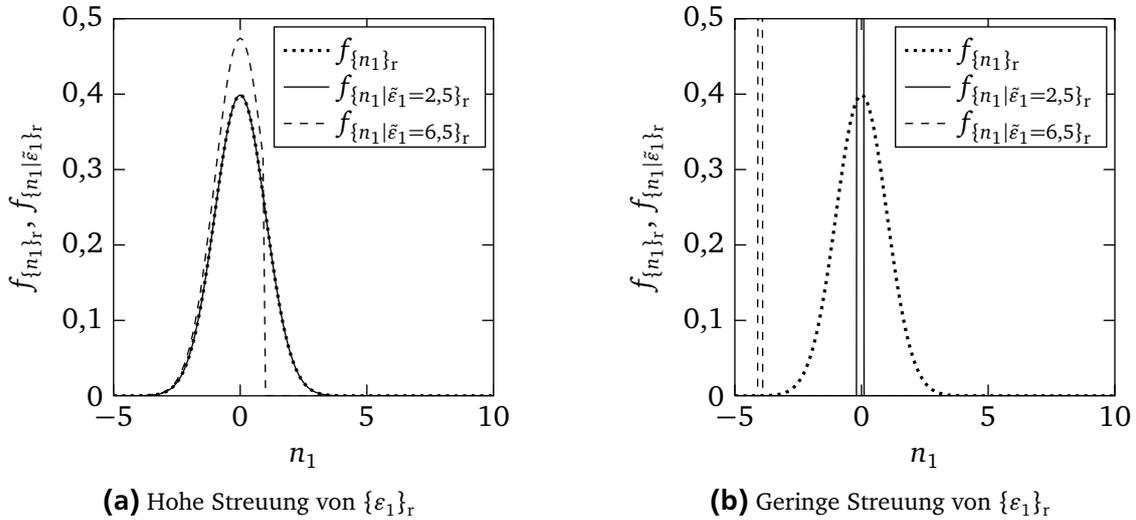


Abbildung 4.3: Vergleich der unbedingten Verteilung $\{n_1\}_r$ mit bedingten Verteilungen $\{n_1|\tilde{\varepsilon}_1\}_r$

$\{\varepsilon_1\}_r$ normalverteilt

Die folgenden Rechnungen nutzen Gleichungen aus [RINNE, 2008, S. 309]. Ein alternativer Rechenweg ist in [KAMEN und SU, 1999, S. 41] angegeben.

Wenn $\{\varepsilon_1\}_r \sim N(\mu_d, \sigma_d)$ und $\{n_1\}_r \sim N(0, \sigma_{n_1}^2)$ gilt⁹, dann ist auch $\{\tilde{\varepsilon}_1\}_r$ normalverteilt mit

$$\{\tilde{\varepsilon}_1\}_r \sim N(\mu_d, \sigma_{n_1}^2 + \sigma_d^2).$$

Um die Eigenschaften der bedingten Verteilung $\{n_1|\tilde{\varepsilon}_1\}_r$ angeben zu können, muss noch der Korrelationskoeffizient ϱ zwischen $\{n_1\}_r$ und $\{\tilde{\varepsilon}_1\}_r$ bekannt sein. Dieser ergibt sich mit den Zwischenschritten

$$\text{Cor}(\{n_1\}_r, \{\tilde{\varepsilon}_1\}_r) = E\{n_1 \cdot \tilde{\varepsilon}_1\}_r = E\{\varepsilon_1 n_1 - n_1^2\}_r = E\{\varepsilon_1\}_r \cdot E\{n_1\}_r - E\{n_1^2\}_r = -E\{n_1^2\}_r = -\sigma_{n_1}^2$$

und $\text{Cov}(\{n_1\}_r, \{\tilde{\varepsilon}_1\}_r) = \text{Cor}(\{n_1\}_r, \{\tilde{\varepsilon}_1\}_r) - \mu_{n_1}\mu_d = -\sigma_{n_1}^2$ zu

$$\varrho = \frac{\text{Cov}(\{n_1\}_r, \{\tilde{\varepsilon}_1\}_r)}{\sigma_{n_1}\sigma_d} = \frac{-\sigma_{n_1}^2}{\sigma_{n_1} \cdot \sqrt{\sigma_{n_1}^2 + \sigma_d^2}}.$$

Die bedingte Dichte zweier korrelierter Normalverteilungen ist wieder die einer Normalverteilung, $\{n_1|\tilde{\varepsilon}_1\}_r \sim N(\mu_{n_1|\tilde{\varepsilon}_1}, \sigma_{n_1|\tilde{\varepsilon}_1}^2)$. Mit den vorliegenden Größen gilt

$$\mu_{n_1|\tilde{\varepsilon}_1} = \varrho \frac{\sigma_{n_1}}{\sigma_d} \cdot (\tilde{\varepsilon}_1 - \mu_d) = -\frac{\sigma_{n_1}^2}{\sigma_{n_1}^2 + \sigma_d^2} \cdot (\tilde{\varepsilon}_1 - \mu_d)$$

und

$$\sigma_{n_1|\tilde{\varepsilon}_1}^2 = \sigma_{n_1}^2 \cdot (1 - \varrho^2) = \frac{\sigma_{n_1}^2 \sigma_d^2}{\sigma_{n_1}^2 + \sigma_d^2}. \quad (4.21)$$

⁹ Auch wenn $\sigma_{n_1}^2 = 1$ definiert ist, bietet es sich an, mit $\sigma_{n_1}^2$ als Variable zu arbeiten, um das Ergebnis besser interpretieren zu können.

Diskussion

Nach Gl. (4.21) wird die Varianz der bedingten Verteilung $\{n_1|\tilde{\varepsilon}_1\}_r$ von der geringeren der Varianzen $\sigma_{n_1}^2$ und σ_d^2 dominiert.

Ist die Varianz der Störung $\{d\}_r$ deutlich kleiner als die des Rauschens $\{n_1\}_r$, $\sigma_d^2 \ll \sigma_{n_1}^2$, so ergibt sich für den Mittelwert $\mu_{n_1|\tilde{\varepsilon}_1} \approx -(\tilde{\varepsilon}_1 - \mu_d)$ und für die Varianz $\sigma_{n_1|\tilde{\varepsilon}_1}^2 \approx \sigma_d^2$. Insbesondere ist für $\sigma_d^2 = 0$, also ein konstantes d , die bedingte Varianz $\sigma_{n_1|\tilde{\varepsilon}_1}^2 = 0$.

Im gegenteiligen Fall, also wenn die Varianz der Störung $\{d\}_r$ deutlich größer als die des Rauschens ist, $\sigma_d^2 \gg \sigma_{n_1}^2$, so gilt $\mu_{n_1|\tilde{\varepsilon}_1} \approx 0$ und $\sigma_{n_1|\tilde{\varepsilon}_1}^2 \approx \sigma_{n_1}^2$. D. h. man nähert sich der unbedingten Verteilung $\{n_1\}_r \sim N(0, \sigma_{n_1}^2)$ an.

Damit erhält man qualitativ dieselben Aussagen, die auch aus dem Beispiel mit dem gleichverteilten $\{\varepsilon_1\}_r$ gewonnen werden konnten, nämlich:

- Für große (gleichmäßige) Streuungen von $\{d\}_r = \{\varepsilon_1\}_r$ nähert sich die bedingte $\{n_1|\tilde{\varepsilon}_1\}_r$ der unbedingten Verteilung $\{n_1\}_r$ an.
- Für geringe Streuungen von $\{d\}_r$ ist auch die Streuung von $\{n_1|\tilde{\varepsilon}_1\}_r$ gering. Falls d konstant ist, ist $\{n_1|\tilde{\varepsilon}_1\}_r$ für jedes $\tilde{\varepsilon}_1$ ebenfalls eine Konstante.
- Die bedingte Verteilung hängt vom Wert $\tilde{\varepsilon}_1$ ab.

Die allgemeine Problematik ist die, dass über die Verteilung von $\{d\}_r$ kaum Aussagen getroffen werden können.

4.5.2 Beispiel Bias

Als Beispiel für die Notwendigkeit der Beachtung der bedingten Verteilung wird hier der Bias der Schätzung des Parameters θ_1 bei $k = 2$ Messwerten behandelt. Die Parameterschätzung wird im Kapitel 6 diskutiert, aber dieses Beispiel ist direkt verständlich und bietet sich sehr gut an, den genannten Punkt darzustellen.

Der Schätzwert der Steigung θ_1 bezüglich u_1 bei $k = 2$ ergibt sich einfach über

$$\hat{\theta}_{1,2} = \frac{\tilde{y}_2 - \tilde{y}_1}{u_{1,2} - u_{1,1}}.$$

bzw.

$$\hat{\vartheta}_{1,2} = \frac{\tilde{\eta}_2 - \tilde{\eta}_1}{u_{1,2} - u_{1,1}}$$

in der normierten Form. Ersetzt man in der normierten Darstellung $\tilde{\eta}_1$ und $\tilde{\eta}_2$ entsprechend durch Gl. (4.8), so erhält man

$$\hat{\vartheta}_{1,2} = \frac{\vartheta_1 \cdot (u_{1,2} - u_{1,1})}{u_{1,2} - u_{1,1}} + \frac{n_2 - n_1}{u_{1,2} - u_{1,1}} = \vartheta_1 + \frac{n_2 - n_1}{u_{1,2} - u_{1,1}}.$$

Der erste Summand entspricht dem wahren Wert des Parameters ϑ_1 . Die Änderung der Stellgröße ist über $u_{1,2} - u_{1,1} = k_I \cdot (y_{\text{soll}} - \tilde{y}_1) = \kappa_I \cdot (\eta_{\text{soll}} - \tilde{\eta}_1)$ gegeben, wobei $\kappa_I = k_I \cdot \sigma_y$ ist. Damit gilt auch $u_{1,2} - u_{1,1} = \kappa_I \cdot (\eta_{\text{soll}} - \eta_1 - n_1)$ und $\hat{\vartheta}_{1,2}$ kann durch

$$\hat{\vartheta}_{1,2} = \vartheta_1 + \frac{n_2}{\kappa_I \cdot (\eta_{\text{soll}} - \eta_1 - n_1)} - \frac{n_1}{\kappa_I \cdot (\eta_{\text{soll}} - \eta_1 - n_1)} \quad (4.22)$$

ausgedrückt werden.

Der Fehler des Schätzwertes ist als

$$\Delta \hat{\vartheta}_{1,2} = \hat{\vartheta}_{1,2} - \vartheta_1$$

definiert. Damit ergibt sich für den Bias als Erwartungswert des Fehlers

$$E\{\Delta \hat{\vartheta}_{1,2}\}_r = E\{n_2\}_r \cdot E\left\{\frac{1}{\kappa_I \cdot (\eta_{\text{soll}} - \eta_1 - n_1)}\right\}_r - E\left\{\frac{n_1}{\kappa_I \cdot (\eta_{\text{soll}} - \eta_1 - n_1)}\right\}_r,$$

wobei die Linearität der Erwartungswertbildung berücksichtigt wurde und auch schon davon Gebrauch gemacht wurde, dass $E\{x \cdot y\}_r = E\{x\}_r \cdot E\{y\}_r$ gilt, wenn $\{x\}_r$ und $\{y\}_r$ unkorreliert sind. Da darüber hinaus $E\{n_k\}_r = 0$ ist, ergibt sich

$$E\{\Delta \hat{\vartheta}_{1,2}\}_r = -E\left\{\frac{n_1}{\kappa_I \cdot (\eta_{\text{soll}} - \eta_1 - n_1)}\right\}_r = -\frac{1}{\kappa_I} \cdot E\left\{\frac{n_1}{\varepsilon_1 - n_1}\right\}_r.$$

Dieser Erwartungswert existiert nicht, da das Integral der dazugehörigen Dichtefunktion nicht konvergiert, was anschaulich darin begründet ist, dass im Nenner $\varepsilon_1 - n_1$ auch den Wert Null annehmen kann [MARSAGLIA, 2006]. In der genannten Quelle ist aber auch beschrieben, dass, wenn der Mittelwert von $\{n_1\}_r$ weit genug von ε_1 entfernt ist, aus praktischer Sicht dennoch sinnvoll ein Erwartungswert angegeben werden kann. Dabei wird die Verteilung $\{n_1\}_r$ so angepasst, dass $\{n_1\}_r^\omega < \varepsilon_1$ ist.

Dies ist, zumindest für kleine ε_1 , hier nicht möglich. Allerdings gibt es andere Erwägungen, die eine Bestimmung eines „praktischen“ Erwartungswertes sinnvoll zulassen. Und zwar werden, wie später in Kapitel 6 vorgestellt, zum einen nur Schätzwerte für $\hat{\vartheta}_1$ verwendet, deren Varianz eine bestimmte Grenze unterschreiten. Zum anderen können noch absolute Grenzen für die Parameterwerte angegeben werden. In der Simulation und den Versuchen zeigt sich, dass schon die erste Maßnahme dazu führt, zu hohe Abweichungen zwischen realen und geschätzten Parametern effektiv zu verhindern.

Der mit einer hier angedeuteten, und später in Abschnitt 6.6 genauer diskutierten Beschränkung bestimmte Erwartungswert wird mit $\bar{E}\{x\}_r$ bezeichnet, so dass der Bias zu

$$\Delta \hat{\vartheta}_{1,\text{mean},2}(\varepsilon_1) = \bar{E}\{\Delta \hat{\vartheta}_{1,2}\}_r = \frac{1}{\kappa_I} \cdot \bar{E}\left\{\frac{n_1}{n_1 - \varepsilon_1}\right\}_r$$

definiert wird. Solange $\{n_1\}_r$ symmetrisch verteilt ist, d. h. $\{n_1\}_r \sim \{-n_1\}_r$ gilt, ist auch $\Delta \hat{\vartheta}_{1,\text{mean},2}(\varepsilon_1)$ symmetrisch bezüglich ε_1 ,

$$\Delta \hat{\vartheta}_{1,\text{mean},2}(\varepsilon_1) = \Delta \hat{\vartheta}_{1,\text{mean},2}(-\varepsilon_1).$$

Wie oben dargelegt, sind $\{n_1\}_r$ und $\{\varepsilon_1\}_r$ unabhängige Zufallsvariablen, womit $\{n_1|\varepsilon_1\}_r \sim \{n_1\}_r \sim N(0, 1)$ gilt. Dafür ist in Abbildung 4.4 der Verlauf des Bias gezeigt.

Bias bei gegebenem Messwert

Sieht man den Messwert $\tilde{\eta}_1$ als feste Größe an, so ergibt sich die Gleichung

$$\hat{\vartheta}_{1,2} = \vartheta_1 + \frac{n_2}{\kappa_I \cdot (\eta_{\text{soll}} - \tilde{\eta}_1)} - \frac{n_1}{\kappa_I \cdot (\eta_{\text{soll}} - \tilde{\eta}_1)}$$

anstelle von Gl. (4.22). Weiter erhält man

$$E\{\Delta \hat{\vartheta}_{1,2}|\tilde{\varepsilon}_1\}_r = E\{n_2|\tilde{\varepsilon}_1\}_r \cdot \frac{1}{\kappa_I \cdot (\eta_{\text{soll}} - \tilde{\eta}_1)} - E\{n_1|\tilde{\varepsilon}_1\}_r \cdot \frac{1}{\kappa_I \cdot (\eta_{\text{soll}} - \tilde{\eta}_1)}$$

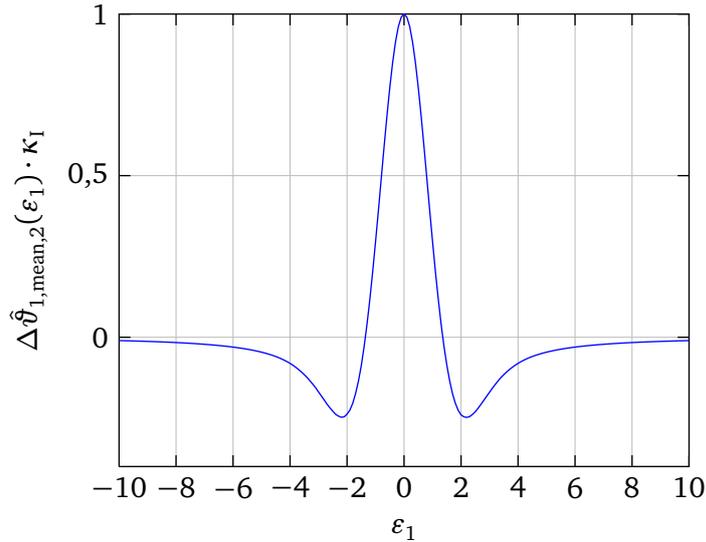


Abbildung 4.4: Bias des ersten Schätzwertes für θ_1 ($k = 2$), normiert auf $\frac{1}{\kappa_1}$

für den Erwartungswert des Bias unter der Voraussetzung, dass $\{\tilde{\varepsilon}_1\}_r = \tilde{\varepsilon}_1$ gilt. $\{n_2\}_r$ ist unabhängig von $\{\tilde{\varepsilon}_1\}_r$ und damit gilt $E\{n_2|\tilde{\varepsilon}_1\}_r = E\{n_2\}_r = 0$, womit sich

$$E\{\Delta\hat{\vartheta}_{1,2}|\tilde{\varepsilon}_1\}_r = -E\{n_1|\tilde{\varepsilon}_1\}_r \cdot \frac{1}{\kappa_1 \cdot \tilde{\varepsilon}_1}$$

ergibt.

Würde man nicht berücksichtigen, dass sich die Verteilung von $\{n_1\}_r$ unter der Bedingung $\{\tilde{\varepsilon}_1\}_r = \tilde{\varepsilon}_1$ wie oben besprochen ändert, so würde man mit $E\{n_1\}_r = 0$ auf das Ergebnis $E\{\Delta\hat{\vartheta}_{1,2}|\tilde{\varepsilon}_1\}_r = 0$ kommen, also dass unabhängig vom ersten Messwert der Bias Null ist.¹⁰

Andererseits wurde auch dargelegt, dass sich $\{n_1|\tilde{\varepsilon}_1\}_r$ der unbedingten Verteilung $\{n_1\}_r$ annähert, je weiter $\{d\}_r$ streut. Dies bedeutet, dass in diesem Fall der Erwartungswert des Bias tatsächlich gegen Null geht.¹¹

4.6 Zusammenfassung

In diesem Kapitel wurden die verwendeten Notationen und Eigenschaften von Zufallsvariablen und Zufallsprozessen kurz zusammenfassend eingeführt. Es wurde dargelegt, welche Eigenschaften für die Störungen $\{d\}_r$ und das Rauschen $\{n_k\}_r$ angenommen werden. Zur leichteren Verallgemeinerung der Ergebnisse wurde eine auf die Standardabweichung des Rauschens normierte Darstellung der Systemgleichung eingeführt.

Zuletzt wurde begründet, dass eine Analyse der Eigenschaften des geregelten Systems unter Berücksichtigung der Störungen und des Rauschens von einem als bekannt vorausgesetzten ε_1 ausgehen muss.

Lediglich für große Schwankungen von d kann auch von $\tilde{\varepsilon}_1$ ausgegangen werden, da näherungsweise $\{n_1\}_r \sim \{n_1|\tilde{\varepsilon}_1\}_r$ gilt, wenn $\text{Var}\{d\}_r \rightarrow \infty$ geht und $\{d\}_r$ beispielsweise normal- oder gleichverteilt ist.

¹⁰ Mit Ausnahme von $\tilde{\varepsilon}_1 = 0$, da in diesem Fall der Schätzwert wegen $u_{1,2} - u_{1,1} = 0$ nicht definiert ist.

¹¹ Für $k = 2$, also den ersten Schätzwert. Für weitere Schätzwerte stellt sich auch unter dieser Voraussetzung ein Bias ein.

5 Regelung des gestörten Systems

In Kapitel 3 wurde die Reglerauslegung für den ungestörten Fall betrachtet. Jetzt wird der Regelkreis aus Abbildung 5.1 untersucht, der durch ein Rauschen n_y gestört ist.

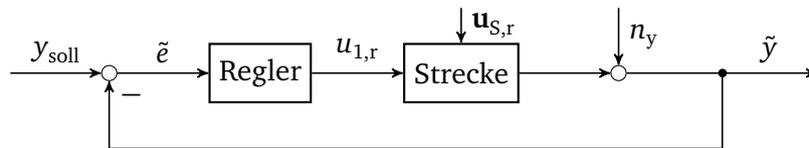


Abbildung 5.1: Regelkreis mit Rauschen

Analog zu dem Vorgehen im ungestörten Fall werden zunächst die Anforderungen an die Regelung in eine dem System und dessen Modellierung angemessene und verwendbare Form gebracht. Dabei werden auch Grenzen aufgezeigt.

Im folgenden Abschnitt 5.2 werden die Standardverfahren zur Regelung stochastischer Systeme kurz vorgestellt. Es wird dargelegt, wieso diese für den speziellen vorliegenden Fall nicht direkt geeignet sind. Jedoch werden grundlegende Ideen übernommen, um ein auf die konkrete Problemstellung angepasstes neues Regelungskonzept zu entwickeln.

Im Hauptteil dieses Kapitels wird der aus Kapitel 3 bekannte I-Regler als Ausgangspunkt für den Reglerentwurf verwendet. Zum einen werden Grenzen für den Parameter k_I angegeben, um die Anforderungen zu erfüllen, zum anderen wird untersucht, wie k_I bzw. $k_{I,k}$ gewählt werden muss, um den Mean-Squared-Error des Ausgangs zu minimieren.

Das Kapitel schließt mit verschiedenen Simulationen.

5.1 Regelziele

Die Regelziele bzw. Anforderungen an den Regler werden in zwei Gruppen aufgeteilt. Zum einen wird das Erreichen des OK-Bereichs betrachtet, wozu die schon in Abschnitt 3.1 behandelten Anforderungen in eine dem Vorhandensein von Rauschen angebrachte Form gebracht werden. Zum anderen wird das Verbleiben im OK-Bereich beschrieben.

Hierbei werden die Anforderungen anhand der gemessenen Größen $\tilde{\eta}$ bzw. $\tilde{\varepsilon}$ betrachtet. Dies ist in dem Beispielproblem motiviert, da das Rauschen nicht nur Messfehler, sondern zum größten Teil tatsächliche Abweichungen der Ausgangsgröße beinhaltet. Zudem erfolgt die Auswertung direkt an der Maschine an einzelnen Fertigteilen und nicht anhand eines Mittelwertes mehrerer Teile, welches dem Wert η bzw. ε entsprechen würde.

5.1.1 Erreichen des OK-Bereichs

Maximales Überschwingen

Die Forderung, dass ein maximales Überschwingen nicht überschritten wird, kann unter der Annahme $\tilde{\varepsilon}_1 > 0$ durch die Ungleichung

$$\tilde{\varepsilon}_2 > -\tilde{\varepsilon}_1 \cdot r_{OS,max} \quad (5.1)$$

beschrieben werden (siehe auch Abbildung 5.2). D. h. ist diese Ungleichung immer erfüllt, so ist die Bedingung eingehalten. Je nachdem, welche Art der Verteilung und welche Varianz der Verteilung für das Rauschen vorliegt bzw. angenommen wird, ist so eine „harte“ Grenze nicht sinnvoll. So wird bei der Normalverteilung immer ein (wenn auch vernachlässigbar kleiner) Anteil außerhalb dieser Grenze liegen.

Die Wahrscheinlichkeit dafür, dass eine Realisierung $\{\tilde{\varepsilon}_2\}_r^\omega$ die Bedingung (5.1) erfüllt, ist über

$$p_{OS,OK} = P(\{\tilde{\varepsilon}_2\}_r^\omega > -\{\tilde{\varepsilon}_1\}_r^\omega \cdot r_{OS,max})$$

gegeben.

Um einen gewissen Anteil an Regelvorgängen zuzulassen, bei denen die Bedingung des maximalen Überschwingens nicht eingehalten wird, wird eine neue Bedingung anhand der Wahrscheinlichkeit formuliert,

$$P(\{\tilde{\varepsilon}_2\}_r^\omega > -\{\tilde{\varepsilon}_1\}_r^\omega \cdot r_{OS,max}) \geq p_{min} \quad (5.2)$$

Dabei legt p_{min} den Anteil an Vorgängen fest, bei denen die Bedingung (5.1) eingehalten werden muss.¹ Ist $p_{min} = 1$, so geht (5.2) in (5.1) über.

Die Bedingung (5.2) versagt aber noch, wenn der Anfangswert $\tilde{\varepsilon}_1$ schon sehr nahe an Null liegt. (Im Extremfall müsste ein negatives α gewählt werden, um sich wieder weiter von Null zu entfernen als man zu Beginn war, damit die Bedingung erfüllt werden kann.) Daher wird die absolute Grenze $-\varepsilon_{OK}$ verwendet, wenn der Betrag $|\tilde{\varepsilon}_1 \cdot r_{OS,max}|$ der relativen Grenze kleiner als der OK-Bereich ε_{OK} ist.

Dies ist in Abbildung 5.2 dargestellt. Im linken Fall ist die gemessene Anfangsabweichung so groß, dass die Grenze $-r_{OS,max} \cdot \tilde{\varepsilon}_{1,a}$ betragsmäßig größer als ε_{OK} ist, und daher ist $-r_{OS,max} \cdot \tilde{\varepsilon}_{1,a}$ die relevante Grenze. Im rechten Fall ist $-r_{OS,max} \cdot \tilde{\varepsilon}_{1,b}$ betragsmäßig kleiner als ε_{OK} , weshalb $-\varepsilon_{OK}$ die relevante Grenze darstellt.

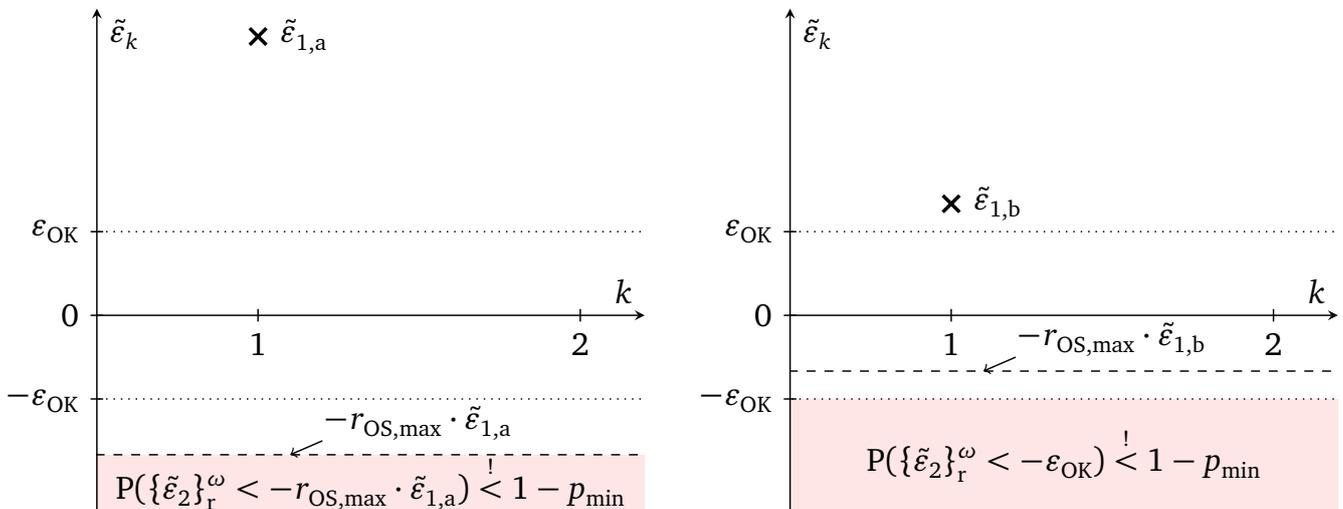


Abbildung 5.2: Zum maximalen Überschwingen

Damit lautet die Bedingung bezüglich des maximalen Überschwingens letztlich

$$\begin{cases} P(\{\tilde{\varepsilon}_2\}_r^\omega > -\{\tilde{\varepsilon}_1\}_r^\omega \cdot r_{OS,max}) \geq p_{min} & \text{für } \tilde{\varepsilon}_1 \cdot r_{OS,max} > \varepsilon_{OK} \\ P(\{\tilde{\varepsilon}_2\}_r^\omega > -\varepsilon_{OK}) \geq p_{min} & \text{sonst,} \end{cases} \quad (5.3)$$

was auch direkt mit bedingten Wahrscheinlichkeiten geschrieben werden könnte:

$$P(\{\tilde{\varepsilon}_2\}_r^\omega > -\{\tilde{\varepsilon}_1\}_r^\omega \cdot r_{OS,max} \mid \tilde{\varepsilon}_1 \cdot r_{OS,max} > \varepsilon_{OK}) \geq p_{min} \\ \vee P(\{\tilde{\varepsilon}_2\}_r^\omega > -\varepsilon_{OK} \mid \tilde{\varepsilon}_1 \cdot r_{OS,max} \leq \varepsilon_{OK}) \geq p_{min} \cdot$$

¹ Hier wird für diese und alle kommenden Forderungen dieselbe Quote p_{min} verlangt, was natürlich nicht zwingend ist.

Minimale Verbesserung

Analoge Überlegungen führen für die minimale Verbesserung auf die Bedingung

$$\begin{cases} P(\{\tilde{\varepsilon}_2\}_r^\omega < \{\tilde{\varepsilon}_1\}_r^\omega \cdot r_{US,\min}) \geq p_{\min} & \text{für } \tilde{\varepsilon}_1 \cdot r_{US,\min} > \varepsilon_{OK} \\ P(\{\tilde{\varepsilon}_2\}_r^\omega < \varepsilon_{OK}) \geq p_{\min} & \text{sonst.} \end{cases} \quad (5.4)$$

Maximale Regelabweichung nach dem k -ten Schritt

Hier wird die Forderung betrachtet, dass nach dem k -ten Regelschritt eine gegebene Regelabweichung $\varepsilon_{\max,k}$ nicht überschritten wird. Die „harte“ Bedingung lautet demnach

$$|\tilde{\varepsilon}_k| < \varepsilon_{\max,k}.$$

Analog zu den oben besprochenen Bedingungen wird diese wieder in einer „weicheren“ Form geschrieben, indem nur mehr gefordert wird, dass ein Anteil von p_{\min} diese Bedingung erfüllen muss:

$$P(|\{\tilde{\varepsilon}_k\}_r^\omega| < \varepsilon_{\max,k}) \geq p_{\min}. \quad (5.5)$$

5.1.2 Verbleiben im OK-Bereich

Der Ausgangswert, der gemessen wird, wenn Regler-OK erreicht ist, wird mit $\tilde{\eta}_{k_{OK}}$ bezeichnet. Dieser Wert entspricht aber nicht dem (mittleren) Wert $\eta_{k_{OK}}$, den das reale System ohne Mess- und Prozessrauschen bei der Eingangsgröße $u_{1,k_{OK}}$ annehmen würde. In Abbildung 5.3 ist beispielhaft gezeigt, dass auch bei einem relativ guten $\tilde{\eta}_{k_{OK}}$ die tatsächliche mittlere Ausgangsgröße $\eta_{k_{OK}}$ weit außerhalb des Toleranzbandes liegen kann. Würde man bei diesem Beispiel weitere Fertigteile vermessen, so würde sich zeigen, dass der Großteil der folgenden Teile nicht OK bzw. \overline{OK} ist.

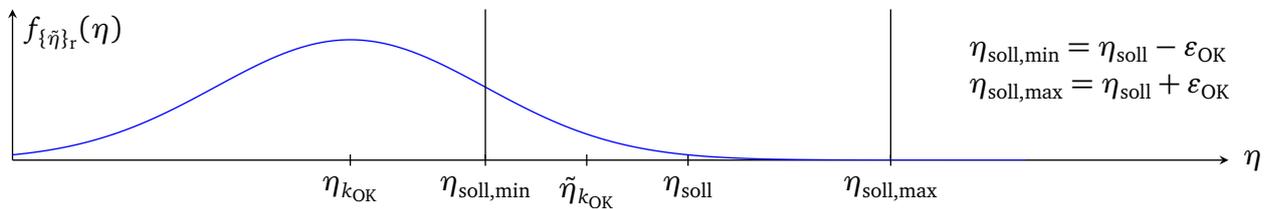


Abbildung 5.3: Zur Erläuterung von $\tilde{\eta}_{k_{OK}}$ und $\eta_{k_{OK}}$

In der Simulation ist $\eta_{k_{OK}}$ bekannt und so lässt sich bei bekannter Verteilung des Rauschens die Wahrscheinlichkeit $p_{\overline{OK}}$ bestimmen, mit der ein weiteres vermessenes Fertigteil \overline{OK} ist,

$$p_{\overline{OK}} = 1 - P(\eta_{soll} - \varepsilon_{OK} \leq \{\tilde{\eta}_k | \eta_{k_{OK}}\}_r^\omega \leq \eta_{soll} + \varepsilon_{OK}).$$

Diese Wahrscheinlichkeit entspricht den in Abbildung 5.4 markierten Flächen.

Unter Annahme einer Standardnormalverteilung $N(0, 1)$ für das Rauschen ergibt sich (siehe auch Anhang A.1)

$$p_{\overline{OK}} = \Phi(\eta_{soll} - \varepsilon_{OK}; \eta_{k_{OK}}, 1) + (1 - \Phi(\eta_{soll} + \varepsilon_{OK}; \eta_{k_{OK}}, 1))$$

bzw. mit den Regelabweichungen² ausgedrückt

$$p_{\overline{OK}} = \Phi(-\varepsilon_{OK}; \varepsilon_{k_{OK}}, 1) + (1 - \Phi(\varepsilon_{OK}; \varepsilon_{k_{OK}}, 1)).$$

² Es ist zu unterscheiden zwischen ε_{OK} , der maximal zulässigen Regelabweichung, und $\varepsilon_{k_{OK}}$, der Regelabweichung bei $k = k_{OK}$.

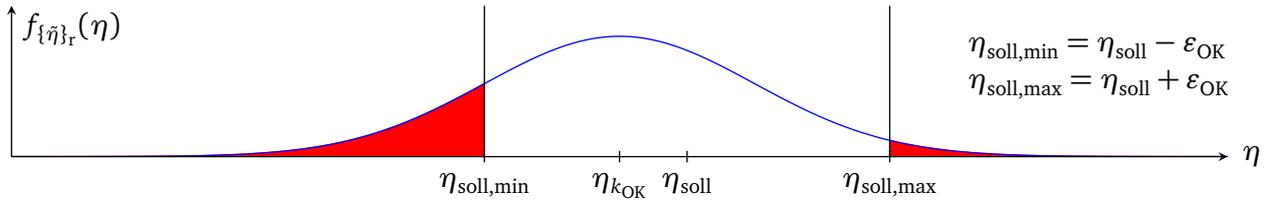


Abbildung 5.4: Zur Erläuterung von $p_{\overline{OK}}$

Der Wert $\eta_{k_{OK}}$ ist nicht deterministisch, sondern stellt eine Realisierung der Zufallsvariablen $\{\eta_{k_{OK}}\}_r$ dar. Entsprechend ist auch $p_{\overline{OK}}$ nur die Realisierung der Zufallsvariablen $\{p_{\overline{OK}}\}_r$.

Für einen gegebenen Testfall und Regler kann die sich ergebende Verteilung für $\{p_{\overline{OK}}\}_r$ simulativ bestimmt werden. (Durch die Abbruchbedingung „OK?“ ist der Prozess unabhängig von seinen sonstigen Eigenschaften nicht-linear.) Diese Verteilung kann grafisch dargestellt werden, wie beispielsweise in Abbildung 5.5.

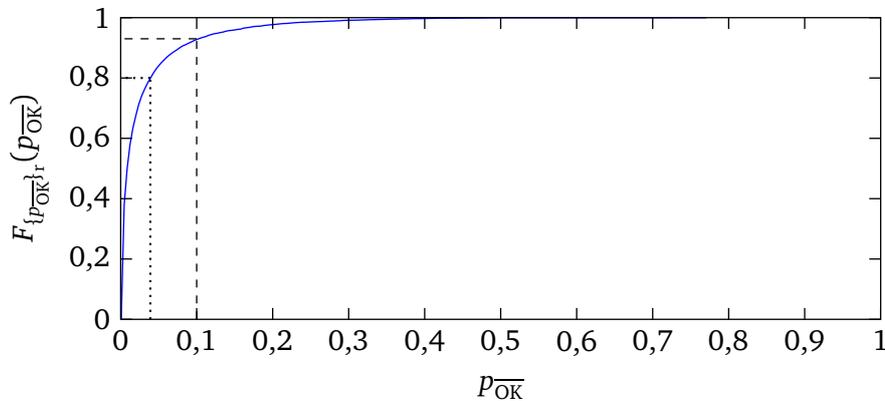


Abbildung 5.5: Beispiel für eine (simulativ bestimmte) Verteilungsfunktion von $\{p_{\overline{OK}}\}_r$

In dieser Abbildung ist mit den gestrichelten Linien beispielhaft für $p_{\overline{OK}} = 10\%$ der Wert $F_{\{p_{\overline{OK}}\}_r}(10\%) \approx 0,93$ abgelesen. Dies bedeutet, dass in 93% der simulierten Einrichtvorgänge nach Erreichen von OK maximal 10% der Fertigteile der Produktionsphase nicht OK sind. Oder andersherum, dass in 7% der Simulationen nach Erreichen von OK mehr als 10% der Fertigteile nicht OK sind. Letztere Aussage wird hier durch die Größe $p_{\overline{OK},>10\%} = 7\%$ ausgedrückt. Allgemein wird hier die Größe

$$p_{\overline{OK},>x\%} = P(\{p_{\overline{OK}}\}_r^\omega > x\%) = 1 - P(\{p_{\overline{OK}}\}_r^\omega \leq x\%) = 1 - F_{\{p_{\overline{OK}}\}_r}(x\%) \quad (5.6)$$

zur Bewertung der Verteilung $\{p_{\overline{OK}}\}_r$ verwendet.

Aus Abbildung 5.5 kann daneben an der gepunkteten Linie auch abgelesen werden, dass in 80% der Fälle $p_{\overline{OK}}$ einen Wert zwischen Null und ca. 4% annimmt. Diese Größe stellt das 80%-Quantil dar und wird hier mit $q_{\overline{OK},80\%}$ bezeichnet. Allgemein ist das Quantil $q_{\overline{OK},x\%}$ über

$$P(\{p_{\overline{OK}}\}_r^\omega \leq q_{\overline{OK},x\%}) = F_{\{p_{\overline{OK}}\}_r}(q_{\overline{OK},x\%}) = x\% \quad (5.7)$$

definiert [STORM, 2001, S. 46].

Diese beiden Größen $p_{\overline{OK},>x\%}$ und $q_{\overline{OK},x\%}$ sind geeignet, die Verteilungsfunktion zu bewerten bzw. um Anforderungen an die Verteilung von $\{p_{\overline{OK}}\}_r$ zu formulieren. So bedeutet

$$p_{\overline{OK},>1\%} \stackrel{!}{\leq} 0,02,$$

dass gefordert wird, dass lediglich nach maximal 2 % der Einrichtvorgänge mehr als 1 % $\overline{\text{OK}}$ -Teile in der Produktionsphase gefertigt werden dürfen. Die Forderung

$$q_{\overline{\text{OK}},99\%} \stackrel{!}{<} 0,01$$

drückt aus, dass nach (mindestens) 99 % der Einrichtvorgänge die $\overline{\text{OK}}$ -Quote der Fertigteile maximal 1 % betragen darf.

Die „Maximalforderung“ nach keinen $\overline{\text{OK}}$ -Teilen kann durch $p_{\overline{\text{OK}},>0\%} \stackrel{!}{=} 0$ bzw. $q_{\overline{\text{OK}},100\%} \stackrel{!}{=} 0$ ausgedrückt werden.

Festlegung von k_{OK}

Der „Zeitpunkt“ k_{OK} bezeichnet jenen Anlauf, der ohne weiteren Stopp in die Produktionsphase übergeht. Standardmäßig wird hierbei der Fall betrachtet, dass der Einregelvorgang wie in Abbildung 2.4 gezeigt abläuft.

Wird demnach ein vermessenes Fertigteil als OK bewertet, so muss ein weiterer Anlauf durchgeführt werden, da die Maschine zwischenzeitlich angehalten wurde. Bei diesem Anlauf wird auch wieder eine Korrektur der Stellgröße durch den Regler vorgenommen. Dieser (weitere) Anlauf wird hier standardmäßig als k_{OK} angesehen.

Allgemeiner könnte aber auch die Fertigung (grundsätzlich, oder zumindest beim letzten Schritt) nicht mehr angehalten werden. Oder es könnten noch bei laufender Maschine Nachführungen der Stellgröße vorgenommen werden.

Bezeichnet man den Schritt, bei dem ein Fertigteil erstmals als OK vermessen wurde, mit k_{OK}^* , so kann allgemeiner

$$k_{\text{OK}} = k_{\text{OK}}^* + \Delta k_{\text{OK}} \tag{5.8}$$

betrachtet werden. Δk_{OK} ist dabei in der Regel eins, wenn nicht explizit anders bestimmt.

Systematische Grenzen der Regelgüte

Minimal erreichbares $p_{\overline{\text{OK}}}$

Der Wert für $p_{\overline{\text{OK}}}$ ist minimal, wenn $\eta_{k_{\text{OK}}} = \eta_{\text{Soll}}$ bzw. $\varepsilon_{k_{\text{OK}}} = 0$ ist. Dieser Wert hängt von der OK-Toleranz und dem Rauschen ab. In Abbildung 5.6 ist das minimal erreichbare $p_{\overline{\text{OK}}}$ über der zulässigen Toleranz ε_{OK} aufgetragen. Bei einer Toleranzgrenze von $\varepsilon_{\text{OK}} = 2$ wäre ein minimales $p_{\overline{\text{OK}}}$ von 4,5 % zu erreichen, bei $\varepsilon_{\text{OK}} = 3$ wäre $p_{\overline{\text{OK}}}$ bestenfalls 0,25 %.

Minimal erreichbares $p_{\overline{\text{OK}}}$ unter Rauschen

Das gerade bestimmte $p_{\overline{\text{OK}},\text{min}}$ gibt die absolute untere Grenze an. Durch das Prozessrauschen, das sich durch die Regelung auch auf ε_k überträgt, wird $\varepsilon_{k_{\text{OK}}}$ aber niemals (sicher) Null sein, so dass dieser Wert nie erreicht werden wird. In diesem Abschnitt werden daher Grenzen angegeben, für die das Rauschen berücksichtigt wird.

Dazu wird zunächst die Frage betrachtet, welche Abweichung $|\varepsilon_k|_{\text{max}}$ maximal zulässig ist, um eine bestimmte Ausschussquote $p_{\overline{\text{OK}},\text{max}}$ bei gegebenen ε_{OK} nicht zu übertreffen. D. h. es wird

$$P(|\{\tilde{\varepsilon}_k\}_r^\omega| > \varepsilon_{\text{OK}}) \stackrel{!}{\leq} p_{\overline{\text{OK}},\text{max}}$$

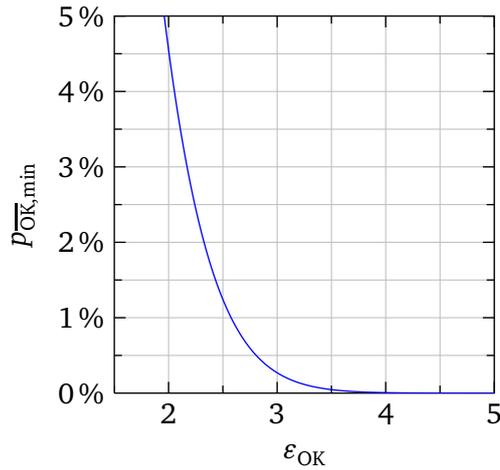


Abbildung 5.6: Minimal erreichbares $p_{\overline{OK}}$

betrachtet. Es gilt hierbei $\{\tilde{\varepsilon}_k\}_r^\omega = \{|\varepsilon_k|_{\max} - n_k\}_r^\omega$, und damit ist $E\{\tilde{\varepsilon}_k\}_r = |\varepsilon_k|_{\max}$ und $\text{Var}\{\tilde{\varepsilon}_k\}_r = 1$. Geht man von einer Normalverteilung aus, so führt dies auf

$$1 - \underbrace{(\Phi(\varepsilon_{OK}; |\varepsilon_k|_{\max}, 1) - \Phi(-\varepsilon_{OK}; |\varepsilon_k|_{\max}, 1))}_{P(\{|\tilde{\varepsilon}_k\}_r^\omega \leq \varepsilon_{OK})} \stackrel{!}{\leq} p_{\overline{OK},\max}, \quad (5.9)$$

was numerisch nach $|\varepsilon_k|_{\max}$ aufgelöst werden kann.

In Abbildung 5.7 sind die sich ergebenden Kurven für verschiedene Werte von $p_{\overline{OK},\max}$ dargestellt. Aus diesen lässt sich beispielsweise ablesen, dass bei einer „Schwankungsbreite“ von $|\varepsilon_k|_{\max} = 2$ und einem $p_{\overline{OK},\max} = 2,5\%$ die Grenze ε_{OK} mindestens 4 betragen muss. Bei $p_{\overline{OK},\max} = 0,1\%$ liegt diese Grenze schon bei 5.

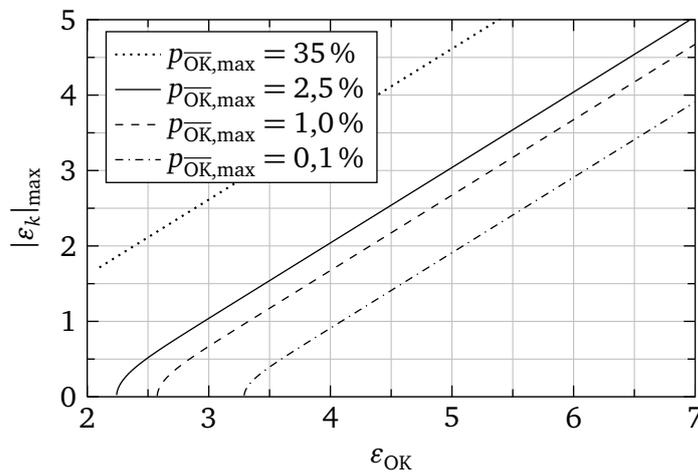


Abbildung 5.7: Maximal zulässige Abweichung $|\varepsilon_k|_{\max}$

Geht man von einer mittelwertfreien Normalverteilung von $\{\varepsilon_{k_{OK}}\}_r$ aus, so kann leicht das Intervall $[-|\varepsilon|_{\max,x\%}, |\varepsilon|_{\max,x\%}]$ bestimmt werden, innerhalb dem $x\%$ aller Werte $\varepsilon_{k_{OK}}$ liegen.³ Mit dem Wert $|\varepsilon|_{\max,x\%}$ kann dann aus Abbildung 5.7 bei gegebenem ε_{OK} der Wert des Quantils $q_{\overline{OK},x\%}$ abgelesen werden. Da $\{\varepsilon_{k_{OK}}\}_r$ im Allgemeinen nicht mittelwertfrei ist, stellt dieser Wert eine untere Grenze für den sich tatsächlich einstellenden Wert dar.

³ $|\varepsilon|_{\max,x\%}$ kann durch numerisches Lösen der Gleichung $-1 + 2 \cdot \Phi(|\varepsilon|_{\max,x\%}; 0, \text{Var}\{\varepsilon_{k_{OK}}\}_r) - x\% = 0$ bestimmt werden, Gl. (A.2).

Als Beispiel wird von $\varepsilon_{OK} = 3$ und $\text{Var}\{\varepsilon_{k_{OK}}\}_r = 1$ ausgegangen. Es liegen 99 % der Werte in dem Intervall $[-2,58, 2,58]$. In Abbildung 5.7 wird $|\varepsilon_k|_{\max} = 2,6$ bei $\varepsilon_{OK} = 3$ von der Kurve für $p_{\overline{OK},\max} = 35\%$ geschnitten. Das bedeutet, dass in 99 % der Einrichtvorgänge ein Wert für $p_{\overline{OK}}$ von maximal 35 % erreicht wird, bzw. dass in 1 % der Fälle der Wert für $p_{\overline{OK}}$ größer als 35 % ist: $q_{\overline{OK},99\%} \approx 35\%$.

Für das 95%-, 99%- und 99,9%-Quantil sind die erreichbaren Werte über ε_{OK} in Abbildung 5.8a aufgetragen.⁴ Hieraus kann beispielsweise abgelesen werden, dass die Forderung $q_{\overline{OK},99\%} \leq 2\%$ bei $\text{Var}\{\varepsilon_{k_{OK}}\}_r = 1$ nur erfüllt werden kann, wenn $\varepsilon_{OK} > 4,6$ ist.

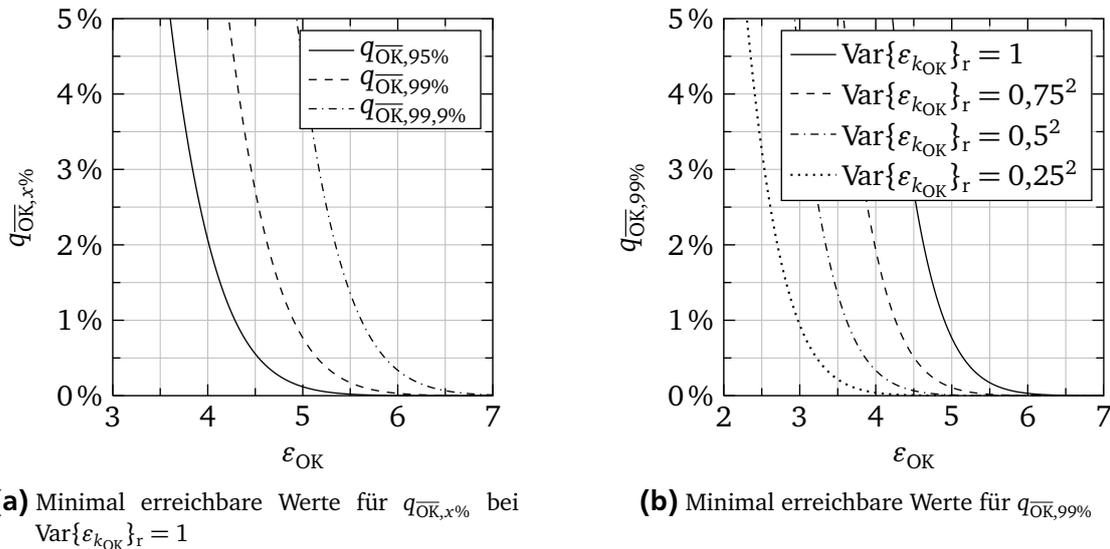


Abbildung 5.8: Zur Erreichbarkeit von $p_{\overline{OK}}$

Zuletzt ist in Abbildung 5.8b die Abhängigkeit der erreichbaren Werte für $q_{\overline{OK},99\%}$ von der Varianz $\text{Var}\{\varepsilon_{k_{OK}}\}_r$ dargestellt.⁵ Hieraus kann z. B. gefolgert werden, dass bei der Forderung $q_{\overline{OK},99\%} \leq 2\%$ und $\varepsilon_{OK} = 4$ die Varianz von $\{\varepsilon_{k_{OK}}\}_r$ maximal $0,75^2$ betragen darf.

Größen zur Bewertung des Reglers

Die Größen zur Bewertung des Reglers sind in Tabelle 5.1 zusammengefasst. Neben $p_{\overline{OK},>x\%}$ und $q_{\overline{OK},x\%}$ sind dort noch der Mittelwert \bar{n}_{RS} und das Quantil $q_{RS,x\%}$ angegeben, welche sich auf die Anzahl der Regelschritte bis Regler-OK beziehen.

5.2 Klassische Ansätze zur Regelung stochastischer Systeme

Dieser Abschnitt gibt einen kurzen Überblick über Standardverfahren aus der Literatur zur Regelung stochastischer Systeme. In Abschnitt 5.2.1 werden diese kurz vorgestellt.

In Abschnitt 5.2.2 wird am Beispiel eines dieser Verfahren die Anwendung gezeigt und dabei erläutert, weshalb diese in der Standardformulierung für das hier betrachtete Problem nicht geeignet sind. Im folgenden Abschnitt dieses Kapitels wird dann, motiviert von diesen Ansätzen, ein neues Regelungskonzept des vorliegenden Systems entwickelt.

⁴ Hierzu wurde für $q_{\overline{OK},95\%}$ in Gl. (5.9) $|\varepsilon_k|_{\max} = |\varepsilon_k|_{\max,95\%}$ gesetzt, womit $p_{\overline{OK},\max}$ gerade $q_{\overline{OK},95\%}$ entspricht: $q_{\overline{OK},95\%} = 1 - (\Phi(\varepsilon_{OK}; |\varepsilon_k|_{\max}, 1) - \Phi(-\varepsilon_{OK}; |\varepsilon_k|_{\max}, 1))$.

⁵ Hierzu wurde $|\varepsilon_k|_{\max,99\%,\sigma_\varepsilon^2}$ durch numerisches Lösen der Gleichung $-1 + 2 \cdot \Phi(|\varepsilon_k|_{\max,99\%,\sigma_\varepsilon^2}; 0, \sigma_\varepsilon^2) - 0,99 = 0$ bestimmt, und dieses dann für $|\varepsilon_k|_{\max}$ in Gl. (5.9) eingesetzt, wodurch $p_{\overline{OK},\max}$ dem gesuchten $q_{\overline{OK},99\%}$ bei $\text{Var}\{\varepsilon_{k_{OK}}\}_r = \sigma_\varepsilon^2$ entspricht.

Tabelle 5.1: Größen zur Bewertung der Einrichtvorgänge durch den Regler

Größe	Beschreibung
\bar{n}_{RS}	Durchschnittliche Anzahl Regelschritte
$q_{RS,x\%}$	Anzahl Regelschritte, nach der mehr als $x\%$ der Anfahrvorgänge abgeschlossen sind. ($q_{RS,100\%}$ ist maximale Anzahl Regelschritte.)
$p_{\overline{OK},>x\%}$	Anteil der Versuche \overline{OK} bei denen nach dem letzten Regelschritt die Wahrscheinlichkeit für \overline{OK} größer als $x\%$ ist. (Gl. (5.6)) ($p_{\overline{OK},>2\%} = 0,05 \rightarrow$ In 5% der Anfahrvorgänge ist zu erwarten, dass mehr als 2% der Fertigteile \overline{OK} sind.)
$q_{\overline{OK},x\%}$	Die Wahrscheinlichkeit für \overline{OK} , die von $x\%$ der Versuche erreicht oder unterschritten wurde. (Gl. (5.7)) ($q_{\overline{OK},99\%} = 0,05 \rightarrow$ In 99% der Anfahrvorgänge ist zu erwarten, dass weniger als 5% der Fertigteile \overline{OK} sind.)

5.2.1 Verfahren

Zur Regelung diskreter Systeme kann von der Differenzgleichung sowie der Zustandsraumdarstellung ausgegangen werden.

Stochastic Optimal Linear Control

Die Regelung stochastischer Systeme im Zustandsraum wird beispielsweise in [MEDITCH, 1969, S. 323ff] behandelt.

Ausgangspunkt ist dabei das lineare zeitvariante System

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{A}_k \mathbf{x}_k + \mathbf{B}_k \mathbf{u}_k + \mathbf{F}_k \mathbf{w}_k \\ \mathbf{y}_k &= \mathbf{C}_k \mathbf{x}_k + \mathbf{v}_k, \end{aligned}$$

wobei \mathbf{w}_k und \mathbf{v}_k jeweils mittelwertfreies weißes Rauschen darstellt.

Das Gütekriterium lautet

$$J^* = E\{J\}_F, \tag{5.10}$$

wobei

$$J = \sum_{k=1}^N (\mathbf{x}_k^T \mathbf{Q}_k \mathbf{x}_k + \mathbf{u}_{k-1}^T \mathbf{R}_{k-1} \mathbf{u}_{k-1}) \tag{5.11}$$

mit symmetrischen, positiv semi-definiten Gewichtsmatrizen \mathbf{Q}_k und \mathbf{R}_{k-1} sowie der betrachteten Anzahl N an Regelschritten ist. [MEDITCH, 1969, S. 326]⁶

⁶ Je nach Autor sind auch andere Formulierungen zu finden, so werden z. B. anstelle der Zustände die Ausgänge bewertet und zusätzlich noch der Wert zum Endzeitpunkt gesondert aufgeführt. Dies ist aber auch alles mit diesem Gütemaß darstellbar.

Seperationsprinzip

Das wesentliche Ergebnis, welches in [MEDITCH, 1969, S. 356] zitiert wird, ist das Seperationsprinzip. Dieses besagt, dass sich die optimale Reglerstruktur, unter den oben gegebenen Voraussetzungen, aus dem am deterministischen System ausgelegten Zustandsregler sowie dem optimalen Filter zusammensetzt, wie in Abbildung 5.9a schematisch dargestellt ist. Damit können beide Teile unabhängig voneinander bestimmt werden.

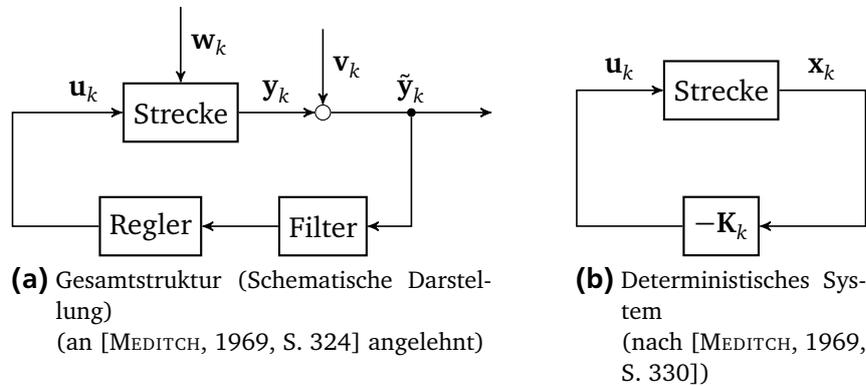


Abbildung 5.9: Zur Linear Stochastic Control

Die Bestimmung der optimalen Zustandsrückführung \mathbf{K}_k erfolgt dabei am deterministischen System, wie in Abbildung 5.9b dargestellt. D. h. die Rückführung

$$\mathbf{u}_k = -\mathbf{K}_k \mathbf{x}_k$$

wird am System

$$\mathbf{x}_{k+1} = \mathbf{A}_k \mathbf{x}_k + \mathbf{B}_k \mathbf{u}_k$$

bezüglich der Gütefunktion (5.11) ausgelegt. Dabei ergibt sich bei einem festen Endzeitpunkt N eine zeitvariante Reglerverstärkung \mathbf{K}_k , die rückwärts bei $k = N - 1$ beginnend bestimmt wird. Dieses \mathbf{K}_k kann mit Hilfe der Gleichungen

$$\mathbf{K}_k = -\left(\mathbf{B}_k^T \mathbf{W}_{k+1} \mathbf{B}_k + \mathbf{R}_k\right)^{-1} \cdot \mathbf{B}_k^T \mathbf{W}_{k+1} \mathbf{A}_k \quad (5.12)$$

$$\mathbf{W}_k = \mathbf{A}_k^T \mathbf{W}_{k+1} \mathbf{A}_k + \mathbf{A}_k^T \mathbf{W}_{k+1} \mathbf{B}_k \mathbf{K}_k + \mathbf{Q}_k \quad (5.13)$$

rekursiv berechnet werden, wobei die Anfangsbedingung $\mathbf{W}_N = \mathbf{Q}_N$ verwendet wird. [MEDITCH, 1969, S. 341]

Für das optimale Filter ergibt sich (unabhängig von der Rückführung \mathbf{K}_k) ein Kalman-Filter mit der zeitvarianten Korrekturmatriix \mathbf{L}_k . Diese kann dabei wie in [MEDITCH, 1969, S. 156ff] beschrieben bestimmt werden.

Die Kombination des optimalen Reglers und des optimalen Filters erfolgt dann so, dass die Rückführung über den Schätzwert $\hat{\mathbf{x}}_k$ des Filters geschlossen wird, d. h. $\mathbf{u}_k = \mathbf{K}_k \hat{\mathbf{x}}_k$.

Weitere Verfahren

Minimum Mean-Squared-Error

Ein Nachteil des bisher vorgestellten Verfahrens ist, dass die Rückführmatrix rückwärts von einem Endzeitpunkt aus bestimmt wird.

In [MEDITCH, 1968] wird ebenfalls für Zustandsraummodelle ein Verfahren vorgestellt, bei dem vorwärts in jedem Schritt der Mean-Squared-Error minimiert wird. Allerdings wird dabei die Voraussetzung gemacht, dass der Anfangswert normalverteilt ist und den Mittelwert Null hat. Damit haben bei der betrachteten Systemklasse auch alle folgenden Zustände den Mittelwert bzw. Erwartungswert Null, womit der Mean-Squared-Error in die Varianz übergeht. Wie im Anschluss noch diskutiert wird, ist die Varianz (alleine) aber kein geeignetes Gütekriterium für den vorliegenden Fall. Jedoch wird die Idee der Minimierung des (richtigen) MSE später wieder aufgenommen werden.

Minimal Variance Control

Bisher wurden Verfahren betrachtet, die sich auf Zustandsraummodelle anwenden lassen. Es existieren aber auch Verfahren, die zur Reglersynthese anhand der Differenzgleichung angewendet werden können, so z. B. das in [ÅSTRÖM, 1970, S. 172ff] vorgestellte „Minimum Variance Control“. Dieses wird mit Erweiterungen, insbesondere der Behandlung von konstanten Störungen und von Null verschiedenen Sollgrößen, auch in [ISERMANN, 1991, S. 13] besprochen.

Bei der Annahme eines weißen Rauschens, das auf den Systemausgang wirkt, wird jeder Minimum Variance Regler Null. D. h. bei weißem Rauschen kann kein Regler die Varianz der Ausgangsgröße reduzieren. [ISERMANN, 1989, S. 19]

Übertragung stochastischer Eigenschaften bei linearen Systemen

Während in diesem Abschnitt bisher Verfahren zur optimalen Regelung von stochastischen Systemen vorgestellt wurden, wird zuletzt noch die Übertragung stochastischer Eigenschaften bei linearen Systemen betrachtet. D. h. es wird die Fragestellung betrachtet, wie sich der Mittelwert und die Varianz einer Eingangs- oder Störgröße auf die Zustände oder den Ausgang eines Systems überträgt. Dabei wird dem Vorgehen aus [BRAMMER und SIFFLING, 1975, S. 220ff] gefolgt. Auf diese Ergebnisse wird später wieder zurückgegriffen werden.

Es wird ein zeitvariantes, diskretes Zustandsraummodell $(\mathbf{A}_k, \mathbf{B}_k, \mathbf{C}_k, \mathbf{D}_k)$ betrachtet. Die Eingangsgröße \mathbf{u}_k ist durch einen, zunächst allgemeinen, Zufallsprozess gegeben, dessen Mittelwerte $\mu_{\mathbf{u}}(k) = E\{\mathbf{u}_k\}_r$ und Kovarianzmatrizen $C_{\mathbf{uu}}(k, m) = E\{(\mathbf{u}_k - \mu_{\mathbf{u}}(k)) \cdot (\mathbf{u}_m - \mu_{\mathbf{u}}(m))^T\}_r$ bekannt seien.

Mittelwerte

Bezüglich der Mittelwerte verhält sich das System genauso wie gegenüber deterministischen Signalen, d. h.

$$\begin{aligned}\mu_{\mathbf{x}}(k+1) &= \mathbf{A}_k \cdot \mu_{\mathbf{x}}(k) + \mathbf{B}_k \cdot \mu_{\mathbf{u}}(k) \\ \mu_{\mathbf{y}}(k) &= \mathbf{C}_k \cdot \mu_{\mathbf{x}}(k) + \mathbf{D}_k \cdot \mu_{\mathbf{u}}(k).\end{aligned}$$

[BRAMMER und SIFFLING, 1975, S. 220, 224]

Kovarianzen

Für die Kovarianzmatrizen der Ein- und Ausgangsgröße gilt allgemein

$$\begin{aligned}C_{\mathbf{xx}}(k+1) &= \mathbf{A}_k \cdot C_{\mathbf{xx}}(k) \cdot \mathbf{A}_k^T + \mathbf{A}_k \cdot C_{\mathbf{xu}}(k) \cdot \mathbf{B}_k^T + \mathbf{B}_k \cdot C_{\mathbf{ux}}(k) \cdot \mathbf{A}_k^T + \mathbf{B}_k \cdot C_{\mathbf{uu}}(k) \cdot \mathbf{B}_k^T \\ C_{\mathbf{yy}}(k) &= \mathbf{C}_k \cdot C_{\mathbf{xx}}(k) \cdot \mathbf{C}_k^T + \mathbf{C}_k \cdot C_{\mathbf{xu}}(k) \cdot \mathbf{D}_k^T + \mathbf{D}_k \cdot C_{\mathbf{ux}}(k) \cdot \mathbf{C}_k^T + \mathbf{D}_k \cdot C_{\mathbf{uu}}(k) \cdot \mathbf{D}_k^T,\end{aligned}$$

[BRAMMER und SIFFLING, 1975, S. 223f]. Zur Berechnung sind neben der Matrix $C_{\mathbf{xx}}(k)$ des vorangehenden Schritts auch die Kenntnis der Kovarianzmatrizen $C_{\mathbf{uu}}(k)$ und $C_{\mathbf{xu}}(k)$ nötig.⁷ Wird der Eingang von

⁷ $C_{\mathbf{ux}}(k) = C_{\mathbf{xu}}^T(k)$

außen vorgegeben und nicht (teilweise) über eine Regelung bestimmt, so hängt $C_{uu}(k)$ nur von den stochastischen Eigenschaften des Eingangsprozesses $\{\mathbf{u}_k\}_r$ ab. Wird eine Regelung unter Störeinfluss $\{\mathbf{n}_k\}_r$ betrachtet, so könnte $C_{uu}(k)$ in Abhängigkeit von $C_{xx}(k)$ oder $C_{yy}(k)$ und $C_{nn}(k)$ angegeben werden. Allerdings ist es in diesem Fall einfacher, zunächst die Zustandsraumdarstellung des geregelten Systems aufzustellen und dann die hier aufgeführten Gleichungen für dieses neue System mit dem Eingang $\mathbf{n}(k)$ anzuwenden.

$C_{xu}(k+1)$ kann über

$$C_{xu}(k+1) = \mathbf{A}_k \cdot C_{xu}(k, k+1) + \mathbf{B}_k \cdot C_{uu}(k, k+1)$$

bestimmt werden, wobei dazu $C_{xu}(k, k+1)$ und $C_{uu}(k, k+1)$ bekannt sein müssen. Dabei ergibt sich $C_{uu}(k, k+1)$ wieder aus dem Eingangsprozess $\{\mathbf{u}_k\}_r$. Ist dieser unkorreliert und mittelwertfrei, so gilt $C_{uu}(k, k+1) \equiv \mathbf{0}$.⁸

Die Matrix $C_{xu}(k, k+1)$ ist ebenfalls identisch Null, wenn $\{\mathbf{x}_k\}_r$ und $\{\mathbf{u}_{k+1}\}_r$ unkorreliert sind. Dies ist sicher der Fall, sofern \mathbf{u}_k von außen vorgegeben wird. Aufgrund der Kausalität kann der Wert \mathbf{u}_{k+1} keinen Einfluss auf \mathbf{x}_k besitzen. Andersherum beeinflusst \mathbf{x} auch nicht \mathbf{u} , wenn \mathbf{u} von außen vorgegeben wird. (Falls \mathbf{u} über einen Regler aus \mathbf{x} bestimmt wird, dann gilt dies nicht!)

Unter der Annahme, dass der Eingangsprozess unkorreliert ist und nicht vom Systemzustand abhängt, vereinfachen sich die Gleichungen für die Kovarianzmatrizen demnach zu

$$\begin{aligned} C_{xx}(k+1) &= \mathbf{A}_k \cdot C_{xx}(k) \cdot \mathbf{A}_k^T + \mathbf{B}_k \cdot C_{uu}(k) \cdot \mathbf{B}_k^T \\ C_{yy}(k) &= \mathbf{C}_k \cdot C_{xx}(k) \cdot \mathbf{C}_k^T + \mathbf{D}_k \cdot C_{uu}(k) \cdot \mathbf{D}_k^T \end{aligned}$$

[BRAMMER und SIFFLING, 1975, S. 223]

5.2.2 Anwendung

Stochastic Optimal Linear Control

Um das Vorgehen nach Abschnitt 5.2.1 anwenden zu können, muss das System in Zustandsform vorliegen. Aufgrund der Anforderungen an den geschlossenen Regelkreis ist auch ein I-Anteil nötig. Ausgangspunkt ist der in Abbildung 5.10a dargestellte Regelkreis, der mit einem PI-Regler geschlossen wurde. Der P-Anteil ist dabei nötig, um das System, [HOLZHÜTER, 2009, S. 9f] folgend, in die in Abbildung 5.10b gezeigte Form bringen zu können.

Aus Abbildung 5.10b können die Gleichungen

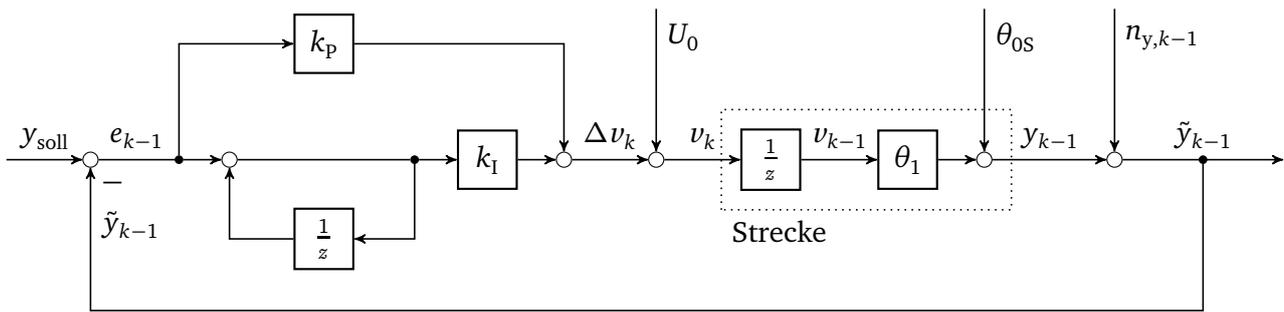
$$\begin{bmatrix} x_{u,k} \\ x_{I,k} \end{bmatrix} = \begin{bmatrix} U_0 + \Delta v_k \\ x_{I,k-1} - \theta_1 \cdot x_{u,k-1} - \theta_{0S} - n_{y,k-1} + y_{\text{soill}} \end{bmatrix}$$

bzw.

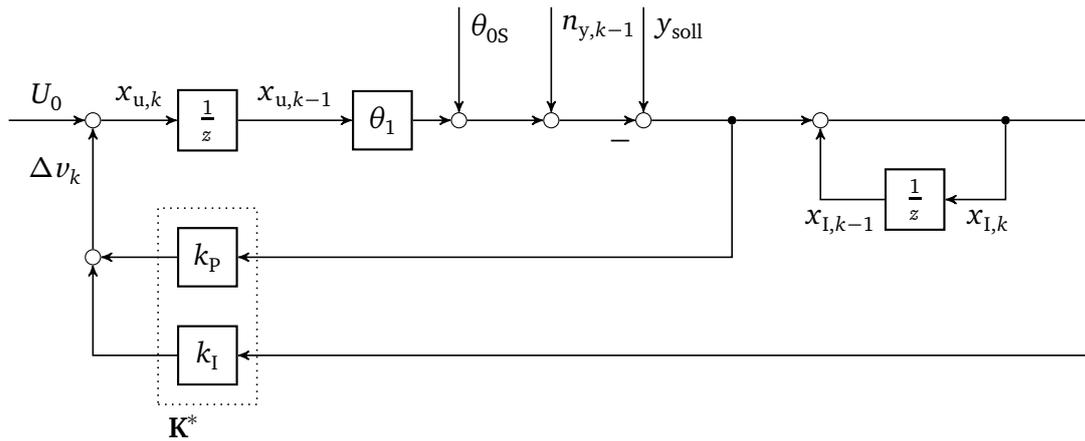
$$\begin{bmatrix} x_{u,k} \\ x_{I,k} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ -\theta_1 & 1 \end{bmatrix} \begin{bmatrix} x_{u,k-1} \\ x_{I,k-1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \Delta v_k + \begin{bmatrix} 1 \\ 0 \end{bmatrix} U_0 - \begin{bmatrix} 0 \\ 1 \end{bmatrix} n_{y,k-1} + \begin{bmatrix} 0 \\ -\theta_{0S} + y_{\text{soill}} \end{bmatrix}$$

abgelesen werden.

⁸ $E\{\mathbf{u}_k \mathbf{u}_{k+1}^T\}_r - \boldsymbol{\mu}_u(k) \boldsymbol{\mu}_u^T(k+1) = E\{\mathbf{u}_k\}_r \cdot E\{\mathbf{u}_{k+1}^T\}_r - \boldsymbol{\mu}_u(k) \boldsymbol{\mu}_u^T(k+1) = 0$, die Unkorreliertheit ist Voraussetzung für das Tauschen des Erwartungswertes und der Multiplikation.



(a) Regelkreis mit PI-Regler



(b) Regelkreis mit PI-Regler in Form einer Zustandsregelung geschrieben

Abbildung 5.10: Blockschaltbilder der mit einem PI-Regler geregelten Strecke

Das Separationsprinzip besagt, dass $n_{y,k-1}$ bei der Bestimmung der Rückführung \mathbf{K}^* weggelassen werden kann, da die Reglerauslegung am deterministischen System geschieht. Wird $n_{y,k-1}$ als Systemrauschen aufgefasst, so muss dieses weiter gar nicht mehr berücksichtigt werden. Die Optimalität der Lösung bleibt dabei erhalten.

Für U_0 ergibt sich aus der Linearisierung des affinen Systems um den Arbeitspunkt $y = y_{\text{soll}}$ der Wert $U_0 = \frac{y_{\text{soll}} - \theta_{0S}}{\theta_1}$. Damit heben sich aber auch gerade die weiteren „Eingänge“ in Abbildung 5.10a bzw. 5.10b auf.

Damit erhält man das linearisierte System

$$\begin{bmatrix} x_{u,k} \\ x_{I,k} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ -\theta_1 & 1 \end{bmatrix} \begin{bmatrix} x_{u,k-1} \\ x_{I,k-1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \Delta v_k.$$

Für dieses System kann ein statischer Zustandsregler $\Delta v_k = \mathbf{K} \mathbf{x}_{k-1}$, wobei \mathbf{K} hier einem liegenden Vektor $\mathbf{K} = \mathbf{k}^T$ entspricht, entworfen werden, aus dem die Verstärkungen des PI-Reglers abgelesen werden können:

$$\mathbf{K} = \begin{bmatrix} k_p/\theta_1 \\ k_i \end{bmatrix}.$$

Da der zweite Zustand den Reglerzustand darstellt, würde dieser im Gütefunktional idealerweise nicht berücksichtigt werden. Denn die Energie, die der Regler benötigt, spielt für diese Anwendung keine Rolle⁹, so dass einzig der Verlauf der ersten Zustandsgröße relevant ist. Wählt man aber entsprechend

$$\mathbf{Q}_k = \begin{bmatrix} q_k & 0 \\ 0 & 0 \end{bmatrix} \quad (5.14)$$

und, aufgrund gleicher Argumentation, auch $\mathbf{R}_{k-1} = 0$, so erhält man aus Gl. (5.12) und (5.13) dass $\mathbf{K}_k \equiv \mathbf{0}$ (und $\mathbf{W}_k = \text{diag}([\sum_{l=k}^N q_k \ 0])$) ist. Der I-Anteil wird damit nicht zurückgeführt und der geschlossene Regelkreis würde damit noch den offenen Integrierer enthalten. (Vgl. dazu auch [LUNZE, 2008, S. 290f] bezüglich der Anforderungen an die Wahl der Gewichtungsmatrix \mathbf{Q} .)

Um dies zu vermeiden, könnte $\mathbf{Q}_k = \text{diag}([q_k \ q_{1,k}])$ gewählt werden. Je größer dabei $q_{1,k}$ ist, umso stärker würde der I-Anteil zurückgeführt werden. Damit ist der Regler jedoch nur optimal bezüglich der neuen, im Sinne des eigentlichen Ziels verfälschter Gewichtungsmatrix $\mathbf{Q}_k = \text{diag}([q_k \ q_{1,k}])$.¹⁰

Interpretation

Mit dem Regler $\mathbf{K}_k \equiv \mathbf{0}$, der sich bei der Gewichtungsmatrix (5.14) ergibt, wird nicht nur der I-Anteil, sondern auch der P-Anteil nicht zurückgeführt. Laut dieser Reglerauslegung wird das Optimum demnach erreicht, wenn der Regelkreis gar nicht geschlossen wird.

Wenn U_0 nicht so gewählt wird, dass es über θ_1 die Größen θ_{0s} und y_{soll} kompensiert, die Vorsteuerung also nicht exakt ist, so bleibt bei θ_{0s} und y_{soll} noch ein „Eingang“ als Störung bestehen. D. h. es handelt sich nicht um eine Anfangsstörung eines Systemzustandes, sondern um eine konstante externe Störung. Auf diese wird der Regler nach dem beschriebenen Verfahren aber nicht ausgelegt.

Bei idealer Kenntnis der Strecke wäre es für das Führungsverhalten natürlich das Beste, überhaupt keinen Regler zu verwenden, sondern den benötigten Eingang einfach vorzusteuern. Und genau das sagt das Ergebnis des Reglerentwurfs aus.

Minimum Variance Control

Bei der Reglerauslegung nach dem Verfahren der Minimum Variance Control erhält man ein gleiches Ergebnis, nämlich dass der Regler Null ist, wenn auch aus anderen Gründen. Beim Stochastic Linear Control ist der deterministische Teil des Reglers Null geworden, während hier das Ergebnis ist, dass die Varianz dann minimal ist, wenn der Regelkreis nicht geschlossen wird [ISERMANN, 1991, S. 19]. Dies wird später auch nochmal gezeigt werden, so dass hier auf die Beschreibung der Berechnung verzichtet wird.

5.2.3 Fazit

Der hier zu entwerfende Regler benötigt einen I-Anteil, was bei dem Entwurf eines Zustandsreglers berücksichtigt werden muss.

Wenn linearisiert wird, so ist der Regler für dieses System unbedingt auch auf das Störverhalten auszulegen, da ein Fehler bei der Bestimmung des Arbeitspunktes für diese Anwendung die wesentliche Fehlerursache ist, in der sich die Störungen widerspiegeln. Alternativ kann der Reglerentwurf direkt am nicht-linearen System durchgeführt werden.

⁹ Die Stellgröße ist am realen Systemen zwar beschränkt, aber alle gewünschten Ausgangswerte sollten mit einer zulässigen Stellgröße erreichbar sein.

¹⁰ Der sich ergebende Regler ist darüber hinaus zeitvariant, so dass vor der Regelung der betrachtete Zeithorizont bekannt sein muss. Aufgrund der wenigen Regelschritte wäre diese Zeitvarianz zu berücksichtigen und es wäre nicht sinnvoll mit $N \rightarrow \infty$ eine konstante Rückführmatrix zu erhalten.

Die Varianz als (alleiniges) Optimierungskriterium zu nehmen, ist bei diesem System nicht zielführend. Dies ist nur sinnvoll, wenn sichergestellt ist, dass der Mittelwert gleich Null ist. Die Varianz würde bei diesem System ansonsten immer minimal sein, wenn der Regelkreis nicht geschlossen wird.

Wie die spätere Diskussion zeigen wird, ist der Ansatz der Minimierung des MSE aber durchaus sinnvoll. Allerdings ist dabei der Mittelwert $\{x_k\}_r$ nicht als Null anzunehmen, wie in den Ansätzen von [ÅSTRÖM, 1970] und [MEDITCH, 1968].

Weiteres Vorgehen

Im Folgenden wird zunächst wieder der in Kapitel 3 beschriebene I-Regler betrachtet. Dabei wird als erstes untersucht, in welchem Bereich eine konstante Reglerverstärkung liegen darf, um die in Abschnitt 5.1.1 formulierten Anforderungen zu erfüllen. Daraufhin werden die Verteilungseigenschaften der stochastischen Ausgangsgröße des geschlossenen Regelkreises betrachtet. Um den MSE zu minimieren, werden verschiedene Variationen des I-Reglers untersucht und besprochen.

5.3 Entwurf anhand der Regelziele bezüglich des Erreichens des OK-Bereichs

In Abschnitt 5.1.1 wurden die Anforderungen an den Regler vorgestellt. In diesem Abschnitt wird der I-Regler aus Kapitel 3 betrachtet und bestimmt, in welchem Bereich eine konstante Reglerverstärkung liegen darf, um diese Anforderungen noch zu erfüllen. Es wird dabei die Anfangsabweichung ε_1 als gegeben vorausgesetzt.

In diesem Abschnitt werden die Zusammenhänge

$$\begin{aligned}\tilde{\varepsilon}_1 &= \varepsilon_1 - n_1 \\ \tilde{\varepsilon}_2 &= (1 - \alpha) \cdot \varepsilon_1 + \alpha n_1 - n_2 \\ \tilde{\varepsilon}_3 &= (1 - \alpha)^2 \cdot \varepsilon_1 + (1 - \alpha) \cdot n_1 + \alpha n_2 - n_3\end{aligned}$$

verwendet, die sich aus der Anwendung der Gleichungen (4.17) und (4.18) ergeben.

Maximales Überschwingen

Die Menge der α , die für ein bestimmtes ε_1 die Bedingung (5.3) erfüllen, wird mit $\mathcal{A}_{OS}(\varepsilon_1)$ bezeichnet¹¹,

$$\mathcal{A}_{OS}(\varepsilon_1) = \left\{ \alpha \left| \begin{array}{ll} \text{P}(\{\tilde{\varepsilon}_2\}_r^\omega > -\{\tilde{\varepsilon}_1\}_r^\omega \cdot r_{OS,max}) \geq p_{min} & \text{für } \tilde{\varepsilon}_1 \cdot r_{OS,max} > \varepsilon_{OK} \\ \text{P}(\{\tilde{\varepsilon}_2\}_r^\omega > -\varepsilon_{OK}) \geq p_{min} & \text{sonst} \end{array} \right. \right\}.$$

Um diese zu berechnen, wird $\text{P}(\{\tilde{\varepsilon}_2\}_r^\omega > -\{\tilde{\varepsilon}_1\}_r^\omega \cdot r_{OS,max})$ zu $\text{P}(\{\tilde{\varepsilon}_2\}_r^\omega + \{\tilde{\varepsilon}_1\}_r^\omega \cdot r_{OS,max} > 0)$ umgestellt. Kürzt man die linke Seite der Ungleichung mit O ab, so erhält man

$$\begin{aligned}O &= \tilde{\varepsilon}_2 + \tilde{\varepsilon}_1 \cdot r_{OS,max} \\ &= (1 - \alpha + r_{OS,max}) \cdot \varepsilon_1 + (\alpha - r_{OS,max}) \cdot n_1 - n_2 > 0.\end{aligned}$$

Damit können für $\{O\}_r$ einfach Erwartungswert und Varianz angegeben werden, $E\{O\}_r = (1 - \alpha + r_{OS,max}) \cdot \varepsilon$ und $\text{Var}\{O\}_r = (\alpha - r_{OS,max})^2 + 1$.

Zur Berechnung muss noch die Bedingung $\tilde{\varepsilon}_1 \cdot r_{OS,max} > \varepsilon_{OK}$ diskutiert werden. Da diese von $\{\tilde{\varepsilon}_1\}_r$ und nicht ε_1 abhängt, kann nicht a priori entschieden werden, welche der beiden Möglichkeiten zu wählen

¹¹ $\tilde{\varepsilon}_1$ hängt von ε_1 ab, und $\tilde{\varepsilon}_2$ hängt von ε_1 und α ab, was der Übersichtlichkeit halber hier nicht explizit angegeben ist.

ist. Jedoch kann diese Bedingung auch so interpretiert werden, dass gerade die zu nehmen ist, bei der die Wahrscheinlichkeit der Erfüllung höher ist.¹²

Damit ergibt sich die gesuchte Menge über

$$\mathcal{A}_{OS} = \mathcal{A}_{OS,1} \cup \mathcal{A}_{OS,2}$$

mit

$$\mathcal{A}_{OS,1} = \{\alpha \mid P(\{O\}_r^\omega > 0) \geq p_{\min}\} \quad \text{und} \quad \mathcal{A}_{OS,2} = \{\alpha \mid P(\{\tilde{\varepsilon}_2\}_r^\omega > -\varepsilon_{OK}) \geq p_{\min}\}.$$

In dem interessierenden Intervall $\alpha \in (0, 2)$ ist dies immer ein zusammenhängendes, nach unten unbeschränktes Gebiet, so dass dieses durch ein Maximum beschrieben werden kann,

$$\alpha_{\max,OS}(\varepsilon_1) = \max \mathcal{A}_{OS}(\varepsilon_1). \quad (5.15)$$

Letztlich berechnet werden die Grenzen durch eine numerische Suche der Nullstellen der Funktionen

$$\begin{aligned} f_{OS1}(\alpha) &= 1 - F_{\{O\}_r}(0) - p_{\min} \quad \text{und} \\ f_{OS2}(\alpha) &= 1 - F_{\{\tilde{\varepsilon}_2\}_r}(-\varepsilon_{OK}) - p_{\min}. \end{aligned}$$

Die Nullstellen entsprechen den α , für die die Bedingung mit genau p_{\min} erfüllt wird.

Minimal geforderte Verbesserung

Ausgehend von der durch die Bedingung (5.4) definierten Menge

$$\mathcal{A}_{US}(\varepsilon_1) = \left\{ \alpha \mid \begin{cases} P(\{\tilde{\varepsilon}_2\}_r^\omega < \{\tilde{\varepsilon}_1\}_r^\omega \cdot r_{US,\min}) \geq p_{\min} & \text{für } \tilde{\varepsilon}_1 \cdot r_{US,\min} > \varepsilon_{OK} \\ P(\{\tilde{\varepsilon}_2\}_r^\omega < \varepsilon_{OK}) \geq p_{\min} & \text{sonst} \end{cases} \right\}$$

erhält man durch eine analoge Vorgehensweise und der Abkürzung

$$\begin{aligned} U &= \tilde{\varepsilon}_2 - \tilde{\varepsilon}_1 \cdot r_{US,\min} \\ &= (1 - \alpha - r_{US,\min}) \cdot \varepsilon_1 + (\alpha + r_{US,\min}) \cdot n_1 - n_2 \end{aligned}$$

die gesuchte Menge

$$\mathcal{A}_{US} = \mathcal{A}_{US,1} \cup \mathcal{A}_{US,2}$$

mit

$$\mathcal{A}_{US,1} = \{\alpha \mid P(\{U\}_r^\omega < 0) \geq p_{\min}\} \quad \text{und} \quad \mathcal{A}_{US,2} = \{\alpha \mid P(\{\tilde{\varepsilon}_2\}_r^\omega < \varepsilon_{OK}) \geq p_{\min}\}.$$

In diesem Fall ist \mathcal{A}_{US} nach oben unbeschränkt, und es gilt daher für die Grenze

$$\alpha_{\min,US}(\varepsilon_1) = \min \mathcal{A}_{US}(\varepsilon_1), \quad (5.16)$$

die wieder durch eine Nullstellensuche der Funktionen

$$f_{US1}(\alpha) = F_{\{U\}_r}(0) - p_{\min} \quad \text{und} \quad f_{US2}(\alpha) = F_{\{\tilde{\varepsilon}_2\}_r}(\varepsilon_{OK}) - p_{\min}$$

berechnet wird.

¹² $P(\{\tilde{\varepsilon}_2\}_r^\omega > a) > P(\{\tilde{\varepsilon}_2\}_r^\omega > b)$ wenn $a < b$.

Maximal zulässige Regelabweichung nach dem k -ten Schritt

Die Bedingung (5.5) wird hier für die ersten beiden Regelschritte, also $k = 2$ und $k = 3$ betrachtet,

$$\mathcal{A}_{k2}(\varepsilon_1) = \left\{ \alpha \mid P(|\{\tilde{\varepsilon}_2\}_r^\omega| < \varepsilon_{\max,k2}) \geq p_{\min} \right\},$$

$$\mathcal{A}_{k3}(\varepsilon_1) = \left\{ \alpha \mid P(|\{\tilde{\varepsilon}_3\}_r^\omega| < \varepsilon_{\max,k3}) \geq p_{\min} \right\}.$$

Mit den Eigenschaften der Verteilung für $\{\tilde{\varepsilon}_2\}_r$ und $\{\tilde{\varepsilon}_3\}_r$, $E\{\tilde{\varepsilon}_2\}_r = (1 - \alpha) \cdot \varepsilon_1$ und $\text{Var}\{\tilde{\varepsilon}_2\}_r = \alpha^2 + 1$ bzw. $E\{\tilde{\varepsilon}_3\}_r = (1 - \alpha)^2 \cdot \varepsilon_1$ und $\text{Var}\{\tilde{\varepsilon}_3\}_r = (1 - \alpha)^2 + \alpha^2 + 1$ können die Mengen bestimmt werden.

In diesen Fällen ist das zulässige Gebiet für α ein zusammenhängendes Intervall, sofern überhaupt für ein gegebenes ε_1 ein zulässiges α existiert, und kann daher mit zwei Grenzen

$$\alpha_{\max,k2}(\varepsilon_1) = \max \mathcal{A}_{k2}(\varepsilon_1) \quad \text{und} \quad (5.17)$$

$$\alpha_{\min,k2}(\varepsilon_1) = \min \mathcal{A}_{k2}(\varepsilon_1) \quad (5.18)$$

bzw.

$$\alpha_{\max,k3}(\varepsilon_1) = \max \mathcal{A}_{k3}(\varepsilon_1) \quad \text{und} \quad (5.19)$$

$$\alpha_{\min,k3}(\varepsilon_1) = \min \mathcal{A}_{k3}(\varepsilon_1), \quad (5.20)$$

beschrieben werden, welche wieder durch die Nullstellen der Funktion

$$f_{k2} = F_{\{\tilde{\varepsilon}_2\}_r}(\varepsilon_{\max,k2}) - F_{\{\tilde{\varepsilon}_2\}_r}(-\varepsilon_{\max,k2}) - p_{\min}$$

(für $k = 3$ entsprechend) gegeben sind.

Gemeinsame Betrachtung aller Bedingungen

Mit den Bedingungen (5.15), (5.17) und (5.19) ist α nach oben begrenzt. Maßgeblich ist dabei der kleinste dieser drei Werte. Entsprechend ist α nach unten durch den Maximalwert von (5.16), (5.18) und (5.20) festgelegt. Somit erhält man

$$\alpha_{\max}(\varepsilon_1) = \min\{\alpha_{\max,OS}(\varepsilon_1), \alpha_{\max,k2}(\varepsilon_1), \alpha_{\max,k3}(\varepsilon_1)\} \quad (5.21)$$

und

$$\alpha_{\min}(\varepsilon_1) = \max\{\alpha_{\min,US}(\varepsilon_1), \alpha_{\min,k2}(\varepsilon_1), \alpha_{\min,k3}(\varepsilon_1)\}. \quad (5.22)$$

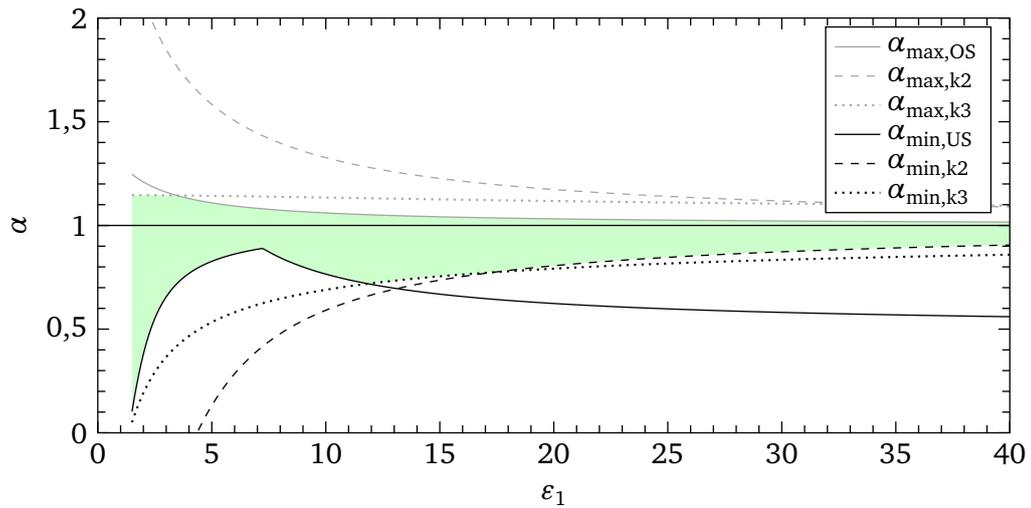
Dadurch ist ein Intervall $[\alpha_{\min}(\varepsilon_1), \alpha_{\max}(\varepsilon_1)]$ gegeben, in dem das tatsächliche α liegen muss, um alle Bedingungen zu erfüllen. In Abbildung 5.11, die im Folgenden besprochen wird, sind die Flächen, die den sich damit ergebenden möglichen Bereichen entsprechen, hervorgehoben.

Diskussion

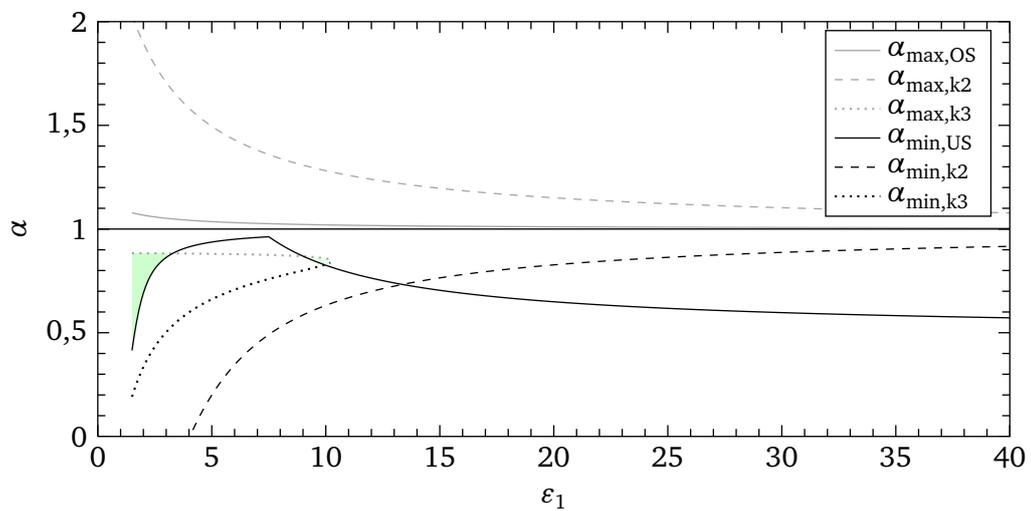
In Abbildung 5.11 sind die sich ergebenden Grenzen für α für verschiedene Parameter dargestellt. Dabei ist immer $r_{OS,\max} = 0$, $r_{US,\min} = 0,5$, $\varepsilon_{\max,k2} = 2\varepsilon_{OK}$ und $\varepsilon_{\max,k3} = \varepsilon_{OK}$ gefordert. Variiert ist die Breite ε_{OK} des OK-Bereichs und p_{\min} .

In Abbildung 5.11a ist ein OK-Bereich von $\varepsilon_{OK} = 3$ und $p_{\min} = 95\%$ gefordert.

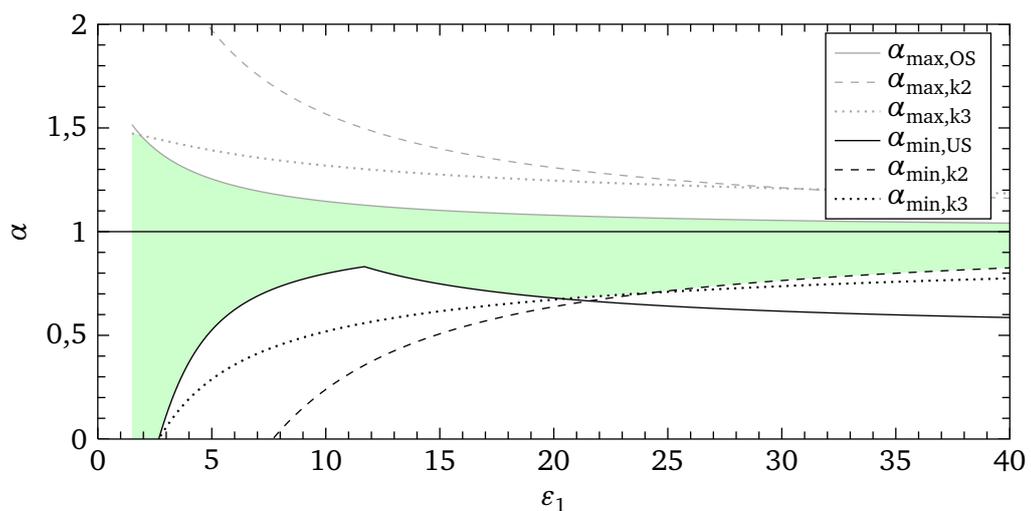
Zunächst fällt auf, dass die Breite des zulässigen Bereichs $[\alpha_{\min}(\varepsilon_1), \alpha_{\max}(\varepsilon_1)]$ stark von dem Anfangswert ε_1 abhängt. Die obere Grenze für α wird – abgesehen im Bereich kleiner ε_1 – von der Forderung dominiert, dass kein Überschwingen auftreten soll. Die untere Grenze wird für kleinere ε_1 von der



(a) $p_{\min} = 95\%$, $\varepsilon_{OK} = 3$



(b) $p_{\min} = 97,5\%$, $\varepsilon_{OK} = 3$



(c) $p_{\min} = 99\%$, $\varepsilon_{OK} = 5$

Abbildung 5.11: Grenzen für α in Abhängigkeit von ε_1 und verschiedene Forderungen

$$r_{OS,\max} = 0, r_{US,\min} = 0,5, \varepsilon_{\max,k2} = 2\varepsilon_{OK}, \varepsilon_{\max,k3} = \varepsilon_{OK}$$

Forderung einer minimalen Verbesserung von 50% und für größere ε_1 von den Forderungen maximal zulässiger Regelabweichungen bei $k = 2$ und $k = 3$ dominiert.

Das Intervall für α , das bis $\varepsilon_1 = 40$ alle Forderungen erfüllt, ist $[0,91, 1,02]$.

Ein OK-Bereich von $\varepsilon_{OK} = 3$ ist schon sehr knapp gewählt, was sich in Abbildung 5.11b zeigt. Dort ist p_{\min} auf 97,5% erhöht. Es zeigt sich, dass die Forderung, dass nach $k = 3$ der OK-Bereich erreicht sein soll, nur für $\varepsilon_1 < 10$ erfüllbar ist. Dabei müsste α aber so gering gewählt werden, dass damit die Bedingung der minimalen Verbesserung meistens nicht mehr erfüllbar ist.

Berücksichtigt man, dass die Standardabweichung nach dem zweiten Regelschritt ($k = 3$) $\sigma_3 = \sqrt{(1 - \alpha)^2 + \alpha^2 + 1}$ beträgt, so lässt sich der geringe zulässige Bereich erklären. Für $\alpha = 1$ ergibt sich $\sigma_3 = \sqrt{2}$, und damit entspricht der OK-Bereich von ± 3 ungefähr dem $\pm 2\sigma$ -Bereich der sich einstellenden Normalverteilung. Im $\pm 2\sigma$ -Bereich um den Mittelwert einer Normalverteilung liegen jedoch nur 95,5% aller Werte, so dass damit auch im Idealfall ($E\{\tilde{\varepsilon}_3\}_r = 0$) die Bedingung mit $p_{\min} = 97,5\%$ nicht erfüllt werden kann. Das minimale σ_3 stellt sich bei $\alpha = 0,5$ ein und beträgt $\sqrt{1,5} \approx 1,2$, womit man sich in einem $\pm 2,5\sigma$ -Bereich befindet, in dem schon knapp 99% aller Werte liegen.

Eine weitere Erhöhung von p_{\min} macht damit keinen Sinn, wenn nicht gleichzeitig der zulässige OK-Bereich vergrößert wird.

In Abbildung 5.11c ist p_{\min} weiter auf 99% erhöht, dafür aber auch ε_{OK} auf fünf gesetzt. Es zeigt sich, dass damit der mögliche Bereich, in dem α liegen kann, deutlich größer wird. Fordert man α unabhängig von ε_1 , so ergibt sich das Intervall $\alpha \in [0,83, 1,04]$.

Unsicherheiten

Im Allgemeinen wird die tatsächliche Streckenverstärkung θ_1 nicht exakt bekannt sein, so dass der Reglerparameter k_1 entsprechen auch nicht exakt so gewählt werden kann, dass sich ein bestimmtes α ergibt. Die relative Abweichung des angenommenen Wertes $\hat{\theta}_1$ wird mit κ_θ bezeichnet, d. h.

$$\hat{\theta}_1 = \theta_1 \cdot \kappa_\theta .$$

Damit gilt

$$\hat{\alpha} = k_1 \hat{\theta}_1 = \alpha \cdot \kappa_\theta$$

für den Zusammenhang zwischen dem gewählten $\hat{\alpha}$ und dem tatsächlich im System wirkenden α . Aus

$$\frac{\hat{\alpha}}{\kappa_\theta} = \alpha \in [\alpha_{\min}, \alpha_{\max}]$$

folgt damit

$$\kappa_\theta \in \left[\frac{\hat{\alpha}}{\alpha_{\max}}, \frac{\hat{\alpha}}{\alpha_{\min}} \right] .$$

Wenn es gleich wahrscheinlich ist, dass man das tatsächliche θ_1 über- wie unterschätzt, dann ist es für eine möglichst gute Robustheit am günstigsten, für $\hat{\alpha}$ den Mittelwert zu wählen, d. h.

$$\hat{\alpha} = \frac{\alpha_{\min} + \alpha_{\max}}{2}$$

zu wählen.¹³ Damit ergibt sich, dass

$$\kappa_\theta \in \left[\frac{1}{2} \cdot \left(\frac{\alpha_{\min}}{\alpha_{\max}} + 1 \right), \frac{1}{2} \cdot \left(\frac{\alpha_{\max}}{\alpha_{\min}} + 1 \right) \right]$$

¹³ Streng genommen ist damit der zulässige Bereich, in dem $\hat{\theta}_1$ um θ_1 liegen darf, nicht ganz symmetrisch bezüglich θ_1 . Aber dies spielt keine größere Rolle.

sein muss.

In Tabelle 5.2 sind die zulässigen Bereiche für κ_θ für zwei verschiedene Forderungen angegeben und damit der zulässige Fehler, den der Schätzwert $\hat{\theta}_1$ haben darf, angegeben.

Tabelle 5.2: Bereich zulässiger Werte für α und κ_θ

Forderung	Zul. Bereich α	$\hat{\alpha}$	Zul. Bereich κ_θ	Zul. Fehler von $\hat{\theta}_1$
$p_{\min} = 95\%$, $\varepsilon_{\text{OK}} = 3$ (Abb. 5.11a)	[0,91, 1,02]	0,965	[0,946, 1,060]	$\approx \pm 5,7\%$
$p_{\min} = 99\%$, $\varepsilon_{\text{OK}} = 5$ (Abb. 5.11c)	[0,83, 1,04]	0,935	[0,899, 1,127]	$\approx \pm 11,4\%$

Fazit

Es lässt sich feststellen, dass der Bereich der zulässigen Werte für α von der Anfangsabweichung ε_1 abhängt. Zwar kann ε_1 nicht gemessen werden, sondern nur $\tilde{\varepsilon}_1$, jedoch enthält auch letzteres eine Information über das (ungefähre) ε_1 . Dies kann bei der Reglerauslegung berücksichtigt werden, indem die Reglerverstärkung von $\tilde{\varepsilon}$ abhängig gemacht wird. Darauf wird später nochmals eingegangen.

Des Weiteren wurde dargelegt, dass die Breite des OK-Bereichs und die Forderung an p_{\min} sinnvoll aufeinander abgestimmt sein müssen.

5.4 Entwurf durch Minimierung des Mean-Squared-Errors

Im vorherigen Abschnitt wurde beschrieben, wie aus den gegebenen Anforderungen für α ein zulässiger Bereich abgeleitet werden kann. Solange der Bereich eine gewisse Breite hat, besteht damit noch ein Freiheitsgrad in der Wahl von α . Zudem muss α auch nicht unbedingt konstant gewählt werden, was weitere Freiheitsgrade ergibt.

Zunächst werden in diesem Abschnitt die Verteilungseigenschaften der Zufallsgrößen $\{\varepsilon_k\}_r$ und $\{\tilde{\varepsilon}_k\}_r$ bei Regelung mit einem konstanten α untersucht. Darauf aufbauend wird begründet, dass der Mean-Squared-Error ein geeignetes Gütekriterium ist, die Freiheitsgrade im Regler festzulegen.

Danach wird die Wahl des Parameters α beim konstanten I-Regler anhand der Minimierung des MSE besprochen. Daraufhin werden noch weitere Varianten des I-Reglers untersucht.

5.4.1 I-Regler mit konstanter Verstärkung k_I – Analyse der Verteilungseigenschaften

Explizite Zeitlösung

Gesucht ist die explizite Lösung η_k des geregelten Systems mit dem Anfangswert η_1 unter Störeinfluss, d. h. eine Darstellung in der Form

$$\eta_k = \eta_k(\eta_1, n_1, n_2, \dots, n_{k-1}).$$

Ausgangspunkt dazu ist die Gleichung (4.12), wobei hier ein konstantes $\alpha_k = \alpha$ angenommen wird,

$$\eta_k = (1 - \alpha)\eta_{k-1} + \alpha \cdot (\eta_{\text{soll}} - n_{k-1}).$$

Daraus ergibt sich (siehe z. B. [LUNZE, 2008, S. 445])

$$\eta_k = (1 - \alpha)^{k-1} \eta_1 + \sum_{k'=1}^{k-1} (1 - \alpha)^{k-k'-1} \alpha \cdot \eta_{\text{soll}} - \sum_{k'=1}^{k-1} (1 - \alpha)^{k-k'-1} \alpha \cdot n_{k'}.$$

In dieser Gleichung beschreibt der erste Summand das Eigenverhalten und die beiden Faltungssummen den Einfluss des „Eingangs“ Sollgröße und des Störeingangs.

Wird angenommen, dass die Sollgröße konstant ist, vereinfacht sich die entsprechende Faltungssumme zu¹⁴

$$\eta_{\text{soll}} \alpha \sum_{k'=1}^{k-1} (1-\alpha)^{k-k'-1} = \eta_{\text{soll}} \alpha \sum_{k'=0}^{k-2} (1-\alpha)^{k'} = \eta_{\text{soll}} \alpha \cdot \frac{1-(1-\alpha)^{k-1}}{1-(1-\alpha)} = \eta_{\text{soll}} - (1-\alpha)^{k-1} \eta_{\text{soll}},$$

und damit lautet die Gleichung für η_k in diesem Fall

$$\eta_k = -(1-\alpha)^{k-1} \cdot (\eta_{\text{soll}} - \eta_1) + \eta_{\text{soll}} - \sum_{k'=1}^{k-1} (1-\alpha)^{k-k'-1} \alpha \cdot n_{k'}. \quad (5.23)$$

Verteilungseigenschaften

Da, wie in Abschnitt 4.5 dargelegt wurde, die Zufallsvariablen $\{\eta_1\}_r$ (bzw. $\{\varepsilon_1\}_r$) und $\{n_k\}_r$ unabhängig sind, ist $\{n_k|\eta_1\}_r = \{n_k\}_r$, und es gilt

$$\{\eta_k\}_r^\omega = -(1-\alpha)^{k-1} \cdot (\eta_{\text{soll}} - \eta_1) + \eta_{\text{soll}} - \sum_{k'=1}^{k-1} (1-\alpha)^{k-k'-1} \alpha \cdot \{n_{k'}\}_r^\omega. \quad (5.24)$$

Für die gemessene Ausgangsgröße $\tilde{\eta}_k$ ergibt sich

$$\{\tilde{\eta}_k\}_r^\omega = \{\eta_k\}_r^\omega + \{n_k\}_r^\omega. \quad (5.25)$$

Dabei ist hervorzuheben, dass in der Gleichung für η_k nur Rauschwerte bis n_{k-1} auftreten, so dass $\{n_k\}_r$ nicht mit $\{\eta_k\}_r$ korreliert ist.

Mit $\{n_k\}_r \sim N(0, 1)$ folgt damit aus Gl. (5.24) der Erwartungswert

$$\begin{aligned} E\{\eta_k\}_r &= -(1-\alpha)^{k-1} \cdot (\eta_{\text{soll}} - \eta_1) + \eta_{\text{soll}} - \sum_{k'=1}^{k-1} (1-\alpha)^{k-k'-1} \alpha \cdot E\{n_{k'}\}_r \\ &= -(1-\alpha)^{k-1} \cdot (\eta_{\text{soll}} - \eta_1) + \eta_{\text{soll}}. \end{aligned} \quad (5.26)$$

sowie weiter aus Gl. (5.25) der Erwartungswert

$$E\{\tilde{\eta}_k\}_r = E\{\eta_k\}_r.$$

Für die Varianz erhält man aus Gl. (5.24)

$$\text{Var}\{\eta_k\}_r = \sum_{k'=1}^{k-1} \left((1-\alpha)^{k-k'-1} \alpha \right)^2 = \alpha^2 \cdot \sum_{k'=1}^{k-1} \left((1-\alpha)^2 \right)^{k-k'-1} = \alpha^2 \cdot \sum_{k'=0}^{k-2} \left((1-\alpha)^2 \right)^{k'}$$

und daraus durch weitere Umformungen

$$\text{Var}\{\eta_k\}_r = \alpha^2 \cdot \sum_{k'=0}^{k-2} \left((1-\alpha)^2 \right)^{k'} = \alpha^2 \cdot \frac{1 - \left((1-\alpha)^2 \right)^{k-1}}{1 - (1-\alpha)^2} = \alpha \cdot \frac{1 - (1-\alpha)^{2k-2}}{2-\alpha}. \quad (5.27)$$

Für die Varianz von $\{\tilde{\eta}_k\}_r$ folgt aus Gl. (5.25), dass diese immer um eins höher ist als $\text{Var}\{\eta_k\}_r$,

$$\text{Var}\{\tilde{\eta}_k\}_r = \text{Var}\{\eta_k\}_r + 1.$$

¹⁴ Dabei wird von der Lösung der endlichen harmonischen Reihe Gebrauch gemacht, siehe z. B. [MERZIGER et al., 2001]. Es kann davon ausgegangen werden, dass $\alpha > 0$ und damit die Bedingung $1 - \alpha \neq 1$ erfüllt ist.

Sonderfälle $\alpha = 1$ und $\alpha = 0$

Vor der weiteren Diskussion werden die Sonderfälle $\alpha = 1$ und $\alpha = 0$ betrachtet. Bei $\alpha = 1$, d. h. der aus deterministischer Sicht optimalen Regelung $k_1 = \frac{1}{\theta_1}$, ergibt sich

$$\text{Var}\{\eta_k\}_r = 1$$

bzw.

$$\text{Var}\{\tilde{\eta}_k\}_r = 2.$$

Diese Werte hängen nicht mehr von k ab, ändern sich also nicht über die Zeit. Die Erwartungswerte für $\{\eta_{\text{soll}} - \eta_k\}_r$ und $\{\eta_{\text{soll}} - \tilde{\eta}_k\}_r$ sind Null. D. h. bei der deterministisch optimalen Regelung wird das Messrauschen am gemessenen Systemausgang $\tilde{\eta}_k$ mit dem Faktor $\sqrt{2}$ verstärkt.

Bei $\alpha = 0$, also keiner Regelung, ergibt sich

$$\text{Var}\{\eta_k\}_r = 0$$

und

$$\text{Var}\{\tilde{\eta}_k\}_r = 1.$$

Der wahre Wert η_k schwankt damit gar nicht mehr und am Ausgang ist gerade das Rauschen n_k zu sehen. Dieser Fall stellt auch die minimale Varianz dar, aber ist natürlich unbrauchbar, da die Regelabweichung im Mittel nicht abgebaut wird, $E\{\eta_{\text{soll}} - \eta_k\}_r = E\{\eta_{\text{soll}} - \tilde{\eta}_k\}_r = \eta_{\text{soll}} - \eta_1$.

Grenzwert

Für andere Werte als 0 und 1 für α ist die Varianz nicht mehr unabhängig von k , konvergiert aber gegen einen von α abhängigen Grenzwert. Dieser lautet

$$\lim_{k \rightarrow \infty} \text{Var}\{\eta_k\}_r = \alpha^2 \cdot \lim_{k \rightarrow \infty} \sum_{k'=0}^{k-2} ((1-\alpha)^2)^{k-k'} = \alpha^2 \cdot \frac{1}{1-(1-\alpha)^2} = \frac{\alpha}{2-\alpha} \quad (5.28)$$

bzw.

$$\lim_{k \rightarrow \infty} \text{Var}\{\tilde{\eta}_k\}_r = 1 + \frac{\alpha}{2-\alpha}$$

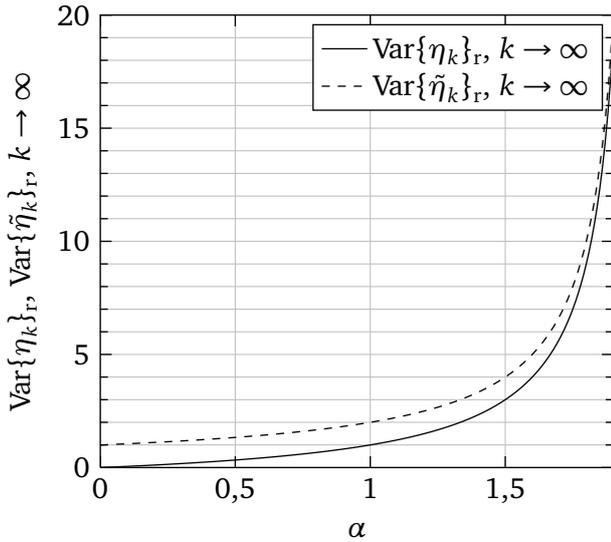
und ist in Abbildung 5.12a über α dargestellt. In Abbildung 5.12b ist die Standardabweichung, also die Wurzel der Varianz gezeigt.

Für $\alpha \rightarrow 2$ geht der Grenzwert gegen Unendlich, da ab $\alpha = 2$ der Regelkreis instabil ist, und damit alle durch das Rauschen auftretenden Störungen mit der Zeit unendlich verstärkt werden.

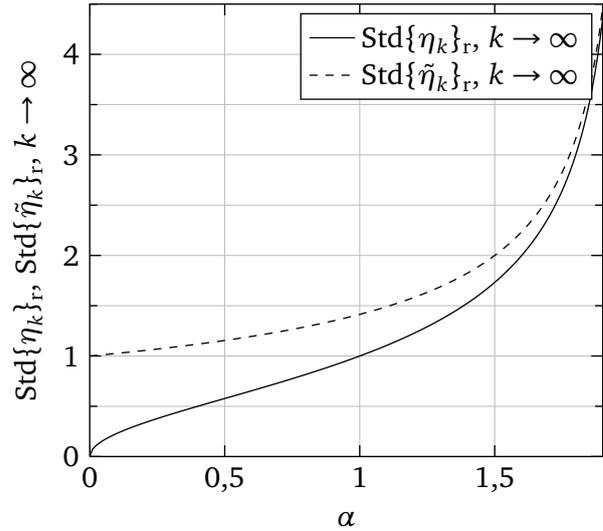
5.4.2 Wahl des MSE als Gütekriterium

In Abschnitt 5.1.2 wurde der Zusammenhang zwischen dem festgelegten Toleranzbereich ε_{OK} und den Forderungen an $q_{\text{OK},x\%}$, sowie der daraus resultierenden maximal zulässigen Varianz $\text{Var}\{\varepsilon_{k_{\text{OK}}}\}_r$ dargelegt.

Die Grenzen in Abbildung 5.8b auf Seite 49 ergeben sich dadurch, dass $E\{\varepsilon_{k_{\text{OK}}}\}_r = 0$ angenommen wurde. (Bei einer größeren mittleren Abweichung würde auch eine Varianz von Null nicht viel helfen, die Regelziele zu erreichen.) In Abbildung 5.7 ist der Zusammenhang zwischen $p_{\text{OK,max}}$, ε_{OK} und der



(a) Varianz



(b) Standardabweichung

Abbildung 5.12: Varianz und Standardabweichung für $k \rightarrow \infty$

maximal zulässigen Abweichung $|\varepsilon_{k_{OK}}|_{\max}$ dargestellt. Daraus kann man ablesen, wie eine niedrigere maximale Abweichung $|\varepsilon_{k_{OK}}|_{\max}$ die Regelgüte verbessert.

In $|\varepsilon_{k_{OK}}|_{\max}$ fließen sowohl die Varianz als auch der Mittelwert von $\{\varepsilon_{k_{OK}}\}_r$ ein. Daher ist es sinnvoll, den Mean-Squared-Error

$$\text{MSE}\{\varepsilon_k\}_r = (\text{E}\{\varepsilon_k\}_r)^2 + \text{Var}\{\varepsilon_k\}_r$$

als Gütekriterium zur Reglerauslegung heranzuziehen.

Es wird im weiteren Abschnitt wie folgt vorgegangen. Zunächst wird ein I-Regler mit konstanter Verstärkung k_I bzw. α betrachtet und abgeleitet, welcher konstante Wert für α optimal im Sinne des MSE wäre.

Danach wird ein I-Regler mit variablen α_k sowie die Erweiterung des Regelkreises bzw. des Reglers mit einem Filter untersucht. Als letzte Variante wird ein I-Regler mit einem von der gemessenen Regelabweichung abhängigen $\alpha(\tilde{\varepsilon}_k)$ behandelt.

Zuletzt werden die in Tabelle 5.1 auf Seite 50 aufgeführten Größen durch numerische Simulationen bestimmt und anhand dieser die verschiedenen Regler abschließend verglichen. Dabei wird sich die Wahl des MSE von $\{\varepsilon_{k_{OK}}\}_r$ als Gütekriterium auch nochmals rechtfertigen.

5.4.3 I-Regler mit konstanter Verstärkung k_I – Bestimmung der optimalen Verstärkung

Wie im letzten Abschnitt besprochen, ist es nicht sinnvoll, den Reglerparameter so zu optimieren, dass die Varianz des tatsächlichen oder des gemessenen Rauschens minimal wird.

Ein Maß, welches die mittlere Abweichung zu einem (beliebigen) Punkt bewertet, ist der Mean-Squared-Error, MSE. Für die Regelabweichung $\eta_{\text{soll}} - \eta_k$ lautet dieser hier mit Gl. (5.26) und (5.27)

$$\begin{aligned} \text{MSE}\{\eta_{\text{soll}} - \eta_k\}_r &= (\text{E}\{\eta_{\text{soll}} - \eta_k\}_r)^2 + \text{Var}\{\eta_k\}_r \\ &= (1 - \alpha)^{2k-2} \cdot (\eta_{\text{soll}} - \eta_1)^2 + \alpha \cdot \frac{1 - (1 - \alpha)^{2k-4}}{2 - \alpha}. \end{aligned} \quad (5.29)$$

Für $\eta_{\text{soll}} - \tilde{\eta}_k$ ist der Wert immer um 1 höher, $\text{MSE}\{\eta_{\text{soll}} - \tilde{\eta}_k\}_r = \text{MSE}\{\eta_{\text{soll}} - \eta_k\}_r + 1$. Damit gilt

$$\arg \min_{0 \leq \alpha < 2} \text{MSE}\{\eta_{\text{soll}} - \tilde{\eta}_k\}_r = \arg \min_{0 \leq \alpha < 2} \text{MSE}\{\eta_{\text{soll}} - \eta_k\}_r$$

und

$$\min_{0 \leq \alpha < 2} \text{MSE}\{\eta_{\text{soll}} - \tilde{\eta}_k\}_r = \min_{0 \leq \alpha < 2} \text{MSE}\{\eta_{\text{soll}} - \eta_k\}_r + 1.$$

Daher reicht es aus, das Minimum von $\text{MSE}\{\eta_{\text{soll}} - \eta_k\}_r$ zu betrachten.

In Gl. (5.29) wird dazu noch $\eta_{\text{soll}} - \eta_1$ durch ε_1 ersetzt, womit man mit

$$\text{MSE}\{\eta_{\text{soll}} - \eta_k\}_r = (1 - \alpha)^{2k-2} \cdot \varepsilon_1^2 + \alpha \cdot \frac{1 - (1 - \alpha)^{2k-4}}{2 - \alpha}$$

eine Darstellung erhält, die nur noch von den bekannten Abkürzungen α und ε_1 sowie der Anzahl der Schritte k abhängt.

Das optimale α hängt damit auch von ε_1 und k ab und wird als

$$\alpha_{\text{opt}}(\varepsilon_1, k) = \arg \min_{\alpha \in [0, 2)} \text{MSE}\{\eta_{\text{soll}} - \eta_k\}_r$$

bezeichnet.

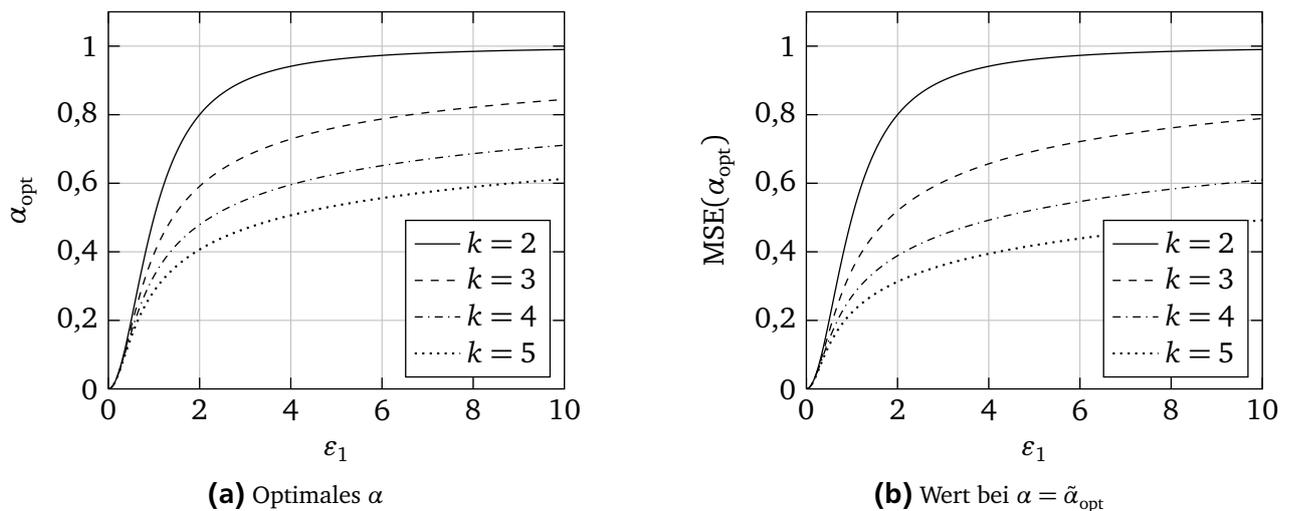


Abbildung 5.13: Optimales α in Abhängigkeit von ε_1 und verschiedenen Werten für k

In Abbildung 5.13a sind die optimalen Werte für α und in Abbildung 5.13b die sich damit ergebenden optimalen Werte des MSE angegeben. Den Graphen sind zwei wesentliche Punkte zu entnehmen. Zum einen ist bei kleinen ε_1 , also bei kleinen Anfangsabweichungen, die Wahl eines geringeren Wertes für α sinnvoll, während bei größeren Abweichungen der optimale Wert für α gegen eins geht. Zum anderen hängt α auch von der Wahl des „Zeitpunktes“ k ab, an dem das Optimum betrachtet wird. Je später dies der Fall ist, desto kleiner ist α zu wählen. Dies liegt darin begründet, dass bei den späteren Regelschritten die Varianz aufgrund eines hohen α dominiert.

In Abbildung 5.14 ist der Verlauf des MSE über α für verschiedene Zeitpunkte aufgetragen. Dabei sind auch die Werte von α markiert, die für die dargestellten „Zeiten“ k optimal sind.

In der oberen Abbildung 5.14a ist der Fall $\varepsilon_1 = 2$ betrachtet. In diesem lässt sich gut erkennen, welchen schlechteren MSE man für die ersten Schritte in Kauf nehmen muss, wenn α bezüglich eines späteren Schrittes $k > 2$ als optimal ausgelegt wird.

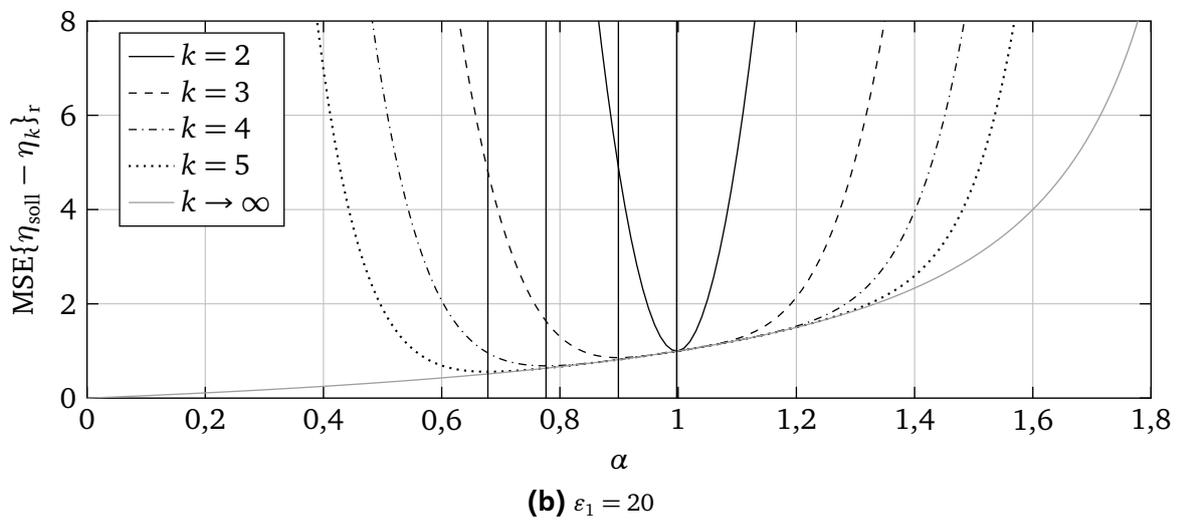
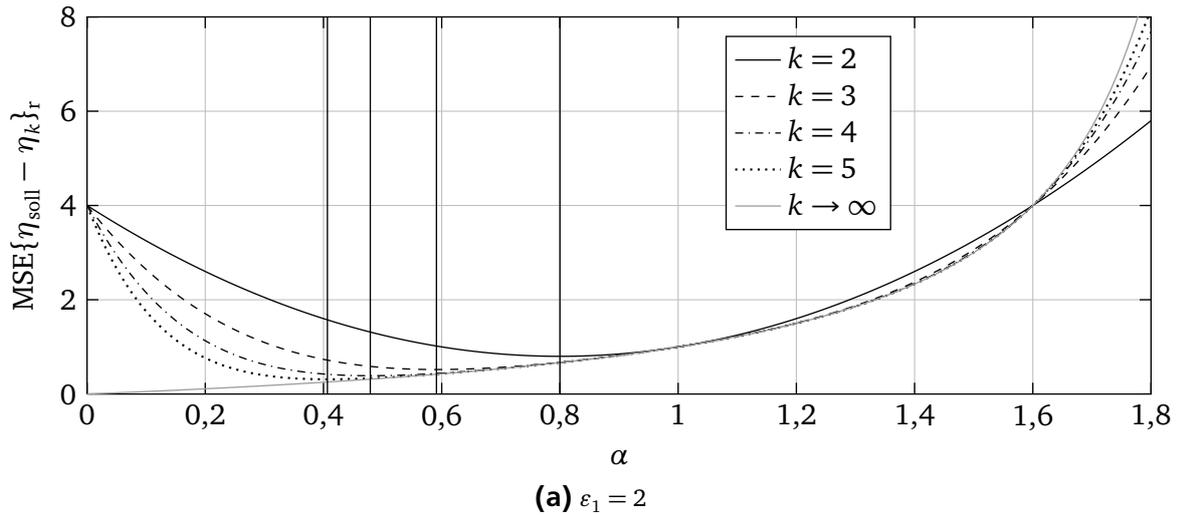


Abbildung 5.14: $MSE\{\eta_{\text{soll}} - \eta_k\}_r$ für $\varepsilon_1 = 2$ und 20 bei konstanter Reglerverstärkung
 (Die senkrechten schwarzen Linien markieren die Werte für α , bei denen für $k = 2, \dots, 5$ jeweils das Minimum vorliegt.)

In der unteren Abbildung 5.14b liegt mit $\varepsilon_1 = 20$ eine deutliche Anfangsabweichung vor. Daher steigen die Verläufe des MSE links und rechts von $\alpha = 1$ schnell sehr stark an.

Je weiter man sich links und rechts von $\alpha = 1$ befindet, desto mehr dominiert eine Abweichung des Erwartungswertes den MSE. Bei kleinen α durch Kriechen, bei großen α durch schlecht gedämpftes Schwingen.

Die Graphen sind so skaliert, dass der MSE nur bis zu einem Wert von acht dargestellt wird. Bezogen auf eine Anforderung wie $\varepsilon_{\text{OK}} = 3$ oder $\varepsilon_{\text{OK}} = 5$, also $\varepsilon_{\text{OK}}^2 = 9$ bzw. $\varepsilon_{\text{OK}}^2 = 25$, erscheint ein MSE von acht nicht unbedingt viel. Dennoch wurde eingangs schon diskutiert, dass auch kleinere Varianzen in Hinblick auf ein möglichst geringes $p_{\overline{\text{OK}},\text{max}}$ als schlecht anzusehen sind.

Diese Art der Darstellung wird auch im Folgenden immer wieder verwendet, um zwei Punkte zu veranschaulichen. Zum einen, wie schnell die Varianz mit steigender Anzahl an Regelschritten absinkt, und zum anderen, wie robust diese Wirkung gegenüber Unsicherheiten bei α ist. Diese kann man bewerten, indem man von $\alpha = 1$ ausgehend den Verlauf der MSE-Kurven nach links und rechts betrachtet. Ein mög-

lichst flacher Verlauf (bei einem insgesamt akzeptabel kleinen MSE) ist dabei als robuster anzusehen als ein steiler Anstieg.

Fazit

Bei fester Reglerverstärkung k_I bzw. α ist es zur Minimierung des MSE über mehrere Schritte sinnvoll, die kleinste zulässige Regelabweichung zu nehmen. Legt man dabei die zulässigen Werte für α zugrunde, die sich für die beispielhaften Anforderungen in Abschnitt 5.3 ergeben haben¹⁵, ist es aber nicht möglich, die Varianz auch nach mehreren Regelschritten effektiv zu reduzieren. So ist bei $\alpha = 0,91$ nur ein minimaler MSE von 0,83, und bei $\alpha = 0,83$ ein minimaler MSE von 0,71 zu erreichen.

5.4.4 I-Regler mit zeitvariabler Verstärkung $k_{I,k}$

Lässt man eine zeitvariable Verstärkung $k_{I,k}$ zu, so stellt sich die Frage, welcher Verlauf optimal ist. Dabei wird hier weiter der MSE der Regelabweichung als zu optimierende Größe angesehen. In Anlehnung an [MEDITCH, 1968] wird so vorgegangen, dass der MSE in jedem Schritt optimiert werden soll.

Aus Gl. (4.16) folgt

$$E\{\varepsilon_k\}_r = (1 - \alpha_{k-1}) \cdot E\{\varepsilon_{k-1}\}_r$$

und

$$\text{Var}\{\varepsilon_k\}_r = (1 - \alpha_{k-1})^2 \cdot \text{Var}\{\varepsilon_{k-1}\}_r + \alpha_{k-1}^2, \quad (5.30)$$

woraus sich direkt

$$\text{MSE}\{\varepsilon_k\}_r = (1 - \alpha_{k-1})^2 \cdot (E\{\varepsilon_{k-1}\}_r)^2 + (1 - \alpha_{k-1})^2 \cdot \text{Var}\{\varepsilon_{k-1}\}_r + \alpha_{k-1}^2$$

ergibt. Ableiten des MSE nach α_{k-1} und Nullsetzen ergibt

$$2\alpha_{k-1} \cdot (E\{\varepsilon_{k-1}\}_r)^2 + 2\alpha_{k-1} \text{Var}\{\varepsilon_{k-1}\}_r + 2\alpha_{k-1} - 2 \cdot (E\{\varepsilon_{k-1}\}_r)^2 - 2 \cdot \text{Var}\{\varepsilon_{k-1}\}_r = 0$$

als notwendige Bedingung für ein Extremum. Diese Gleichung hat für α_{k-1} die einzige Lösung

$$\alpha_{k-1} = \frac{(E\{\varepsilon_{k-1}\}_r)^2 + \text{Var}\{\varepsilon_{k-1}\}_r}{1 + (E\{\varepsilon_{k-1}\}_r)^2 + \text{Var}\{\varepsilon_{k-1}\}_r}, \quad (5.31)$$

wobei es sich um das Minimum handeln muss, da der MSE für $\alpha \rightarrow \pm\infty$ unbeschränkt ist.

Geht man für die Rechnung von einem bekannten ε_1 aus, so ergibt sich für das α_1 des ersten Schritts mit $E\{\varepsilon_1\}_r = \varepsilon_1$ und $\text{Var}\{\varepsilon_1\}_r = 0$ der Wert

$$\alpha_1 = \frac{\varepsilon_1^2}{1 + \varepsilon_1^2}. \quad (5.32)$$

Der Verlauf von α_1 über ε_1 ist identisch mit dem des optimalen α für $k = 2$ in Abbildung 5.13a. Es ist dort zu erkennen, dass dieser Wert schnell gegen eins geht. Für $\varepsilon_1 = 3$ beträgt α_1 beispielsweise schon 0,9 und für $\varepsilon_1 = 5$ liegt das Optimum bei 0,96. Daher wird $\alpha_1 = 1$ gesetzt.

¹⁵ Das minimale α liegt dabei – je nach Forderung – bei 0,91 oder 0,83.

Mit der Wahl $\alpha_1 = 1$ folgt aus Gl. (4.16), dass $E\{\varepsilon_k\}_r = 0$ für alle $k \geq 2$ gilt. Somit ergibt sich nach Gl. (5.31)

$$\alpha_{k-1} = \frac{\text{Var}\{\varepsilon_{k-1}\}_r}{\text{Var}\{\varepsilon_{k-1}\}_r + 1}$$

als optimaler Wert der Reglerverstärkung. Setzt man dieses α_{k-1} wieder in Gl. (5.30) ein, so erhält man

$$\text{Var}\{\varepsilon_k\}_r = \frac{\text{Var}\{\varepsilon_{k-1}\}_r}{\text{Var}\{\varepsilon_{k-1}\}_r + 1}$$

als rekursive Gleichung der optimalen (minimalen) Varianz im k -ten Schritt. Dabei bestimmt sich die optimale Reglerverstärkung nach der gleich aufgebauten Gleichung

$$\alpha_k = \frac{\alpha_{k-1}}{\alpha_{k-1} + 1},$$

bzw. für $\alpha_1 = 1$ auch direkt mit

$$\alpha_k = \frac{1}{k}.$$

In Tabelle 5.3 sind die Werte für $\text{Var}\{\varepsilon_k\}_r$ und α_k in Abhängigkeit von k dargestellt.

Tabelle 5.3: Optimales α_k mit minimalen $\text{Var}\{\varepsilon_k\}_r$ (für hohe Anfangsabweichungen $\varepsilon_1 \rightarrow \infty$)

k	1	2	3	4	5	6
α_k	1	0,5	0,33	0,25	0,2	0,17
$\text{Var}\{\varepsilon_k\}_r$	—	1	0,5	0,33	0,25	0,2

Es wurde der Anfangswert $\alpha_1 = 1$ gesetzt, obwohl dies nur für große ε_1 dem optimalen Wert entspricht. Daher wird hier noch der Fehler bzw. die Verschlechterung des MSE betrachtet, der dadurch in Kauf genommen wird. Dazu wird die Größe

$$\Delta\text{MSE}_{\text{rel},k} = \frac{\text{MSE}\{\varepsilon_k\}_r|_{\alpha_1=1} - \text{MSE}\{\varepsilon_k\}_r|_{\text{opt } \alpha_1}}{\text{MSE}\{\varepsilon_k\}_r|_{\text{opt } \alpha_1}}$$

verwendet, wobei $\text{MSE}\{\varepsilon_k\}_r|_{\text{opt } \alpha_1}$ für den MSE steht, der sich mit dem optimalen Startwert für α_1 nach Gl. (5.32) ergibt. $\Delta\text{MSE}_{\text{rel},k}$ gibt also an, um wie viel der MSE im k -ten Schritt bei der Wahl von $\alpha_1 = 1$ gegenüber dem optimalen Startwert schlechter ist, wobei dies auf den Optimalwert des k -ten Schrittes bezogen ist. In Tabelle 5.4 ist für verschiedene, kleinere Anfangswerte ε_1 diese Größe angegeben.

Tabelle 5.4: $\Delta\text{MSE}_{\text{rel},k}$

k	2	3	4	5	6
$\varepsilon_1 = 2$	25,0 %	12,5 %	8,3 %	6,3 %	5 %
$\varepsilon_1 = 3$	11,1 %	5,6 %	3,7 %	2,8 %	2,2 %
$\varepsilon_1 = 5$	4,0 %	2,0 %	1,3 %	1,0 %	0,8 %

Es zeigt sich, dass der Abstand zum optimalen MSE umso größer ist, je geringer ε_1 ist. Allerdings baut sich dieser Abstand mit steigender Schrittzahl auch schnell ab.

Robustheit

Während bei dem bisher vorgestellten Regler mittels α nur der Streckenparameter ϑ_1 im Entwurf verwendet wurde, wird in der allgemeinen Form des optimalen, zeitvarianten Reglers auch die bezogene Regelabweichung ε_1 bzw. damit die Varianz σ_y^2 benötigt.

Da hier jedoch nur die suboptimale Lösung weiter betrachtet wird, indem unabhängig von ε_1 der Anfangswert $\alpha_1 = 1$ verwendet wird, ergeben sich die Werte α_k unabhängig von der Varianz σ_y^2 und damit auch unabhängig von einem möglichen Fehler bei dem angenommenen Wert für diese Varianz.¹⁶

Somit bleibt ein Fehler im angenommenen Streckenparameter $\hat{\vartheta}_{1,r}$,

$$\hat{\vartheta}_{1,r} = \vartheta_1 \cdot \kappa_\theta$$

zu beachten. Dazu muss im Weiteren zwischen der Reglerverstärkung $\hat{\alpha}_k$, die der Regler vorzugeben „meint“, und der Reglerverstärkung α_k , die tatsächlich wirkt, unterschieden werden.

Der Regler wird nach dem in diesem Abschnitt beschriebenen Vorgehen ausgelegt, so dass

$$\hat{\alpha}_k = \frac{1}{k}$$

gilt. Damit wirkt im Regelkreis tatsächlich eine Reglerverstärkung von

$$\alpha_k = \hat{\alpha}_k \cdot \frac{1}{\kappa_\theta} = \frac{1}{k \cdot \kappa_\theta}.$$

Es interessiert, wie die Verteilungseigenschaften der tatsächlichen Regelabweichung $\{\varepsilon_k\}_r$ in Abhängigkeit des Fehlers κ_θ aussehen. Dazu werden Erwartungswert und Varianz von $\{\varepsilon_k\}_r$ für verschiedene Verläufe von α_k , die sich durch verschiedene Werte für κ_θ aus $\hat{\alpha}_k$ ergeben, über

$$E\{\varepsilon_k\}_r = (1 - \alpha_{k-1}) \cdot E\{\varepsilon_{k-1}\}_r \quad \text{und} \quad \text{Var}\{\varepsilon_k\}_r = (1 - \alpha_{k-1})^2 \cdot \text{Var}\{\varepsilon_{k-1}\}_r + \alpha_{k-1}^2$$

mit den Startwerten

$$E\{\varepsilon_2\}_r = (1 - \alpha_1) \cdot \tilde{\varepsilon}_1 \quad \text{und} \quad \text{Var}\{\varepsilon_2\}_r = 1$$

bestimmt. Daraus kann für jedes κ_θ der $\text{MSE}\{\varepsilon_k\}_r$ bestimmt werden. In Abbildung 5.15 ist der MSE für $\varepsilon_1 = 20$ und verschiedene Schritte k grafisch dargestellt.

Es ist gerade der Kehrwert gewählt, da damit die Verläufe mit denen aus Abbildung 5.14 und 5.20 vergleichbar sind, die über α aufgetragen sind.

Es zeigt sich im Vergleich zu Abbildung 5.14b über alle dargestellten Schritte $k \geq 3$ ein engerer „Kessel“ um $\kappa_\theta = 1$, was auf eine etwas geringere Robustheit gegenüber Parameterunsicherheiten hinweist. Allerdings zeigt sich, dass der MSE auch bei $\kappa_\theta = 1$ unter den Wert von eins abfällt, im Gegensatz zur Regelung mit konstantem α , worin der große Vorteil dieses Ansatzes liegt.

Fazit

Mit diesem Ansatz ergibt sich zwar das insgesamt, d. h. über viele Einrichtvorgänge, optimale Regelverhalten im Sinne des MSE, jedoch treten bei der praktischen Anwendung Probleme auf.

¹⁶ Wenn die optimale Lösung bezüglich einer gegebenen Regelabweichung e_1 bestimmt werden soll, dann entspricht ein Fehler in $\hat{\sigma}_y$ einem Fehler in ε_1 , womit α_k bezüglich eines falschen Startwertes bestimmt wird.

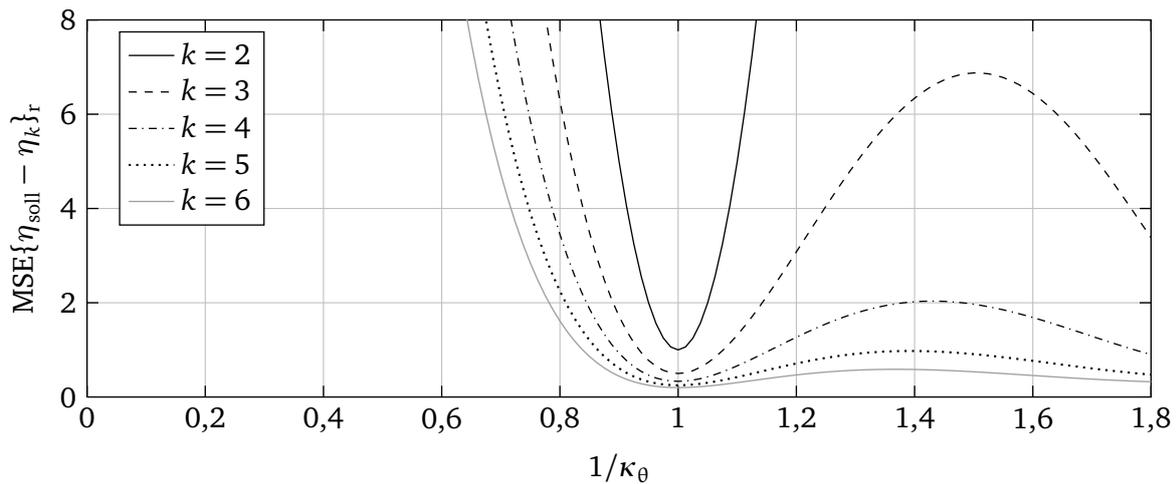


Abbildung 5.15: $\text{MSE}\{\eta_{\text{soll}} - \eta_k\}_r$ für $\varepsilon_1 = 20$ bei zeitabhängigen $\hat{\alpha}_k$

Zum einen sinkt die Reglerverstärkung mit steigender Schrittzahl k schnell auf sehr kleine Werte ab, so dass die nicht auszuschließenden sprungartigen Störungen nur noch sehr langsam ausgeregelt werden können.

Zum anderen werden auch bei dem optimalen Verlauf des MSE große Abweichungen auftreten, wenn auch mit geringer Wahrscheinlichkeit. Diese werden aufgrund der schon erwähnten starken Verkleinerung der Reglerverstärkung nur noch sehr schlecht ausgeregelt, was praktisch nicht zu tolerieren ist.

Dies widerspricht aber nicht den Ergebnissen, dass der genannte Verlauf von α_k optimal im Sinne des MSE ist. Der MSE bewertet eine Schar verschiedener Einrichtvorgänge, nicht aber die verschiedenen Regelschritte eines einzelnen Einrichtvorgangs.

Möglichkeiten, damit umzugehen, werden am Ende des nächsten Abschnitts besprochen, da bei den Filtern ähnliche Probleme auftreten.

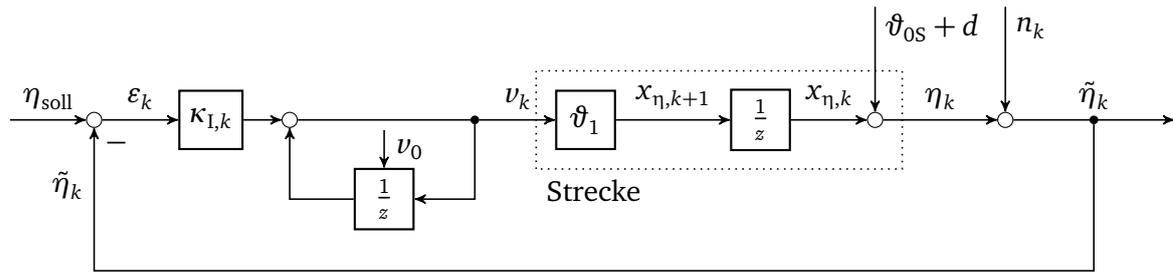
5.4.5 I-Regler mit Filter

Wird dem System ein Filter hinzugefügt, so lässt sich die Lösung nicht mehr einfach in expliziter Form angeben bzw. wird dann die Darstellung als Differenzgleichung schnell unübersichtlich. Daher wird in diesem Abschnitt mit der Zustandsraumdarstellung gearbeitet.

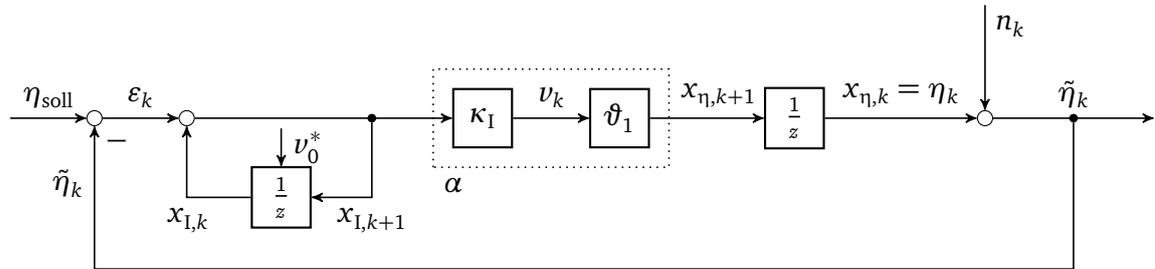
Dazu wird einführend der Regelkreis mit dem einfachen I-Regler in eine Zustandsraumdarstellung überführt, um insbesondere die Wahl der Anfangswerte der Kovarianzmatrizen darzulegen. Weitere Betrachtungen werden an diesem Modell nicht mehr durchgeführt, da sich die Ergebnisse mit denen aus Abschnitt 5.4.3 decken. Es werden dann Zustandsraumdarstellungen unter Berücksichtigung eines Filters abgeleitet und anhand derer die Wirkung des Filters erläutert.

Formulierung als Zustandsraummodell

In Abbildung 5.16a ist das Blockschaltbild des Regelkreises in normierten Größen nach Gl. (4.8) und (4.9) dargestellt, wobei v_0 den Anfangswert des Integrator des Reglers darstellt.



(a) Regelkreis nach Gl. (4.8) und (4.9)



(b) Zeitinvariantes κ_I hinter den Integrator gezogen

Abbildung 5.16: Blockschaltbilder der mit einem I-Regler geregelten Strecke

Geht man davon aus, dass κ_I zeitinvariant ist, so kann dieser Faktor hinter den Summiererblock des Reglers verschoben werden, womit sich die Darstellung in Abbildung 5.16b ergibt. Aus dieser kann die linearisierte Zustandsraumdarstellung inklusive I-Regelung abgelesen werden. Diese lautet

$$\begin{bmatrix} x_{\eta,k+1} \\ x_{I,k+1} \end{bmatrix} = \begin{bmatrix} -\alpha & \alpha \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_{\eta,k} \\ x_{I,k} \end{bmatrix} + \begin{bmatrix} -\alpha \\ -1 \end{bmatrix} n_k + \begin{bmatrix} \alpha \\ 1 \end{bmatrix} \eta_{\text{soll}}$$

$$\eta_k = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_{\eta,k} \\ x_{I,k} \end{bmatrix},$$

wobei wieder die Abkürzung $\alpha = k_I \theta_1 = \kappa_I \vartheta_1$ verwendet ist. Der konstante „Eingang“ $\vartheta_{0S} + d$ ist nicht mehr dargestellt, da dieser auch durch eine entsprechende Wahl der Anfangswerte für $x_{\eta,k}$ und den Integrator des Reglers dargestellt werden kann.

Diese Zustandsraumdarstellung entspricht nicht einer Minimalrealisierung, da sie nicht steuerbar ist. Steuerbarkeit ist aber keine Voraussetzung zur Anwendung der Gleichungen zur Übertragung stochastischer Eigenschaften aus Abschnitt 5.2.1, Seite 52, so dass diese dennoch in dieser Form verwendet wird. Dies hat auch den Vorteil, dass die Zustände besser interpretierbar bleiben.

Anfangswerte der Kovarianzmatrizen

Soll für $k = 1$ die Ausgangsgröße η_1 einen festen Wert annehmen, so ist für $k = 1$ die Größe η_1 keine Zufallsvariable und die Anfangswerte der Kovarianzmatrizen vereinfachen sich stark.

Für den ersten Zustand gilt $x_{\eta,1} = \eta_1$. Der Zustand $x_{I,1}$ entspricht v_0/κ_I , also der Eingangsgröße, die η_1 erzeugt (hat). Damit ist $x_{I,1} = \eta_1/\alpha$ und der gesamte Zustandsvektor für $k = 1$ deterministisch, und es gilt $C_{xx}(1) = \mathbf{0}$.

Für den Rauscheingang gilt $C_{nn}(k) = 1$. Da die Zustände $x_{\eta,k}$ und $x_{I,k}$ und das Rauschen für $k = 1$ nicht korreliert sind, gilt auch $C_{xn}(1) = \mathbf{0}$. Für alle weiteren Zeitpunkte k sind die Zustände dann zwar ebenfalls Zufallsvariablen, jedoch immer nur mit $\mathbf{n}_{k'}$, $k' = 1, \dots, k - 1$, korreliert, so dass auch $C_{xn}(k \geq 2) = \mathbf{0}$ gilt.

Vorteile der Darstellung

Zwar kann in dieser Form die Varianz und der MSE für einen Schritt k nicht mehr geschlossen angegeben werden, allerdings können hiermit auch kompliziertere Systeme behandelt werden, für die die Angabe der Transitionsmatrix nicht einfach möglich ist.

Art und Positionierung des Filters

Im Folgenden wird ein einfacher Moving-Average-Filter betrachtet. Dieser hat Vorteile dahingehend, dass er einfach strukturvariabel aufgebaut werden kann, was zu Beginn der Regelung nötig ist, um lange Einschwingvorgänge zu vermeiden. Auch zeigt sich eine interessante Verbindung zur im vorherigen Abschnitt besprochenen Auslegung auf einen optimalen MSE.

Auf der anderen Seite wirkt dieser Filter, wie gezeigt wird, destabilisierend, wenn er als einfacher Filter für die Messgröße implementiert wird. Er muss daher in den Regler „integriert“ werden. Auch dann zeigt sich eine gewisse Neigung zum Schwingen, welche aber hier nicht wesentlich ist. Andere Filter könnten diesbezüglich ein besseres Verhalten zeigen, darauf wird hier aber nicht weiter eingegangen. Das hier gezeigte Vorgehen zu Untersuchung der Eigenschaften des Regelkreises mit Filter lässt sich natürlich direkt auf andere Filter übertragen.

Moving-Average-Filter

Ein M -stufiger Moving-Average-Filter ist durch die Gleichung

$$z'_k = \sum_{l=0}^{M-1} b_l z_{k-l}$$

gegeben [ISERMANN, 1991, S. 9], wobei z_k das Eingangs- und z'_k das Ausgangssignal bezeichnet. Hier werden alle Gewichte $b_l = \frac{1}{M}$ gesetzt, womit der Filter einfach den Mittelwert der letzten M Werte zurückgibt. In der Zustandsraumdarstellung lautet dieser Filter

$$\begin{bmatrix} x_{F,1,k+1} \\ x_{F,2,k+1} \\ \vdots \\ x_{F,M-2,k+1} \\ x_{F,M-1,k+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \dots & \mathbf{0} \\ & 0 & \ddots & \ddots & \\ & & \ddots & \ddots & 0 \\ & & & 0 & 1 \\ \mathbf{0} & & & & 0 \end{bmatrix} \begin{bmatrix} x_{F,1,k} \\ x_{F,2,k} \\ \vdots \\ x_{F,M-2,k} \\ x_{F,M-1,k} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} z_k \quad (5.33)$$

$$z'_k = \begin{bmatrix} b_{M-1} & b_{M-2} & \dots & b_2 & b_1 \end{bmatrix} \mathbf{x}_{F,k} + b_0 z_k \quad (5.34)$$

Zu Beginn kann der Filter natürlich noch nicht vollständig gefüllt sein, da noch nicht genügend Eingangswerte vorliegen. Man könnte zwar den Filter zu Beginn vollständig mit dem ersten Eingangswert füllen, allerdings würde dies das Gewicht dieses ersten Wertes überhöhen und ein längeres Einschwingverhalten wäre die Folge. Daher wird von einem strukturvariablen Filter ausgegangen, dessen Ordnung m zu Beginn von eins auf die nominelle Filterordnung M anwächst.

Am einfachsten kann dies dadurch erreicht werden, dass die Ausgangs- und Durchgangsmatrix des Filters zeitvariabel zu

$$\mathbf{C}_{F,k} = \left[\mathbf{0}_{1 \times M-m_k} \underbrace{\begin{bmatrix} \frac{1}{m_k} & \dots & \frac{1}{m_k} \end{bmatrix}}_{m_k-1 \text{ mal}} \right]$$

bzw.

$$\mathbf{D}_{F,k} = \begin{bmatrix} 1 \\ m_k \end{bmatrix}$$

gewählt wird, wobei

$$m_k = \min(k, M) \quad (5.35)$$

gesetzt wird. Somit findet bei $k = m_1 = 1$ keine Filterung statt, und für $k \geq M$ besitzt der Filter die maximale Ordnung.¹⁷

Positionierung im Rückwärtszweig

In Abbildung 5.17 ist eine Konfiguration gezeigt, bei der der Filter im Rückwärtszweig sitzt und damit das gemessene Ausgangssignal glättet.

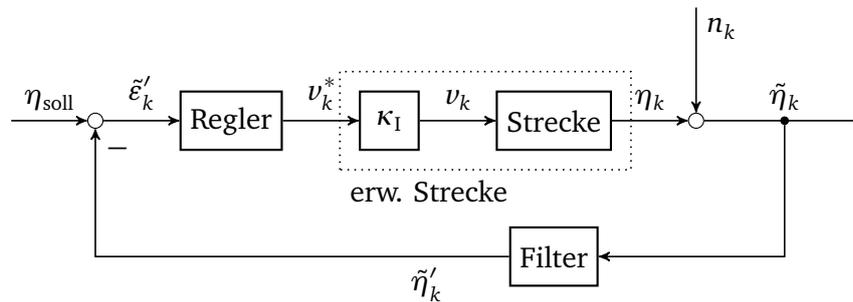


Abbildung 5.17: Filterung der Messgröße im Rückwärtszweig

Der (zeitinvariante) Faktor κ_1 ist dabei hinter den Regler gezogen, um diesen in der erweiterten Strecke mit dem Faktor ϑ_1 der Strecke zu α zusammenfassen zu können. Die Matrizen der Zustandsraumdarstellung $(\mathbf{A}_S, \mathbf{B}_S, \mathbf{C}_S, \mathbf{0})$ der erweiterten Strecke lauten damit

$$\mathbf{A}_S = 0 \quad \mathbf{B}_S = \alpha \quad \mathbf{C}_S = 1,$$

während der Regler (ohne k_1^*) in Zustandsraumdarstellung $(\mathbf{A}_R, \mathbf{B}_R, \mathbf{C}_R, \mathbf{D}_R)$ durch die Matrizen

$$\mathbf{A}_R = 1 \quad \mathbf{B}_R = 1 \quad \mathbf{C}_R = 1 \quad \mathbf{D}_R = 1$$

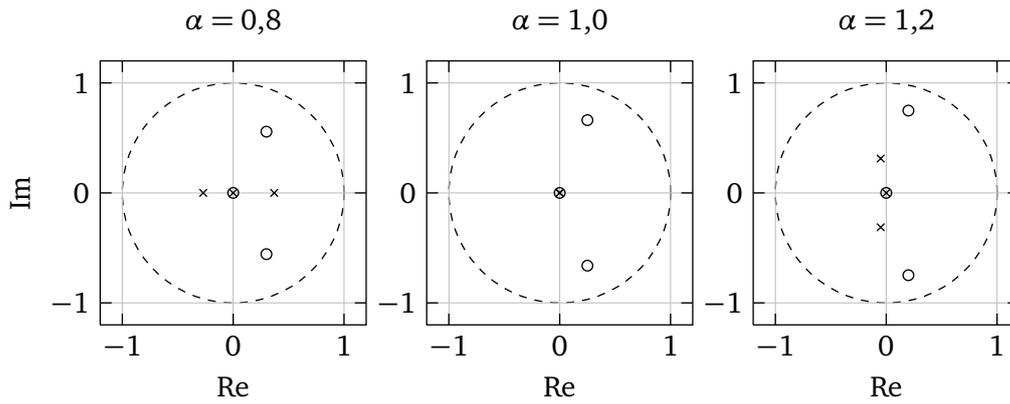
beschrieben wird. Mit den aus (5.33, 5.34) ablesbaren Matrizen $(\mathbf{A}_F, \mathbf{B}_F, \mathbf{C}_{F,k}, \mathbf{D}_{F,k})$ des Filters kann aus Abbildung 5.17 damit

$$\begin{bmatrix} x_{S,k+1} \\ x_{R,k+1} \\ x_{F,k+1} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_S - \mathbf{B}_S \mathbf{D}_R \mathbf{D}_{F,k} \mathbf{C}_S & \mathbf{B}_S \mathbf{C}_R & -\mathbf{B}_S \mathbf{D}_R \mathbf{C}_{F,k} \\ -\mathbf{B}_R \mathbf{D}_{F,k} \mathbf{C}_S & \mathbf{A}_R & -\mathbf{B}_R \mathbf{C}_{F,k} \\ \mathbf{B}_F \mathbf{C}_S & \mathbf{0} & \mathbf{A}_F \end{bmatrix} \begin{bmatrix} x_{S,k} \\ x_{R,k} \\ x_{F,k} \end{bmatrix} + \begin{bmatrix} -\mathbf{B}_S \mathbf{D}_R \mathbf{D}_{F,k} & \mathbf{B}_S \mathbf{D}_R \\ -\mathbf{B}_R \mathbf{D}_{F,k} & \mathbf{B}_R \\ \mathbf{B}_F & \mathbf{0} \end{bmatrix} \begin{bmatrix} n_k \\ \eta_{\text{soll}} \end{bmatrix}$$

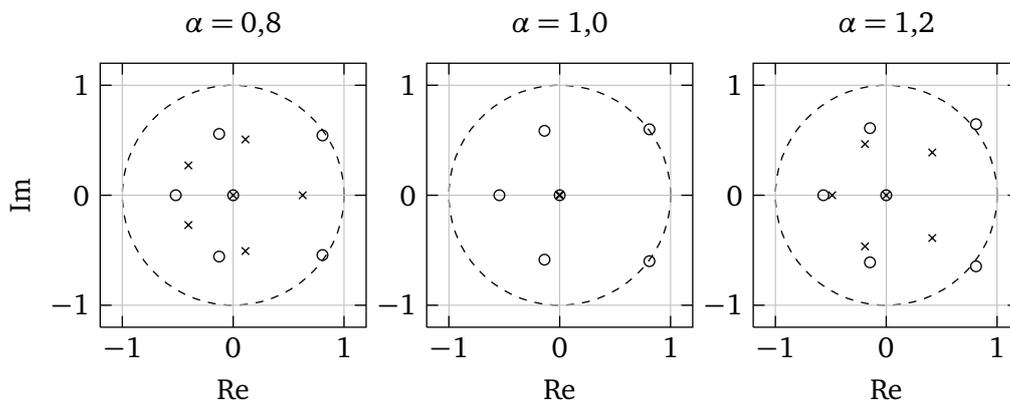
$$\eta_k = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \mathbf{x}_k$$

als Zustandsraumdarstellung des geregelten Systems abgelesen werden. Untersucht man dieses System näher, so erkennt man, dass der gewählte Filter an dieser Stelle destabilisierend wirkt. So besitzt das System, wie in Abbildung 5.18b zu sehen, ab einer Filterordnung $M = 5$ schon am „Nennpunkt“ $\alpha = 1$ Pole außerhalb des Einheitskreises, ist also instabil. Auch bei niedrigeren Filterordnungen sind die Pole nur schwach gedämpft, siehe Abbildung 5.18a. Daher wird auf die weitere Betrachtung des Filters im Rückwärtszweig verzichtet.

¹⁷ Anstelle eines wirklich strukturvariablen Filters, dessen Systemordnung anwächst, liegt damit ein zeitvariantes System fester Struktur vor. Die „variable“ Struktur zeigt sich bei einem festen $m \neq M$ darin, dass $M - m$ Eigenwerte nicht beobachtbar sind.



(a) Filterordnung $M = 2$



(b) Filterordnung $M = 5$

Abbildung 5.18: Lage der Pole der geregelten Systeme mit Filtern
(Kreise: Filter im Rückwärtszweig, Kreuze: Filter innerhalb Regler)

Positionierung innerhalb des Reglers im Vorwärtszweig

Wie gerade festgestellt wurde, ist ein Filter im Rückwärtszweig nicht geeignet. Damit sind auch alle Moving-Average-Filter ungeeignet, die an beliebiger Stelle zwischen Regler und Strecke im Regelkreis angeordnet sind, da dies für die Stabilität des Systems keinen Unterschied macht.

Eine weitere Variante, die sich auch stabilitätstechnisch von dem Filter im Rückwärtszweig unterscheidet, ist in Abbildung 5.19 dargestellt. Hier befindet sich der Filter innerhalb der Rückführung des Reglers, womit auch von einem neuen Regler gesprochen werden könnte. Es wird sich zeigen, dass diese Anordnung (bis $k = M$) genau das Verhalten ergibt, welches auch die Auslegung nach dem optimalen MSE zeigt, wodurch diese Betrachtung motiviert ist.

Für dieses System lässt sich die Zustandsraumdarstellung

$$\begin{bmatrix} x_{S,k+1} \\ x_{I,k+1} \\ x_{F,k+1} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_S - \mathbf{B}_S \mathbf{D}_{F,k} \mathbf{C}_S & \mathbf{B}_S \mathbf{D}_{F,k} & \mathbf{B}_S \mathbf{C}_{F,k} \\ -\mathbf{D}_{F,k} \mathbf{C}_S & \mathbf{D}_{F,k} & \mathbf{C}_{F,k} \\ -\mathbf{B}_F \mathbf{C}_S & \mathbf{B}_F & \mathbf{A}_F \end{bmatrix} \begin{bmatrix} x_{S,k} \\ x_{I,k} \\ x_{F,k} \end{bmatrix} + \begin{bmatrix} -\mathbf{B}_S \mathbf{C}_{F,k} & \mathbf{B}_S \mathbf{D}_{F,k} \\ -\mathbf{C}_{F,k} & \mathbf{D}_{F,k} \\ -\mathbf{B}_F & \mathbf{B}_F \end{bmatrix} \begin{bmatrix} n_k \\ \eta_{\text{soll}} \end{bmatrix}$$

$$\eta_k = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \mathbf{x}_k$$

ablesen. Eine Analyse der Pollagen zeigt ein deutlich günstigeres Verhalten, so dass diese Konfiguration prinzipiell geeignet ist, wenn auch für größere α eine deutliche Neigung zum Schwingen besteht.

Die Anfangswerte der Kovarianzmatrizen lauten auch hier $\mathbf{C}_{xx}(1) = \mathbf{0}$ und $\mathbf{C}_{xn}(k) = 0$.

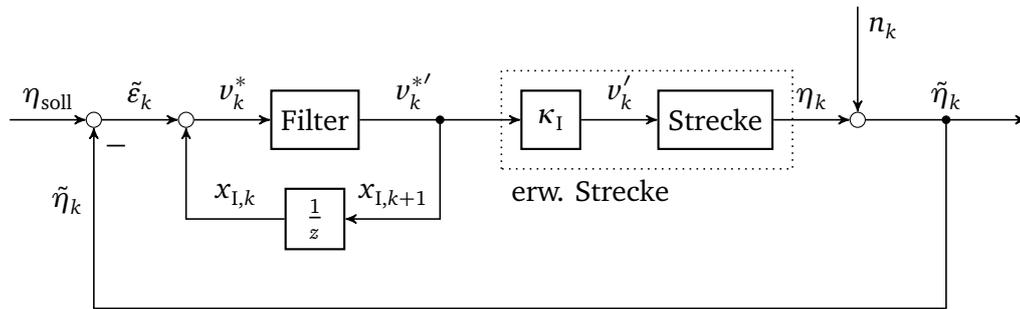


Abbildung 5.19: Filterung im Regler

Ergebnisse

In Abbildung 5.20 sind die sich ergebenden MSE über α aufgetragen. Es sind insgesamt drei verschiedene Fälle gezeigt, die im Folgenden diskutiert werden.

Zeitvariable Filterordnung

Zuerst wird der Fall betrachtet, dass sich die Ordnung des Filters wie zuvor durch Gl. (5.35) beschrieben aufbaut. Dies bedeutet, dass der erste Wert der Eingangsgröße derselbe ist, wie er sich auch im Fall ohne Filter ergeben würde. Ab dem zweiten Wert weichen die Werte dann von den ungefilterten ab.

Das Ergebnis ist in Abbildung 5.20a zu sehen. Im Vergleich mit Abbildung 5.15 erkennt man, dass die Verläufe des MSE identisch mit denen sind, die sich bei der Wahl des optimalen α_k ergeben. (Bis die maximale Filterordnung erreicht ist.)

Zeitvariable, verzögerte Filterordnung

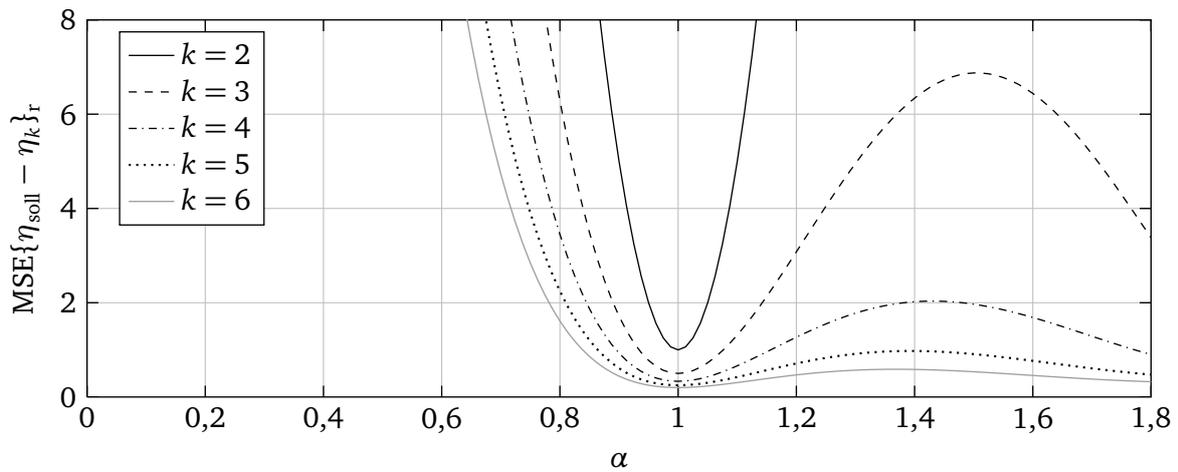
Da der zeitvariable Filter und das optimale α_k zu Beginn identisch sind, haben auch beide das gleiche Problem bezüglich der Robustheit, d. h. das Optimum ist sehr schmal ausgeprägt. In der Interpretation als Filter beruht das darin, dass die erste vom Regler bestimmte Stellgröße $u_{1,2}$ bzw. v_2 für zu kleine oder zu große α entsprechend „falsch“ ist, und dieser ungünstige Wert zunächst noch einen relativ großen Einfluss bei der Mittelwertbildung erhält. (Bei einem falsch angenommenen α sind natürlich alle Korrekturen des Reglers nicht optimal, nur ist der Effekt beim ersten Regelschritt am größten, da dort die Regelabweichung noch entsprechend hoch ist.)

Um dieses Verhalten zu verbessern, wird der Filter erst einen Schritt später aktiviert, d. h. für die Filterordnung gilt

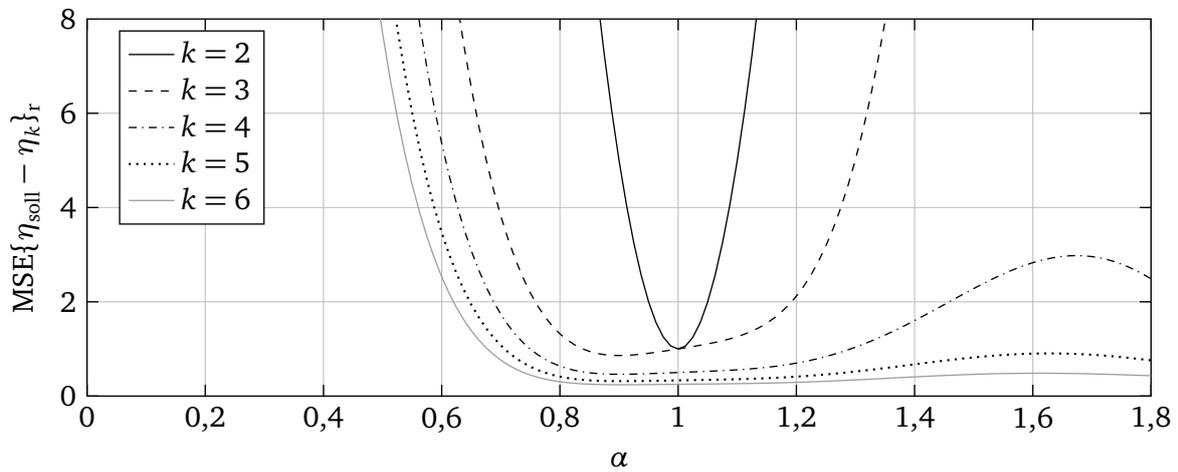
$$m_k = \max(\min(k - 1, M), 1) .$$

In Abbildung 5.20b ist das Ergebnis dargestellt. Es zeigt sich eine deutliche Verbreiterung des optimalen Bereichs. Allerdings auch wieder verbunden mit dem Nachteil, dass für $k = 3$ der MSE im Nominalpunkt $\alpha = 1$ höher ist.

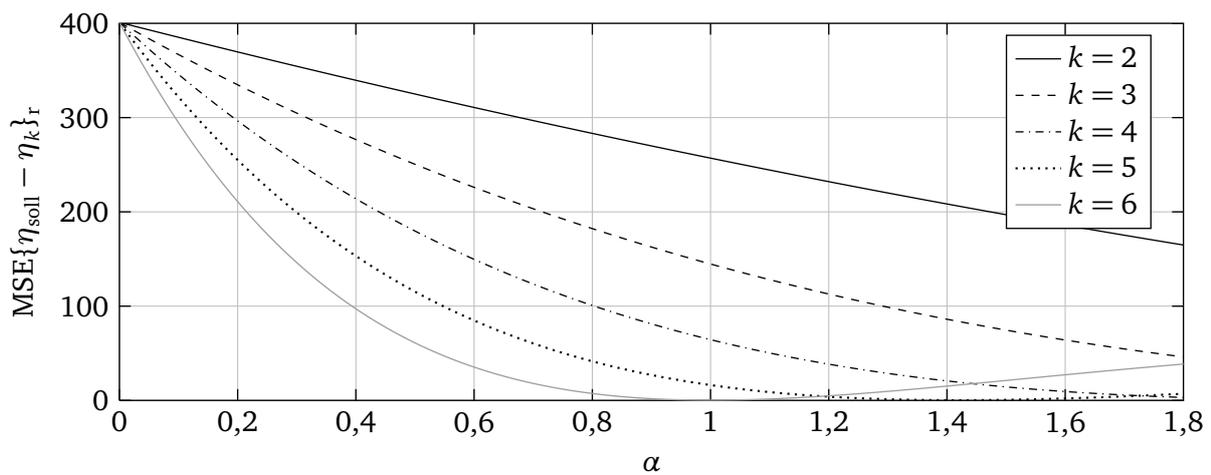
Bei höheren α setzt die stabilisierende Wirkung des Filters auch erst einen Schritt später ein, so dass rechts von $\alpha = 1$ der MSE insbesondere für $k = 3$ noch höher ist als im Fall ohne den Offset. Dafür ist das Kriechen links von $\alpha = 1$ weniger stark ausgeprägt, was in der Summe zu einem etwas robusteren Verhalten führt.



(a) Zeitvariable Filterordnung



(b) Zeitvariable, verzögerte Filterordnung



(c) Volle Filterordnung

Abbildung 5.20: $MSE\{\eta_{\text{soll}} - \eta_k\}_r$ für $\varepsilon_1 = 20$ bei Filterung ($M = 5$)

Volle Filterordnung

In Abbildung 5.20c ist zuletzt noch der Fall dargestellt, dass die volle Filterordnung von hier $M = 5$ aktiv ist. Dies steht für den Fall, dass bei vollständig gefülltem Filter eine sprungförmige Störung auftritt. Es zeigt sich deutlich das Kriechen. So werden selbst dann, wenn α exakt bekannt wäre, $M = 5$ Regelschritte benötigt, um die Regelabweichung abzubauen.

Diskussion der prinzipiellen Problematik der Filterung und Gegenmaßnahmen

Um eine grundsätzliche Problematik der Filterung aufzuzeigen, werden sämtliche Rauschanteile, welche sonst in n_y zusammengefasst sind, wieder genauer betrachtet. Das Rauschen n_y beinhaltet zum einen das eigentliche Messrauschen n_M , welches keinen Einfluss auf die tatsächliche Eigenschaft des Fertigteils hat, und daher nicht ausgeregelt werden soll. Zum anderen enthält es das Prozessrauschen n_p , welches von Fertigteil zu Fertigteil auftritt und ebenfalls nicht ausgeregelt werden soll. Dazu kommen mit d_y noch eine sprungförmige Störung, die Änderungen des Parameters θ_0 darstellt. Letztere soll möglichst schnell ausgeregelt werden. In Abbildung 5.21 sind die hier beschriebenen Arten von Störungen und Rauschen einzeln eingetragen.

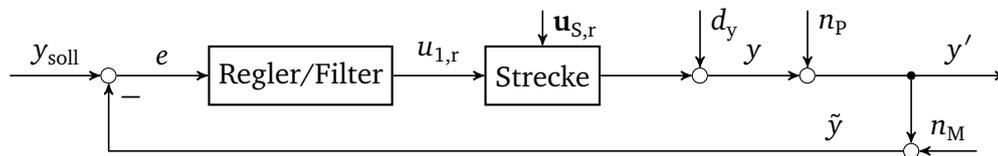


Abbildung 5.21: Regelkreis mit Störungen und Filter

Auf das eigentliche Messrauschen braucht, zumindest bei dem behandelten Beispielprozess, nicht weiter eingegangen zu werden, da dieses in der Regel unkritisch ist. (Es sei denn, eine Messung ist beispielsweise durch einen Fingerabdruck auf einem Kontrollfeld verfälscht.)

Wie in Abbildung 5.21 deutlich wird, greifen das Prozessrauschen n_p sowie die (sprungförmigen) Störungen d_y an der gleichen Stelle an, so dass durch einen Filter kein unterschiedliches Verhalten bezüglich n_p und d_y erreicht werden kann.

Daneben soll einer Änderung der Sollgröße schnell, d. h. möglichst in einem Schritt, gefolgt werden. Anders als bei d_y hat man zwar bezüglich der Übertragung von y_{soll} auf y gewisse Freiheitsgrade, wenn man einen Filter verwendet, der auch in den Rückwärtszweig gesetzt werden kann und y_{soll} damit an einer anderen Stelle als n_p angreift. Auch wäre es möglich, eine Struktur mit Vorsteuerung zu verwenden. Aber wenn im Schritt nach Änderung der Sollgröße dieser nicht exakt gefolgt wird, dann wirkt dieser Fehler wie ein d_y .

Prinzipiell kann ein Filter natürlich das hochfrequente Rauschen n_p filtern und die als stückweise konstant angenommen Signale y_{soll} und d_y durchlassen. Allerdings werden dabei die Sprünge dieser Signale zwangsweise geglättet. Und da diese Störungen in möglichst einem Schritt ausgeregelt werden sollen, liegt darin die Problematik der Filterauslegung.

Bedingte Filterung

Das gerade beschriebene Problem kann dadurch angegangen werden, dass eine bedingte Filterung verwendet wird. D. h. der Filter ist nur unter bestimmten Bedingungen aktiv. Da, wie begründet, die Störungen, die gefiltert werden sollen und die Störungen, die ausgeregelt werden sollen, nicht durch die Frequenz unterschieden werden können, wird die Höhe der Störung als Kriterium herangezogen.

Für die Aktivierung und Deaktivierung des Filters werden folgende Bedingungen verwendet:

- **Aktivierung:**
Die Regelabweichung liegt innerhalb einer definierten Toleranzgrenze, z. B. e_{OK} UND die Regelabweichung hat ihr Vorzeichen geändert.¹⁸
- **Deaktivierung:**
Die Regelabweichung liegt außerhalb der definierten Toleranzgrenze ODER y_{soll} hat sich geändert ODER u_s hat sich geändert.

In Abbildung 5.22 ist der Filter mit dem Aktivierungsblock dargestellt. Dabei bedeutet ein inaktiver Filter, dass die Filterordnung M auf eins gesetzt wird. Bei jeder neuen Aktivierung wird ein Reset des Filters durchgeführt.

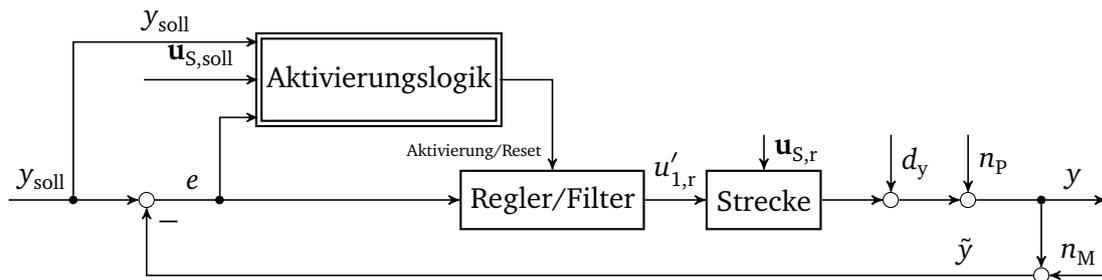


Abbildung 5.22: Regelkreis mit Störungen und bedingter Filterung

5.4.6 I-Regler mit abweichungsabhängiger Verstärkung $k_I(\tilde{\epsilon})$

In den vorangehenden Abschnitten wurde dargelegt, dass eine zeitabhängige Reglerverstärkung bzw. eine Filterung den MSE bzw. die Varianz senken können. Allerdings muss dabei immer geprüft werden, ob eine unvorhergesehene Störung aufgetreten ist, um zu verhindern, dass keine Korrekturen mehr vorgenommen werden, obwohl dies nötig ist.

Daher soll in diesem Abschnitt ein weiterer Ansatz, der auch in anderen Regelungen des Projektpartners verwendet wird, betrachtet werden. Bei diesem ist der Verstärkungsfaktor α ebenfalls nicht konstant, allerdings hängt dieser nicht von der „Zeit“ k , sondern der aktuell gemessenen Regelabweichung $\tilde{\epsilon}_k$ ab. Damit soll erreicht werden, dass bei größeren Abweichungen und Störungen immer eine schnelle Ausregelung erfolgt, während aber nahe dem Sollwert das Messrauschen weniger verstärkt wird.

Ein prinzipieller Verlauf für das abweichungsabhängige α ist in Abbildung 5.23 gezeigt. Wesentlich ist, dass bei niedrigen Regelabweichungen die Verstärkung gering ist, während diese zu höheren Regelabweichungen hin ansteigt.

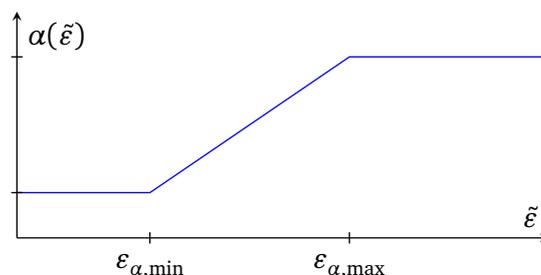


Abbildung 5.23: Variables $\alpha(\tilde{\epsilon})$

¹⁸ Die zweite Bedingung sorgt dafür, dass bei einem zu kleinen $\hat{\theta}_{1,r}$ das ohnehin schon vorhandene Kriechen nicht verstärkt wird.

Festlegen der Funktion $\alpha(\tilde{\varepsilon}_k)$

Im Folgenden wird anhand der Forderungen, die Abbildung 5.11c zugrunde liegen, der Weg beschrieben, einen geeigneten Verlauf für $\alpha(\tilde{\varepsilon})$ festzulegen.

Der Graph in Abbildung 5.24 ist ähnlich dem unteren Teil des Graphen aus Abbildung 5.11c, nur das dieser hier bezüglich der gemessenen Regelabweichung $\tilde{\varepsilon}_1$ erstellt wurde und nur die Bedingungen berücksichtigt sind, die α nach unten begrenzen und sich auf den ersten Regelschritt ($k = 2$) beziehen. Die sich aus diesen beiden verbleibenden Bedingungen ergebende Grenze wurde noch mit 1,1 multipliziert, um eine gewisse Robustheit gegenüber Fehler im angenommenen θ_1 zu erreichen.

Damit im ersten Schritt alle Bedingungen erfüllt werden, muss $\alpha(\tilde{\varepsilon}_k)$ so gewählt werden, dass dessen Funktion die mit 1,1 multiplizierte Grenze nicht schneidet.

Unter dieser Voraussetzung sind beispielhaft zwei mögliche Verläufe für $\alpha(\tilde{\varepsilon}_1)$ gezeigt. Die Funktion $\alpha_a(\tilde{\varepsilon}_1)$ ist so gewählt, dass die Reglerverstärkung für eine gemessene Regelabweichung von null ebenfalls null ist und dann so ansteigt, dass die gerade beschriebene Grenze tangiert wird, womit eine Verstärkung von eins bei $\tilde{\varepsilon}_1 = 6$ erreicht wird. Für gemessene Regelabweichungen von größer als fünfzehn wird die Reglerverstärkung wieder etwas abgesenkt, damit diese etwas zentraler im zulässigen Bereich verläuft.

Die zweite Funktion $\alpha_b(\tilde{\varepsilon}_1)$ unterscheidet sich von $\alpha_a(\tilde{\varepsilon}_1)$ dahingehend, dass sie für $\tilde{\varepsilon}_1 = 0$ bei einer Reglerverstärkung von $\alpha = 0,5$ beginnt.

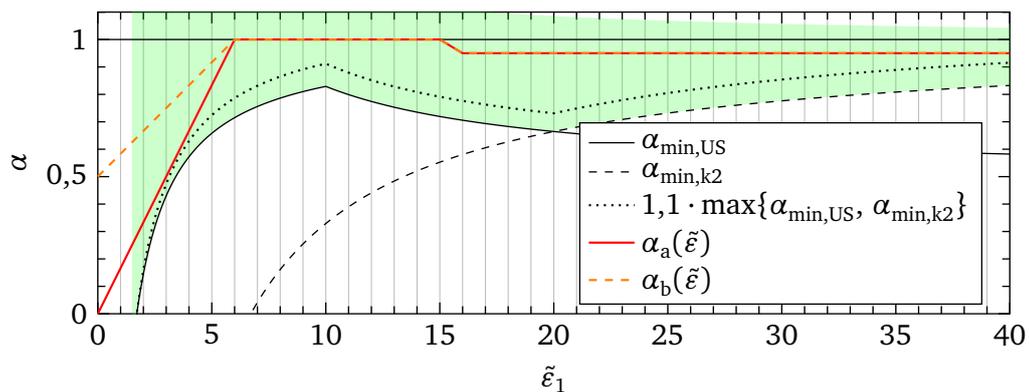


Abbildung 5.24: Variables $\alpha(\tilde{\varepsilon})$

Simulative Bestimmung der Verteilungseigenschaften

Dadurch, dass erstens der Reglerparameter vom Systemzustand und zweitens die Kennlinie nicht linear ist, können die Verteilungseigenschaften weder direkt noch schrittweise analytisch berechnet werden. In diesem Fall sind Simulationen notwendig.

Hierbei wird von Gl. (4.17) ausgegangen, wobei α_{k-1} durch $\alpha(\tilde{\varepsilon}_{k-1})$ ersetzt wird,

$$\tilde{\varepsilon}_k = (1 - \alpha(\tilde{\varepsilon}_{k-1})) \cdot \tilde{\varepsilon}_{k-1} + n_{k-1} - n_k.$$

Ausgangspunkt ist aber auch hier weiterhin ε_1 und nicht $\tilde{\varepsilon}_1$. Entsprechend ist für den ersten Schritt $\tilde{\varepsilon}_1 = \varepsilon_1 - n_1$ zu setzen.

In Abbildung 5.25 sind die aufgrund von 10 000 Simulationen je Punkt, d. h. Wert von ε_1 , geschätzten Werte des MSE über der anfänglichen Regelabweichung aufgetragen.

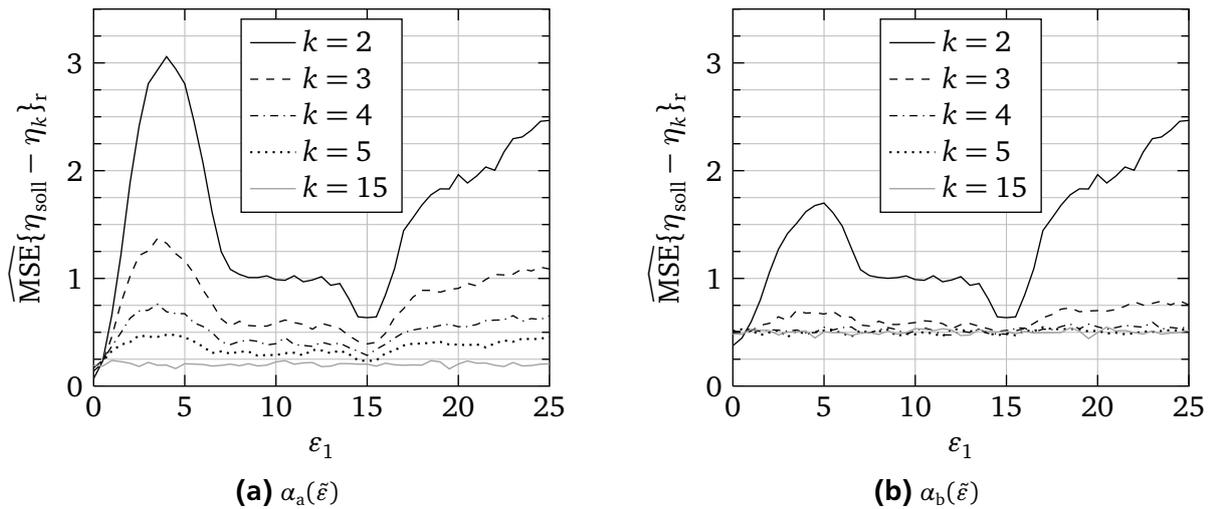


Abbildung 5.25: MSE in Abhängigkeit der Anfangsabweichung für verschiedene Funktionen für α (Schätzwerte aufgrund 10 000 Simulationen je Datenpunkt)

Diskussion der Darstellung über ε_1

Beim Betrachten der Abbildungen 5.25a und 5.25b ist zu beachten, dass die gezeigten Verläufe über ε_1 und nicht $\tilde{\varepsilon}_1$ dargestellt sind. Damit ergeben sich um die Punkte $\varepsilon_1 = 6$ und $\varepsilon = 15$ Effekte, die einer kurzen Erläuterung bedürfen.

Im Bereich um $\varepsilon_1 = 10$ hat der MSE nach dem ersten Schritt ($k = 2$) den Wert eins. Dieser wird nur durch die Varianz verursacht, da in diesem Bereich (meistens) $\alpha = 1$ gilt und der Mittelwert von ε_2 damit Null ist.

Würde man die Verläufe über $\tilde{\varepsilon}_1$ darstellen, so wären die Übergänge in die Gebiete einer niedrigeren Reglerverstärkung bei $\tilde{\varepsilon}_1 = 6$ und $\tilde{\varepsilon} = 15$ auch „hart“, d. h. zwischen 6 und 15 wären der MSE genau (im Sinne der simulativen Bestimmung) eins. Da über ε_1 geplottet ist, sind diese Bereiche etwas „verschliffen“.

Im Bereich um $\varepsilon_1 = 6$ treten zwei Effekte zu Tage. Wenn bei $\varepsilon_1 = 6$ ein Wert für $\tilde{\varepsilon}_1$ gemessen wird, der geringer als sechs ist, wird demnach eine Reglerverstärkung kleiner als eins gewählt, die auf eine gemessene Regelabweichung $\tilde{\varepsilon}_1$ angewandt wird, die kleiner als die tatsächliche Regelabweichung ε_1 ist. Somit wird zu wenig korrigiert. Für Werte größer als sechs dagegen wird die Reglerverstärkung von eins auf eine gemessene Regelabweichung angewandt, die größer als die tatsächliche Regelabweichung ist. Damit tritt ein gewisses Überschwingen auf. Nimmt man beide Effekte zusammen, so erklärt sich, dass die Verteilung von ε_2 etwas auseinandergezogen wird, und die Varianz damit ansteigt. Als zweiter Effekt tritt eine systematische Regelabweichung für $\varepsilon_1 < 6$ auf, da die Reglerverstärkung kleiner als eins ist, welches ebenfalls den MSE erhöht. Dies erklärt die Überhöhung des MSE bei geringen Regelabweichungen und wenigen Schritten k .

Bei $\varepsilon_1 = 15$ tritt ein gegenteiliger Effekt auf, da die Reglerverstärkung mit einer höheren gemessenen Regelabweichung $\tilde{\varepsilon}_1$ sinkt. Daher ergibt sich in diesem Bereich eine etwas engere Verteilung, welche sich in einer geringeren Varianz und damit auch geringerem MSE zeigt. (Für größere ε_1 steigt die Varianz schnell wieder auf den zu $\alpha = 0,95$ gehörenden konstanten Wert von 0,90.)

Für $\varepsilon_1 > 15$ steigt der MSE wieder an, da sich mit $\alpha = 0,95$ eine zur gemessenen Regelabweichung proportionale Regelabweichung ε_2 nach dem ersten Schritt einstellt.

Vergleich der Reglerverstärkungen $\alpha_a(\tilde{\varepsilon}_1)$ und $\alpha_b(\tilde{\varepsilon}_1)$

Vergleicht man die Ergebnisse, die sich mit den Funktionen $\alpha_a(\tilde{\varepsilon}_1)$ und $\alpha_b(\tilde{\varepsilon}_1)$ ergeben, so erkennt man, dass der Endwert des MSE, als welcher hier der Verlauf für $k = 15$ angesehen wird, bei $\alpha_a(\tilde{\varepsilon}_1)$ knapp weniger als halb so groß wie der Endwert bei Verwendung von $\alpha_b(\tilde{\varepsilon}_1)$ ist, nämlich ca. 0,2 anstelle von 0,5.

Dafür dauert es länger bis der MSE auf den Endwert konvergiert. So sind bei $\alpha_a(\tilde{\varepsilon}_1)$ noch bei $k = 5$ deutliche Abweichungen zum Endwert erkennbar, während bei $\alpha_b(\tilde{\varepsilon}_1)$ der Endwert schon bei $k = 4$ erreicht ist. Auf der anderen Seite ist bei $\alpha_a(\tilde{\varepsilon}_1)$ der MSE für $k = 4$ bis auf den Bereich kleiner ε_1 nicht nennenswert schlechter oder sogar besser als bei $\alpha_b(\tilde{\varepsilon}_1)$. Nur bei $\varepsilon_1 < 6$ und $k < 5$ zeigt sich ein deutlicherer Vorteil von $\alpha_b(\tilde{\varepsilon}_1)$ gegenüber $\alpha_a(\tilde{\varepsilon}_1)$.

Vergleich mit den anderen Regleransätzen

Der Endwert des MSE lautet bei $\alpha_a(\tilde{\varepsilon}_1)$ ca. 0,2 und bei $\alpha_b(\tilde{\varepsilon}_1)$ ungefähr 0,5. Mit Gl. (5.28) ist die Varianz $\text{Var}\{\eta_k\}_r$ für $k \rightarrow \infty$ bei einem festen α gegeben. Die Varianz von η entspricht immer der von ε , so dass durch Umstellen dieser Gleichung das konstante α_{const} bestimmt werden kann, welches eine gegebene Varianz im Grenzfall $k \rightarrow \infty$, hier jetzt kurz als $\text{Var}\{\varepsilon_\infty\}_r$ bezeichnet, erzeugt,

$$\alpha_{\text{const}} = \frac{2 \cdot \text{Var}\{\varepsilon_\infty\}_r}{\text{Var}\{\varepsilon_\infty\}_r + 1}.$$

In Tabelle 5.5 ist dieser Zusammenhang für ein paar Werte angegeben. Anhand dieser lässt sich erkennen, dass unter Verwendung der Funktion α_a letztlich ein Rauschen erreicht werden kann, welches einem konstanten α von 0,33 entsprechen würde. Ein festes α von 0,33 wäre unmöglich zu wählen wie direkt aus Abbildung 5.24 ersichtlich ist. Der Endwert des MSE von 0,5 bei $\alpha_b(\tilde{\varepsilon}_1)$ entspricht noch einem konstanten α von 0,67.

Tabelle 5.5: Zusammenhang zwischen einem konstanten α und $\text{Var}\{\varepsilon_\infty\}_r$

α_{const}	1	0,9	0,67	0,5	0,41	0,33	0
$\text{Var}\{\varepsilon_\infty\}_r$	1	0,82	0,5	0,33	0,26	0,20	0

Bei der Regelung mit konstantem α könnte bei gleicher Reserve lediglich ein konstantes α von 0,9 gewählt werden (siehe Abbildung 5.24), was eine Varianz von 0,82 zur Folge hätte. Dazu wäre aufgrund des Kriechens der MSE im ersten Schritt noch zusätzlich höher (Abbildung 5.14b).

In Abbildung 5.26 ist die „Erfüllungsquote“ der verschiedenen Anforderungen über den Kehrwert von κ_θ dargestellt. Insbesondere ist die Bedingung „Regelabweichung Schritt 2“ (\hat{p}_{k3}) interessant, da diese bei der Festlegung des Verlaufs von $\alpha(\tilde{\varepsilon})$ überhaupt nicht berücksichtigt wurde. Die weiteren Bedingungen sind dagegen schon beachtet worden, allerdings unter der Annahme, dass $\tilde{\varepsilon}_1$ bekannt ist, so dass auch diese einer Überprüfung bedürfen.

In Abbildung 5.27 ist für beide Funktionen der MSE der Regelabweichung nach den ersten fünf Regelschritten ($k = 2, \dots, 6$) dargestellt. Dabei ist dieser über dem Kehrwert des relativen κ_θ aufgetragen, um die Robustheit beurteilen zu können.

Durch die Verringerung der Reglerverstärkung für $\tilde{\varepsilon}_1 > 15$ liegt der Optimalpunkt bei $1/\kappa_\theta = 1/0,95 \approx 1,05$. Dort sinkt der MSE für $\alpha_a(\tilde{\varepsilon}_k)$ und den gezeigten k vergleichbar mit dem optimalen zeitvarianten Regler bzw. dem entsprechenden Filter (siehe Abbildung 5.15 und 5.20a). Bewegt man sich etwas von dem optimalen Punkt weg, so steigt der MSE etwas langsamer an als bei den genannten Alternativen.

Bei $\alpha_b(\tilde{\varepsilon}_1)$ ist die Robustheit gegenüber Unsicherheiten in κ_θ etwas besser, d. h. der MSE verläuft abseits des Optimalpunktes insbesondere für $k = 3$ und $k = 4$ unterhalb der entsprechenden Verläufe für $\alpha_a(\tilde{\varepsilon}_k)$.

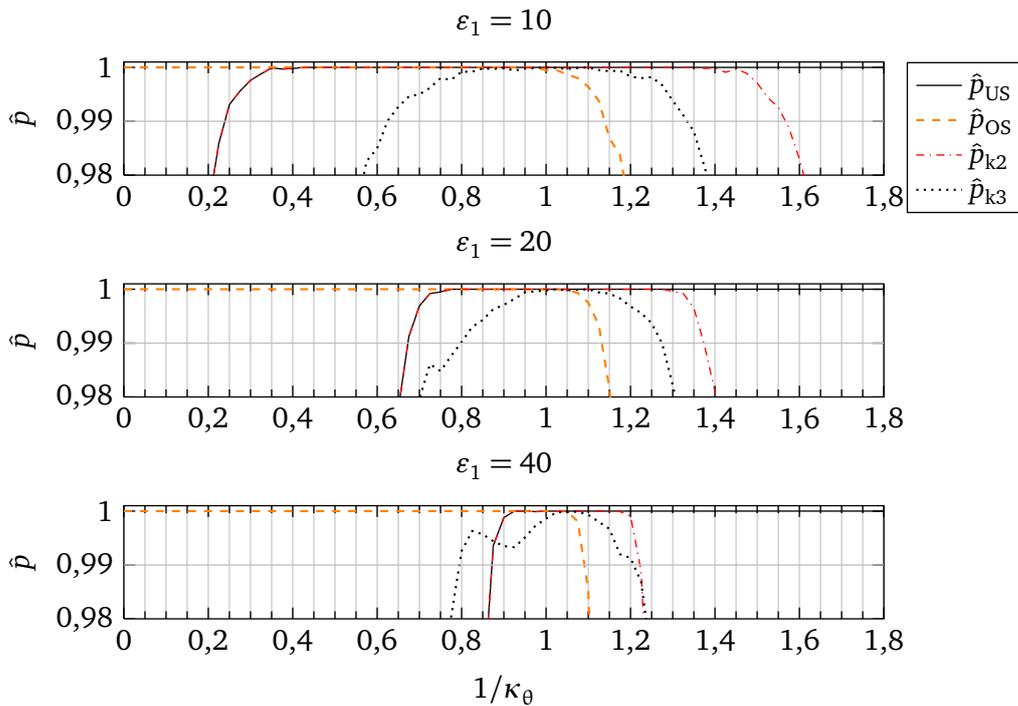


Abbildung 5.26: Erfüllungsquote der verschiedenen Bedingungen für $\alpha_a(\tilde{\varepsilon})$ (Schätzwerte anhand 10 000 Simulationen je Punkt)

Der Vorteil des Reglers mit abweichungsabhängiger Regelverstärkung ist aber, dass eine Reset-Logik nicht zwingend notwendig ist, wie dies bei der Variante mit Filter und dem zeitvarianten Regler der Fall ist. Vorteile kann eine zusätzliche Logik jedoch auch hier bringen, wie im Folgenden gezeigt wird.

Erweiterung

Aufgrund der Absenkung der Reglerverstärkung bei geringen gemessenen Regelabweichungen neigt dieser Regler etwas zum Kriechen, was sich in den höheren MSE-Werten für $k = 2, \dots, 5$ in den Bereichen $\varepsilon_1 < 7$ und $\varepsilon_1 > 15$ in den Abbildungen 5.25a und 5.25b zeigt.

Dies kann vermieden werden, wenn der Regler um eine Logik ähnlich dem Filter aus Abbildung 5.22 erweitert wird. Diese Logik ist hier derart aufzubauen, dass bei einem neuen Einrichtungsvorgang (und nach Änderung der Sollgröße oder der sekundären Eingänge) die Absenkung der Reglerverstärkung bei geringen Regelabweichungen erst dann zugelassen wird, wenn sich das Vorzeichen der Regelabweichung einmal geändert hat.

Die um diese Logik erweiterte Funktion $\alpha_a(\tilde{\varepsilon}_k)$ wird mit

$$\alpha_a^*(\tilde{\varepsilon}_k)$$

bezeichnet.

In Abbildung 5.28a ist der MSE analog zu Abbildung 5.25 über der Anfangsabweichung ε_1 aufgetragen. Es ist direkt ersichtlich, dass die Erhöhung des MSE bei niedrigen Anfangsabweichungen vollständig verschwunden ist. Bei höheren Anfangsabweichungen ($\varepsilon_1 > 15$) nimmt der MSE für $k = 2$ immer noch zu, da die Absenkung der Reglerverstärkung bei hohen Regelabweichungen immer aktiv ist, jedoch ist für $k \geq 3$ keine Überhöhung des MSE mehr auszumachen.

Bei einem ε_1 von 20 beträgt der MSE bei $\alpha_a(\tilde{\varepsilon}_k)$ und $k = 3$ ungefähr 1, während der bei $\alpha_a^*(\tilde{\varepsilon}_k)$ etwa 0,75 beträgt. Dies mag als gering erscheinen, jedoch kann dieser Unterschied stark ins Gewicht fallen, wenn das Verbleiben im OK-Bereich bei relativ engen Toleranzen gefordert ist.

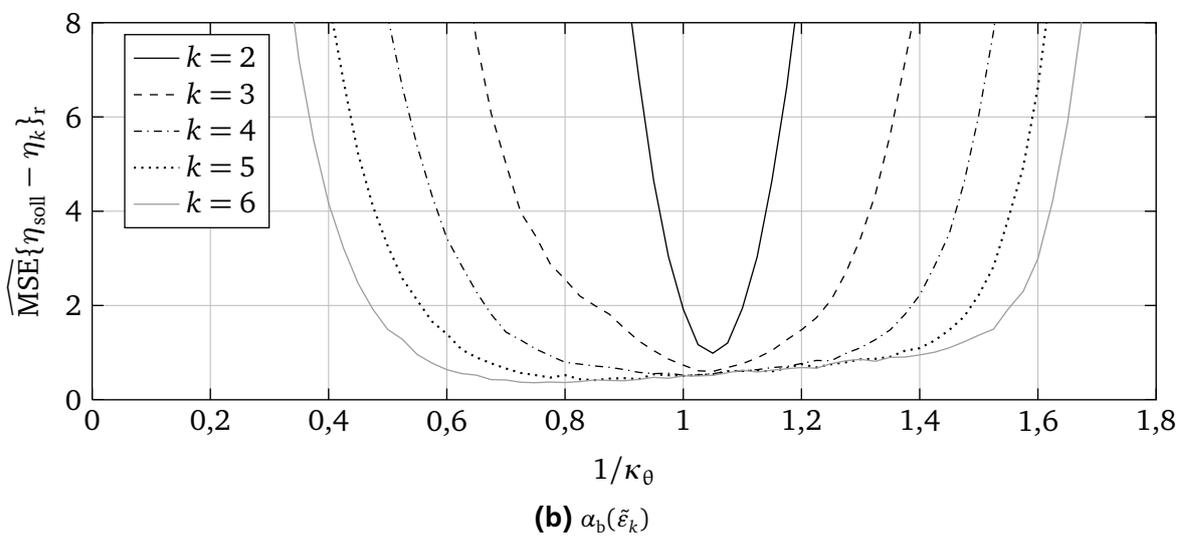
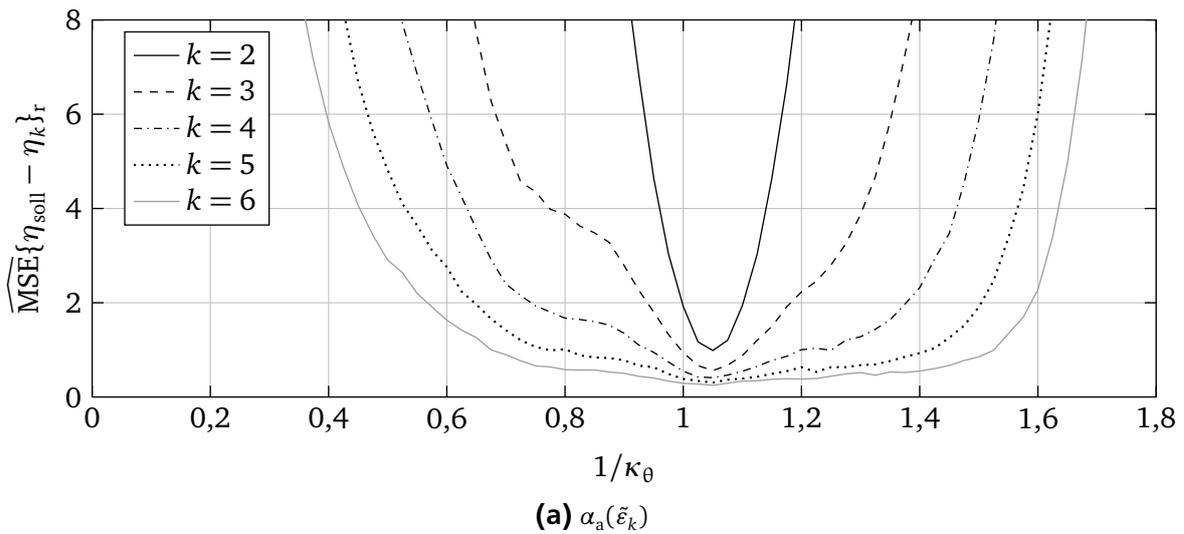


Abbildung 5.27: MSE der Regelabweichung mit $\alpha_a(\tilde{\varepsilon})$ und $\alpha_b(\tilde{\varepsilon})$ bei der Anfangsabweichung $\varepsilon_1 = 20$ (Schätzwerte anhand 10 000 Simulationen)

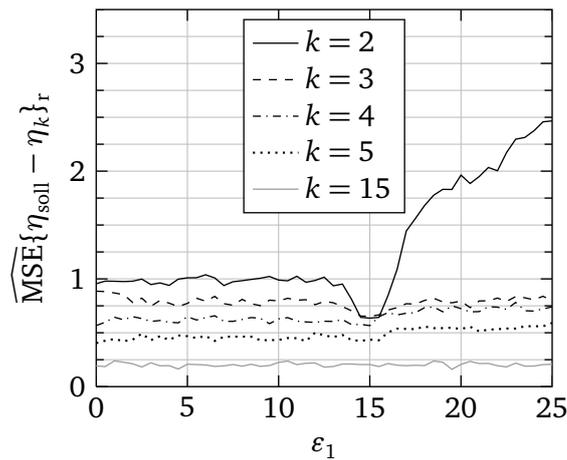
In Abbildung 5.28b lässt sich noch erkennen, dass die Kurven für $k \geq 3$ und $1/\kappa_\theta < 1$ deutlich flacher als in Abbildung 5.27a verlaufen, so dass auch die Robustheit zunimmt.

Nachteilig an dieser Erweiterung ist die zusätzlich notwendige Logik, die bei der Implementierung beachtet werden muss.

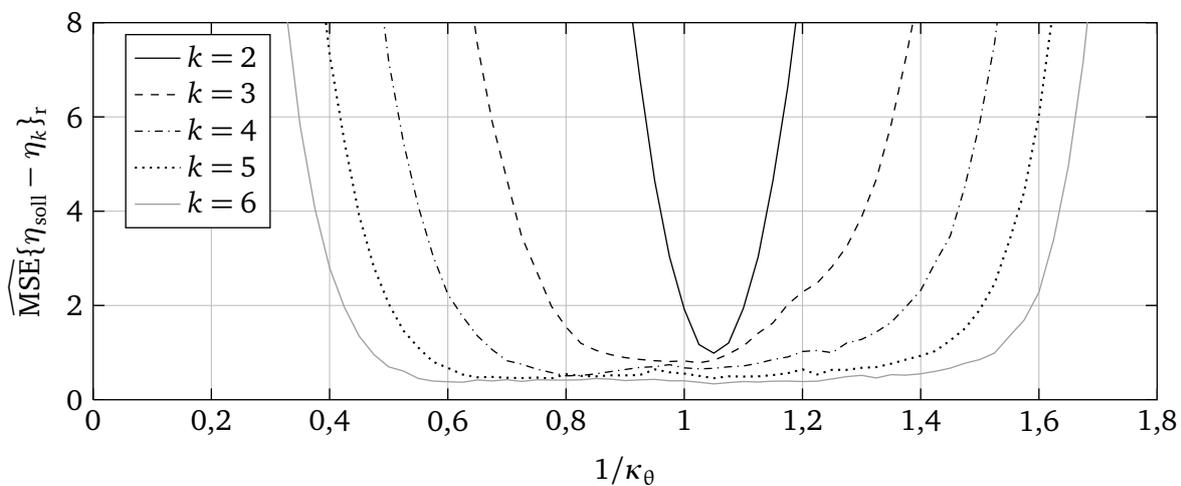
5.4.7 Fazit

Ein I-Regler mit fester Reglerverstärkung α kann zwar die in Abschnitt 5.1.1 formulierten Anforderungen erfüllen, jedoch wird das Rauschen am Ausgang verstärkt.

Dieses kann entweder mit einem I-Regler mit zeitvarianter Reglerverstärkung oder unter Zuhilfenahme eines Filters effektiv reduziert werden, jedoch wird dabei eine Reset-Logik benötigt, die bei zu großen Regelabweichungen und Änderungen der Sollgröße oder der sekundären Eingänge den Filter bzw. den zeitvarianten Regler auf den Ausgangswert zurücksetzt, damit ein Kriechen verhindert wird.



(a) $\kappa_\theta = 1$



(b) Anfangsabweichung $\varepsilon_1 = 20$

Abbildung 5.28: MSE der Regelabweichung mit $\alpha_a^*(\tilde{\varepsilon})$ (Schätzwerte anhand 10 000 Simulationen)

Alternativ kann ein I-Regler verwendet werden, der eine abweichungsabhängige Reglerverstärkung besitzt. Dieser kombiniert eine geringe Verstärkung des Rauschens mit einer guten Robustheit gegen sprungförmige Störungen.

5.5 Verbleiben im OK-Bereich

5.5.1 Testfall

Der hier vorgestellte Testfall ist ein Einrichtungsvorgang mit einer Anfangsabweichung von $\varepsilon_1 = 20$. Die Systemparameter sowie die Parameterschätzungen, die zur Berechnung der Reglerparameter zugrunde liegen, sind in Tabelle 5.6 zusammengefasst.

Der Regler nimmt mit $\hat{\vartheta}_{1,r} = 1,1 \cdot \vartheta_1$ einen etwas zu hohen Wert für ϑ_1 an, so dass die (nominelle) Reglerverstärkung etwas zu niedrig angesetzt wird. Bezüglich des Kriechens ist dies der kritischere Fall, weshalb dieser gewählt wurde. Bezüglich der Varianz und damit dem Verbleiben im OK-Bereich wäre eine Unterschätzung von ϑ_1 ungünstiger, worauf an der entsprechenden Stelle hingewiesen werden wird. Auf ein weiteres, eigenes Beispiel mit unterschätztem ϑ_1 wird jedoch verzichtet.

Tabelle 5.6: Testfall

Parameter	Strecke
ϑ_0	50
ϑ_1	0,5
$u_{1,N}$	50
n_A	$N(0, 1)$
n_S	$N(0, 0,5^2)$ oder $N(0, 1)$
Parameterschätzung für Regler	
$\hat{\vartheta}_{1,r}$	$1,1 \cdot \vartheta_1$
Startwert für Eingangsgröße	
$u_{1,1} = 10$	$\Rightarrow \eta_1 = 30$

Als zulässige Toleranzgrenzen für den OK-Bereich werden $\varepsilon_{OK} = 3$ und $\varepsilon_{OK} = 5$ betrachtet.

Für das Rauschen n_S in der Produktionsphase werden wie in der Tabelle angegeben zwei Varianten betrachtet, die sich in ihrer Varianz $\sigma_S^2 = \text{Var}\{n_S\}_r$ unterscheiden.

5.5.2 Simulationen

Von dem beschriebenen Testfall sind für jede Reglerkonfiguration 10 000 Simulationen durchgeführt worden. In Abbildung 5.29a sind die Verteilung von $\{\tilde{\eta}_k\}_r$ zu den simulierten Zeitpunkten k als (stehende) Histogramme dargestellt. Dazu ist die Achse für $\tilde{\eta}_k$ in Intervalle zerlegt. Zu jedem k zeigt die Breite des Balkens an, wie groß der Anteil der Simulationen ist, in denen $\tilde{\eta}$ in dem jeweiligen Intervall liegt. Dazu sind noch zwei Anmerkungen zu machen: Die Breite der Balken ist für jedes k einzeln so normiert, dass der breiteste Balken die Breite 0,9 besitzt. Außerdem ist die minimale Balkenbreite mit 0,025 festgelegt, d. h. auch kleinere Balken werden auf diesen Wert vergrößert. Somit sind auch einzelne Ausreißer bei einer sehr großen Anzahl an Simulationen sichtbar.

Im Plot von $\tilde{\eta}$ markiert die durchgezogene schwarze Linie den Sollwert von $\eta_{\text{soll}} = 50$ und die gestrichelten Linien die ± 3 - bzw. ± 5 -Toleranzgrenze.

Beim Betrachten von Abbildung 5.29a lässt sich gut erkennen, wie zu Beginn die Streuungen größer sind als am Ende. Dies kommt daher, dass wie in Abbildung 2.4 dargestellt, zunächst das reglerrelevante Rauschen n_A zugrunde gelegt wird, welches mit dem Beginn des Fortdrucks durch das kleinere Rauschen im Fortdruck n_S ersetzt wird.

In Tabelle 5.7 sind die oben beschriebenen Kennzahlen, die aus den Simulationsergebnissen ermittelt wurden, zusammengefasst. Dabei stammen die Daten der ersten Zeile aus der auch in Abbildung 5.29a dargestellten Simulation für $\varepsilon_{OK} = 3$, während für die zweite und dritte Zeile die Varianz des Rauschens in der Produktionsphase höher angesetzt wurde, nämlich genau wie das Rauschen n_A . In der dritten Zeile wurde dazu noch die Toleranzgrenze auf $\varepsilon_{OK} = 5$ erhöht.

Tabelle 5.7: Kennzahlen zur Reglerbewertung, $\alpha_a = 0,95$, aus je 10 000 Simulationen bestimmt

ε_{OK}	σ_S	\bar{n}_{RS}	$q_{RS,99\%}$	$q_{RS,100\%}$	$P_{\overline{OK}, >2\%}$	$P_{\overline{OK}, >1\%}$	$q_{\overline{OK}, 99\%}$	$q_{\overline{OK}, 100\%}$
3	0,5	1,4	2	4	4,1 %	5,5 %	14 %	93 %
3	1	1,4	2	4	34 %	51 %	29 %	77 %
5	1	1	2	2	0,1 %	0,4 %	0,5 %	10 %

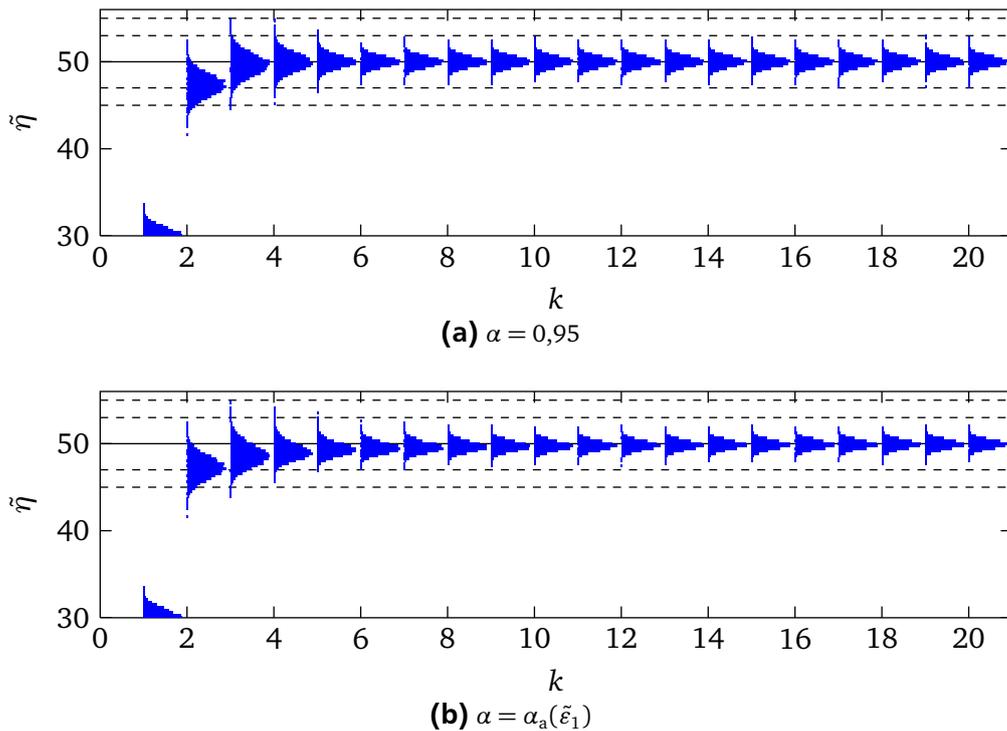


Abbildung 5.29: Ergebnis von 10 000 Simulationen des Testfalls ($\sigma_S = 0,5$, $\epsilon_{OK} = 3$)

Die Anzahl der Schritte, bis OK erreicht ist, hängt nicht vom Rauschen n_S ab, so dass die beiden ersten Zeilen in Tabelle 5.7 in den entsprechend Spalten die gleichen Werte besitzen. So hat es im Schnitt 1,4 Regelschritte gedauert, bis OK erreicht wurde. In 99% der Fälle wurden maximal zwei Schritte benötigt ($q_{RS,99\%} = 2$) und maximal waren vier Regelschritte nötig ($q_{RS,100\%} = 4$).

Die letzten vier Spalten beschreiben die Verteilung von $\{p_{\overline{OK}}\}_r$, die für die ersten beiden Zeilen der Tabelle auch in Abbildung 5.30 dargestellt ist. Die Bedeutung von $p_{\overline{OK}}$ sowie den darauf aufbauenden Kennzahlen wurde in Abschnitt 5.1.2 erläutert.

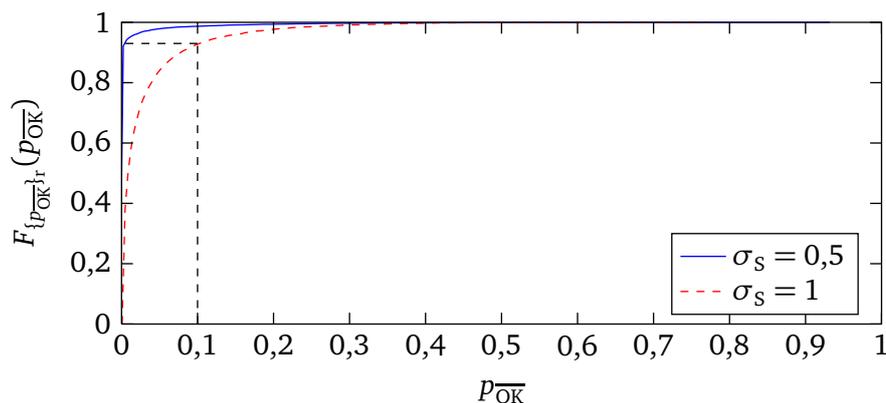


Abbildung 5.30: Verteilung von $\{p_{\overline{OK}}\}_r$ ($\sigma_S = 0,5$, $\epsilon_{OK} = 3$, über 10 000 Simulationen bestimmt)

In Abbildung 5.30 ist für die zu $\sigma_S = 1$ gehörende Verteilungsfunktion auch beispielhaft für $p_{\overline{OK}} = 10\%$ der Wert $F_{\{p_{\overline{OK}}\}_r}(10\%) \approx 0,93$ abgelesen, womit sich $p_{\overline{OK}, > 10\%} = 7\%$ ergibt. (Siehe auch Abschnitt 5.1.2.)

Ein Wert von $p_{\overline{OK}} \geq 10\%$ sollte als sehr schlecht angesehen werden, so dass in Tabelle 5.7 die Werte für die etwas strengeren Grenzen 1% und 2% angegeben sind. Es lässt sich deutlich erkennen, dass diese Zahlen für das hohe Produktionsrauschen $\sigma_S = 1$ deutlich schlechter sind. So werden nach 51% der

Einrichtvorgänge mehr als 1 % $\overline{\text{OK}}$ -Teile gefertigt, während dies bei $\sigma_S = 0,5$ lediglich in ca. 5,5 % der Einrichtvorgänge der Fall ist.

Eine weitere Größe sind die Quantile $q_{\overline{\text{OK}},x\%}$. So sagt der Wert $q_{\overline{\text{OK}},99\%} = 29\%$ aus, dass in 99 % der Simulationen nach Abschluss der Einrichtvorgänge maximal 29 % der produzierten Teile $\overline{\text{OK}}$ sind. $q_{\overline{\text{OK}},100\%} = 77\%$ gibt entsprechend den schlechtesten Wert für $p_{\overline{\text{OK}}}$ an, der in allen 10 000 Simulationen ermittelt wurde.

Anhand der letzten Zeile in Tabelle 5.7 lässt sich erkennen, dass bei einer Toleranzgrenze von $\varepsilon_{\text{OK}} = 5$ die Einrichtung deutlich weniger kritisch ist.

Im Weiteren wird der Fall $\varepsilon_{\text{OK}} = 1$ und $\sigma_S = 1$ betrachtet, da dies der kritischste Fall ist, und sich die Unterschiede der Regelkonzepte am deutlichsten zeigen.

Vergleich der Regler

In den Abbildungen 5.31a und 5.31b sind für verschiedene Regler die Verläufe der Werte für $q_{\overline{\text{OK}},99\%}$ bzw. $p_{\overline{\text{OK}},>1\%}$ über Δk_{OK} aufgetragen. Dabei bezeichnet Δk_{OK} die in Abschnitt 5.1.2 auf Seite 47 beschriebene Größe.

$\Delta k_{\text{OK}} = 0$ bezeichnet den Anlauf, in dem erstmals OK erreicht wurde, $\Delta k_{\text{OK}} = 1$ dann den darauffolgenden Anlauf sowie $\Delta k_{\text{OK}} \geq 2$ die folgenden Nachführungen in der Produktionsphase, d. h. ohne neuen Anlauf. Wird davon ausgegangen, dass die Maschine jedes Mal angehalten wird, bevor OK gemessen wurde, dann ist $\Delta k_{\text{OK}} = 1$ der erste relevante Wert. $\Delta k_{\text{OK}} = 0$ wäre nur dann ein sinnvoller Wert, wenn die Maschine grundsätzlich zur Messung nicht angehalten wird bzw. der Bediener vorausgesehen hat, dass mit diesem Schritt OK erreicht wird und die Maschine durchlaufen ließ.

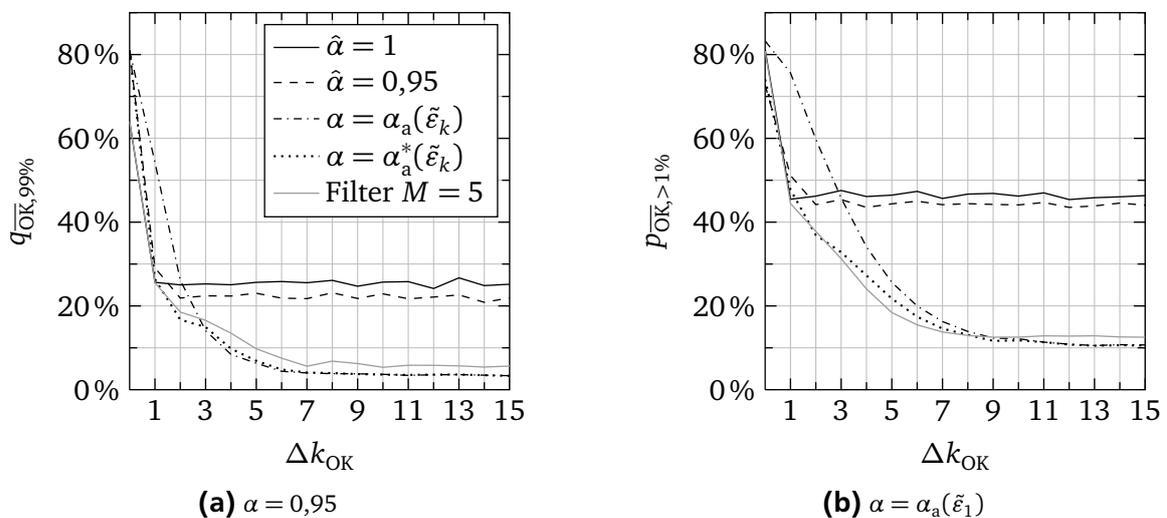


Abbildung 5.31: Vergleich der verschiedenen Regler ($\sigma_S = 0,5$, $\varepsilon_{\text{OK}} = 3$, über 10 000 Simulationen bestimmt)

Betrachtet man zunächst die Regler mit fester Verstärkung α , so fällt auf, dass diese Regler bei beiden Kenngrößen schon bei $\Delta k_{\text{OK}} = 1$ (ungefähr) ihren letztlichen Endwert erreichen. Diese Endwerte liegen für beide Regler mit festem α deutlich über denen der übrigen dargestellten Regler, wobei der Regler mit $\hat{\alpha} = 0,95$ ein wenig besser als der Regler mit $\hat{\alpha} = 1$ abschneidet.

Mit $\hat{\alpha} = 1$ ergibt sich durch den angenommenen Fehler in $\hat{\vartheta}_1$ ein wahres α von 0,91. Daraus ergibt sich für $k \rightarrow \infty$ ein $\text{Var}\{\varepsilon_{\infty}\}_T$ von ca. 0,835 (Gl. (5.28)), woraus sich nach dem auf Seite 49 in der Fußnote beschriebenen Vorgehen ungefähr 20 % als untere Grenze für $q_{\overline{\text{OK}},99\%}$ berechnen lässt, was sich mit den Ergebnissen in Abbildung 5.31a deckt.

Wäre $\hat{\alpha} = \alpha = 1$, so würde diese untere Grenze bei 35 % liegen. Wäre dagegen $\hat{\vartheta}_{1,r} = 0,9 \cdot \vartheta_1$, so ergäbe sich mit $\hat{\alpha} = 1$ über $\alpha = 1,1$ zunächst $\text{Var}\{\varepsilon_{\infty}\}_r = 1,22$ und damit 56 % als untere Grenze von $q_{\overline{\text{OK}},99\%}$! Bezüglich dieser Werte ist, wie bei der Beschreibung des Testfalls in Abschnitt 5.5.1 schon erwähnt, ein zu niedrig geschätztes θ_1 und damit ein zu hohes α deutlich kritischer.

Dieses Verhalten der festeingestellten Regler spiegelt sich auch in Abbildung 5.14b wider. Für $\alpha = 0,9$ ist der MSE für $k = 2$ relativ hoch, aber ab $k = 3$ ist kaum ein Unterschied zum Grenzwert des MSE für $k \rightarrow \infty$ auszumachen. In der Hälfte der Simulationen ist mit $k = 2$ auch der OK-Bereich schon erreicht, so dass der hohe Wert des MSE bei $k = 2$ die hohen – und damit schlechten – Werte der Größen $q_{\overline{\text{OK}},99\%}$ und $p_{\overline{\text{OK}},>1\%}$ für $\Delta k_{\text{OK}} = 0$ erklärt. Ab den nächsten Regelschritten ist der MSE bei $\alpha = 0,9$ näherungsweise konstant, und so ändern sich auch $q_{\overline{\text{OK}},99\%}$ und $p_{\overline{\text{OK}},>1\%}$ nicht mehr, wenn man noch Nachführungen durchführt.

Für $\hat{\alpha} = 0,95$ ergibt sich $\alpha = 0,86$. Damit lässt sich in Abbildung 5.14b erkennen, dass hier auch bei $k = 3$ der MSE ein Stück von seinem Endwert entfernt ist. Entsprechend verbessern sich die Kenngrößen bei der ersten Nachführung ($\Delta k_{\text{OK}} = 2$) noch etwas, um dann konstant zu bleiben.

Bei $\Delta k_{\text{OK}} = 1$ sind der Regler mit Filter sowie der erweiterte abweichungsabhängige Regler $\alpha_a^*(\tilde{\varepsilon}_k)$ ungefähr gleich gut wie der Regler mit $\hat{\alpha} = 1$, verbessern sich aber bis ca. $\Delta k_{\text{OK}} = 8$ stetig und erreichen deutlich bessere Werte für $q_{\overline{\text{OK}},99\%}$ und $p_{\overline{\text{OK}},>1\%}$.

Der abweichungsabhängige Regler $\alpha_a(\tilde{\varepsilon}_k)$ ohne die erweiterte Logik erreicht bei vielen Nachführungen letztlich auch die Werte des erweiterten Reglers $\alpha_a^*(\tilde{\varepsilon}_k)$, jedoch ist er bei $\Delta k_{\text{OK}} = 2$ und 3 (für $q_{\overline{\text{OK}},99\%}$) bzw. $\Delta k_{\text{OK}} = 2, \dots, 7$ (für $p_{\overline{\text{OK}},>1\%}$) zum Teil deutlich schlechter als der erweiterte Regler. Dies liegt an dem Kriechen, welches auch in Abbildung 5.29b deutlich zu sehen ist. In Abbildung 5.27a zeigt sich dieses Verhalten im nur langsamen Abbau des MSE bei $1/\kappa_{\theta} = 0,91$. Bei der erweiterten Variante existiert dies nicht, wie in Abbildung 5.28b zu sehen ist, was sich auch durch die Simulationsergebnisse bestätigt.

5.5.3 Fazit

Zunächst kann festgehalten werden, dass sich die Wahl des MSE als Gütekriterium bestätigt hat. Es hat sich darüber hinaus gezeigt, dass sich schon kleine Unterschiede beim MSE deutlich auf die Verteilung von $\{p_{\overline{\text{OK}}}\}_r$ auswirken können, wenn knappe Anforderungen an die Toleranz ε_{OK} gestellt sind.

Wenn sich die Anforderungen ohne Nachführungen, d. h. $\Delta k_{\text{OK}} = 1$, erfüllen lassen, dann bringt ein Regler mit Filter oder abweichungsabhängiger Verstärkung bezüglich der Anforderungen keinen Vorteil gegenüber einem Regler mit festem α , jedoch würde sich auch in diesem Fall die Varianz von $\{\varepsilon_k\}_r$ über die weiteren Nachführungen verringern, was sich auch in kleineren Stellgrößenänderungen niederschlägt.

Können die Anforderungen nicht mit einem festeingestellten Regler erreicht werden, so ist ein Regler mit abweichungsabhängiger Reglerverstärkung oder ein Filter zu verwenden. Dabei sind dann aber auch immer zwingend Nachführungen nötig.

Ist abzusehen, dass die Maschine mit dem folgenden Regelschritt den OK-Bereich erreicht und lässt man diese daher durchlaufen, so hat das Beispiel gezeigt, dass dann (bei $\Delta k_{\text{OK}} = 0$) das Verbleiben im OK-Bereich deutlich schlechter ausgeprägt sein kann.

5.6 Zusammenfassung

In diesem Kapitel wurde die Regelung des gestörten Systems behandelt. Dazu wurden zunächst die Regelziele in eine entsprechend angepassten Form gebracht, wobei unterschieden wurde zwischen Regelzielen, die das Erreichen des OK-Bereichs betreffen (z. B. Anzahl Regelschritte, maximales Überschwin-

gen), und solchen, die den Verbleib im OK-Bereich, d. h. die OK-Quote bzw. $\overline{\text{OK}}$ -Quote der Fertigteile in der Produktionsphase beschreiben.

Es wurden kurz Standardverfahren aus der Literatur zur Regelung stochastischer Systeme vorgestellt, und dargelegt, weshalb diese nicht (direkt) verwendet wurden.

Für den Reglerentwurf wurde zunächst die Reglerverstärkung des aus Kapitel 3 bekannten I-Reglers betrachtet und ein Bereich für diese bestimmt, mit der die Regelziele bezüglich des Erreichens des OK-Bereichs erfüllt werden.

Danach wurde der MSE als Gütekriterium für den Reglerentwurf betrachtet. Dieses wurde bei der Untersuchung verschiedener Varianten des I-Reglers verwendet. Zuletzt wurde das Verhalten der verschiedenen Regler bezüglich des Verbleibs im OK-Bereich bewertet und damit auch die Wahl des MSE als Gütekriterium bestätigt.

Damit wurde dann eine Bewertung der verschiedenen Regler möglich und es konnten die Einflüsse verschiedener Faktoren auf das Erreichen der Regelziele dargelegt werden.

6 Online-Identifikation und Adaption

6.1 Einführung

In den Kapiteln 3 und 5 wurde die Reglerauslegung bei (zumindest in Grenzen) bekannten Streckenparametern behandelt. Dabei wurde insbesondere für θ_1 ein Wertebereich angenommen, in dem dieser Parameter liegt. Natürlich ist der damit entworfene Regler nicht optimal für alle zulässigen Werte für θ_1 . Insbesondere wenn für das vom Regler angenommene $\hat{\theta}_{1,r} < 0,5 \cdot \theta_1$ gilt, ist der geschlossene Regelkreis für $\hat{\alpha} = 1$ instabil. Da dies bei (zunächst noch) unbekanntem Werkzeug- und Rohstoffvarianten durchaus auftreten kann, ist eine Adaption des Reglers nicht nur zur Verbesserung des Einregelverhaltens wünschenswert, sondern zur Sicherstellung der Stabilität zwingend notwendig. Auch verbessern sich die Kompensation der sekundären Eingangsgrößen und die Voreinstellung durch eine Adaption, wobei hier noch θ_s und θ_0 berücksichtigt werden muss.

Der Regelkreis mit Adaption hat den in Abbildung 6.1 gezeigten Aufbau. Dabei ist der Regler unverändert aus Kapitel 5 übernommen. Die Reglerparameter $\hat{\theta}_{1,r}$, $\hat{\theta}_{s,r}$ und $\hat{\theta}_{0,r}$ sind jedoch nicht mehr fest, sondern werden von dem Block Adaption anhand der Messungen $(\tilde{u}_{1,k}, \tilde{\mathbf{u}}_{s,k}, \tilde{y}_k)$ bzw. $(u_{1,k}, \mathbf{u}_{s,k}, \tilde{y}_k)$ während der Regelvorgänge bestimmt. Wie diese Online-Identifikation bzw. das Lernen durchgeführt wird, ist Thema dieses Kapitels.

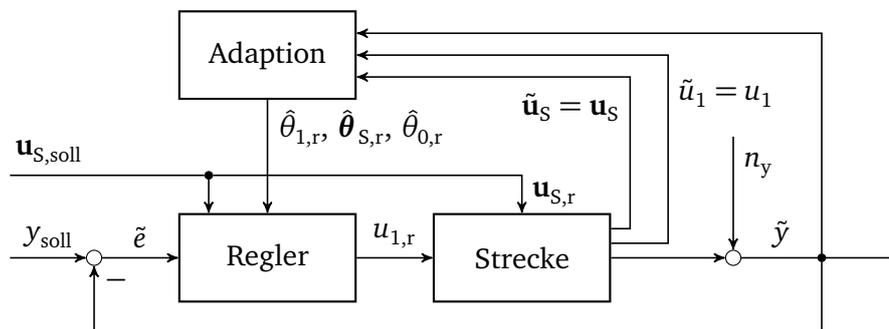


Abbildung 6.1: Regelkreis mit Adaption

Für die Identifikation kann unterschieden werden, ob die Parameter

- während eines Auftrags (Einricht- und Produktionsphase) oder
- über mehrere Aufträge

bestimmt werden.

In der Einrichtphase kann i. d. R. davon ausgegangen werden, dass die Parameter, insbesondere auch θ_0 konstant sind.¹ Je nach Dauer des Auftrags könnte in der Produktionsphase schon eine Drift des Parameters θ_0 auftreten. Ausreißer können natürlich auch während der Einrichtphase auftreten.

Über mehrere Aufträge können sich alle Parameter ändern. Insbesondere θ_0 kann deutlich springen. Dies muss entsprechend berücksichtigt werden. Für die weiteren Parameter wird maximal ein leichter Drift über mehrere Aufträge angenommen. Liegt ein System vor, bei dem auch hier Sprünge vorkommen können, so ist ein Lernen der Parameter über mehrere Aufträge schlicht nicht sinnvoll.

¹ Für Fälle, in denen dennoch ein Sprung von θ_0 auftritt, siehe Abschnitt 6.7.2.

Weiteres Vorgehen

Im Anschluss wird die Identifikationsaufgabe kurz beschrieben und die Benennung der Größen erläutert. Im Abschnitt 6.2 wird eine Übersicht über die relevante Literatur gegeben.

Daraufhin wird in Abschnitt 6.3 die Identifikation mithilfe des Least-Squares-Verfahrens allgemein vorgestellt und dessen Verwendung begründet.

In Abschnitt 6.4 wird dann die Schätzung während eines Auftrages behandelt, wobei die Thematik der Identifikation im geschlossenen Regelkreis zunächst zurückgestellt wird.

Es folgt die Schätzung über mehrere Aufträge in Abschnitt 6.5.

Die Folgen, die durch die Identifikation im geschlossenen Regelkreis auftreten, werden dann in Abschnitt 6.6 betrachtet.

Die praktischen Versuche haben gezeigt, dass Erweiterungen des Algorithmus notwendig bzw. wünschenswert sind. Die aufgetretenen Probleme und deren Lösung werden in Abschnitt 6.7 behandelt.

Der ausgewählte Algorithmus wird in Abschnitt 6.8 zusammengefasst und in Abschnitt 6.9 in den Kontext der Adaptiven Regelungen eingeordnet. Vor der Zusammenfassung dieses Kapitels wird in Abschnitt 6.10 beispielhaft eine Messung an dem Beispielsystem Druckmaschine vorgestellt.

6.1.1 Identifikationsaufgabe

Ziel ist es, die Parameter θ_i der Funktion

$$y_k = \theta_0 + \theta_1 \cdot (u_{1,k} - u_{1,N}) + \boldsymbol{\theta}_S^T \cdot (\mathbf{u}_{S,k} - \mathbf{u}_{S,N})$$

anhand der bei den Eingangsgrößen $u_{1,k}$ und $\mathbf{u}_{S,k}$ gemessenen Ausgangswerten \tilde{y}_k zu bestimmen, um diese dann zur Regelung zu verwenden. Dabei sind die Messungen durch ein Rauschen gestört, d. h.

$$\tilde{y}_k = y_k + n_{y,k}.$$

Für die Diskussion wird im Folgenden meist davon ausgegangen, dass $\{n_{y,k}\}_r \sim N(0, \sigma_y^2)$ und unkorreliert ist. Damit gilt für die Messwerte

$$\{\tilde{y}_k\}_r \sim N(y_k, \sigma_y^2).$$

Betrachtet man das Regelgesetz (3.14) und dabei insbesondere die Kompensation der sekundären Eingänge, so wird deutlich, dass entweder die auch oben auftretenden Parameter $\theta_1, \dots, \theta_p$ benötigt werden, oder alternativ θ_1 und q_2, \dots, q_p , wobei q_i für die Quotienten

$$q_i = \frac{\theta_i}{\theta_1}, \quad i = 2, \dots, p$$

steht. Für eine gute Kompensation der sekundären Eingänge sind die Verteilungseigenschaften der q_i auch wesentlicher als die von θ_i und θ_1 alleine.

6.1.2 Benennung der Größen

Die gemessene Ausgangsgröße wird mit \tilde{y} bezeichnet. Die Eingangsgrößen u_1 und \mathbf{u}_S sind in Abbildung 6.1 ebenfalls mit Tilden gekennzeichnet ($\tilde{u}_1, \tilde{\mathbf{u}}_S$). Es handelt sich dabei aber nicht um Messgrößen.

Die Motivation dieser Kennzeichnung liegt darin, dass der Bediener der Maschine diese Werte manuell ändern können soll. Damit müssen auch diese Daten wieder an die Regelung übertragen werden, so dass diese aus dieser Sicht zu den „Messgrößen“ gezählt werden. Im Weiteren werden hier aber ausschließlich die Bezeichnungen u_1 und \mathbf{u}_s verwendet.

Die Schätzwerte, die der Algorithmus während eines Auftrags aus den aktuellen Messdaten berechnet, werden mit einem Dach markiert, also $\hat{\theta}_0$, $\hat{\theta}_1$ und $\hat{\theta}_s$. Diese werden aber nicht unbedingt in dieser Form für die Regelung verwendet. Die Parameterwerte, die an den Regler übergeben werden, sind neben dem Dach noch mit dem Index „r“ gekennzeichnet, d. h. $\hat{\theta}_{0,r}$, $\hat{\theta}_{1,r}$ und $\hat{\theta}_{s,r}$.

In Abschnitt 6.5 treten noch gemittelte Schätzwerte auf. Diese besitzen den Index „m“, also $\hat{\theta}_{0,m}$, $\hat{\theta}_{1,m}$ und $\hat{\theta}_{s,m}$.

Die „Zeitvariable“ wird mit k und k' bezeichnet, wobei letztere Bezeichnung für Summenbildungen verwendet wird. So werden zur Schätzung des Parameters θ_k nach dem k -ten Schritt die Eingangs- und Messwerte der Schritte $k' = 1, \dots, k$ verwendet. Der Index k wird an den auftretenden Matrizen und Vektoren weggelassen, wenn Verwechslungen ausgeschlossen sind und es der besseren Lesbarkeit dient, z. B. $\Psi_k = \Psi$.

Bei der Betrachtung der Identifikation über mehrere Aufträge wird die „Zeitvariable“ mit l und l' bezeichnet.

6.2 Literatur

Wie beschrieben besteht die Aufgabe darin, aus Messdaten und der gegebenen Modellstruktur die Parameter des Systems zu bestimmen. Regelungstechnisch wird dies als parametrische Systemidentifikation bezeichnet [ISERMANN und MÜNCHHOF, 2011, S. 18].

Die Systemidentifikation ist Thema vieler Lehrbücher und anderer Veröffentlichungen. Beispiele für Grundlagenwerke sind [ISERMANN und MÜNCHHOF, 2011], [LJUNG, 1999]. Dabei liegt das Augenmerk auf dynamischen Systemen, wobei aber auch statische Systeme (Kennfeldermittlung) behandelt werden, z. B. [ISERMANN und MÜNCHHOF, 2011, S. 501ff], [NELLES, 1999, S. 8ff, 38ff].

Bei den genannten Quellen liegt der Schwerpunkt darauf, kompliziertere statische Zusammenhänge mit einfachen Funktionen zu approximieren oder überhaupt erst zu identifizieren und nicht darauf, die Koeffizienten eines bekannten Zusammenhangs mit wenigen Messungen zu bestimmen.

Die Beschreibung statischer Zusammenhänge durch gegebene (mögliche) Einflussgrößen ist auch Thema der Regressionsrechnung, die u. A. in der Wirtschaftslehre verbreitet Anwendung findet. Beispiele für Lehrbücher, die im Folgenden auch verwendet werden, sind [RAWLINGS et al., 1998], [FREUND et al., 2006] und [WOOLDRIDGE, 2003].

Die Aufgabe, den Einfluss sehr vieler möglicher Einflussgrößen auf ein System zu untersuchen und relevante Abhängigkeiten zu identifizieren, fällt ebenfalls in diesen Bereich und wird mit „Design of Experiments“ bezeichnet, siehe z. B. [HINKELMANN und KEMPTHORNE, 2008].²

Die starke Abhängigkeit zwischen den Eingangsgrößen, auch als Multikollinearität bezeichnet, ist ein Thema, mit dem sich weniger die regelungstechnische Literatur als die Literatur zur Regressionsrechnung beschäftigt.

² Diese grundlegende Identifikation des Prozesses, d. h. die Bestimmung der grundsätzlichen Einflüsse der Roh- und Hilfsstoffe, der Prozessparameter sowie der Umweltbedingungen und anderer (messbarer) Störungen auf die Ausgangsgröße, ist auch Voraussetzung, um die in dieser Arbeit beschriebene Methodik einsetzen zu können. Entsprechend wurde im Rahmen des eingangs beschriebenen Projektes geeignetes Vorwissen verwendet, siehe z. B. [BERTI et al., 2010, S. 10ff], sowie eigene Versuche zur Identifikation durchgeführt, [LENZ, 2012, S. 15ff]. Im Rahmen dieser Arbeit wird darauf jedoch nicht weiter eingegangen.

Ein Aspekt, der dagegen eher in der regelungstechnischen Literatur zu finden ist, ist die Problematik, die aufgrund der Identifikation innerhalb eines geschlossenen Regelkreises auftritt. Dies führt, wie später gezeigt wird, zu einem Bias der geschätzten Parameter.

6.3 Schätzung über die Methode der kleinsten Quadrate

6.3.1 Schätzgleichung

Die Herleitung des LS-Verfahrens ist beispielsweise in [ISERMANN und MÜNCHHOF, 2011, S. 203ff] nachzulesen und im Anhang B.1 kurz dargestellt. Hier werden nur die wichtigsten Ergebnisse zusammengefasst.

Ausgangspunkt ist die Modelgleichung

$$y_{k'} = \boldsymbol{\psi}_{k'}^T \boldsymbol{\theta} \quad (6.1)$$

mit

$$\boldsymbol{\psi}_{k'}^T = \begin{bmatrix} 1 & u_{1,k'} - u_{1,N} & \cdots & u_{p,k'} - u_{p,N} \end{bmatrix} \quad \text{und} \\ \boldsymbol{\theta} = \begin{bmatrix} \theta_0 & \theta_1 & \cdots & \theta_p \end{bmatrix}^T .$$

Für k Messpunkte können die entstehenden k Gleichungen in der Form

$$\mathbf{y}_k = \boldsymbol{\Psi}_k \boldsymbol{\theta} \quad (6.2)$$

mit

$$\mathbf{y}_k = \begin{bmatrix} y_1 \\ \vdots \\ y_k \end{bmatrix} \quad \text{und} \quad \boldsymbol{\Psi}_k = \begin{bmatrix} \boldsymbol{\psi}_1^T \\ \vdots \\ \boldsymbol{\psi}_k^T \end{bmatrix} = \begin{bmatrix} 1 & u_{1,1} - u_{1,N} & \cdots & u_{p,1} - u_{p,N} \\ \vdots & \vdots & & \vdots \\ 1 & u_{1,k} - u_{1,N} & \cdots & u_{p,k} - u_{p,N} \end{bmatrix}$$

geschrieben werden.

Aufgrund des Messrauschens kann jedoch nie $y_{k'}$, sondern immer nur die gestörte Größe $\tilde{y}_{k'}$ gemessen werden,

$$\tilde{\mathbf{y}}_k = \boldsymbol{\Psi}_k \boldsymbol{\theta}_k + \mathbf{n}_{y,k} .$$

Gesucht ist ein Schätzwert $\hat{\boldsymbol{\theta}}_k$ für den Parametervektor, der den Fehler

$$\hat{\mathbf{e}}_k = \tilde{\mathbf{y}}_k - \boldsymbol{\Psi}_k \hat{\boldsymbol{\theta}}_k = \tilde{\mathbf{y}}_k - \hat{\mathbf{y}}_k$$

im quadratischen Sinne minimiert, d. h. $\sum \hat{e}_{k'}^2 = \hat{\mathbf{e}}_k^T \hat{\mathbf{e}}_k \rightarrow \min$.

Dieser Schätzwert $\hat{\boldsymbol{\theta}}_k$ ist durch

$$\hat{\boldsymbol{\theta}}_k = (\boldsymbol{\Psi}_k^T \boldsymbol{\Psi}_k)^{-1} \boldsymbol{\Psi}_k^T \tilde{\mathbf{y}}_k . \quad (6.3)$$

gegeben.

Bezeichnet man $\mathbf{L}_k = \boldsymbol{\Psi}_k^T \boldsymbol{\Psi}_k$ als Informationsmatrix ([KRAUS, 1986, S. 30]), und verwendet man die Abkürzung $\mathbf{r}_k = \boldsymbol{\Psi}_k^T \tilde{\mathbf{y}}_k$, so lässt sich die Schätzgleichung auch als

$$\hat{\boldsymbol{\theta}}_k = \mathbf{L}_k^{-1} \mathbf{r}_k \quad (6.4)$$

schreiben.

Die Informationsmatrix \mathbf{L}_k ist symmetrisch und mindestens positiv semidefinit, d. h. es gilt $\mathbf{x}^T \mathbf{L}_k \mathbf{x} \geq 0$ für alle Vektoren \mathbf{x} .³

Ist \mathbf{L}_k symmetrisch und positiv definit, so ist auch die Inverse \mathbf{L}_k^{-1} symmetrisch⁴ und positiv definit⁵. Aus der positiven Definitheit folgt auch, dass alle Diagonaleinträge der Matrizen \mathbf{L}_k und \mathbf{L}_k^{-1} positiv sind⁶.

6.3.2 Eigenschaften

Um weitere Eigenschaften, insbesondere bezüglich der geschätzten Parameter $\hat{\boldsymbol{\theta}}$ angeben zu können, müssen Annahmen getroffen werden. Die folgende Aufzählung fasst die „klassischen“ Annahmen zusammen. Anschließend werden verschiedene Eigenschaften vorgestellt, für die zum Teil jedoch nicht alle unten genannten Annahmen zutreffen müssen. [WOOLDRIDGE, 2003, S. 790ff]

Annahmen

A1: $y = \boldsymbol{\psi}^T \boldsymbol{\theta}$:

Das ungestörte System wird durch die Gleichung (6.1) vollständig beschrieben.

A2: $E\{\mathbf{n}_y\}_r = \mathbf{0}$ und nicht mit $\boldsymbol{\Psi}$ korreliert:

Das Rauschen hat den Erwartungswert Null und ist nicht mit den Eingangsgrößen korreliert.

A3: $\text{rg}(\boldsymbol{\Psi}) = p + 1$ ($\boldsymbol{\Psi}$ besitzt vollen Spaltenrang):

Keine Eingangsgröße ist konstant und es existieren keine exakten linearen Abhängigkeiten unter den Eingangsgrößen.

A4: $\text{Cov}(\{\mathbf{n}_y\}_r, \{\mathbf{n}_y\}_r) = \sigma_y^2 \cdot \mathbf{I}$:

Die Varianz des Rauschens ist konstant σ_y^2 und das Rauschen ist unkorreliert

A5: $\{n_{y,k}\}_r \sim N(0, \sigma_y^2)$:

Das Rauschen ist normalverteilt

Die wesentlichen Eigenschaften von Schätzern sind Bias, Konsistenz und Effizienz. Die Definitionen können bspw. in [ISERMANN und MÜNCHHOF, 2011, S. 687f] nachgeschlagen werden.

Biasfreiheit

Unter Bias wird die Abweichung des Erwartungswerts der Schätzwerte $\hat{\boldsymbol{\theta}}$ vom tatsächlichen Wert $\boldsymbol{\theta}$ verstanden, $\mathbf{b} = E\{\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}\}_r$.

Unter den Annahmen A1, A2 und A3 ist der Schätzwert $\hat{\boldsymbol{\theta}}$ biasfrei, $E\{\hat{\boldsymbol{\theta}}\}_r = \boldsymbol{\theta}$. Dies gilt für jedes beliebige k und nicht nur für $k \rightarrow \infty$.

Konsistenz

Ein Schätzer wird als konsistent bezeichnet, wenn der Schätzwert für $k \rightarrow \infty$ fast sicher bzw. mit der Wahrscheinlichkeit eins dem wahren Wert entspricht,

$$P\left(\lim_{k \rightarrow \infty} \hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta} = \mathbf{0}\right) = 1.$$

³ Jede Matrix $\mathbf{M}^T \mathbf{M}$ ist positiv semidefinit, da $\mathbf{x}^T \mathbf{M}^T \mathbf{M} \mathbf{x} = \mathbf{y}^T \mathbf{y}$ und $\mathbf{y}^T \mathbf{y} \geq 0 \forall \mathbf{y}$. (Zur Definition und den hier genannten Eigenschaften symmetrischer und positiv (semi)definiten Matrizen siehe auch [HORN und JOHNSON, 1985, S. 169, 396ff].)

⁴ $\mathbf{I} = (\mathbf{L} \cdot \mathbf{L}^{-1})^T = (\mathbf{L}^{-1})^T \cdot \mathbf{L}^T = (\mathbf{L}^{-1})^T \cdot \mathbf{L} \Rightarrow (\mathbf{L}^{-1})^T = \mathbf{L}^{-1}$.

⁵ Es gelte $\mathbf{x}^T \mathbf{L} \mathbf{x} \geq 0$. Mit $\mathbf{x} = \mathbf{L}^{-1} \mathbf{y}$ gilt für alle \mathbf{y} , dass $\mathbf{x}^T \mathbf{L} \mathbf{x} = \mathbf{y}^T \mathbf{L}^{-1} \mathbf{L} \mathbf{L}^{-1} \mathbf{y} = \mathbf{y}^T \mathbf{L}^{-1} \mathbf{y} \geq 0$.

⁶ Mit dem i -ten Einheitsvektor \mathbf{e}_i gilt $\mathbf{e}_i^T \mathbf{L} \mathbf{e}_i = (\mathbf{L})_{ii} \geq 0$.

Ein Schätzer wird als „konsistent im quadratischen Mittel“ bezeichnet, wenn

$$\lim_{k \rightarrow \infty} E\{(\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}) \cdot (\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta})^T\}_r = \mathbf{0}$$

gilt. In dieser Arbeit wird immer letztere Definition verwendet, und daher wird im Weiteren auch kurz „konsistent“ für „konsistent im quadratischen Mittel“ geschrieben.

Ein konsistenter Schätzer ist (zumindest) asymptotisch biasfrei und die Varianz geht für $k \rightarrow \infty$ gegen Null.

Unter den Annahmen A1 bis A4 ist der LS-Schätzer konsistent.

Effizienz

Unter Effizienz wird die Eigenschaft verstanden, dass ein Schätzer die minimale Varianz innerhalb einer bestimmten Gruppe von Schätzern besitzt.

Unter den Annahmen A1 bis A4 besitzt der LS-Schätzer die geringste Varianz⁷ aller linearen, biasfreien Schätzer (Gauß-Markov-Theorem, [WOOLDRIDGE, 2003, S. 792]). Dabei werden unter linearen Schätzern alle Schätzer verstanden, deren Schätzgleichung in der Form $\hat{\boldsymbol{\theta}} = \mathbf{M}(\boldsymbol{\Psi}) \cdot \tilde{\mathbf{y}}$ geschrieben werden kann, wobei die Einträge der Matrix \mathbf{M} Funktionen der Eingangsgrößen sein können, aber unabhängig von den gemessenen Ausgangsgrößen $\tilde{\mathbf{y}}$ sein müssen.

Wird die Annahme A5 hinzugenommen, so lässt sich zeigen, dass die Varianz des LS-Schätzers gerade der Cramer-Rao-Grenze entspricht, welche die geringste Varianz aller (nicht nur linearen) biasfreien Schätzer darstellt. Somit kann kein biasfreier Schätzer besser als der LS-Schätzer sein und dieser ist damit effizient. [WOOLDRIDGE, 2003, S. 794f]⁸

Varianz der Schätzwerte

Unter den Annahmen A1 bis A4 gilt für die Kovarianzmatrix $\mathbf{P} = \text{Cov}(\{\hat{\boldsymbol{\theta}}\}_r, \{\hat{\boldsymbol{\theta}}\}_r)$ der Schätzwerte

$$\mathbf{P} = \sigma_y^2 \cdot \mathbf{L}^{-1}. \quad (6.5)$$

Die Varianz der Schätzwerte ist eine wesentliche Größe, so dass auf diese noch öfters eingegangen werden wird. Die prinzipiellen Einflussfaktoren auf die Varianz werden in Abschnitt 6.3.3 diskutiert.

Verteilung der Schätzwerte

Unter den Annahmen A1 bis A5 sind die Schätzwerte $\hat{\boldsymbol{\theta}}$ ebenfalls normalverteilt.

Verlauf der Varianz für wenige Messungen

Wie gerade beschrieben ist der LS-Schätzer konsistent, d. h. die Varianz der Schätzwerte geht gegen Null, wenn die Anzahl der Messpunkte k gegen unendlich geht. Dies ist natürlich eine sinnvolle Eigenschaft für einen Schätzer. Im vorliegenden Fall ist jedoch besonders auch das Verhalten bei wenigen Messungen von Interesse. Insbesondere ist es wesentlich, dass die Varianz der Schätzwerte mit wachsender Anzahl an Messpunkten monoton abnimmt, d. h. es soll

$$\text{Var}\{\hat{\theta}_{i,k+1}\}_r \leq \text{Var}\{\hat{\theta}_{i,k}\}_r, \quad i = 1, \dots, p \quad (6.6)$$

⁷ Dies bedeutet, dass die Varianz jedes Schätzwertes θ_i minimal ist.

⁸ Für die Herleitung der Cramer-Rao-Grenze siehe z. B. [HÄNSLER, 2001, S. 434ff].

gelten, wobei p die Anzahl der zu schätzenden Parameter ist.

Die praktische Bedeutung liegt darin, dass damit jede neue Messung ohne weitere Prüfungen mit den bestehenden Messungen kombiniert werden kann. Wäre dies nicht gegeben, so müsste man die Messpunkte einzeln vorhalten und in jedem Schritt prüfen, welche Kombination aus den bisherigen Messpunkten die optimale, d. h. minimale Varianz ergibt.

Das Gl. (6.6) erfüllt ist, kann man zum Einen mit der Eigenschaft der Effizienz begründen. Dazu betrachtet man zwei Schätzer für $k + 1$ Messungen. Den normalen LS-Schätzer, und einen modifizierten LS-Schätzer, der nur die ersten k Messwerte verwendet. Da der normale LS-Schätzer effizient ist, kann es keinen besseren Schätzer bei $k + 1$ Messungen geben, insbesondere kann der modifizierte LS-Schätzer nicht besser sein, d. h. seine Schätzwerte können keine geringere Varianz haben als die des normalen LS-Schätzers. Demnach kann die Varianz bei dem normalen LS-Schätzer nicht steigen, wenn die Anzahl der Messpunkte erhöht wird.

Alternativ kann man Gl. (6.6) auch wie folgt beweisen, ohne dass ein Rückgriff auf die Eigenschaft der Effizienz erfolgen muss.

Unter Annahme einer konstanten Varianz σ_y^2 des Messrauschens gilt

$$\text{Var}\{\hat{\theta}_{i,k}\}_r = \sigma_y^2 \cdot (\mathbf{L}_k^{-1})_{ii}, \quad i = 1, \dots, p,$$

und damit ist

$$(\mathbf{L}_{k+1}^{-1})_{ii} \leq (\mathbf{L}_k^{-1})_{ii}, \quad i = 1, \dots, p,$$

zu zeigen.

Für die Informationsmatrix \mathbf{L}_{k+1} zur Messung $k + 1$ gilt

$$\mathbf{L}_{k+1} = \mathbf{L}_k + \boldsymbol{\psi}_{k+1} \boldsymbol{\psi}_{k+1}^T.$$

Daraus kann die Inverse nach der Sherman-Morrison-Formel ([ZURMÜHL und FALK, 1984, S. 310]) angegeben werden,

$$\mathbf{L}_{k+1}^{-1} = \left(\mathbf{L}_k + \boldsymbol{\psi}_{k+1} \boldsymbol{\psi}_{k+1}^T \right)^{-1} = \mathbf{L}_k^{-1} - \frac{\mathbf{L}_k^{-1} \boldsymbol{\psi}_{k+1} \boldsymbol{\psi}_{k+1}^T \mathbf{L}_k^{-1}}{1 + \boldsymbol{\psi}_{k+1}^T \mathbf{L}_k^{-1} \boldsymbol{\psi}_{k+1}}.$$

Diese Gleichung ist auch Bestandteil des rekursiven LS-Verfahrens.

Formal erhält man die interessierenden Diagonaleinträge durch die Multiplikation mit den entsprechenden Einheitsvektoren von links und von rechts,

$$(\mathbf{L}_{k+1}^{-1})_{ii} = \mathbf{e}_i^T \mathbf{L}_{k+1}^{-1} \mathbf{e}_i.$$

Die i -te Spalte von \mathbf{L}_k^{-1} wird mit $\mathbf{p}_{k,i}$ bezeichnet, d. h. $\mathbf{L}_k^{-1} \mathbf{e}_i = \mathbf{p}_{k,i}$. Aufgrund der Symmetrie von \mathbf{L}_k^{-1} gilt damit auch $\mathbf{e}_i^T \mathbf{L}_k^{-1} = \mathbf{p}_{k,i}^T$, d. h. die i -te Zeile ist die Transponierte der i -ten Spalte. Damit ergibt sich für die Diagonalelemente der Ausdruck

$$(\mathbf{L}_{k+1}^{-1})_{ii} = (\mathbf{L}_k^{-1})_{ii} - \frac{\mathbf{p}_{k,i}^T \boldsymbol{\psi}_{k+1} \boldsymbol{\psi}_{k+1}^T \mathbf{p}_{k,i}}{1 + \boldsymbol{\psi}_{k+1}^T \mathbf{L}_k^{-1} \boldsymbol{\psi}_{k+1}}, \quad i = 1, \dots, p.$$

Da \mathbf{L}_k^{-1} positiv definit ist, ist der Nenner des Bruchs immer positiv. Der Zähler $\mathbf{p}_{k,i}^T \boldsymbol{\psi}_{k+1} \boldsymbol{\psi}_{k+1}^T \mathbf{p}_{k,i}$ ist ebenfalls immer größer oder gleich Null, da $\boldsymbol{\psi}_{k+1} \boldsymbol{\psi}_{k+1}^T$ positiv semidefinit ist. Damit ergibt sich der gesuchte Zusammenhang

$$0 \leq (\mathbf{L}_{k+1}^{-1})_{ii} \leq (\mathbf{L}_k^{-1})_{ii}, \quad i = 1, \dots, p,$$

wobei die zusätzlich angegebene erste Bedingung daher kommt, dass \mathbf{L}_{k+1} und damit auch \mathbf{L}_{k+1}^{-1} immer positiv definit ist und damit nie negative Diagonalelemente auftreten können.

6.3.3 Varianz

Die Kovarianzmatrix $\mathbf{P} = \text{Cov}(\{\hat{\theta}\}_r, \{\hat{\theta}\}_r)$ lautet $\mathbf{P} = \sigma_y^2 \cdot \mathbf{L}^{-1}$ (Gl. (6.5)). Die Varianzen $\text{Var}\{\hat{\theta}_j\}_r$ der Schätzwerte sind damit die Einträge auf der Diagonalen von \mathbf{P} , $\text{Var}\{\hat{\theta}_j\}_r = (\mathbf{P})_{jj}$.

Direkt ersichtlich und anschaulich ist, dass die Varianzen der Schätzwerte umso größer sind, je größer die Varianz σ_y^2 des Rauschens ist.

In [WOOLDRIDGE, 2003, S. 96] werden die Einflussfaktoren auf die Varianz der Schätzwerte anhand der Darstellung

$$\text{Var}\{\hat{\theta}_j\}_r = \frac{\sigma_y^2}{\text{SST}_j \cdot (1 - R_j^2)} \quad (6.7)$$

besprochen. Dies wird hier zusammenfassend wiedergegeben. In Gl. (6.7) ist

$$\text{SST}_j = \sum_{k'=1}^k (u_{j,k'} - \bar{u}_j)^2 \quad (6.8)$$

die totale, quadratische Variation des Eingangs u_j (\bar{u}_j ist der arithmetische Mittelwert des Eingangs u_j). Für die Größe R_j^2 wird eine „Schätzung“ $\hat{u}_{j,k'}$ der Eingangsgröße $u_{j,k'}$ über die verbleibenden Eingangsgrößen sowie einer Konstanten durchgeführt. D. h. es werden Parameter $\hat{\vartheta}_i$ so bestimmt, dass der „Schätzwert“

$$\hat{u}_{j,k'} = \hat{\vartheta}_0 + \hat{\vartheta}_1 u_{1,k'} + \dots + \hat{\vartheta}_{j-1} u_{j-1,k'} + \hat{\vartheta}_{j+1} u_{j+1,k'} + \dots + \hat{\vartheta}_p u_{p,k'}$$

den quadratischen Abstand $\sum (u_{j,k'} - \hat{u}_{j,k'})^2$ minimiert. Dann kann R_j^2 als das Quadrat des empirischen Korrelationskoeffizienten von $u_{j,k'}$ und $\hat{u}_{j,k'}$ geschrieben werden ([WOOLDRIDGE, 2003, S. 81]),

$$R_j^2 = \left(\frac{\sum_{k'=1}^k (u_{j,k'} - \bar{u}_j) \cdot (\hat{u}_{j,k'} - \bar{\hat{u}}_j)}{\sqrt{\sum_{k'=1}^k (u_{j,k'} - \bar{u}_j)^2 \cdot \sum_{k'=1}^k (\hat{u}_{j,k'} - \bar{\hat{u}}_j)^2}} \right)^2. \quad (6.9)$$

Aus Gl. (6.7) und (6.8) kann geschlossen werden, dass eine hohe Variation der Werte eines Eingangs u_j die Schätzung des dazugehörigen Parameters θ_j begünstigt. Alternativ zu hohen Änderungen der Eingangsgröße u_j kann SST_j dadurch erhöht werden, dass die Anzahl k der Messungen erhöht wird.

Kann $u_{j,k'}$ vollständig durch die anderen Eingangsgrößen $u_{i,k'}$, $i \neq j$, „erklärt“ werden, so wäre $R_j^2 = 1$ und die Varianz nach Gl. (6.7) nicht definiert. In diesem Fall würde ein exakter linearer Zusammenhang zwischen den Eingangsgrößen existieren, was nach der Voraussetzungen A3 ausgeschlossen ist. Je geringer ein Zusammenhang zwischen den Eingangsgrößen ist, desto mehr nähert sich der Wert für R_j^2 der Null an, und desto geringer wird die Varianz von $\hat{\theta}_j$.

Besteht keine exakte, aber eine hohe lineare Abhängigkeit der Eingangsgrößen untereinander, so spricht man von „Multikollinearität“. Diese Problematik und Möglichkeiten damit umzugehen sind ein Standardthema in der Literatur zur Regressionsanalyse.

6.3.4 Bewertung des Schätzers für das vorliegende Schätzproblem

Die Annahme A2 ist nicht vollständig zu erfüllen. Dass das Rauschen mittelwertfrei ist, ist zwar häufig – so wie auch in dem vorliegenden Beispielsystem – zutreffend, aber durch die Regelung ist das Rauschen

mit den Eingangswerten zwangsweise korreliert. Damit ist der Schätzer im Allgemeinen nicht mehr biasfrei. Auch ist der LS-Schätzer nicht mehr konsistent, wenn die Annahme A2 nicht erfüllt ist. Dies bedeutet, dass der Bias auch für $k \rightarrow \infty$ nicht verschwindet. [WOOLDRIDGE, 2003, S. 169]

Eine weitere Problematik ist die Multikollinearität. Diese liegt hier vor, da die Regelung bzw. Steuerung die primäre und sekundäre Eingangsgröße gerade so ändert, dass in Summe möglichst keine Wirkung auf den Ausgang entsteht. Damit liegen aber gerade Eingangsdaten vor, die für eine Identifikation der Parameter θ ungünstig sind.

Auch ist die Anzahl der Messwerte hier sehr gering, so dass mit hohen Varianzen zu rechnen ist.

Die genannten Punkte stellen gewisse Einschränkungen bei der Verwendung des LS-Schätzers dar. Dennoch wird dieser für die Schätzung verwendet, da die genannten Punkte gelöst bzw. abgeschwächt werden können.

So wird sich zeigen, dass das Problem der Korrelation der Eingangswerte mit dem Rauschen beherrschbar ist. Die Multikollinearität der Eingangsgrößen ist prinzipiell nicht zu beseitigen, aber es werden Maßnahmen vorgestellt, mit denen die Varianzen der relevanten Größen möglichst niedrig gehalten werden können.

6.4 Identifikation während eines Auftrags

In diesem Abschnitt wird die Parameteridentifikation während eines Auftrags betrachtet. Es wird dazu das in Abschnitt 6.3 beschriebene LS-Verfahren als Grundlage verwendet.

In den Ablaufplänen zu den Verfahren sowie manchen Diskussionen wird ein Prozess zugrunde gelegt, der nur einen sekundären Eingang u_2 besitzt. Dies erleichtert die Übersicht und die Erweiterung auf Prozesse mit mehreren sekundären Eingängen ist in der Regel einfach möglich.

Bei der Durchführung der Schätzung nach den Gleichungen des angegebenen Abschnitts ist zunächst lediglich zu beachten, dass ein Parameter θ_i nur dann geschätzt werden kann, wenn sich der entsprechende Eingang u_i während des aktuellen Auftrags auch geändert hat. Genauer gesagt müssen sich die verschiedenen Eingänge linear unabhängig verändern.

Speziell bedeutet das, dass der primäre Parameter frühestens nach dem ersten Regelschritt geschätzt werden kann, wenn der Regler die Eingangsgröße verändert hat. Sollte sich zum ersten Regelschritt auch die sekundäre Eingangsgröße geändert haben, so muss ein weiterer Schritt abgewartet werden.

Solange der sekundäre Eingang u_2 sich nicht ändert können nur θ_0 und θ_1 geschätzt werden. Dazu kann, wie in Abschnitt 6.4.4 beschrieben wird, auf die entsprechenden Einträge in \mathbf{L} und \mathbf{r} zurückgegriffen werden, d. h. es müssen keine zusätzlichen Daten gespeichert werden.

Mathematisch ausgedrückt bedeutet diese Bedingung, dass die Matrix \mathbf{L} (bzw. $\mathbf{L}_{[1\ 2],[1\ 2]}$, solange sich u_2 nicht geändert hat) invertierbar sein muss, deren Rang also maximal ist. Praktisch ist es unter Umständen einfacher zu prüfen, ob sich die Eingangsgrößen geändert haben, anstatt eine Rangbestimmung durchzuführen.

Begriffe

Mit „Drei-Parameter-Schätzung“ wird die Schätzung der Parameter θ_0 , θ_1 und θ_2 mit dem beschriebenen LS-Verfahren bezeichnet.

Unter einer „Zwei-Parameter-Schätzung“ von θ_1 oder θ_2 – welcher Parameter gemeint ist wird immer erwähnt bzw. ist eindeutig aus dem Kontext ersichtlich – wird die Schätzung der Parameter θ_0 und θ_1 , bzw. θ_0 und θ_2 mit dem LS-Verfahren verstanden, wobei für θ_2 bzw. θ_1 ein fester Wert eingesetzt wird. (Auch wenn der Schätzwert für θ_0 aus der Zwei-Parameter-Schätzung nicht verwendet werden soll, so

muss dieser zwingend mitgeschätzt werden, da der interessierende Schätzwert $\hat{\theta}_1$ bzw. $\hat{\theta}_2$ ansonsten unbrauchbar ist.) Dies wird in Abschnitt 6.4.4 genauer betrachtet.

6.4.1 Überprüfen der Varianz

Wird die Least-Squares-Methode zur Schätzung der Parameter verwendet, dann kann – wie oben schon öfters erläutert – die Varianz der Schätzwerte bei bekannter Varianz des Messrauschens angegeben werden. Dies kann und sollte genutzt werden, um zu verhindern, dass zu ungenaue Parameter für die Regelung verwendet werden.

Die berechneten Varianzen für die Schätzwerte stimmen zwar nur, wenn die Annahmen A1 bis A4 erfüllt sind, doch auch wenn diese Punkte hier nicht (ideal) erfüllt werden, zeigen die Simulationen und Versuche, dass die berechneten Varianzen dennoch einen sehr guten Anhaltswert für die Güte der Schätzung liefern.

Die unbekannte Varianz σ_y^2 könnte man in der Gleichung

$$\mathbf{P} = \sigma_y^2 \cdot (\Psi^T \Psi)^{-1}$$

durch eine aus den Messdaten geschätzte Varianz $\hat{\sigma}_y^2$ ersetzen. Allerdings wäre diese Schätzung aufgrund der wenigen Messwerte sehr ungenau. Zudem muss die Störung auch nicht einer Normalverteilung entsprechen. Auf die Schätzung der Varianz des Prozess- und Messrauschens wird nochmals in Abschnitt 6.4.6 eingegangen.

Trotz dieser theoretischen Einschränkungen stellt diese Methode mit einem sinnvoll gewählten σ_y^2 eine effektive, wenn auch heuristische Methode dar, die Güte der Schätzung zu bewerten.

Verwenden der Parameter für die Regelung

Für jeden Parameter θ_i ist eine maximal zulässige Varianz $\sigma_{\hat{\theta}_{i,r,\max}}^2$ festzulegen. Wenn die Varianz des Schätzwertes $\hat{\theta}_{i,k}$ eines Parameters geringer als diese Grenze ist, so wird dieser für die Regelung verwendet. Ansonsten wird der bisherige Wert für die Regelung beibehalten,

$$\hat{\theta}_{i,r,k} = \begin{cases} \hat{\theta}_{i,k} & \text{wenn } \sigma_{\hat{\theta}_{i,k}}^2 \leq \sigma_{\hat{\theta}_{i,r,\max}}^2 \\ \hat{\theta}_{i,r,k-1} & \text{sonst.} \end{cases}$$

Dies ist im nächsten Abschnitt auch in Abbildung 6.3 dargestellt.

Wahl der Varianzgrenzen

Zum einen können die Grenzen anhand der Verteilungsfunktion der Normalverteilung festgelegt werden. So liegen bei einer standardverteilten Zufallsgröße 95 % der Werte in einem Bereich von $\pm 2\sigma$ um den Mittelwert. Soll eine Abweichung vom realen Wert um mehr als 20 % weniger als 5 % wahrscheinlich sein, so muss also $2\sigma_{\hat{\theta}_{i,r,\max}} = 0,2 \cdot \theta_i$ gelten. θ_i ist zwar nicht genau bekannt, aber mit einem angenommenen Wert $\tilde{\theta}_i$, der die Größenordnung richtig wiedergibt, erhält man so beispielsweise

$$\sigma_{\hat{\theta}_{i,r,\max}} = \frac{1}{10} \tilde{\theta}_i.$$

Zum anderen können die Grenzen von schon verwendeten Schätzalgorithmen abgeleitet werden. Wenn z. B. bekannt ist, dass eine Schätzung der Steigung bei einer einzelnen Differenz der Eingangsgröße ab $\Delta u_{i,\min}$ zu guten Ergebnissen führt, dann kann aus

$$\sigma_{\hat{\theta}_{i,r,\max}}^2 = \sigma_y^2 \frac{2}{\Delta u_{i,\min}^2}$$

das maximal zulässige $\sigma_{\hat{\theta}_{1,r,\max}}^2$ bestimmt werden.

Da die angenommene Varianz der Störung durchaus fehlerbehaftet sein kann und auch die Annahme einer Normalverteilung beim realen System nicht erfüllt sein muss, stellen die so bestimmten Grenzen nur Anhaltspunkte dar. In Versuchen sind die Grenzen dann gegebenenfalls noch anzupassen.

In Abbildung 6.2 ist der Einfluss der Varianzgrenze beispielhaft dargestellt, wobei der angenommene Parameterwert $\hat{\theta}_{1,r}$ zu Beginn doppelt so hoch wie der reale ist. Damit ergibt sich ohne Adaption ein kriechendes Verhalten. In Abbildung 6.2a ist die Grenze $\sigma_{\hat{\theta}_{1,r,\max}}$ auf ein Zehntel des (in der Simulation bekannten) realen Parameterwerts gesetzt. Erst bei $k = 4$, d. h. dem dritten Regelschritt ist die Variation des Eingangs so groß geworden, dass θ_1 teilweise geschätzt werden konnte. Und erst bei $k = 5$ ist dieser dann in allen Versuchen geschätzt worden. Entsprechend stark ist das kriechende Verhalten ausgeprägt.

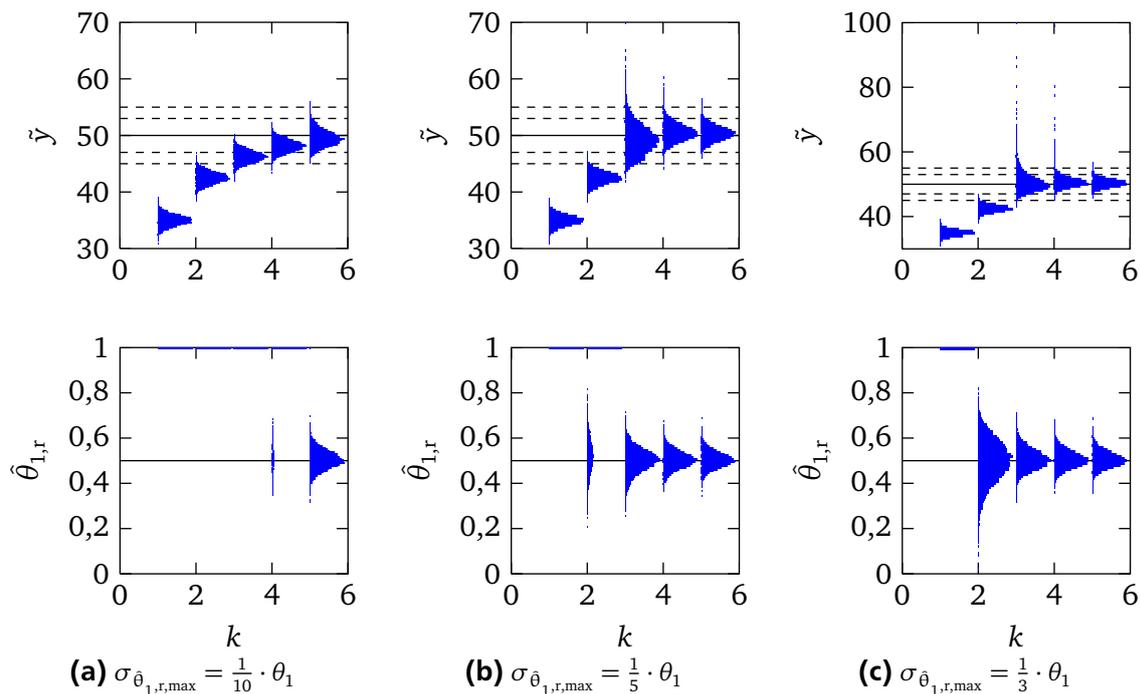


Abbildung 6.2: Beispiele für verschiedene Varianzgrenzen (Daten aus je 10 000 Simulationen)

In Abbildung 6.2b ist $\sigma_{\hat{\theta}_{1,r,\max}}$ auf ein Fünftel des Parameterwerts gesetzt. Hier wird schon bei $k = 2$, also dem frühestmöglichen Zeitpunkt begonnen, θ_1 zu schätzen. Allerdings ist die Streuung von $\{\hat{\theta}_{1,r,k}\}_r$ zunächst deutlich größer, was sich bei $\{\check{y}_k\}_r$ auch bei $k = 3$ und $k = 4$ in vereinzelt deutlichen Überschwingern zeigt. Insgesamt wird der Sollwert aber zügiger erreicht.

Zuletzt ist in Abbildung 6.2c der Fall betrachtet, dass $\sigma_{\hat{\theta}_{1,r,\max}}$ nur einem Drittel des Parameterwerts entspricht. Hier wird θ_1 in allen Simulationen schon bei $k = 2$ geschätzt, jedoch mit einer sehr großen Streuung, die vereinzelt zu extremen Überschwingern führt. (Diese wären teilweise noch größer, wenn die Eingangsgröße u_1 nicht beschränkt wäre.)

Bei $k = 5$ ist die Verteilung von $\{\hat{\theta}_{1,r,k}\}_r$ in allen drei Fällen ungefähr gleich, da auch die Daten ungefähr dieselben waren. Die Varianzgrenze hat demnach keinen Einfluss auf die erreichbare Güte der Schätzung, sie lehnt nur (wahrscheinlich) zu schlechte Schätzungen ab.

Schätzung des primären Parameters

Durch die Überprüfung der Varianzen (und der später vorgestellten Mittelung der Schätzungen aus mehreren Aufträgen sowie weiterer Prüfungen) können andere Parameter $\hat{\theta}_{r,i}$, $i = 1, \dots, p$, für die Regelung

verwendet werden, als die Parameter $\hat{\theta}_i$, die aus den Daten des aktuellen Auftrags mit dem LS-Schätzer bestimmt wurden.

Der Schätzwert $\hat{\theta}_0$ für den Parameter θ_0 ist zwar biasfrei, allerdings hat es dennoch Vorteile, diesen bezüglich der für die Regelung verwendeten Parameterwerte $\hat{\theta}_{1,r}$ und $\hat{\theta}_{s,r}$ zu berechnen (der dann als $\hat{\theta}_{0,r}$ bezeichnet wird), auch wenn diese im Allgemeinen nicht den wahren Parameterwerten θ_1 bzw. $\theta_{s,r}$ entsprechen und damit $\hat{\theta}_{0,r}$ nicht biasfrei ist. Siehe dazu auch Abschnitt 6.4.4. Nach den Gleichungen dieses Abschnitts wird $\hat{\theta}_{0,r}$ über

$$\hat{\theta}_{0,r} = \mathbf{L}_{1,1}^{-1} \cdot (\mathbf{r}_1 - \Psi_{1,[2:(p+1)]}) \cdot \begin{bmatrix} \hat{\theta}_{1,r} & \hat{\theta}_{r,s}^T \end{bmatrix}^T \quad (6.10)$$

bestimmt.

Es lassen sich für beide Möglichkeiten (Schätzung über „normales“ LS bzw. Schätzung bezüglich $\hat{\theta}_{1,r}$ und $\hat{\theta}_{s,r}$) Situationen konstruieren, in denen die eine ein besseres Verhalten als die andere liefert. Jedoch hat die Schätzung bezüglich $\hat{\theta}_{1,r}$ und $\hat{\theta}_{s,r}$ den Vorteil, dass bei einem Folgeauftrag mit gleicher Sollgröße und gleichen sekundären Eingangsgrößen der Vorsteuerungswert des primären Parameters ungefähr dem Wert entspricht, den dieser im eingeregelter Zustand des Vorauftrags angenommen hat. Dies entspricht damit dem Vorgehen, welches ein Bediener auch von sich aus anwenden würde.

6.4.2 Implementierung

Iterative Berechnung der Größen

Wenn die Schätzung online erfolgen soll, dann kommt zu jedem Abtastschritt ein neuer Datensatz hinzu. Es müssen aber nicht in jedem Abtastschritt Produkte immer größer werdender Matrizen berechnet werden, sondern man kann die Berechnung iterativ ausführen.

Die Matrix Ψ und der Vektor \mathbf{y} werden hier nur in den Kombinationen

$$\mathbf{L}_k = \Psi_k^T \Psi_k$$

und

$$\mathbf{r}_k = \Psi_k^T \mathbf{r}_k$$

verwendet. Diese können iterativ über

$$\mathbf{L}_k = \begin{cases} \mathbf{0} & k = 0 \\ \mathbf{L}_{k-1} + \psi_k \psi_k^T & k > 0 \end{cases}$$

und

$$\mathbf{r}_k = \begin{cases} \mathbf{0} & k = 0 \\ \mathbf{r}_{k-1} + \tilde{y}_k \psi_k & k > 0 \end{cases}$$

berechnet werden (siehe auch Anhang B.3).

Zum Bestimmen der Parameter über

$$\hat{\theta}_k = \mathbf{L}_k^{-1} \mathbf{r}_k$$

muss also in jedem Schritt (wenn alle Eingänge ausreichend angeregt wurden) eine $(p+1) \times (p+1)$ -Matrix invertiert bzw. ein lineares Gleichungssystem $(p+1)$ -ter Ordnung gelöst werden.

Für den Fall $p = 2$ ergibt sich der in Abbildung 6.3 dargestellte Ablaufplan. Solange noch keine Änderung der sekundären Eingangsgröße u_2 vorliegt, kann θ_2 auch noch nicht geschätzt werden. Die Schätzung von θ_1 erfolgt dann über eine Zwei-Parameter-Schätzung. Siehe dazu auch Abschnitt 6.4.4.

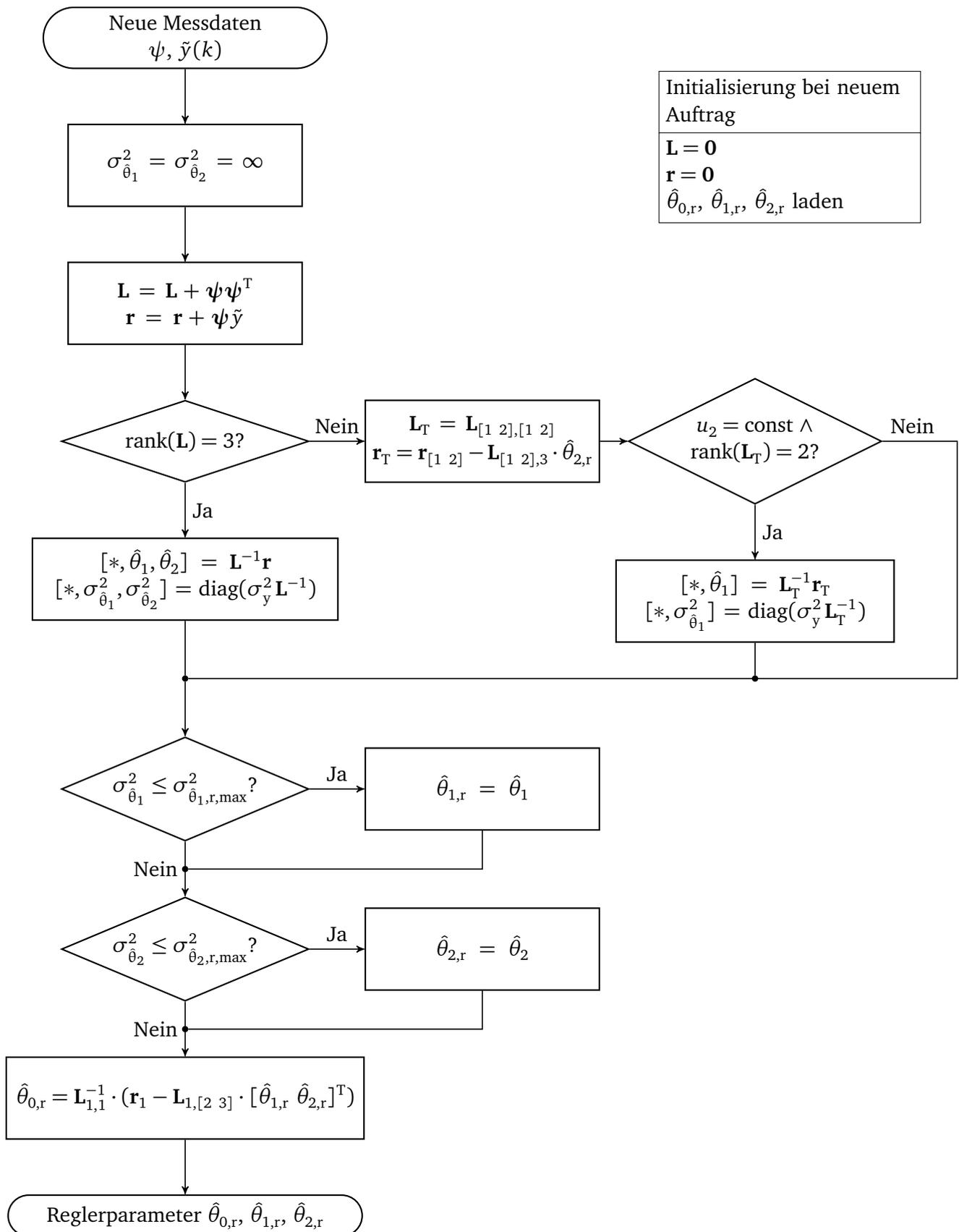


Abbildung 6.3: Ablaufplan Drei-Parameter-Schätzung

Identifikation bezüglich des Nennpunkts

In Abbildung 6.4 ist beispielhaft dargestellt, weshalb es günstiger ist, die Identifikation bezüglich der Nenngrößen, d. h. dem Zusammenhang

$$y_{k'} = \theta_0 + \theta_1 \cdot (u_{1,k'} - u_{1,N}) + \theta_2 \cdot (u_{2,k'} - u_{2,N}) + \dots$$

durchzuführen, und nicht den Zusammenhang

$$y_{k'} = \theta_{00} + \theta_1 \cdot u_{1,k'} + \theta_2 \cdot u_{2,k'} + \dots$$

zugrunde zu legen.

Dazu ist in Abbildung 6.4 für den Fall $p = 1$ die Schätzung der Geraden mit zwei Messungen bei $u_1 = 40$ und 60 dargestellt. Die durchgezogene Linie entspricht dem tatsächlichen Zusammenhang. Durch das Messrauschen entstehen bei den beiden Eingangswerten von 40 und 60 Unsicherheiten, die durch die kleinen senkrechten Striche dargestellt sind. Der Bereich, in dem die damit geschätzte Gerade liegt, ist durch den grauen Bereich angedeutet. Es ist zu erkennen, dass dieser Bereich bei $u_1 = 50$ deutlich kleiner ist als bei $u_1 = 0$. D. h. die Varianz des Schätzwertes $\{\hat{\theta}_{00}\}_r$ wäre deutlich größer als die des Schätzwertes $\{\hat{\theta}_0\}_r$. Da man sich in der Anwendung immer um den Nennpunkt aufhalten wird, ist die Varianz von $\{\hat{\theta}_0\}_r$ aussagekräftiger. (Vgl. dazu auch die Bemerkung von [FREUND et al., 2006, S. 42, 50] bezüglich der Interpretierbarkeit von $\hat{\theta}_0$.)

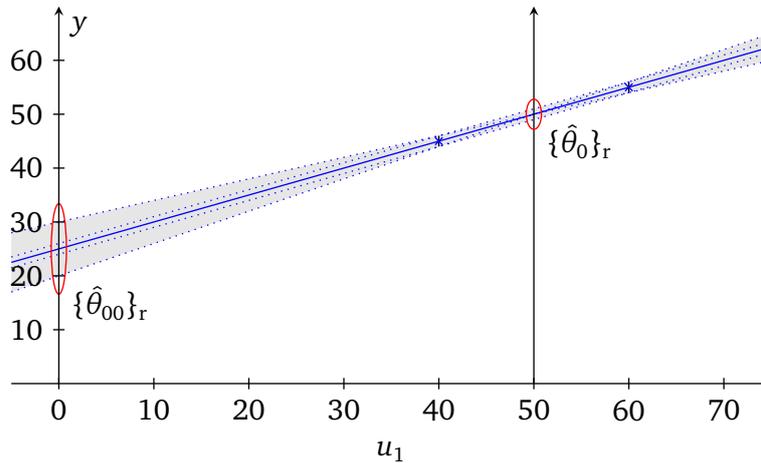


Abbildung 6.4: Vergleich von θ_0 und θ_{00}

Skalierung der Eingangsgrößen

Um Einträge gleicher Größenordnung in der L -Matrix und damit günstigere Bedingungen für die numerische Berechnung zu erhalten, werden die Eingangsgrößen normiert. Dazu wird die Gleichung

$$y_{k'} = \theta_0 + \theta_1 \cdot (u_{1,k'} - u_{1,N}) + \theta_2 \cdot (u_{2,k'} - u_{2,N})$$

in

$$y_{k'} = s_0 \cdot \frac{\theta_0}{s_0} + (u_{1,k'} - u_{1,N}) \cdot s_1 \cdot \frac{\theta_1}{s_1} + (u_{2,k'} - u_{2,N}) \cdot s_2 \cdot \frac{\theta_2}{s_2}$$

umgeschrieben. Nun werden die Größen $\frac{\theta_0}{s_0}$, $\frac{\theta_1}{s_1}$ und $\frac{\theta_2}{s_2}$ als neue Parameter θ'_0 , θ'_1 und θ'_2 aufgefasst. Damit wird der Parametervektor

$$\boldsymbol{\theta}' = [\theta'_0 \quad \theta'_1 \quad \theta'_2]^T$$

gesucht, wobei der Eingangsvektor

$$\boldsymbol{\psi}'_{k'} = [s_0 \quad (u_{1,k'} - u_{1,N}) \cdot s_1 \quad (u_{2,k'} - u_{2,N}) \cdot s_2]$$

lautet.

Algorithmen zur numerischen Lösung

Im Ablaufplan in Abbildung 6.3 ist sowie der Gleichung (6.4) bzw. (6.3) wird der Schätzwert über die Inverse der Matrix $\mathbf{L} = \Psi^T \Psi$ bestimmt, $\hat{\boldsymbol{\theta}} = \mathbf{L}^{-1} \mathbf{r}$. Dies ist jedoch nicht „wörtlich“ zu verstehen, sondern der Vektor der Schätzwerte wird numerisch über die Lösung des Gleichungssystems

$$\mathbf{L} \hat{\boldsymbol{\theta}} = \mathbf{r},$$

der sogenannten Normalgleichung, bestimmt. Diese Lösung wird, da \mathbf{L} , wenn Ψ vollen Spaltenrang hat, symmetrisch und positiv definit ist, über eine Cholesky-Zerlegung von \mathbf{L} berechnet. [DAHMEN und REUSKEN, 2008, S. 127] Dabei kann die Inverse von \mathbf{L} , die zur Bestimmung der Varianzen der Schätzwerte benötigt wird, parallel zur Lösung $\hat{\boldsymbol{\theta}}$ berechnet werden, d. h. es muss nur einmal die Cholesky-Zerlegung durchgeführt werden. [DAHMEN und REUSKEN, 2008, S. 80]

Grundsätzlich ist bekannt, dass die Bestimmung der Lösung des LS-Problems über die Berechnung der Lösung von $\mathbf{L} \hat{\boldsymbol{\theta}} = \mathbf{r}$ bzw. $\Psi^T \Psi \hat{\boldsymbol{\theta}} = \Psi^T \tilde{\mathbf{y}}$ numerisch ungünstig sein kann und bei schlecht konditionierten Matrizen \mathbf{L} zu Problemen führt, siehe z. B. [DAHMEN und REUSKEN, 2008, S. 128]. Darüber hinaus wird an der genannten Stelle auch darauf hingewiesen, dass schon die Bestimmung der Matrix $\mathbf{L} = \Psi^T \Psi$ für große k aufgrund von Rundungsfehlern kritisch sein kann.

Prinzipiell wird in der Literatur häufig empfohlen, die Lösung des LS-Problem direkt, basierend auf einer QR-Zerlegung der Eingangsmatrix Ψ zu bestimmen, siehe z. B. [DAHMEN und REUSKEN, 2008, S. 129], [VON FINCKENSTEIN et al., 2002, S. 222].

Für den vorliegenden Zweck ist die Bestimmung der Lösung des LS-Problems über die Normalgleichung jedoch unkritisch. Dies liegt daran, dass ein schlecht konditioniertes \mathbf{L} darauf hinweist, dass zumindest ein Parameter nur sehr schlecht, d. h. mit hoher Varianz geschätzt werden kann. Es besteht damit keine Notwendigkeit, die Schätzwerte bei einer schlecht konditionierten Matrix \mathbf{L} zu bestimmen, da diese dann ohnehin nicht berücksichtigt werden dürfen. Es muss also nur sichergestellt sein, dass die schlechte Kondition von \mathbf{L} erkannt wird, bevor die Schätzwerte bestimmt werden.

Der Rang einer Matrix kann numerisch über die Berechnung der Singulärwerte bestimmt werden, siehe z. B. [DAHMEN und REUSKEN, 2008, S. 150]. Die berechneten Singulärwerte können hierbei auch genutzt werden, bei zu schlecht konditionierten Matrizen von der Berechnung der Schätzwerte abzusehen. In diesem Fall ist die Bedingung $\text{rank}(\mathbf{L}) = 3$ nicht nur so zu verstehen, dass \mathbf{L} vollen Rang hat, sondern auch eine genügend große Anregung der Eingänge vorliegt, um die Lösung des Gleichungssystems $\mathbf{L} \hat{\boldsymbol{\theta}} = \mathbf{r}$ sowie die Inverse von \mathbf{L} numerisch sicher bestimmen zu können.⁹

Anstelle der numerischen Bestimmung des Ranges könnte auch überprüft werden, ob sich die Eingangsgrößen genügend (linear unabhängig) geändert haben, wobei dafür geeignete Kriterien zu definieren sind.

Implementierung als RLS

Der Least-Squares-Algorithmus kann auch in einer rekursiven Form implementiert werden (siehe Abschnitt B.4). In dieser muss anstelle der in jedem Schritt notwendigen Invertierung der Matrix $(\Psi^T \Psi)^{-1}$, die hier die Dimension $(p + 1 \times p + 1)$ besitzt, bzw. der Lösung eines linearen Gleichungssystems mit $p + 1$ Unbekannten, nur eine skalare Division durchgeführt werden. Mathematisch sind die beiden Verfahren äquivalent, sofern die Matrizen $(\Psi^T \Psi)^{-1}$ und \mathbf{P} nach dem ersten Schritt identisch sind. Dies kann näherungsweise durch die Wahl sehr großer Einträge im Anfangswert \mathbf{P}_0 oder durch die Berechnung des erstens Schritts über das „normale“ Least-Squares-Verfahren und das Fortsetzen im RLS-Verfahren erreicht werden.

⁹ Bei den zu dieser Untersuchung durchgeführten Simulationen und Messungen wurde die untere Grenze für die Singulärwerte auf $1 \cdot 10^{-10}$ gesetzt, d. h. kleinere Singulärwerte wurden bei der Rangbestimmung als Null angesehen. Damit sind bei der Verwendung von Gleitkommazahlen mit 64 bit Länge keine Probleme mit der Lösung von $\mathbf{L} \hat{\boldsymbol{\theta}} = \mathbf{r}$ bzw. der Inversion von \mathbf{L} aufgetreten.

Dem Vorteil des geringeren Rechenaufwands stehen hier aber auch Nachteile gegenüber, die durch die Notwendigkeit entstehen, Anfangswerte für $\hat{\theta}$ und \mathbf{P} anzugeben. Es werden dazu in [ISERMANN und MÜNCHHOF, 2011, S. 272] verschiedene Möglichkeiten angegeben. Existieren sinnvolle Schätzungen aus a-priori-Wissen (z. B. aus vorherigen Versuchen), so können diese verwendet werden. Dies ist hier aber nicht möglich, da zum einen für θ_0 keine genaue a-priori-Schätzung angegeben werden kann, und zum anderen mit diesem Verfahren viele Messwerte benötigt werden, um einen falschen Anfangswert zu korrigieren. Eine weitere Methode wäre, $\mathbf{P}_0 = \alpha \mathbf{I}$ mit großem α zu setzen. Im Grenzfall $\alpha \rightarrow \infty$ nähert man sich damit dem normalen LS-Verfahren an, unabhängig von dem Anfangswert $\hat{\theta}_0$. Aber für endliche α werden wieder mehrere Messwerte benötigt, um die gewünschten Werte zu erhalten. Auch treten Probleme auf, wenn Parameter zu sekundären Eingängen erst im späteren Verlauf eines Auftrags mitgeschätzt werden können.

Letztlich bleibt nur die Möglichkeit, das RLS-Verfahren mit einem normalen LS-Verfahren zu starten. Dies muss während eines Auftrags auch jedes Mal neu gemacht werden, wenn ein weiterer Eingang sich ändert und damit ein zusätzlicher Parameter geschätzt werden kann. Damit erhält man mit dem RLS-Verfahren immer dieselben Werte wie mit dem LS-Verfahren.

Da damit aber ohnehin Speicher für all die Daten vorgehalten werden muss, die auch für das normale LS-Verfahren benötigt werden, und auch die entsprechenden Gleichungssysteme gelöst werden müssen, wenn auch nicht so häufig, wird auf das RLS-Verfahren vollständig verzichtet.

6.4.3 Einflussfaktoren auf die Varianz

In Abschnitt 6.3.3 wurde schon auf die Faktoren eingegangen, die die Varianz beeinflussen. Dies sind neben der Varianz des Messrauschens die Anzahl der Messpunkte, die Variation in den Eingangsgrößen und der Grad der Abhängigkeit der Eingangsgrößen untereinander. Da insbesondere der letzte Punkt eine große Rolle spielt, soll dieser zunächst nochmals an Beispielen erläutert werden. Im Anschluss daran wird dargelegt, wieso dies für den vorliegenden Fall wesentlich ist. Es wird weiterhin kurz dargestellt, warum die Methoden aus der Literatur für den Umgang mit der Multikollinearität für diesen Einsatzzweck nicht geeignet sind.

Für zwei Eingangsgrößen u_1 und u_2 sind in Tabelle 6.1 bzw. Abbildung 6.5 vier Beispiele für mögliche Eingangsdaten gegeben. Diese unterscheiden sich neben der Anzahl der Messpunkte vor Allem im Grad der Korrelation zwischen den beiden Eingangsgrößen u_1 und u_2 .

Tabelle 6.1: Eingangsdaten zu Abbildung 6.5

<p>(a) Ideale Korrelation (Abbildung 6.5a)</p> <table style="width: 100%; border-collapse: collapse; border-top: 1px solid black; border-bottom: 1px solid black;"> <tr> <td style="padding: 5px;">u_1</td> <td style="padding: 5px;">31</td> <td style="padding: 5px;">35</td> <td style="padding: 5px;">39</td> </tr> <tr> <td style="padding: 5px;">u_2</td> <td style="padding: 5px;">9</td> <td style="padding: 5px;">12</td> <td style="padding: 5px;">15</td> </tr> </table>	u_1	31	35	39	u_2	9	12	15	<p>(b) Starke Korrelation (Abbildung 6.5b)</p> <table style="width: 100%; border-collapse: collapse; border-top: 1px solid black; border-bottom: 1px solid black;"> <tr> <td style="padding: 5px;">u_1</td> <td style="padding: 5px;">33</td> <td style="padding: 5px;">31</td> <td style="padding: 5px;">35</td> <td style="padding: 5px;">33</td> <td style="padding: 5px;">34</td> <td style="padding: 5px;">36</td> <td style="padding: 5px;">37</td> </tr> <tr> <td style="padding: 5px;">u_2</td> <td style="padding: 5px;">9</td> <td style="padding: 5px;">9</td> <td style="padding: 5px;">12</td> <td style="padding: 5px;">12</td> <td style="padding: 5px;">12</td> <td style="padding: 5px;">15</td> <td style="padding: 5px;">15</td> </tr> </table>	u_1	33	31	35	33	34	36	37	u_2	9	9	12	12	12	15	15
u_1	31	35	39																						
u_2	9	12	15																						
u_1	33	31	35	33	34	36	37																		
u_2	9	9	12	12	12	15	15																		
<p>(c) Schwache Korrelation (Abbildung 6.5c)</p> <table style="width: 100%; border-collapse: collapse; border-top: 1px solid black; border-bottom: 1px solid black;"> <tr> <td style="padding: 5px;">u_1</td> <td style="padding: 5px;">37</td> <td style="padding: 5px;">31</td> <td style="padding: 5px;">37</td> </tr> <tr> <td style="padding: 5px;">u_2</td> <td style="padding: 5px;">9</td> <td style="padding: 5px;">9</td> <td style="padding: 5px;">15</td> </tr> </table>	u_1	37	31	37	u_2	9	9	15	<p>(d) Mittlere Korrelation (Abbildung 6.5d)</p> <table style="width: 100%; border-collapse: collapse; border-top: 1px solid black; border-bottom: 1px solid black;"> <tr> <td style="padding: 5px;">u_1</td> <td style="padding: 5px;">24</td> <td style="padding: 5px;">31</td> <td style="padding: 5px;">37</td> </tr> <tr> <td style="padding: 5px;">u_2</td> <td style="padding: 5px;">9</td> <td style="padding: 5px;">9</td> <td style="padding: 5px;">15</td> </tr> </table>	u_1	24	31	37	u_2	9	9	15								
u_1	37	31	37																						
u_2	9	9	15																						
u_1	24	31	37																						
u_2	9	9	15																						

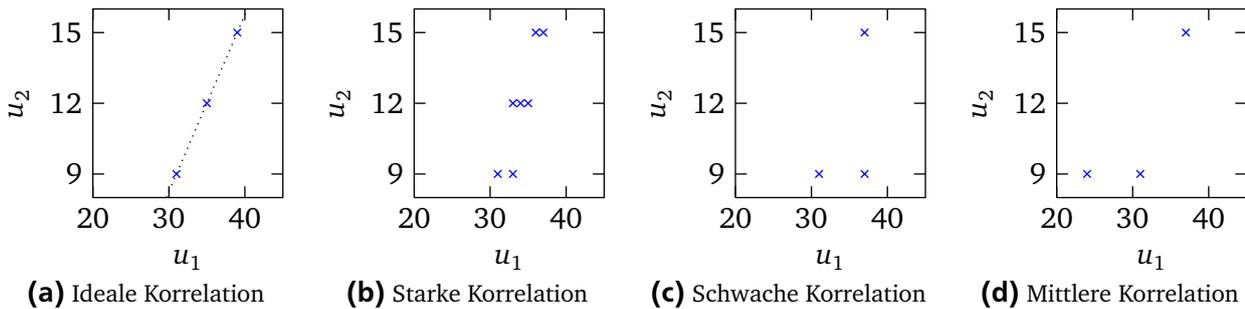


Abbildung 6.5: Beispiele für Eingangsdaten bei zwei Eingangsgrößen u_1 und u_2

Ideale Korrelation

Die Daten in Abbildung 6.5a sind so konstruiert, dass u_2 um drei zunimmt, wenn u_1 um vier zunimmt. Die beiden Folgen sind daher ideal korreliert. Damit kann z. B. $u_{2,k'}$ über

$$u_{2,k'} = 12 + \frac{4}{3} \cdot (u_{1,k'} - 35)$$

durch $u_{1,k'}$ ausgedrückt werden, so dass neben der Konstanten Eins nur eine Eingangsgröße vorliegt. Damit können aber keine drei Parameter geschätzt werden und es ist anschaulich klar, dass in diesem Fall die Matrix $\Psi^T \Psi$ singulär ist, also nicht invertiert werden kann. (Bei idealer Korrelation könnten auch beliebig viele Messpunkte vorliegen, $\Psi^T \Psi$ bleibt immer singulär.)

Tabelle 6.2: Varianzen zu den Eingangsdaten aus Abbildung 6.5

Eingangsdaten	$\sigma_{\hat{\theta}_1}^2 / \sigma_y^2$	$\sigma_{\hat{\theta}_2}^2 / \sigma_y^2$
6.5a	$\Psi^T \Psi$ singulär	
6.5b	0,217	0,150
6.5b, nur u_1	0,040	—
6.5b, nur u_2	—	0,028
6.5c	0,056	0,056
6.5c, nur u_2	—	0,042
6.5d	0,041	0,144
6.5d, nur u_2	—	0,042

Starke Korrelation

In Abbildung 6.5b sind sieben Datenpunkte dargestellt, die noch relativ gut auf einer Geraden in der u_1 - u_2 -Ebene liegen. Die sich damit ergebenden Varianzen sind in Tabelle 6.2 aufgeführt. Ebenfalls dort angegeben sind die Varianzen für $\{\hat{\theta}_1\}_r$ und $\{\hat{\theta}_2\}_r$ die sich ergeben würden, wenn man jeweils nur den Parameter θ_1 bzw. θ_2 schätzt und den jeweils anderen als bekannt annimmt. Es wird dabei deutlich, dass die Varianzen der stark korrelierten Folge deutlich größer, also schlechter sind, als es die Einzelfolgen für $u_{1,k'}$ und $u_{2,k'}$ alleine „hergeben“ würden.

Schwache Korrelation

In Abbildung 6.5c sind nur drei Datenpunkte vorhanden. Diese sind aber so gelegt, dass diese quasi das einhüllende Rechteck der Punkte aus Abbildung 6.5b bilden. Vergleicht man die sich ergebenden

Varianzen in Tabelle 6.2 mit denen, die sich bei der starken Korrelation ergeben, so zeigt sich, dass die Varianzen im Fall der schwachen Korrelation deutlich geringer sind, obwohl nur $k = 3$ anstatt $k = 7$ Datenpunkte vorliegen.

Die alleinige Schätzung von θ_2 bei einem als bekannt angenommenen θ_1 bringt immer noch einen Vorteil, der aber nicht mehr so deutlich wie in dem Fall zuvor ausfällt.

Mittlere Korrelation

In Abbildung 6.5d sind ebenfalls nur drei Datenpunkte vorhanden. Im Vergleich zu Abbildung 6.5c ist der rechte Punkt bei $u_2 = 9$ nach links gerückt. Die Korrelation zwischen u_1 und u_2 nimmt damit wieder zu, anschaulich lassen sich diese Punkte wieder besser durch eine Gerade verbinden. Tabelle 6.2 zeigt zwar, dass die Schätzung von θ_1 besser wird, was hier auf die Verdoppelung des Wertebereichs von u_1 zurückzuführen ist. Allerdings wird θ_2 deutlich schlechter geschätzt.

Hier würde es wieder einen deutlichen Vorteil bringen, wenn θ_1 als bekannt vorausgesetzt werden und θ_2 alleine geschätzt werden könnte.

Fazit

Da nach dem Regelgesetz Gl. (3.13) bzw. Gl. (3.14) $u_{1,k}$ bei einer Änderung der sekundären Eingänge $\mathbf{u}_{S,k}$ gerade so gewählt wird, dass deren Wirkung auf y_k möglichst kompensiert wird, sind die verschiedenen Eingänge bei dieser Anwendung tendenziell stark korreliert. Dies würde ohne weitere Maßnahmen zu sehr hohen Varianzen der Schätzwerte der sekundären Eingänge führen.

Behandlung der Multikollinearität in der Literatur

Der Thematik der starken Abhängigkeiten zwischen verschiedenen Eingangsgrößen, der sogenannten Multikollinearität, widmen sich [FREUND et al., 2006, S. 177ff] und [RAWLINGS et al., 1998, S. 369ff, 433ff, 463ff] jeweils in mindestens einem Kapitel.

Zum einen werden dort Größen eingeführt, um den Grad der Multikollinearität zu analysieren, sowie Maßnahmen beschrieben, deren Einfluss zu vermindern. Ein Verfahren zur Analyse ist die Hauptkomponentenanalyse. Dabei wird anhand der Eigenwerte der Matrix $\Psi^T\Psi$ festgestellt, welche durch die Eigenvektoren beschriebenen Kombinationen der Eingangsgrößen problematisch bezüglich der Multikollinearität sind. [FREUND et al., 2006, S. 192]

Eine Gegenmaßnahme wäre eine geeignete Umdefinition der Eingangsgrößen, so dass sich unabhängige oder zumindest weniger abhängige neue Eingangsgrößen ergeben. Dies kann aufgrund von Wissen über den Prozess oder auf Basis der Hauptkomponentenanalyse erfolgen. Allerdings ändern sich dabei die Eigenschaften des gesamten geschätzten Zusammenhanges nicht. [FREUND et al., 2006, S. 199, 205f] Die hohe Varianz wird lediglich auf die weniger wesentlichen Variablen „abgeschoben“.

Alternativ wird vorgeschlagen, anstelle des biasfreien LS-Schätzers einen biasbehafteten Schätzer zu verwenden, der dafür eine geringere Varianz und insgesamt einen geringeren MSE besitzt.

In [FREUND et al., 2006, S. 216ff] wird dazu die Ridge-Regression und die „Incomplete Principal Component Regression“ vorgestellt. Erstere Methode geht von der Schätzgleichung $\hat{\theta} = (\Psi^T\Psi + c \cdot \mathbf{I})^{-1} \cdot \Psi^T\tilde{\mathbf{y}}$ aus, wobei c passend gewählt werden muss. Dabei ist diese Wahl jedoch nicht ohne weiteres automatisierbar und der Einfluss des gewählten c ist auch nicht gut vorherzusagen. Bei der „Incomplete Principal Component Regression“ werden die Hauptkomponenten mit den niedrigsten Eigenwerten weggelassen.

Bei der hier vorliegenden Aufgabenstellung ist es aber das Ziel, sowohl auf einzelne Sollgrößensprünge als auch auf Änderungen der sekundären Eingangsgrößen möglichst gut reagieren zu können. D. h. es werden (neben θ_0) immer p (verschiedene, unabhängige) Parameter benötigt, und es ist sicher, dass

keiner vernachlässigt werden kann. Auch würde die Wahl, welche Kombination der Eingangsgrößen einen hohen Einfluss hat und welche weniger auch von dem aktuellen Auftrag abhängen.

Damit lassen sich diese (speziellen) Methoden hier nicht anwenden. Es wird aber die Möglichkeit weiter verfolgt, eine (allgemeine) biasbehaftete Schätzung für die sekundären Parameter durchzuführen. Dabei wird hier die primäre Eingangsgröße auf dem aktuellen zur Regelung verwendeten Schätzwert festgehalten, der nicht unbedingt (vollständig) aus den Daten des aktuellen Auftrags gewonnen wird. Es wird sich zeigen, dass dieses Verfahren sehr gute Ergebnisse liefert.

6.4.4 Least-Squares-Schätzung bei teilweise festgehaltenen Parametern

Wenn aufgrund der hohen Korreliertheit der Eingangsgrößen bzw. der Multikollinearität die Varianzen der Schätzwerte sehr hoch werden, dann stellt sich die Frage, ob es sinnvoll wäre, nur einen Teil θ_e der Parameter $\theta^T = \begin{bmatrix} \theta_e^T & \theta_f^T \end{bmatrix}$ zu schätzen¹⁰ (siehe oben).

Bezeichnet man die Schätzwerte der reduzierten Schätzung mit $\hat{\theta}_e$ und die Schätzwerte der entsprechenden Parameter aus der „vollständigen“ Schätzung mit $(\hat{\theta})_e$, so werden in [RAWLINGS et al., 1998, S. 208f] folgende Aussagen für den Fall zitiert, dass $\hat{\theta}_f = \mathbf{0}$ gesetzt wird:

- Die Schätzwerte $\hat{\theta}_e$ sind biasbehaftet, sofern die tatsächlichen Parameter θ_f nicht Null sind oder bestimmte Anforderungen an die Anregung $\mathbf{u}_{f,k}$ erfüllt werden, auf die weiter unten eingegangen wird.
- Die Varianzen $\text{Var}\{\hat{\theta}_e\}_r$ sind maximal so groß wie die entsprechenden Varianzen $\text{Var}\{(\hat{\theta})_e\}_r$. In der Regel sind diese sogar kleiner. (siehe Abschnitt B.5)
- Unter bestimmten Bedingungen kann der $\text{MSE}\{\hat{\theta}_e - \theta_e\}_r$ damit kleiner sein als $\text{MSE}\{(\hat{\theta})_e - \theta_e\}_r$.
- Außerdem besitzt der Schätzwert $\hat{\sigma}_y^2$, der sich aus den Residuen ergibt, bei der θ_e -Schätzung einen positiven Bias.

In [RAWLINGS et al., 1998] wird die Regression für den Fall betrachtet, dass von vorliegenden Daten durch Regression ein Modell abgeleitet werden soll. Der hier behandelte Fall unterscheidet sich davon dahingehend, dass für das gleiche System mehrere Regressionen mit anderen Messdaten vorgenommen werden. Dadurch kann ein gewisses Vorwissen vorausgesetzt werden.

Daher werden, anders als in [RAWLINGS et al., 1998], die Terme zu den Parametern θ_f nicht einfach aus der Gleichung gestrichen, sondern es werden für θ_f feste Werte $\hat{\theta}_f$ angenommen.

Es wird der Fall betrachtet, dass nur noch die Parameter θ_e geschätzt werden sollen, während für die Parameter θ_f ein fester Wert $\hat{\theta}_f$ vorgegeben wird, der aus vorherigen Schätzungen oder Vorwissen über das System stammt. Die Gleichung, die der Schätzung zugrunde liegt, lautet in diesem Fall

$$\underbrace{\tilde{y}_{k'} - \hat{\theta}_f^T \cdot \mathbf{u}_{f,k'}}_{\tilde{y}_{f,k'}} = \theta_{e|f}^T \cdot \begin{bmatrix} 1 \\ \mathbf{u}_{e,k'} \end{bmatrix} + n_{y,k'} \quad (6.11)$$

Dabei ist $\tilde{y}_{f,k'} = \tilde{y}_{k'} - \hat{\theta}_f^T \cdot \mathbf{u}_{f,k'}$ die modifizierte Messgröße zur Schätzung der Parameter θ_e . Der konstante Anteil θ_{00} der Systemgleichung wird immer den zu schätzenden Parametern θ_e zugeschlagen.

Der Eingangsvektor lautet in dem beschriebenen Fall $\boldsymbol{\psi}_{e,k'}^T = \begin{bmatrix} 1 & \mathbf{u}_{e,k'}^T \end{bmatrix}$. Fasst man diese für $k' = 1, \dots, k$ zur Matrix $\boldsymbol{\Psi}_e$ sowie die $\tilde{y}_{f,k'}$ zu $\tilde{\mathbf{y}}_f$ zusammen, so lauten die k Messgleichungen in vektorieller Form

$$\tilde{\mathbf{y}}_f = \boldsymbol{\Psi}_e \theta_{e|f} + \mathbf{n}_y$$

¹⁰ Die Indizes stehen für „estimated“ bzw. „fixed“.

Daraus ergibt sich für den Schätzwert $\hat{\boldsymbol{\theta}}_{\text{elf}}$ der Ausdruck

$$\hat{\boldsymbol{\theta}}_{\text{elf}} = \mathbf{L}_e^{-1} \mathbf{r}_{\text{elf}} \quad (6.12)$$

mit $\mathbf{L}_e = \boldsymbol{\Psi}_e^T \boldsymbol{\Psi}_e$ und $\mathbf{r}_{\text{elf}} = \boldsymbol{\Psi}_e^T \tilde{\mathbf{y}}_f$.

Zunächst wird nun angegeben, wie sich die in der Schätzgleichung verwendeten Matrizen aus den Matrizen $\boldsymbol{\Psi}^T \boldsymbol{\Psi}$ und $\boldsymbol{\Psi}^T \tilde{\mathbf{y}}$ (bzw. \mathbf{L} und \mathbf{r}) des Gesamtsystems erhalten lassen. Im Anschluss werden die Auswirkungen von Fehlern bei $\hat{\boldsymbol{\theta}}_f$ auf die Schätzwerte von $\boldsymbol{\theta}_e$ untersucht.

Berechnung aus Matrizen des Gesamtsystems

Der einfacheren Schreibweise halber wird angenommen, dass die zu schätzenden Parameter im Gesamtparametervektor $\boldsymbol{\theta}$ vor den festgehaltenen Parametern liegen, d. h. $\boldsymbol{\theta}^T = [\boldsymbol{\theta}_e^T \quad \boldsymbol{\theta}_f^T]$. Dies stellt keine Einschränkung dar, da dies durch ein einfaches Umsortieren, welches sich auch auf \mathbf{L} und \mathbf{r} anwenden lässt, immer erreicht werden kann.

Des Weiteren wird auch $\boldsymbol{\psi}_{f,k'}^T = \mathbf{u}_{f,k'}^T$ definiert, obwohl es sich bei \mathbf{u}_f^T nicht um Eingangsgrößen des neuen Schätzproblems handelt. Die Vektoren $\boldsymbol{\psi}_{e,k'}^T$ und $\boldsymbol{\psi}_{f,k'}^T$ aller Zeitpunkte $k' = 1, \dots, k$ werden auch hier zu den Matrizen $\boldsymbol{\Psi}_e$ und $\boldsymbol{\Psi}_f$ zusammengefasst. Da $\boldsymbol{\psi}_{k'}^T = [\boldsymbol{\psi}_{e,k'}^T \quad \boldsymbol{\psi}_{f,k'}^T]$ ist, gilt auch $\boldsymbol{\Psi} = [\boldsymbol{\Psi}_e \quad \boldsymbol{\Psi}_f]$.

Die Informationsmatrix \mathbf{L} des Gesamtsystems kann dann als

$$\mathbf{L} = \boldsymbol{\Psi}^T \boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\Psi}_e^T \\ \boldsymbol{\Psi}_f^T \end{bmatrix} [\boldsymbol{\Psi}_e \quad \boldsymbol{\Psi}_f] = \begin{bmatrix} \boldsymbol{\Psi}_e^T \boldsymbol{\Psi}_e & \boldsymbol{\Psi}_e^T \boldsymbol{\Psi}_f \\ \boldsymbol{\Psi}_f^T \boldsymbol{\Psi}_e & \boldsymbol{\Psi}_f^T \boldsymbol{\Psi}_f \end{bmatrix} = \begin{bmatrix} (\mathbf{L})_e & (\mathbf{L})_{\text{ef}} \\ (\mathbf{L})_{\text{fe}} & (\mathbf{L})_f \end{bmatrix} \quad (6.13)$$

geschrieben werden, woraus sich mit der angegebenen Aufteilung

$$\mathbf{L}_e = (\mathbf{L})_e \quad (6.14)$$

ergibt. Für den \mathbf{r} -Vektor des Gesamtsystems ergibt sich

$$\mathbf{r} = \boldsymbol{\Psi}^T \mathbf{y} = \begin{bmatrix} \boldsymbol{\Psi}_e^T \tilde{\mathbf{y}} \\ \boldsymbol{\Psi}_f^T \tilde{\mathbf{y}} \end{bmatrix} = \begin{bmatrix} (\mathbf{r})_e \\ (\mathbf{r})_f \end{bmatrix}. \quad (6.15)$$

Der Ausdruck $\hat{\boldsymbol{\theta}}_f^T \cdot \mathbf{u}_{f,k'}$ in der Gleichung (6.11) für den modifizierten Ausgang \tilde{y}_f ist ein Skalar, so dass dieser auch transponiert werden kann, was

$$\tilde{y}_{f,k'} = \tilde{y}_{k'} - \hat{\boldsymbol{\theta}}_f^T \cdot \mathbf{u}_{f,k'} = \tilde{y}_{k'} - \mathbf{u}_{f,k'}^T \cdot \hat{\boldsymbol{\theta}}_f$$

ergibt. Damit lautet der Vektor $\tilde{\mathbf{y}}_f$

$$\tilde{\mathbf{y}}_f = \tilde{\mathbf{y}} - \boldsymbol{\Psi}_f \cdot \hat{\boldsymbol{\theta}}_f. \quad (6.16)$$

Dieser wird in den oben angegebenen Ausdruck für \mathbf{r}_{elf} eingesetzt,

$$\mathbf{r}_{\text{elf}} = \boldsymbol{\Psi}_e^T \tilde{\mathbf{y}}_f = \boldsymbol{\Psi}_e^T \tilde{\mathbf{y}} - \boldsymbol{\Psi}_e^T \cdot \boldsymbol{\Psi}_f \cdot \hat{\boldsymbol{\theta}}_f.$$

Der erste Summand entspricht $(\mathbf{r})_e$ und $\boldsymbol{\Psi}_e^T \cdot \boldsymbol{\Psi}_f$ ist gerade $(\mathbf{L})_{\text{ef}}$, so dass sich

$$\mathbf{r}_{\text{elf}} = (\mathbf{r})_e - (\mathbf{L})_{\text{ef}} \cdot \hat{\boldsymbol{\theta}}_f \quad (6.17)$$

ergibt.

Bias

Der angenommene Wert $\hat{\theta}_f$ für θ_f wird im Allgemeinen von diesem abweichen, d. h.

$$\hat{\theta}_f = \theta_f + \Delta\hat{\theta}_f.$$

Wird dies in Gl. (6.16) eingesetzt,

$$\begin{aligned}\tilde{y}_f &= \tilde{y} - \Psi_f \cdot (\theta_f + \Delta\hat{\theta}_f) \\ &= \tilde{y} - \Psi_f \theta_f - \Psi_f \cdot \Delta\hat{\theta}_f,\end{aligned}$$

so ergibt sich

$$\begin{aligned}\hat{\theta}_{\text{elf}} &= (\Psi_e^T \Psi_e)^{-1} \Psi_e^T \tilde{y}_e \\ &= \underbrace{(\Psi_e^T \Psi_e)^{-1} \Psi_e^T \cdot (\tilde{y} - \Psi_f \theta_f)}_{\hat{\theta}_e} - (\Psi_e^T \Psi_e)^{-1} \Psi_e^T \Psi_f \cdot \Delta\hat{\theta}_f\end{aligned}\quad (6.18)$$

für den Schätzwert für θ_e . Der erste Summand $\hat{\theta}_e$ ist ein biasfreier Schätzwert für θ_e , da die „Anpassung“ $-\Psi_f \theta_f$ der Messgröße mit den wahren Werten θ_f erfolgt.

Der zweite Summand stellt den systematischen Schätzfehler bzw. Bias aufgrund des Fehlers $\Delta\hat{\theta}_f$ dar,

$$\Delta\hat{\theta}_{\text{elf}} = \hat{\theta}_{\text{elf}} - \hat{\theta}_e = -(\Psi_e^T \Psi_e)^{-1} \Psi_e^T \Psi_f \cdot \Delta\hat{\theta}_f. \quad (6.19)$$

Die Matrix $\Psi_e^T \Psi_f$ entspricht gerade wieder $(L)_{\text{ef}}$, womit der Fehler $\Delta\hat{\theta}_{\text{elf}}$ aus den schon vorhandenen Daten abgeschätzt werden kann, wenn sinnvolle Abschätzungen für $\Delta\hat{\theta}_f$ existieren.

Wenn die Eingänge \mathbf{u}_f zu allen festgehaltenen Parametern $\hat{\theta}_f$ konstant sind, so darf dies anschaulich nur einen Fehler bei der Schätzung der Konstanten θ_{00} erzeugen, nicht aber bei den Parametern zu den Eingängen \mathbf{u}_e . Dies soll hier als Sonderfall noch kurz betrachtet werden. Die Informationsmatrix lautet für diesen Fall

$$\Psi_e^T \Psi_e = \begin{bmatrix} k & \sum \mathbf{u}_{e,k'}^T \\ \sum \mathbf{u}_{e,k'} & \sum \mathbf{u}_{e,k'} \mathbf{u}_{e,k'}^T \end{bmatrix}.$$

Diese Blockmatrix wird nach der in der Fußnote¹¹ angegebenen Gleichung invertiert, wobei die entsprechenden Blöcke hier $\mathbf{A} = k$, $\mathbf{B} = \sum \mathbf{u}_{e,k'}^T$, $\mathbf{C} = \mathbf{B}^T$ und $\mathbf{D} = \sum \mathbf{u}_{e,k'} \mathbf{u}_{e,k'}^T$ lauten. Darüber hinaus wird $\mathbf{b} = \mathbf{B} = \sum \mathbf{u}_{e,k'}^T$ gesetzt und die Abkürzung $\mathbf{S}'_{\text{uu}} = \mathbf{D} - \mathbf{b}^T \mathbf{b} / k = \sum \mathbf{u}_{e,k'} \mathbf{u}_{e,k'}^T - \frac{1}{k} \cdot \sum \mathbf{u}_{e,k'} \cdot \sum \mathbf{u}_{e,k'}^T$ verwendet. Damit erhält man

$$(\Psi_e^T \Psi_e)^{-1} = \begin{bmatrix} \frac{1}{k} + \frac{1}{k} \mathbf{b} \cdot (\mathbf{S}'_{\text{uu}})^{-1} \cdot \mathbf{b}^T \frac{1}{k} & -\frac{1}{k} \mathbf{b} \cdot (\mathbf{S}'_{\text{uu}})^{-1} \\ -(\mathbf{S}'_{\text{uu}})^{-1} \cdot \mathbf{b}^T \frac{1}{k} & (\mathbf{S}'_{\text{uu}})^{-1} \end{bmatrix}.$$

Für das Produkt $\Psi_e^T \Psi_f$ ergibt sich

$$\Psi_e^T \Psi_f = \sum \left(\begin{bmatrix} 1 \\ \mathbf{u}_{e,k'} \end{bmatrix} \cdot \mathbf{u}_f^T \right) = \begin{bmatrix} k \\ \sum \mathbf{u}_{e,k'} \end{bmatrix} \cdot \mathbf{u}_f^T = \begin{bmatrix} k \\ \mathbf{b}^T \end{bmatrix} \cdot \mathbf{u}_f^T,$$

¹¹ $\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1} \mathbf{B} \mathbf{D}_{\text{red}}^{-1} \mathbf{C} \mathbf{A}^{-1} & -\mathbf{A}^{-1} \mathbf{B} \mathbf{D}_{\text{red}}^{-1} \\ -\mathbf{D}_{\text{red}}^{-1} \mathbf{C} \mathbf{A}^{-1} & \mathbf{D}_{\text{red}}^{-1} \end{bmatrix}$ mit $\mathbf{D}_{\text{red}} = \mathbf{D} - \mathbf{C} \mathbf{A}^{-1} \mathbf{B}$ [ZURMÜHL und FALK, 1984, S. 303].

wobei von der Voraussetzung, dass \mathbf{u}_f^T konstant ist, Gebrauch gemacht wurde. Damit ergibt sich für den Fehler

$$\Delta \hat{\boldsymbol{\theta}}_{\text{elf}} = \begin{bmatrix} -\mathbf{u}_f^T \cdot \Delta \hat{\boldsymbol{\theta}}_f \\ \mathbf{0} \end{bmatrix}.$$

Wie erwartet, ist dieser für alle Parameter bis auf die Konstante gleich Null.

Etwas allgemeiner können die festgehaltenen Parameter $\hat{\boldsymbol{\theta}}_f$ auch aufgeteilt werden in die Parameter $\hat{\boldsymbol{\theta}}_{f,v}$, deren zugehörige Eingänge $\mathbf{u}_{f,v,k'}$ sich während der Schätzung ändern, und die Parameter $\hat{\boldsymbol{\theta}}_{f,c}$, deren Eingänge $\mathbf{u}_{f,c}$ konstant sind, $\hat{\boldsymbol{\theta}}_f^T = \begin{bmatrix} \hat{\boldsymbol{\theta}}_{f,v}^T & \hat{\boldsymbol{\theta}}_{f,c}^T \end{bmatrix}$. Damit ergibt sich auch der entsprechend aufgeteilte Eingangsvektor $\mathbf{u}_{f,k'}^T = \begin{bmatrix} \mathbf{u}_{f,v,k'}^T & \mathbf{u}_{f,c}^T \end{bmatrix}$.

Damit lautet

$$\Psi_e^T \Psi_f = \sum \left(\begin{bmatrix} 1 \\ \mathbf{u}_{e,k'} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{u}_{f,v,k'}^T & \mathbf{u}_{f,c}^T \end{bmatrix} \right) = \begin{bmatrix} \sum \mathbf{u}_{f,v,k'}^T & \sum k \cdot \mathbf{u}_{f,c}^T \\ \sum \mathbf{u}_{e,k'} \mathbf{u}_{f,v,k'}^T & \sum \mathbf{u}_{e,k'} \cdot \mathbf{u}_{f,c}^T \end{bmatrix}$$

und es ergibt sich weiter

$$(\Psi_e^T \Psi_e)^{-1} \Psi_e^T \Psi_f \cdot \Delta \hat{\boldsymbol{\theta}}_f = \begin{bmatrix} * & \mathbf{u}_{f,c}^T \\ * & \mathbf{0} \end{bmatrix} \cdot \begin{bmatrix} \Delta \hat{\boldsymbol{\theta}}_{f,v} \\ \Delta \hat{\boldsymbol{\theta}}_{f,c} \end{bmatrix},$$

wobei * als Platzhalter für hier nicht weiter interessierende Ausdrücke steht. Wesentlich ist, dass ein Fehler in den Parametern $\Delta \hat{\boldsymbol{\theta}}_{f,c}$, die zu den konstanten Eingängen gehören, sich nicht auf den Fehler der Schätzwerte $\hat{\boldsymbol{\theta}}_e$ (bis auf θ_{00}) auswirken.

Anwendung

Der gerade beschriebene Fehler kann groß sein, so dass es nicht ohne weiteres möglich ist, die bei festgehaltenen Parametern bestimmten Schätzwerte zu verwenden. Im Fall der sekundären Parameter θ_2, \dots kommt es aber weniger auf deren Absolutwert, als vielmehr auf das Verhältnis zu θ_1 an. Daher ist es hier sinnvoll, diesen Parameter auch alleine zu schätzen. Dies wird im Abschnitt 6.4.5 betrachtet.

Bei der Schätzung des primären Parameters θ_1 kommt es dagegen auf den genauen Wert an. Hier kann dafür ausgenutzt werden, dass die sekundären Parameter abschnittsweise konstant sind. Dies wird zunächst betrachtet.

Phasenweises Schätzen von θ_1

Es ist oben gezeigt, dass eine Schätzung des primären Parameters θ_1 mit einer Zwei-Parameter-Schätzung bei konstanter sekundärer Eingangsgröße u_2 auch dann biasfrei ist, wenn dabei ein falsches $\hat{\theta}_2$ angenommen wird. Damit stellt sich die Frage, ob man zu besseren Ergebnissen, d. h. einer geringeren Varianz käme, wenn man die Einrichtphase in Phasen konstanter sekundärer Eingänge aufteilt und für jede dieser Phasen eine Zwei-Parameter-Schätzung des primären Parameters θ_1 durchführt. Diese einzelnen Messungen werden dann optimal im Sinne der geringsten Varianz gemittelt.

Ein solches Vorgehen ist in Abbildung 6.6 als Ablaufplan dargestellt. Die Zwei-Parameter-Schätzung wird immer neu gestartet, wenn sich der sekundäre Eingang geändert hat. Der Subskript „Z“ bezeichnet die Schätzung aus der aktuellen Phase eines konstanten sekundären Eingangs. Mit „Zm“ werden die Werte bezeichnet, die aus den vorherigen Phasen bestimmt wurden, und „Z+“ schließlich ist die optimale Mittelung der Werte aus den vorherigen Phasen und der aktuellen Phase.

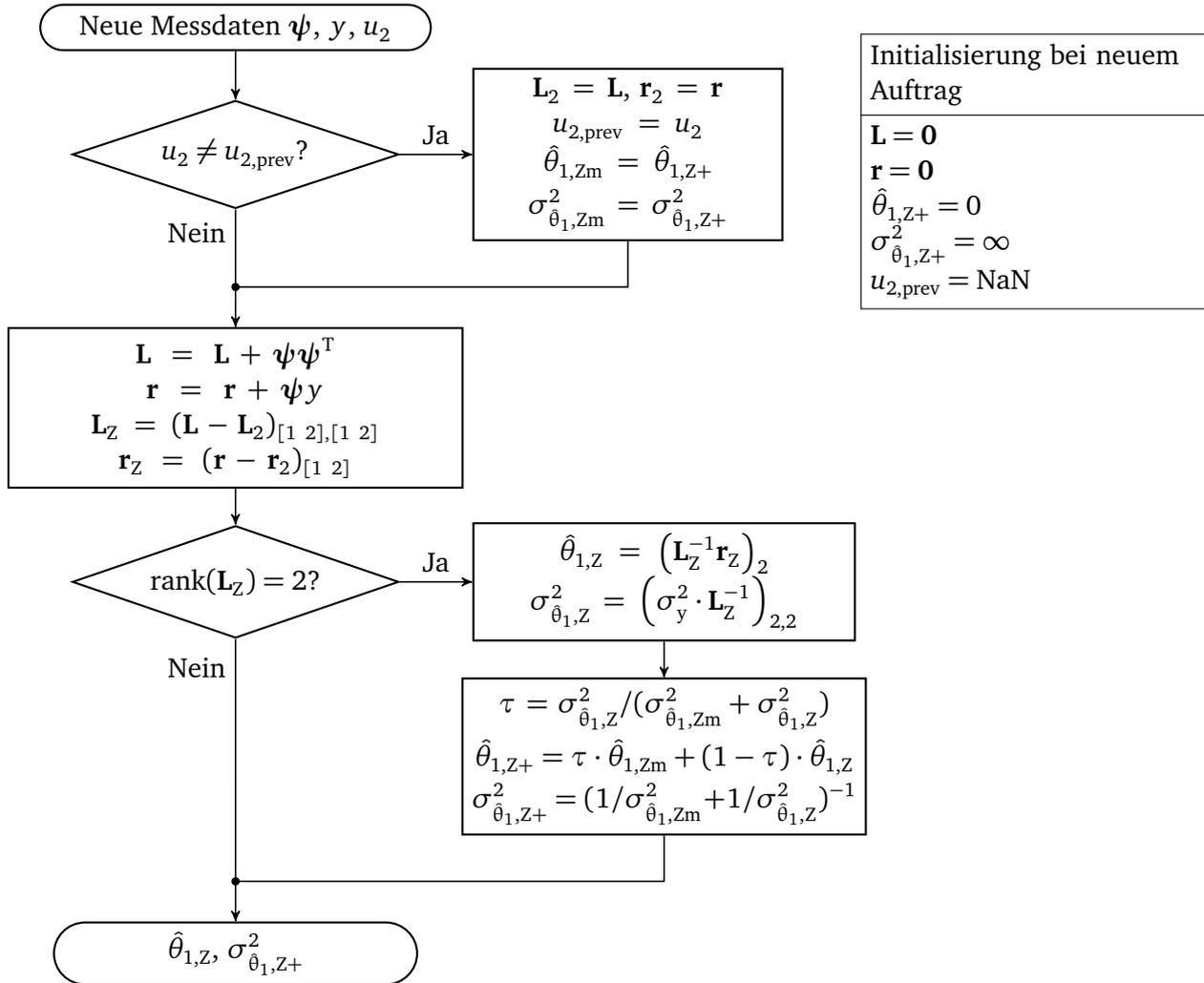


Abbildung 6.6: Phasenweises Schätzen von θ_1

Mit dem gezeigten Vorgehen ergibt sich ein biasfreier Schätzer für θ_1 . Da aber der normale LS-Schätzer effizient bezüglich aller biasfreier Schätzer ist, bedeutet das, dass die Varianz des Schätzwertes aus Abbildung 6.6 immer genauso groß oder größer wie die des LS-Schätzwertes ist, d. h.

$$\sigma_{\hat{\theta}_1}^2 \leq \sigma_{\hat{\theta}_{1,Z+}}^2.$$

Genauer lässt sich zeigen, dass die Varianz des Schätzwertes $\hat{\theta}_1$ der Drei-Parameter-Schätzung bei einer einzigen Änderung der sekundären Eingangsgröße genauso groß wie die Varianz bei der Zwei-Parameter-Schätzung ist. Bei mehreren Änderungen der sekundären Eingangsgröße ist die Drei-Parameter-Schätzung im Allgemeinen besser, d. h. deren Varianz für $\hat{\theta}_1$ geringer.

Damit ist eine Schätzung nach Abbildung 6.6 nicht sinnvoll, sondern der Algorithmus in Abbildung 6.3 stellt schon das Optimum im Sinne der Varianz dar.

6.4.5 Schätzen der sekundären Parameter bei festem primären Parameter

Verwendet man in Gl. (6.18) die Abkürzung

$$\mathbf{C}_{\text{elf}} = (\Psi_e^T \Psi_e)^{-1} \Psi_e^T \Psi_f, \tag{6.20}$$

so lautet diese

$$\hat{\theta}_{e|f} = \hat{\theta}_e - \mathbf{C}_{e|f} \cdot (\hat{\theta}_f - \theta_f).$$

Würde man nicht von den Parametern $\hat{\theta}_f$, sondern $\hat{\theta}'_f$ ausgehen, so würde sich damit natürlich ein anderer Schätzwert

$$\hat{\theta}'_{e|f} = \hat{\theta}_e - \mathbf{C}_{e|f} \cdot (\hat{\theta}'_f - \theta_f)$$

ergeben. Subtrahiert man die letzten beiden Gleichungen, so erhält man nach Umstellen

$$\hat{\theta}'_{e|f} = \hat{\theta}_{e|f} - \mathbf{C}_{e|f} \cdot (\hat{\theta}'_f - \hat{\theta}_f). \quad (6.21)$$

Geht man abermals von einem anderen Parametersatz $\hat{\theta}''_f$ aus, so kann der dazu gehörende Schätzwert $\hat{\theta}''_{e|f}$ für θ_e nach

$$\hat{\theta}''_{e|f} = \hat{\theta}_{e|f} - \mathbf{C}_{e|f} \cdot (\hat{\theta}''_f - \hat{\theta}_f) = \hat{\theta}'_{e|f} - \mathbf{C}_{e|f} \cdot (\hat{\theta}''_f - \hat{\theta}'_f)$$

aus $\hat{\theta}_{e|f}$ oder $\hat{\theta}'_{e|f}$ bestimmt werden.

Die Matrix $\mathbf{C}_{e|f}$ stellt also einen Korrekturfaktor dar, mit dessen Hilfe die Schätzung von θ_e nachträglich an ein neues $\hat{\theta}_f$ angepasst werden kann.

Der Nutzen dieser Möglichkeit der nachträglichen Korrektur des Bezugswertes $\hat{\theta}_f$ liegt darin, dass damit die Schätzwerte $\hat{\theta}_{e|f,l'}$ aller Aufträge l' an den immer besser werdenden Schätzwert für θ_f angepasst werden können, ohne die Schätzung mit den alten Messdaten und dem neuen $\hat{\theta}_f$ zu wiederholen. Geht man davon aus, dass mit $l' \rightarrow \infty$ auch $\hat{\theta}_{f,l'} \rightarrow \theta_f$ geht, so werden dadurch alle korrigierten Schätzwerte $\hat{\theta}_{e|f,l'}$ ebenfalls biasfrei und durch die Mittelung erhält man einen asymptotisch biasfreien Schätzwert für $\hat{\theta}_f$ mit immer geringer werdender Varianz.

Für diese Korrektur müssen keine Messdaten, d. h. die Informationsmatrizen $\mathbf{L}_{l'}$ sowie die Vektoren $\mathbf{r}_{l'}$ der einzelnen Aufträge vorgehalten werden, sondern lediglich die Korrekturmatrizen $\mathbf{C}_{e|f,l'}$. Genauer müssen nur die Zeilen aus $\mathbf{C}_{e|f,l'}$ gespeichert werden, die zu den Parametern gehören, deren Schätzwerte korrigiert werden sollen. So muss die Konstante θ_0 sowohl bei der kompletten Schätzung wie auch bei der Schätzung von θ_f mitgeschätzt werden, es wird aber nur der Wert aus der Gesamtschätzung verwendet. Somit muss die erste Zeile jedes $\mathbf{C}_{e|f,l'}$ nicht gespeichert werden. Besteht θ_e neben θ_0 nur aus einem weiteren Parameter, so muss nur ein Skalar gespeichert werden.

Im Folgenden wird die Mittelung der Schätzwerte der einzelnen Aufträge etwas genauer betrachtet und gezeigt, dass man mit noch weniger zu speichernden Daten auskommen kann.

Mittelung über mehrere Aufträge

Später wird in Abschnitt 6.5.2 besprochen, wie die Schätzwerte verschiedener Aufträge sinnvoll gemittelt werden, um zu besseren Ergebnissen zu kommen. Der gemittelte Schätzvektor nach dem l -ten Auftrag wird mit $\hat{\theta}_{e|f,m}(l)$ bezeichnet.¹² Dann kann diese Mittelung iterativ in der Form

$$\hat{\theta}_{e|f,m}(l+1) = \frac{g(l) \cdot \hat{\theta}_{e|f,m}(l) + \hat{\theta}_{e|f}(l+1)}{g(l) + 1} \quad \hat{\theta}_{e|f,m}(1) = \hat{\theta}_{e|f}(1)$$

¹² Abweichend zum übrigen Text wird in diesem Abschnitt die Unabhängige nicht im Index, sondern in Klammern angegeben, um zu lange Indizes zu vermeiden.

dargestellt werden, wobei $g(l)$ geeignete Gewichte sind, die im genannten Abschnitt beschrieben werden. Für die folgende Diskussion spielen die genauen Werte für $g(l)$ keine Rolle.

Nach dem $l + 1$ -ten Auftrag sollen die geschätzten Werte für $\hat{\theta}_{\text{elf}}(l')$ gemittelt werden. Dies soll bezüglich des Wertes $\hat{\theta}_f(l + 1)$ geschehen, da davon ausgegangen wird, dass dies die beste verfügbare Schätzung für θ_f ist. D. h. die bisherigen Schätzwerte $\hat{\theta}_{\text{elf}}(l')$, $l' = 1, \dots, l$, müssen vor der Mittelung auf den neuen Bezugswert $\hat{\theta}_f(l + 1)$ umgerechnet werden. Die umgerechneten Werte werden mit $\hat{\theta}'_{\text{elf}}(l')$ bezeichnet, so dass diese Umrechnung schematisch als

$$\begin{array}{ccc} \hat{\theta}_{\text{elf}}(1), \mathbf{C}_{\text{elf}}(1), \hat{\theta}_f(1) & \xrightarrow{\hat{\theta}_f(l+1)} & \hat{\theta}'_{\text{elf}}(1) \\ \hat{\theta}_{\text{elf}}(2), \mathbf{C}_{\text{elf}}(2), \hat{\theta}_f(2) & \longrightarrow & \hat{\theta}'_{\text{elf}}(2) \\ \vdots & & \vdots \\ \hat{\theta}_{\text{elf}}(l), \mathbf{C}_{\text{elf}}(l), \hat{\theta}_f(l) & \longrightarrow & \hat{\theta}'_{\text{elf}}(l) \end{array}$$

dargestellt werden kann. Mit diesen umgerechneten Werten kann nun der Mittelwert $\hat{\theta}'_{\text{elf},m}(l)$ bestimmt werden. Dieses geschieht über die iterative Berechnung von $\hat{\theta}''_{\text{elf},m}(l, l + 1)$. Allgemein soll $\hat{\theta}''_{\text{elf},m}(L, M)$ den Mittelwert bezeichnen, der sich aus den ersten L Aufträgen ergibt, wenn diese auf den Wert $\hat{\theta}_f$ des M -ten Auftrags bezogen werden. Damit gilt $\hat{\theta}'_{\text{elf},m}(l) = \hat{\theta}''_{\text{elf},m}(l, l + 1)$.

$\hat{\theta}''_{\text{elf},m}(l, l + 1)$ wird ausgehend von

$$\hat{\theta}''_{\text{elf},m}(1, l + 1) = \hat{\theta}'_{\text{elf}}(1) = \hat{\theta}_{\text{elf}}(1) - \mathbf{C}_{\text{elf}}(1) \cdot (\hat{\theta}_f(l + 1) - \hat{\theta}_f(1))$$

über

$$\hat{\theta}''_{\text{elf},m}(l' + 1, l + 1) = \frac{g(l') \cdot \hat{\theta}''_{\text{elf},m}(l', l + 1) + \hat{\theta}'_{\text{elf}}(l' + 1)}{g(l') + 1}, \quad \text{für } l' = 1, \dots, l - 1$$

bzw.

$$\hat{\theta}''_{\text{elf},m}(l' + 1, l + 1) = \frac{g(l') \cdot \hat{\theta}''_{\text{elf},m}(l', l + 1) + \hat{\theta}_{\text{elf}}(l' + 1) - \mathbf{C}_{\text{elf}}(l' + 1) \cdot (\hat{\theta}_f(l + 1) - \hat{\theta}_f(l' + 1))}{g(l') + 1}, \quad \text{für } l' = 1, \dots, l - 1 \quad (6.22)$$

bestimmt.

Der Wert $\hat{\theta}''_{\text{elf},m}(l, l + 1)$ stellt also jenen Mittelwert dar, der sich auch ergeben hätte, wenn von Beginn an mit $\hat{\theta}_f(l + 1)$ als Bezugswert gerechnet worden wäre.

Zuletzt ergibt sich

$$\hat{\theta}_{\text{elf},m}(l + 1) = \frac{g(l) \cdot \hat{\theta}''_{\text{elf},m}(l, l + 1) + \hat{\theta}_{\text{elf}}(l + 1)}{g(l) + 1} \quad (6.23)$$

als neuer Mittelwert nach dem $l + 1$ -ten Auftrag.

Um diesen Algorithmus anzuwenden, müssten also für jeden Auftrag l' die Größen $\hat{\theta}_{\text{elf}}(l')$, $\mathbf{C}_{\text{elf}}(l')$ und $\hat{\theta}_f(l')$ sowie das Gewicht $g(l')$ vorgehalten werden.¹³ Dies stellt schon einen Vorteil gegenüber der Möglichkeit dar, alle bisherigen Schätzungen anhand Gl. (6.12) zu wiederholen, wozu die L-Matrizen und r-Vektoren aller Aufträge behalten werden müssten.

Es kann aber parallel zur iterativen Mittelung der Schätzwerte auch ein Korrekturfaktor bestimmt werden, der sich auf den Mittelwert bezieht. Damit entfällt die Notwendigkeit, Daten zu einzelnen Aufträgen zu speichern. Dies wird im Folgenden gezeigt.

¹³ Allgemein wird sich das Gewicht für jeden geschätzten Parameter unterscheiden. Dies spielt aber keine Rolle, da die Mittelung ohnehin später parameterweise erfolgen wird.

Mittlung der Korrekturfaktoren

Erfolgt die Mittlung der Schätzwerte iterativ nach

$$\hat{\theta}_{\text{elf},m}(l+1) = \frac{g(l) \cdot \hat{\theta}'_{\text{elf},m}(l) + \hat{\theta}_{\text{elf}}(l+1)}{g(l) + 1}, \quad (6.24)$$

wobei $\hat{\theta}'_{\text{elf},m}(l)$ der auf $\hat{\theta}_f(l+1)$ angepasste (gewichtete) Mittelwert der ersten l Aufträge ist,

$$\hat{\theta}'_{\text{elf},m}(l) = \hat{\theta}_{\text{elf},m}(l) - \mathbf{C}_{\text{elf},m}(l) \cdot (\hat{\theta}_f(l+1) - \hat{\theta}_f(l)), \quad (6.25)$$

und wird die Korrekturmatrix $\mathbf{C}_{\text{elf},m}(l)$ durch eine analoge Mittlung

$$\mathbf{C}_{\text{elf},m}(l+1) = \frac{g(l) \cdot \mathbf{C}_{\text{elf},m}(l) + \mathbf{C}_{\text{elf}}(l+1)}{g(l) + 1} \quad (6.26)$$

bestimmt, so entspricht $\hat{\theta}_{\text{elf},m}(l+1)$ dem Mittelwert nach Gl. (6.23), den man erhalten würde, wenn man vor allen Schritten sämtliche Schätzwerte auf $\hat{\theta}_f(l+1)$ korrigiert. Bzw., wenn man Gl. (6.23) mit (6.24) vergleicht, gilt $\hat{\theta}''_{\text{elf},m}(l, l+1) = \hat{\theta}'_{\text{elf},m}(l)$. Hierbei ist zu beachten, dass diese Gleichheit nur für das Argument l gilt, nicht aber für alle l' , denn $\hat{\theta}''_{\text{elf},m}(l', l+1)$ ist immer auf $\hat{\theta}_f(l+1)$ bezogen, während sich $\hat{\theta}'_{\text{elf},m}(l')$ auf $\hat{\theta}_f(l'+1)$ bezieht.

Der Beweis ist in Anhang C.1 zu finden. Damit ist gezeigt, dass zur Korrektur der Schätzwerte auf ein neues $\hat{\theta}_f$ nur eine einzige Korrekturmatrix \mathbf{C}_{elf} gespeichert werden muss, unabhängig von der Anzahl der Aufträge bzw. Einzelschätzwerte.

Alternative Darstellung der Mittlung

Eine alternative Darstellung der gewichteten Mittlung, die später benötigt wird, ergibt sich aus

$$\begin{aligned} \hat{\theta}_{\text{elf},m}(l+1) &= \frac{g(l) \cdot \hat{\theta}'_{\text{elf},m}(l) + \hat{\theta}_{\text{elf}}(l+1)}{g(l) + 1} = \frac{g(l)}{g(l) + 1} \cdot \hat{\theta}'_{\text{elf},m}(l) + \frac{1}{g(l) + 1} \cdot \hat{\theta}_{\text{elf}}(l+1) \\ &= \left(1 - \frac{1}{g(l) + 1}\right) \cdot \hat{\theta}'_{\text{elf},m}(l) + \frac{1}{g(l) + 1} \cdot \hat{\theta}_{\text{elf}}(l+1) \\ &= (1 - \gamma(l)) \cdot \hat{\theta}'_{\text{elf},m}(l) + \gamma(l) \cdot \hat{\theta}_{\text{elf}}(l+1) \end{aligned} \quad (6.27)$$

mit $\gamma(l) = \frac{1}{g(l)+1}$. Entsprechend kann dann die Korrekturmatrix über

$$\mathbf{C}_{\text{elf},m}(l+1) = (1 - \gamma(l)) \cdot \mathbf{C}_{\text{elf},m}(l) + \gamma(l) \cdot \mathbf{C}_{\text{elf}}(l+1) \quad (6.28)$$

gemittelt werden.

Diskussion

Im Allgemeinen wird nicht immer dieselbe Aufteilung in θ_e und θ_f vorliegen.

Diejenigen Parameter, die zu Eingängen gehören, die sich bei einem Auftrag nicht geändert haben, können prinzipbedingt nicht geschätzt werden und müssen daher den θ_f zugeordnet werden. Allerdings wurde auch schon dargelegt, dass dann ein Fehler in diesen Parametern keinen Fehler der übrigen Parameterschätzung erzeugt (mit Ausnahme von $\hat{\theta}_0$).

Wenn nicht sichergestellt ist, dass alle Eingänge bei jedem Auftrag genügend angeregt werden, so muss die Mittelung und Korrektur parameterweise geschehen. Dabei fallen aber auch nicht mehr Daten an. So wird nicht eine Korrekturmatrix gespeichert, sondern mehrere Korrekturvektoren.

Wenn die Vereinbarung getroffen wird, dass ein sekundärer Parameter immer dann geschätzt wird, wenn sich der dazugehörige Eingang ändert und der primäre Eingang dabei immer festgehalten wird, dann bleibt die Handhabung relativ einfach. Da dann ein falsch angenommener Parameter bei einem sekundären Eingang nie einen Schätzfehler bei einem anderen sekundären Eingang verursacht, ist ein Fehler des Parameterschätzwertes $\hat{\theta}_1$ des primären Eingangs die einzige Fehlerquelle und damit besteht der Korrekturvektor für jeden sekundären Eingang nur aus einem Skalar.

Wenn zwei sekundäre Eingänge während eines Auftrags zu ausschließlich gleichen Schritten geändert werden, dann muss einer der beiden Parameter für die Schätzung festgehalten werden. Beispielhaft seien u_2 und u_3 zwei Eingänge, die sich zusammen geändert haben, und der Parameter θ_3 wurde zur Schätzung des Parameters θ_2 auf dem Wert $\hat{\theta}_{3,r}$ festgehalten. Dann muss zu dem Schätzwert $\hat{\theta}_2$ noch ein weiterer Korrekturfaktor $c_{2|3,m}$ gespeichert werden, mit dem $\hat{\theta}_2$ bei einer Änderungen von $\hat{\theta}_{3,r}$ angepasst werden kann. Werden danach Aufträge durchgeführt, bei denen beide Eingänge u_2 und u_3 unabhängig verändert werden und damit beide Parameter geschätzt werden können, oder der Eingang u_3 gar nicht verändert wird, so wird $c_{2|3,m}$ entsprechend der gegebenen Gleichung bei der Mittelung von $\hat{\theta}_2$ angepasst, wobei der Korrekturwert dieser Aufträge $c_{2|3} = 0$ ist.

Auf den Unterschied der hier vorgestellten Schätzung mit festgehaltenen primären Parametern und der normalen LS-Schätzung aller Parameter bezüglich der eigentlich interessierenden Werte \hat{q}_i wird nochmals in Anhang D eingegangen. Dabei wird auch noch eine weitere Variante zur Schätzung von q_i vorgestellt und in den Vergleich einbezogen. Dazu ist es aber sinnvoll, zunächst in Abschnitt 6.5 die Identifikation über mehrere Aufträge zu besprechen.

6.4.6 Schätzen der Varianz des Prozess- und Messrauschens

Die Varianz des Messrauschens bei einer Least-Squares-Schätzung wird biasfrei über den in Gl. (B.4) beschriebenen Schätzer,

$$\hat{\sigma}_{y,k}^2 = \frac{1}{k-p-1} \cdot \tilde{\mathbf{y}}_k^T \cdot \left(\mathbf{I} - \Psi_k \cdot (\Psi_k^T \Psi_k)^{-1} \cdot \Psi_k^T \right) \cdot \tilde{\mathbf{y}}_k,$$

geschätzt, wobei $p+1$ die Anzahl der Freiheitsgrade der Schätzung ist [WOOLDRIDGE, 2003, S. 793]. In diesem Fall ist $p+1 = 2$, sofern nur zwei Parameter, oder $p+1 = 3$, sofern drei Parameter geschätzt werden.

Durch Ausmultiplizieren erhält man

$$\tilde{\mathbf{y}}_k^T \cdot \left(\mathbf{I} - \Psi_k \cdot (\Psi_k^T \Psi_k)^{-1} \cdot \Psi_k^T \right) \cdot \tilde{\mathbf{y}}_k = \underbrace{\tilde{\mathbf{y}}_k^T \tilde{\mathbf{y}}_k}_{\mathbf{r}_k^T} - \underbrace{\tilde{\mathbf{y}}_k^T \Psi_k}_{\mathbf{L}_k^{-1}} \cdot \overbrace{\left(\Psi_k^T \Psi_k \right)^{-1}}^{\theta_k} \cdot \underbrace{\Psi_k^T \tilde{\mathbf{y}}_k}_{\mathbf{r}_k}.$$

Die Faktoren des rechten Summanden können wie angegeben über die iterativ bestimmten Größen \mathbf{L}_k und \mathbf{r}_k ausgedrückt werden. Der erste Summand

$$Y_k = \tilde{\mathbf{y}}_k^T \tilde{\mathbf{y}}_k$$

kann ebenfalls nach

$$Y_k = \begin{cases} 0 & k = 0 \\ Y_{k-1} + \tilde{y}_k^2 & k > 0 \end{cases}$$

iterativ berechnet werden. Somit erhält man

$$\tilde{\mathbf{y}}_k^T \cdot \left(\mathbf{I} - \Psi_k \cdot (\Psi_k^T \Psi_k)^{-1} \Psi_k^T \right) \cdot \tilde{\mathbf{y}}_k = Y_k - \mathbf{r}_k^T \mathbf{L}_k^{-1} \mathbf{r}_k = Y_k - \mathbf{r}_k^T \hat{\boldsymbol{\theta}}_k$$

und damit

$$\hat{\sigma}_{y,k}^2 = \frac{1}{k-p-1} \cdot (Y_k - \mathbf{r}_k^T \mathbf{L}_k^{-1} \mathbf{r}_k). \quad (6.29)$$

Zur Bestimmung des Schätzwertes für die Varianz muss also nur ein weiterer Skalar gespeichert werden.

Varianz des Schätzwertes der Varianz

Die Varianz des Schätzwertes der Varianz ist für wenige Messungen noch sehr groß, so dass dieser nicht sehr aussagekräftig ist. Der Schätzwert für die Varianz wird dazu verwendet, Messdaten zu verwerfen, wenn die Maschine bei einem Auftrag zu sehr rauscht oder Sprünge in θ_0 auftreten (siehe Abschnitt 6.7.2). Aufgrund der hohen Varianz des Schätzwertes $\hat{\sigma}_{y,k}^2$ bei wenigen Messungen würde eine festen Grenze für die zulässige Varianz häufig „auslösen“, obwohl die Messung an sich in Ordnung ist. Man könnte dem mit einer variablen Grenze begegnen. Es hat sich anhand von Simulationen und den Versuchsdaten aber als sinnvoll herausgestellt, nicht $p = 1$ bzw. $p = 2$ zu setzen, wie es für eine biasfreie Schätzung notwendig wäre, sondern mit

$$\hat{\sigma}_{y,k}^2 = \frac{1}{k-1} \cdot (Y_k - \mathbf{r}_k^T \mathbf{L}_k^{-1} \mathbf{r}_k). \quad (6.30)$$

einen biasbehafteten Schätzer zu verwenden. Damit erhält man für wenige Messungen zwar einen Bias auf die Schätzung (für $k \rightarrow \infty$ geht dieser gegen Null), d. h. man schätzt die Varianz als zu gering. Dies entspricht quasi einer variablen Grenze für die zulässige geschätzte Varianz $\hat{\sigma}_{y,k}^2$. In Abschnitt 6.7.2 wird diese Wahl nochmals an Simulationsdaten begründet und gerechtfertigt.

Damit erhält man den in Abbildung 6.7 dargestellten Ablauf zur Ermittlung der geschätzten Varianz.

Verwenden des Schätzwertes der Varianz

Der Schätzwert der Varianz wird später dazu verwendet werden, die Messdaten auf zu starkes Rauschen und Ausreißer zu prüfen (Abschnitt 6.7.2). Auch könnte dieser Wert als Merkmal für Diagnosefunktionen bzw. Fehlererkennung für den Prozess Verwendung finden.

Prinzipiell wäre es auch denkbar, den Schätzwert $\hat{\sigma}_{y,k}^2$ zur Adaption des angenommenen Wertes für σ_y^2 zu verwenden. Davon wird im Rahmen dieser Untersuchung aber abgesehen, da dieser Wert bei den relativ wenigen Messungen während eines Auftrags nicht sehr sicher geschätzt werden kann und darüber hinaus die Gefahr besteht, dass man einer sich langsam verschlechternden Maschine folgt. In diesem Fall wäre es besser, dem Bediener mitzuteilen, dass Probleme vorliegen.

6.4.7 Weitere Schätzalgorithmen

Der Least-Squares-Schätzer kann durch eine Gewichtung der einzelnen Messwerte oder einen Vergessensfaktor erweitert werden. Auch wäre der Einsatz eines Kalmanfilters denkbar. Für den hier betrachteten Fall des Schätzens während eines Auftrags spielen diese Modifikationen aber eine weniger große Rolle.

Ein prinzipiell interessante Variante wäre die Methode der „Total Least Squares“, da sich damit auch Eingangsstörungen explizit berücksichtigen lassen [ISERMANN und MÜNCHHOF, 2011, S. 297ff]. Allerdings müssten die Parameter des Eingangsrauschens dazu bekannt sein und die Methode ist mathematisch deutlich aufwendiger umzusetzen.

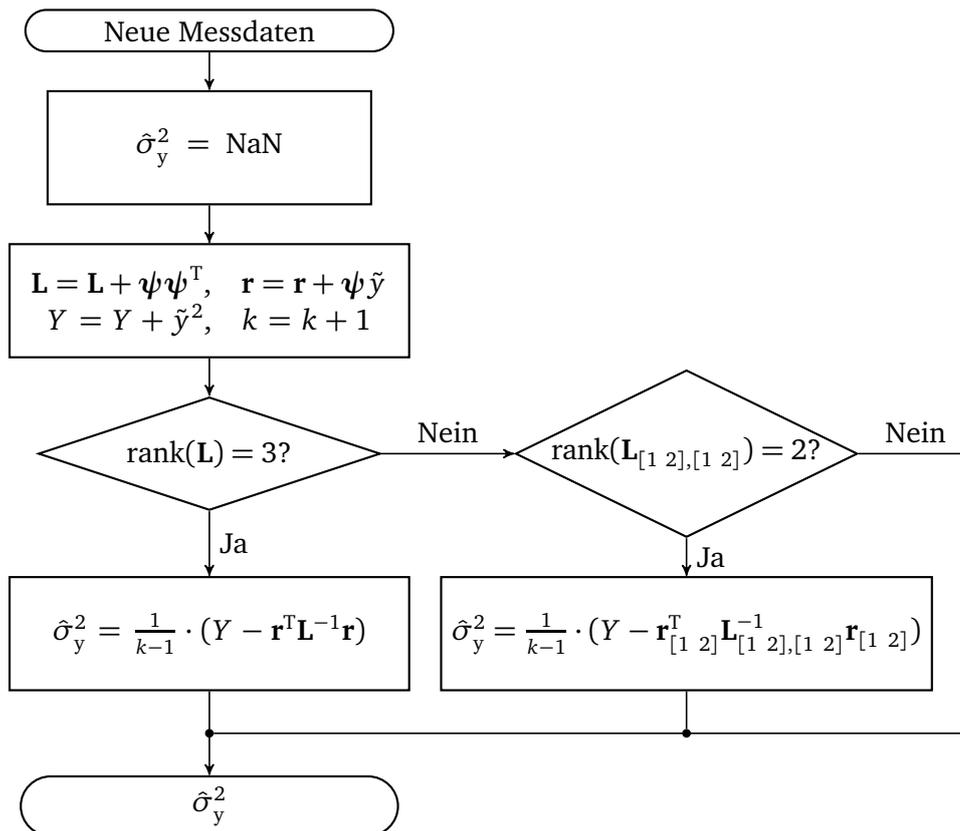


Abbildung 6.7: Bestimmen des Schätzwerts der Varianz

6.5 Identifikation über mehrere Aufträge

Bei der Identifikation über mehrere Aufträge besteht die hauptsächliche Problematik darin, dass die Parameter – allen voran θ_0 – von Auftrag zu Auftrag schwanken können.

Es wurden zwei Vorgehensweisen untersucht, nämlich

- das Fortführen einer einzelnen Schätzung über mehrere Aufträge sowie
- die Mittelung der für jeden Auftrag einzeln geschätzten Parameter.

Zunächst wird in Abschnitt 6.5.1 kurz auf die Verfahren eingegangen, die auf einer fortzuführenden Schätzung beruhen. Da sich mit diesen keine zufriedenstellenden Ergebnisse erzielen lassen, wird anschließend in Abschnitt 6.5.2 die Mittelung der einzeln geschätzten Parameter behandelt.

6.5.1 Fortführen eines Schätzalgorithmus über mehrere Aufträge

Zunächst wird die

- Zusammenführung der Messungen mit Korrektur des Offsets (θ_0)

behandelt. Darüber hinaus gibt es auch Varianten des Least-Squares-Verfahrens, die für zeitvariante Systeme geeignet sind. In [ISERMANN und MÜNCHHOF, 2011, S. 335ff] werden dazu folgende Verfahren angegeben:

- Exponentielles Vergessen mit konstantem Vergessensfaktor λ ,
- Exponentielles Vergessen mit variablen Vergessensfaktor λ_k und

- Direkte Beeinflussung der Kovarianzmatrix.

Eine weitere Möglichkeit ist die Verwendung eines

- „Extended-Kalmanfilters“.

Zusammenführung der Messungen mit Korrektur des Offsets

In Abbildung 6.8a sind (konstruierte) Messdaten eines Auftrags dargestellt. Dazu ist mit der schwarzen gestrichelten Linie die reale Funktion und mit der durchgezogenen schwarzen Linie die auf den Messpunkten beruhende Schätzung der Funktion dargestellt. In Abbildung 6.8b ist selbiges für einen zweiten Auftrag, in Grau, dargestellt. (Auf die gepunktete schwarze Linie wird später eingegangen.) Beim grauen Auftrag ist im Vergleich zum schwarzen Auftrag der Parameter θ_0 des realen Systems deutlich kleiner, so dass die Gerade tiefer verläuft. Der Parameter θ_1 , und damit die Steigung der Geraden, ist in beiden Fällen gleich.

Die Messpunkte des grauen Auftrags liegen sehr dicht beieinander, so dass keine zuverlässige Schätzung möglich ist, wie es sich hier auch zeigt.

Es stellt sich die Frage, ob durch die Kombination der beiden Messwertsätze eine bessere Schätzung möglich ist. Offensichtlich reicht es nicht, die Messwertsätze einfach aneinanderzuhängen, wie in Abbildung 6.8c gezeigt. Dabei würde hier das θ_1 viel zu gering geschätzt werden.

Wenn der Unterschied der Parameter θ_0 von beiden Aufträgen exakt bekannt wäre, so könnten die Messwerte des ersten Auftrags entsprechend korrigiert werden (siehe Anhang B.6), so dass sich das in Abbildung 6.8d gezeigte Ergebnis ergibt. In diesem Fall verbessert sich die Schätzung der Steigung θ_1 im Vergleich zum Auftrag 1 in Abbildung 6.8a.

In der Praxis ist aber die Abweichung der Parameter θ_0 nicht bekannt und muss aus den Messdaten geschätzt werden. Dazu wird mit den Daten von Auftrag 2 der Parameter $\theta'_{0,2}$ geschätzt, wobei dazu der Schätzwert $\hat{\theta}_{1,1}$ für die Steigung des Auftrags 1 verwendet wurde. Damit ergibt sich die in Abbildung 6.8b dargestellte gepunktete schwarze Linie. Es ist deutlich zu sehen, dass bei $u_{1,N} = 50$ ein gewisser Abstand zur grauen gestrichelten Linie besteht, der Schätzwert $\hat{\theta}'_{0,2}$ demnach fehlerhaft ist. Bildet man die Differenz des Schätzwertes $\hat{\theta}'_{0,2}$ und $\hat{\theta}_{0,1}$ und korrigiert bzw. verschiebt man damit die Messwerte aus Auftrag 1, so erhält man das in Abbildung 6.8e Gezeigte. Die Schätzung ist deutlich besser als bei der einfachen Kombination der Messungen, aber schlechter als bei der idealen Korrektur. Bezüglich der Schätzung aus Auftrag 1 ist man, bezogen auf die Steigung, etwas schlechter geworden.

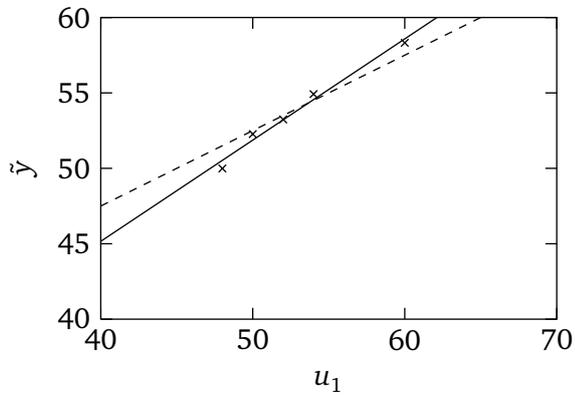
Aus diesem einfachen Beispiel können folgende Punkte festgehalten werden:

- Die Messwerte verschiedener Aufträge dürfen nicht einfach aneinandergehängt werden, da dies zu völlig falschen Ergebnissen führen kann. Dies ist auch nochmals in Abbildung 6.9 an dem Testfall mit einem Sprung in θ_0 im vierten Auftrag gezeigt.
- Eine Anpassung der zurückliegenden Messungen an das aktuelle θ_0 lässt sich nicht sicher durchführen, da hierfür nur Schätzwerte vorliegen.

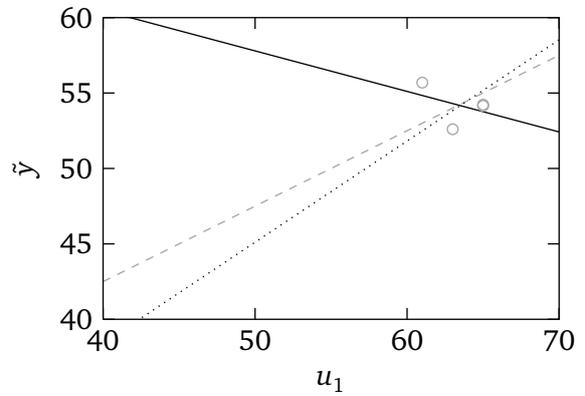
Erweiterungen der Least-Squares-Methode

Konstanter oder variabler Vergessensfaktor

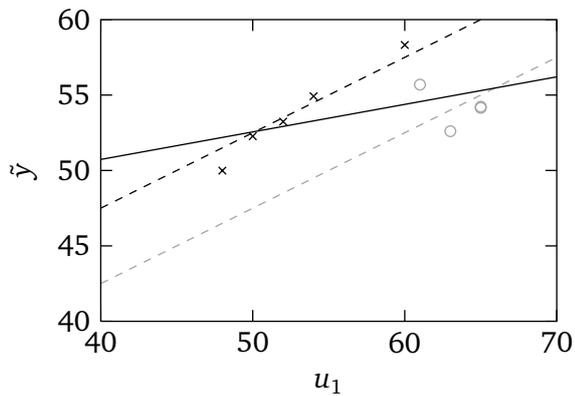
Durch die Verwendung eines Vergessensfaktors wird erreicht, dass die weiter zurückliegenden Messungen immer weniger Einfluss auf den Schätzwert haben. Dadurch verfälschen diese immer weniger den Schätzwert, wenn sich die Systemparameter mit der Zeit ändern.



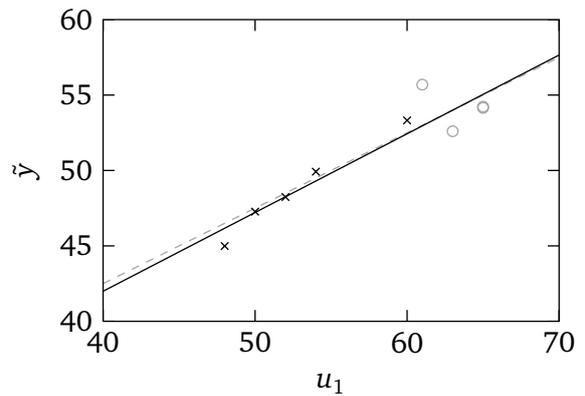
(a) Auftrag 1



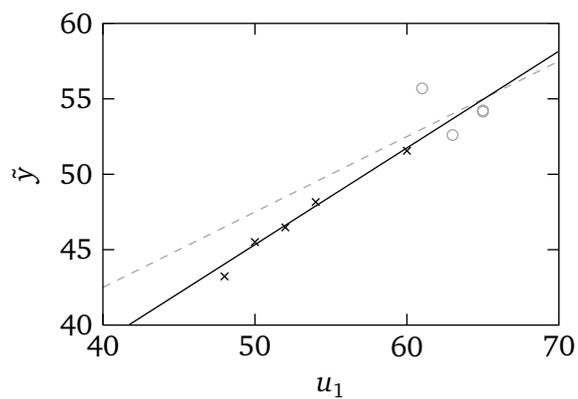
(b) Auftrag 2



(c) Auftrag 1 + Auftrag 2



(d) Auftrag 1 + Auftrag 2 ideal korrigiert



(e) Auftrag 1 + Auftrag 2 real korrigiert

Abbildung 6.8: Zur Identifikation über mehrere Aufträge

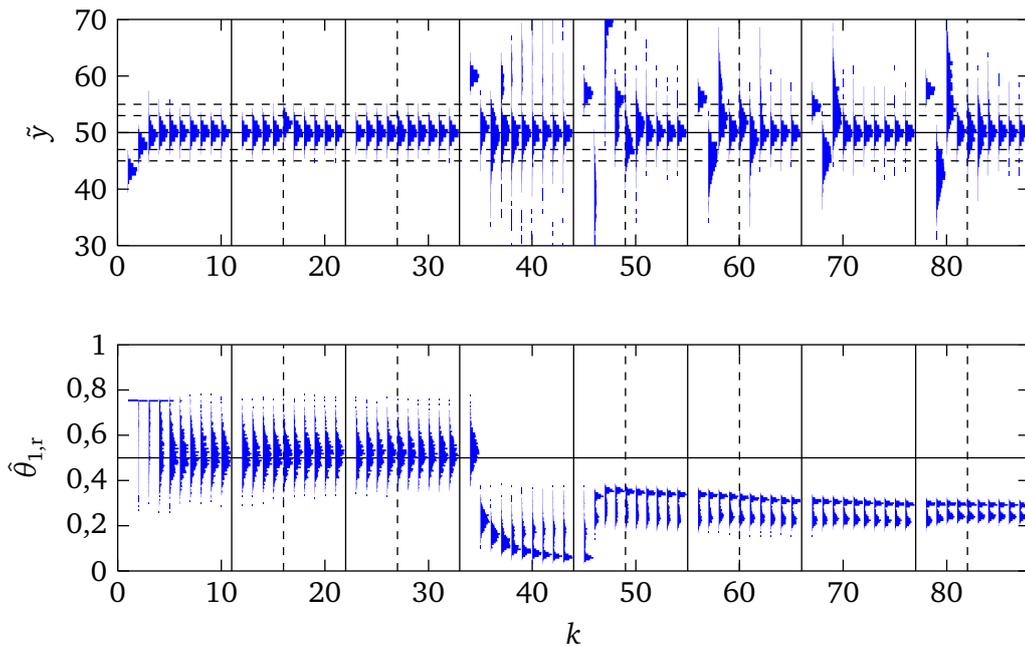


Abbildung 6.9: Eine LS-Schätzung über mehrere Aufträge mit Sprung in θ_0 (Die senkrechten schwarzen Striche markieren jeweils den Beginn eines neuen Auftrags, ab dem vierten Auftrag ist θ_0 um 10 höher.) (1 000 Simulationen)

Ein Sprung eines Parameterwerts kann dabei natürlich nicht ideal erkannt werden. Dadurch, dass die Messwerte vor dem Sprung zunächst noch stark berücksichtigt werden, werden die Parameter nach dem Sprung zunächst falsch geschätzt.

So ist dieses Verfahren, wie auch in [ISERMANN und MÜNCHHOF, 2011] beschrieben, nur für Systeme mit langsamen Änderungen der Parameter geeignet.

Die Fehler, die in den Schätzungen von θ_1 und θ_2 durch die möglichen Sprünge von θ_0 auftreten können, können sehr groß sein, so dass ein rasches Vergessen notwendig ist. Damit werden aber die alten Messungen zu schnell verworfen. Bei der Auslegung des Vergessensfaktors ist auch problematisch, dass die Anzahl der Messpunkte je Auftrag nicht gleich ist.

Auch können die Varianzen der Schätzwerte wieder steigen, d. h. schlechter werden, wenn viele Messungen über mehrere Aufträge ohne große Änderungen der Eingangsgrößen durchgeführt werden. Während Phasen konstanter Eingänge müsste der Vergessensfaktor daher auf eins gesetzt werden (und damit zu dem normalen LS-Verfahren zurückgekehrt werden), damit keine singulären Matrizen auftreten bzw. starkes Rauschen der Schätzwerte vermieden wird.

Direkte Beeinflussung der Kovarianzmatrix

Für Systeme mit schnell veränderlichen Parametern wird in [ISERMANN und MÜNCHHOF, 2011, S. 341ff] die direkte Beeinflussung der Kovarianzmatrix vorgeschlagen. Für dieses Verfahren muss die rekursive Least-Squares-Methode (siehe Anhang B.4) verwendet werden.

Die „Kovarianzmatrix“ \mathbf{P} ist gleich der Matrix $\mathbf{\Psi}^T \mathbf{\Psi}$. (Sie entspricht bis auf den Faktor σ_y^2 der tatsächlichen Kovarianzmatrix.) Anhand dieser und den neuen Eingangswerten ψ_{k+1} wird ein Korrekturfaktor

$$\gamma_k = \frac{\mathbf{P}_k \psi_{k+1}}{\psi_{k+1}^T \mathbf{P}_k \psi_{k+1} + 1}$$

bestimmt, mit dem der Schätzvektor und die Kovarianzmatrix korrigiert wird,

$$\hat{\boldsymbol{\theta}}_{k+1} = \hat{\boldsymbol{\theta}}_k + \boldsymbol{\gamma}_k \cdot \underbrace{(\tilde{y}_{k+1} - \hat{y}_{k+1|k})}_{\boldsymbol{\psi}_{k+1}^T \hat{\boldsymbol{\theta}}_k} \quad \text{und}$$

$$\mathbf{P}_{k+1} = (\mathbf{I} - \boldsymbol{\gamma}_k \boldsymbol{\psi}_{k+1}^T) \cdot \mathbf{P}_k.$$

In die Korrektur des Schätzvektors fließt neben dem Korrekturfaktor die Differenz zwischen dem Messwert \tilde{y}_{k+1} und dem mit dem bisherigen Parametervektor $\hat{\boldsymbol{\theta}}_k$ bestimmten Schätzwert $\hat{y}_{k+1|k}$ ein. Die Korrektur von \mathbf{P} ist unabhängig vom Messwert.

Beim normalen Least-Squares-Verfahren gilt

$$\mathbf{P} = (\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} = \frac{1}{\sigma_y^2} \cdot \begin{bmatrix} \sigma_{\hat{\theta}_0}^2 & \text{Cov}(\{\hat{\theta}_0\}_r, \{\hat{\theta}_1\}_r) \\ \text{Cov}(\{\hat{\theta}_0\}_r, \{\hat{\theta}_1\}_r) & \sigma_{\hat{\theta}_1}^2 \end{bmatrix}$$

bzw. $\mathbf{P} = (\boldsymbol{\Psi}^T \mathbf{W} \boldsymbol{\Psi})^{-1}$ beim Least-Squares-Verfahren mit Gewichtung oder Vergessensfaktor, wobei im letztgenannten Fall die Interpretation der Matrix \mathbf{P} als Kovarianzmatrix nicht mehr direkt möglich ist. Insbesondere wenn die Matrix \mathbf{P} per Hand beeinflusst wird entsprechen die Einträge dieser Matrix nicht mehr den Varianzen der Schätzwerte, auch wenn diese Interpretation immer noch gut den Einfluss der manuell geänderten Werte in \mathbf{P} erklärt. Im Folgenden werden am Beispiel der Schätzung von θ_0 und θ_1 die Möglichkeiten und Grenzen der Einflussnahme auf \mathbf{P} dargelegt. In diesem Fall hat \mathbf{P} die allgemeine Struktur

$$\mathbf{P} = \begin{bmatrix} s_0 & c_{01} \\ c_{01} & s_1 \end{bmatrix}.$$

Während man mit dem Vergessensfaktor nur Messungen gewichtet, hat man hierbei auch die Möglichkeit, gezielt auf einzelne Parameter Einfluss zu nehmen. Dazu wird die Gleichung für \mathbf{P}_{k+1} um einen Summanden $\mathbf{Q} = \text{diag}([q_{\theta_0} \ q_{\theta_1}])$ erweitert,

$$\mathbf{P}_{k+1} = (\mathbf{I} - \boldsymbol{\gamma}_k \boldsymbol{\psi}_{k+1}^T) \cdot \mathbf{P}_k + \begin{bmatrix} q_{\theta_0} & \\ & q_{\theta_1} \end{bmatrix}.$$

Wird q_{θ_0} sehr groß gewählt, dann wird damit eine große Unsicherheit des bisher bestimmten Parameters $\hat{\theta}_0$ ausgedrückt und folglich dieser stark korrigiert. Für $q_{\theta_0} \rightarrow \infty$ wird

$$\boldsymbol{\gamma}_k = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

und damit

$$\hat{\boldsymbol{\theta}}_{k+1} = \hat{\boldsymbol{\theta}}_k + \begin{bmatrix} 1 \\ 0 \end{bmatrix} (y_{k+1} - \boldsymbol{\psi}_{k+1}^T \hat{\boldsymbol{\theta}}_k) = \hat{\boldsymbol{\theta}}_k + \begin{bmatrix} 1 \\ 0 \end{bmatrix} (y_{k+1} - \hat{y}_{k+1|k}).$$

Aus der letzten Zeile ist zu sehen, dass in diesem Fall der Schätzwert des zweiten Parameters θ_1 überhaupt nicht korrigiert wird und die gesamte Abweichung mit dem ersten Parameter vollständig erklärt wird. „Vollständig“ bedeutet dabei, dass nach der Korrektur der letzte Messwert ($\boldsymbol{\psi}_{k+1}, y_{k+1}$) mit dem neuen Parametervektor $\hat{\boldsymbol{\theta}}_{k+1}$ exakt getroffen wird.

Für \mathbf{P}_{k+1} ergibt sich bei $q_{\theta_0} \rightarrow \infty$

$$\mathbf{P}_{k+1} = \begin{bmatrix} 1 + s_1 v_{k+1}^2 & -s_1 v_{k+1} \\ -s_1 v_{k+1} & s_1 \end{bmatrix}.$$

Der Eintrag $(\mathbf{P})_{1,1}$ ist damit direkt wieder relativ klein.

Zusammengefasst lässt sich sagen, dass ein hoher Eintrag für q_{θ_0} dazu führt, dass

- der Wert $\hat{\theta}_0$ so korrigiert wird, dass dieser die neue Messung erklärt und
- der Eintrag s_0 nach dieser Messung wieder relativ klein ist, so dass sich $\hat{\theta}_0$ nur wieder langsam ändern wird.

Dies ist ähnlich dem Verhalten, wenn man schon nach einem Schritt die Messdatensätze mit einer Off-setkorrektur zusammenfassen würde. Dabei ist wieder schwer zu sagen, wie sich ein Fehler, der damit möglicherweise gemacht wird, auf die folgenden Schätzwerte auswirkt. Damit darf q_{θ_0} nicht zu groß gewählt werden. Wählt man q_{θ_0} andererseits zu klein, so dauert es länger, bis nach einem Sprung in den Parametern diese wieder richtig gelernt werden. In der Zwischenzeit können auch grob falsche Parameter auftreten.

Somit kann festgehalten werden, dass es zwar über die direkte Beeinflussung der Kovarianzmatrix möglich ist, einzelne Parameter als sicher bzw. unsicher festzulegen und eine entsprechende Korrektur zu erreichen. Dieses Verfahren ist aber hier, mit sprunghaften Änderungen von θ_0 und sehr wenigen Messungen, nicht geeignet.

Zudem hat man bei den rekursiven Verfahren auch immer das Problem, die Matrix \mathbf{P} zu initialisieren. Es hat sich gezeigt, dass die richtige Wahl der Startwerte wesentlich für eine schnelle Konvergenz ist. So ist diese idealerweise über das bekannte Least-Squares-Verfahren zu bestimmen. Da in der Regel aber zunächst nur θ_0 und θ_1 zu schätzen sind und θ_2 erst später dazukommt, wird dies auch wieder aufwendiger.

Extended-Kalmanfilter

Im Prinzip entspricht der Extended-Kalmanfilter in diesem speziellen Fall einem gewichteten (rekursiven) Least-Squares-Verfahren, bei dem die Kovarianzmatrix in jedem Schritt direkt beeinflusst wird. Dabei wird in jedem Schritt die Kovarianzmatrix \mathbf{Q}_k des Zustandsrauschens zu \mathbf{P} hinzuaddiert. Dieses kann, wie angedeutet, auch zeitvariant sein. (Siehe auch Abschnitt B.7.)

Für den Gewichtungsfaktor r_k ist dabei die Varianz des Messrauschens zu wählen, also $r_k = \sigma_y^2$. Es kann aber auch mit $r_k = 1$ gearbeitet werden, nur sind dann \mathbf{Q}_k und der Anfangswert \mathbf{P}_1 entsprechend darauf anzupassen.

Da dieses Verfahren damit identisch zu der gerade besprochenen Variante des rekursiven Least-Squares-Verfahren ist, gilt auch das dort gesagte. Der Kalmanfilter liefert damit nur noch die theoretische Grundlage für eine andere Interpretation.

6.5.2 Mittelung der für jeden Auftrag einzeln geschätzten Parameter

Wie im vorherigen Abschnitt dargelegt wurde, ist es aufgrund der möglichen Sprünge von θ_0 nicht möglich, eine einzige Schätzung über mehrere Aufträge durchzuführen. Schätzt man die Parameter dagegen für jeden Auftrag einzeln, so wird das Problem mit den Sprüngen von θ_0 umgangen. Es stellt sich damit aber die Frage, wie die Schätzwerte aus den einzelnen Aufträgen gemittelt werden sollen. Dieser Abschnitt widmet sich dieser Frage.

Dazu wird einleitend die „einfache Mittelung“ und anschließend die „varianz-optimale Mittelung“ der Schätzwerte beschrieben. Beide Verfahren haben den Nachteil, dass diese ungeeignet sind, wenn sich die zu schätzenden Parameter mit der Zeit ändern. Für diesen Fall ist eine „Mittelung mit Vergessensfaktor“ notwendig. Zuletzt wird eine „Kombination aus varianz-optimaler Mittelung und Mittelung mit Vergessensfaktor“ betrachtet.

Im darauffolgenden Abschnitt 6.5.3 wird das Verhalten bei Sprüngen von θ_1 und θ_2 anhand von Simulationen besprochen.

Bevor mit der Beschreibung der verschiedenen Möglichkeiten zur Mittelung begonnen wird, werden noch das prinzipielle Vorgehen und die Benennung der Variablen erläutert.

Prinzipielles Vorgehen, Benennung der Variablen

Es werden die Schätzwerte $\hat{\theta}_{i,l'}$, $i \geq 1$, verschiedener Aufträge l' jeweils für sich gemittelt. Aufgrund der starken Schwankungen wird für θ_0 immer nur der Wert des aktuellen Auftrags berücksichtigt. Im Folgenden wird zunächst der Parameter θ_1 zum primären Eingang behandelt, für die Parameter θ_i , $i \geq 2$, zu den sekundären Eingängen ergibt sich ein weitestgehend analoges Vorgehen. Auf Unterschiede wird am Ende eingegangen.

In Abbildung 6.10 ist vereinfacht die Mittelung eines Parameters während eines Auftrages dargestellt. Mit dem Index „m“ werden die Mittelwerte der geschätzten Parameter bezeichnet, die jeweils am *Ende* eines Auftrags gültig sind und abgespeichert werden.

Wenn ein neuer Auftrag begonnen wird, wird die Variable $\hat{\theta}_{1,m}$ geladen, welche den gemittelten Schätzwert für θ_1 aus den bisherigen Aufträgen enthält. Dieser wird zunächst in die Variable $\hat{\theta}_{1,+}$ kopiert.¹⁴

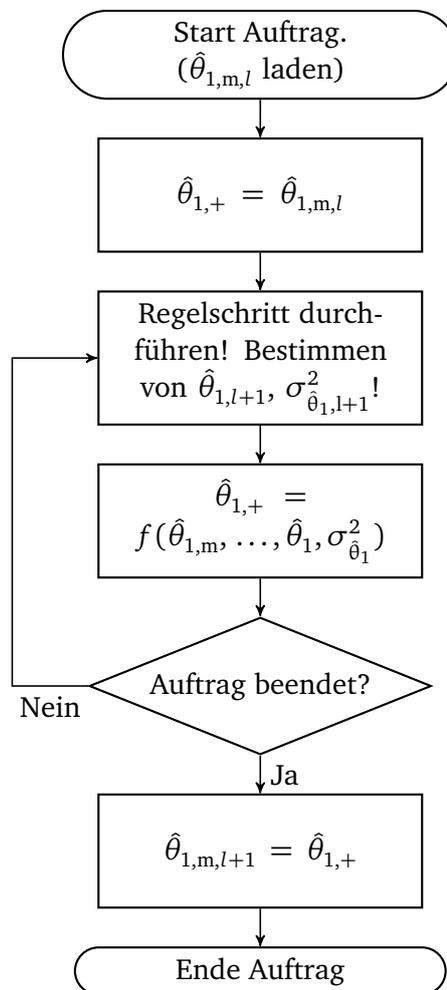


Abbildung 6.10: Zur Indizierung der Variablen bei der Mittelung

¹⁴ Im später gezeigten Algorithmus wird $\hat{\theta}_{1,+}$ nur unter bestimmten Bedingungen aktualisiert, was in dieser prinzipiellen Darstellung nicht gezeigt ist. Mit dem Kopieren wird $\hat{\theta}_{1,+}$ auf einen definierten Wert gesetzt und kann ohne weitere Abfrage nach Beenden des Auftrags in $\hat{\theta}_{1,m}$ zurückgeschrieben werden.

Wird ein Regelschritt durchgeführt, so wird nur mit den Messdaten des aktuellen Auftrags ein Schätzwert $\hat{\theta}_1$ und dessen Varianz $\sigma_{\hat{\theta}_1}^2$ bestimmt. Mit diesen Werten und dem bisherigen gemittelten Schätzwert $\hat{\theta}_{1,m}$ sowie weiteren, hier nicht aufgeführten Werten, wird ein neuer gemittelter Schätzwert $\hat{\theta}_{1,+}$ bestimmt. Dieser wird in der mit „+“ indizierten Variablen gespeichert, da noch weitere Regelschritte in diesem Auftrag folgen können und der mit „m“ indizierte Mittelwert daher nicht überschrieben werden darf. Dies ist darin begründet, dass die Mittelung nur einmal am Ende jedes Auftrags durchgeführt werden soll. Da aber möglichst früh schon etwaige bessere Schätzwerte zur Regelung verwendet werden sollen, wäre es ungünstig, bis zum Ende des aktuellen Auftrags damit zu warten. Daher werden während eines Auftrags die „+“-Werte erzeugt. Der letzte dieser Werte wird nach Beenden des Auftrags in die „m“-Variable kopiert und für den nächsten Auftrag gespeichert.

Einfache Mittelung

Es liegen l Schätzwerte $\hat{\theta}_{1,l'}$ vor, die jeweils der Verteilung

$$\{\hat{\theta}_{1,l'}\}_r \sim N(\theta_1, \sigma_{\hat{\theta}_{1,l'}}^2)$$

entsprechen und unkorreliert sind. Der Mittelwert der Schätzwerte 1 bis l ist über

$$\hat{\theta}_{1,m,l} = \frac{1}{l} \sum_{l'=1}^l \hat{\theta}_{1,l'}$$

gegeben. Dieser ist nach

$$\{\hat{\theta}_{1,m,l}\}_r \sim N\left(\theta_1, \sum_{l'=1}^l \frac{\sigma_{\hat{\theta}_{1,l'}}^2}{l^2}\right) \sim N(\theta_1, \sigma_{\hat{\theta}_{1,l}}^2).$$

verteilt. Aus dem Auftrag $l + 1$ wird ein Parameter

$$\{\hat{\theta}_{1,l+1}\}_r \sim N(\theta_1, \sigma_{\hat{\theta}_{1,l+1}}^2)$$

geschätzt. Bezieht man diesen in die Mittelung ein,

$$\hat{\theta}_{1,m,l+1} = \frac{1}{l+1} \sum_{l'=1}^{l+1} \hat{\theta}_{1,l'} = \frac{l \cdot \hat{\theta}_{1,m,l} + \hat{\theta}_{1,l+1}}{l+1},$$

so ist $\hat{\theta}_{1,m,l+1}$ eine Zufallsvariable mit der Verteilung

$$\{\hat{\theta}_{1,m,l+1}\}_r \sim N\left(\theta_1, \sum_{l'=1}^{l+1} \frac{\sigma_{\hat{\theta}_{1,l'}}^2}{(l+1)^2}\right) \sim N\left(\theta_1, \frac{l^2}{(l+1)^2} \sigma_{\hat{\theta}_{1,m,l}}^2 + \frac{1}{(l+1)^2} \sigma_{\hat{\theta}_{1,l+1}}^2\right).$$

Würde man alle Schätzwerte ohne Überprüfung für die Mittelung berücksichtigen, so würde die Varianz des Mittelwertes und damit dessen Güte von den Einzelschätzwerten dominiert werden, die eine hohe Varianz haben. Damit dürfen neue Schätzwerte $\hat{\theta}_{1,l'}$ nur dann berücksichtigt werden, wenn deren Varianz unter einer festzulegenden Grenze liegt. Solch eine Grenze kann über verschiedene Bedingungen bestimmt werden.

Eine mögliche Bedingung wäre zu fordern, dass nur Schätzwerte $\hat{\theta}_{1,l}$ berücksichtigt werden, deren Varianz geringer als $\sigma_{\hat{\theta}_{1,m,\max}}^2$ ist,

$$\sigma_{\hat{\theta}_{1,l}}^2 < \sigma_{\hat{\theta}_{1,m,\max}}^2. \quad (6.31)$$

Damit würde

$$\sigma_{\hat{\theta}_{1,m,l+1}}^2 < \sigma_{\hat{\theta}_{1,m,\max}}^2 \cdot \frac{1}{l+1}$$

gelten. Wenn beispielsweise $\sigma_{\hat{\theta}_{1,m,\max}}^2 = 3 \cdot \sigma_{\hat{\theta}_{1,r,\max}}^2$ gesetzt wird, dann wäre nach maximal $l+1 = 3$ Mittelungen die Varianz von $\sigma_{\hat{\theta}_{1,m,l+1}}^2$ geringer als die für die Regelung maximal zulässige Varianz $\sigma_{\hat{\theta}_{1,r,\max}}^2$.

Allerdings ist damit weder sichergestellt, dass die Varianz des gemittelten Wertes monoton fällt, noch dass die Varianz des gemittelten Wertes geringer ist als die des aktuellen Einzelschätzwertes.

Varianz-optimale Mittelung

Die Zufallsvariablen $\hat{\theta}_{1,m,l}$ und $\hat{\theta}_{1,l+1}$ mit den Varianzen $\sigma_{\hat{\theta}_{1,m,l}}^2$ bzw. $\sigma_{\hat{\theta}_{1,l+1}}^2$ werden optimal im Sinne der kleinsten Varianz über die Gleichung

$$\hat{\theta}_{1,m,l+1} = (1 - \gamma_{1,l}) \cdot \hat{\theta}_{1,m,l} + \gamma_{1,l} \cdot \hat{\theta}_{1,l+1} \quad (6.32)$$

mit

$$\gamma_{1,l} = \frac{\sigma_{\hat{\theta}_{1,m,l}}^2}{\sigma_{\hat{\theta}_{1,m,l}}^2 + \sigma_{\hat{\theta}_{1,l+1}}^2} \quad (6.33)$$

gemittelt (siehe Abschnitt A.3). Für die Varianz der gewichteten Summe ergibt sich damit

$$\sigma_{\hat{\theta}_{1,m,l+1}}^2 = (1 - \gamma_{1,l})^2 \cdot \sigma_{\hat{\theta}_{1,m,l}}^2 + \gamma_{1,l}^2 \cdot \sigma_{\hat{\theta}_{1,l+1}}^2 = (1 - \gamma_{1,l}) \cdot \sigma_{\hat{\theta}_{1,m,l}}^2. \quad (6.34)$$

Diese Gleichungen entsprechen gerade dem Kalman-Filter zur Identifikation des einzigen Zustandes $\theta_{1,m,l}$ des durch die Zustandsgleichung $\theta_{1,m,l+1} = \theta_{1,m,l}$ und Ausgangsgleichung $y = \theta_{1,m,l} + r_l$ gegebenen Systems, wobei für das Messrauschen $\text{Var}\{r_l\}_r = \sigma_{\hat{\theta}_{1,l}}^2$ gilt. (D. h. für die Gleichungen in Anhang B.7 gilt $\mathbf{A} = 1$, $\mathbf{B} = 0$, $\mathbf{C} = 1$ und $\mathbf{D} = 0$ sowie $\mathbf{Q} = 0$ und $\mathbf{R} = \sigma_{\hat{\theta}_{1,l}}^2$.)

Hierbei ist auch sichergestellt, dass die Varianz des gemittelten Schätzwertes monoton fällt und dass diese immer kleiner ist als die des aktuellen (und aller vorherigen) Einzelschätzwertes.

Mittelung mit Vergessensfaktor

Bisher wurde angenommen, dass θ_1 über alle Aufträge, bei denen die $\hat{\theta}_{1,l}$ ermittelt wurden, konstant ist. Im Allgemeinen wird θ_1 aber auch vom aktuellen Auftrag abhängen, also $\theta_1 = \theta_{1,l'}$. Damit ergibt sich (bei der einfachen Mittelung) die Verteilung

$$\{\hat{\theta}_{1,m,l+1}\}_r \sim N \left(\frac{1}{l+1} \sum_{l'=1}^{l+1} \theta_{1,l'}, \sum_{l'=1}^{l+1} \frac{\sigma_{\hat{\theta}_{1,l'}}^2}{l+1^2} \right)$$

für den gemittelten Wert. Auf die Varianz hat ein veränderliches $\theta_{1,l'}$ also keinen Einfluss, allerdings auf den Mittelwert von $\{\hat{\theta}_{1,m,l+1}\}_r$. Dieser ist im Allgemeinen ungleich dem Ist-Wert von $\theta_{1,l+1}$, so dass der Schätzer biasbehaftet ist.

Um diesem Problem zu begegnen, kann ein Vergessensfaktor λ_1 ($0 < \lambda_1 < 1$) bei der Mittelung verwendet werden, mit dem die Werte $\hat{\theta}_{1,l}$ umso weniger gewichtet werden, je weiter sie zurückliegen (vgl. [ISERMANN und MÜNCHHOF, 2011, S. 335], [KRAUS, 1986, S. 48ff]). Der Wert $\hat{\theta}_{1,m,l+1}$ wird dabei nach

$$\hat{\theta}_{1,m,l+1} = \frac{\sum_{l'=1}^{l+1} \lambda_1^{l+1-l'} \hat{\theta}_{1,l'}}{\sum_{l'=1}^{l+1} \lambda_1^{l+1-l'}} \quad (6.35)$$

berechnet. In rekursiver Form, d. h. zur Berechnung von $\hat{\theta}_{1,m,l+1}$ aus $\hat{\theta}_{1,m,l}$ und $\hat{\theta}_{1,l+1}$, kann die Gleichung

$$\hat{\theta}_{1,m,l+1} = \frac{\lambda_1 g_{1,m,l} \cdot \hat{\theta}_{1,m,l} + \hat{\theta}_{1,l+1}}{\lambda_1 g_{1,m,l} + 1} \quad (6.36)$$

verwendet werden, wobei

$$g_{1,m,l} = \sum_{l'=1}^l \lambda_1^{l-l'} = \frac{1 - \lambda_1^l}{1 - \lambda_1}$$

hier das „Gewicht“ der ersten l Messungen ist (siehe Abschnitt C.4), das sich auch über

$$g_{1,m,l+1} = \lambda_1 g_{1,m,l} + 1 \quad (6.37)$$

rekursiv berechnen lässt.¹⁵ Damit ergibt sich

$$\sigma_{\hat{\theta}_{1,m,l+1}}^2 = \frac{\lambda_1^2 g_{1,m,l}^2}{(\lambda_1 g_{1,m,l} + 1)^2} \sigma_{\hat{\theta}_{1,m,l}}^2 + \frac{1}{(\lambda_1 g_{1,m,l} + 1)^2} \sigma_{\hat{\theta}_{1,l+1}}^2 \quad (6.38)$$

als Varianz des gewichteten mittleren Schätzwerts. Je kleiner λ_1 gewählt ist, desto mehr dominiert die Varianz des letzten Schätzwerts $\hat{\theta}_{1,l+1}$ die Varianz des gemittelten Schätzwerts.

Aus der nicht-rekursiven Darstellung ergibt sich der Erwartungswert

$$E\{\hat{\theta}_{1,m,l+1}\}_r = \frac{\sum_{l'=1}^{l+1} \lambda_1^{l+1-l'} \hat{\theta}_{1,l'}}{\sum_{l'=1}^{l+1} \lambda_1^{l+1-l'}}$$

des Schätzers. Während sich $\theta_{1,l}$ ändert, ist auch der Schätzer mit Vergessensfaktor nicht biasfrei. Wenn der Parameter θ_1 nach einer Änderung aber wieder konstant bleibt, dann nähert sich der Bias bei $\lambda_1 < 1$ mit den folgenden Mittelungen wieder relativ schnell der Null.

Wie bei der einfachen Mittelung sollten Einzelschätzwerte nur berücksichtigt werden, wenn deren Varianz einen Maximalwert $\sigma_{\hat{\theta}_{1,m,\max}}^2$ nicht überschreiten, da sonst Schätzungen mit sehr großer Varianz den Mittelwert zumindest zeitweise dominieren würden.

Man kann dabei auch die Möglichkeit nutzen, die einzelnen $\hat{\theta}_{1,l}$ über die Bedingung

$$\hat{\theta}_{1,l} \stackrel{!}{\in} [\theta_{1,\min}, \theta_{1,\max}]$$

auf Plausibilität zu prüfen, bevor sie in die Mittelung einbezogen werden. Dadurch werden Ausreißer von der Mittelung ausgenommen. Allerdings kann dadurch auch ein Bias entstehen, wenn die Grenzen asymmetrisch bezogen auf den wahren Wert von θ_1 liegen.

¹⁵ Die rekursive Berechnung ist vor Allem wichtig, wenn λ_1 zeitvariant gewählt wird.

Grenzwert der Varianz

Während die Varianz des gemittelten Wertes im Fall von $\lambda_1 = 1$ gegen Null strebt, so ist dies im Fall $\lambda_1 < 1$ nicht der Fall, wie im Folgenden kurz gezeigt wird. Aus Gl. (6.35) ergibt sich

$$\sigma_{\hat{\theta}_{1,m,l+1}}^2 = \frac{\sum_{l'=1}^{l+1} \left(\lambda_1^{l+1-l'} \right)^2 \sigma_{\hat{\theta}_{1,l'}}^2}{\left(\sum_{l'=1}^{l+1} \lambda_1^{l+1-l'} \right)^2}$$

für die Varianz des gemittelten Schätzwertes. Die Größe $\sigma_{\hat{\theta}_{1,\min}}^2$ sei die untere Grenze für die Varianz der Einzelschätzwerte, d. h. $\sigma_{\hat{\theta}_{1,l'}}^2 > \sigma_{\hat{\theta}_{1,\min}}^2$, womit sich

$$\sigma_{\hat{\theta}_{1,m,l+1}}^2 > \frac{\sum_{l'=1}^{l+1} \left(\lambda_1^{l+1-l'} \right)^2}{\left(\sum_{l'=1}^{l+1} \lambda_1^{l+1-l'} \right)^2} \sigma_{\hat{\theta}_{1,\min}}^2$$

ergibt. Für die Summe im Nenner gilt

$$\sum_{l'=1}^{l+1} \lambda_1^{l+1-l'} = \sum_{l'=0}^l \lambda_1^{l'} \stackrel{\lambda_1 \neq 1}{=} \frac{1 - \lambda_1^{l+1}}{1 - \lambda_1}$$

und für die Summe im Zähler gilt

$$\sum_{l'=1}^{l+1} \left(\lambda_1^{l+1-l'} \right)^2 = \sum_{l'=0}^l \lambda_1^{2l'} = \sum_{l'=0}^l (\lambda_1^2)^{l'} \stackrel{\lambda_1 \neq 1}{=} \frac{1 - (\lambda_1^2)^{l+1}}{1 - \lambda_1^2},$$

womit sich

$$\sigma_{\hat{\theta}_{1,m,l+1}}^2 > \frac{1 - (\lambda_1^2)^{l+1}}{1 - \lambda_1^2} \cdot \left(\frac{1 - \lambda_1}{1 - \lambda_1^{l+1}} \right)^2 \cdot \sigma_{\hat{\theta}_{1,\min}}^2 \quad (6.39)$$

ergibt. Für $l \rightarrow \infty$ erhält man damit

$$\lim_{l \rightarrow \infty} \sigma_{\hat{\theta}_{1,m,l+1}}^2 > \sigma_{\hat{\theta}_{1,m,\infty}}^2 = \frac{(1 - \lambda_1)^2}{1 - \lambda_1^2} \sigma_{\hat{\theta}_{1,\min}}^2$$

als untere Grenze für die Varianz.

Für $\lambda_1 = 0,9$ ergibt sich damit beispielsweise $\sigma_{\hat{\theta}_{1,m,\infty}}^2 = 0,053 \cdot \sigma_{\hat{\theta}_{1,\min}}^2$ bzw. $\sigma_{\hat{\theta}_{1,m,\infty}} = 0,23 \cdot \sigma_{\hat{\theta}_{1,\min}}$ und für $\lambda_1 = 0,75$ ergibt sich $\sigma_{\hat{\theta}_{1,m,\infty}}^2 = 0,14 \cdot \sigma_{\hat{\theta}_{1,\min}}^2$ bzw. $\sigma_{\hat{\theta}_{1,m,\infty}} = 0,38 \cdot \sigma_{\hat{\theta}_{1,\min}}$.

Beschreibung des Algorithmus

Aus dem bisher Gesagten folgt der in Abbildung 6.11 dargestellte Algorithmus.

Mit dem Starten eines neuen Auftrags werden die für die vorliegenden Auftragseigenschaften gültigen Werte geladen. Dies sind die Daten zur Mittelung der Parameter $\hat{\theta}_{1,m}$ und $\sigma_{\hat{\theta}_{1,m}}^2$ sowie den zuletzt für die Regelung verwendeten Schätzwert $\hat{\theta}_{1,r}$.

Als erstes werden die zur Mittelung gehörenden Daten in die Variablen mit dem „+“-Index kopiert. (Dieser Schritt ermöglicht am Ende ein Aktualisieren der „m“-Werte, ohne zu überprüfen ob bei dem Auftrag überhaupt eine Mittelung durchgeführt wurde, also die „+“-Werte berechnet wurden.)

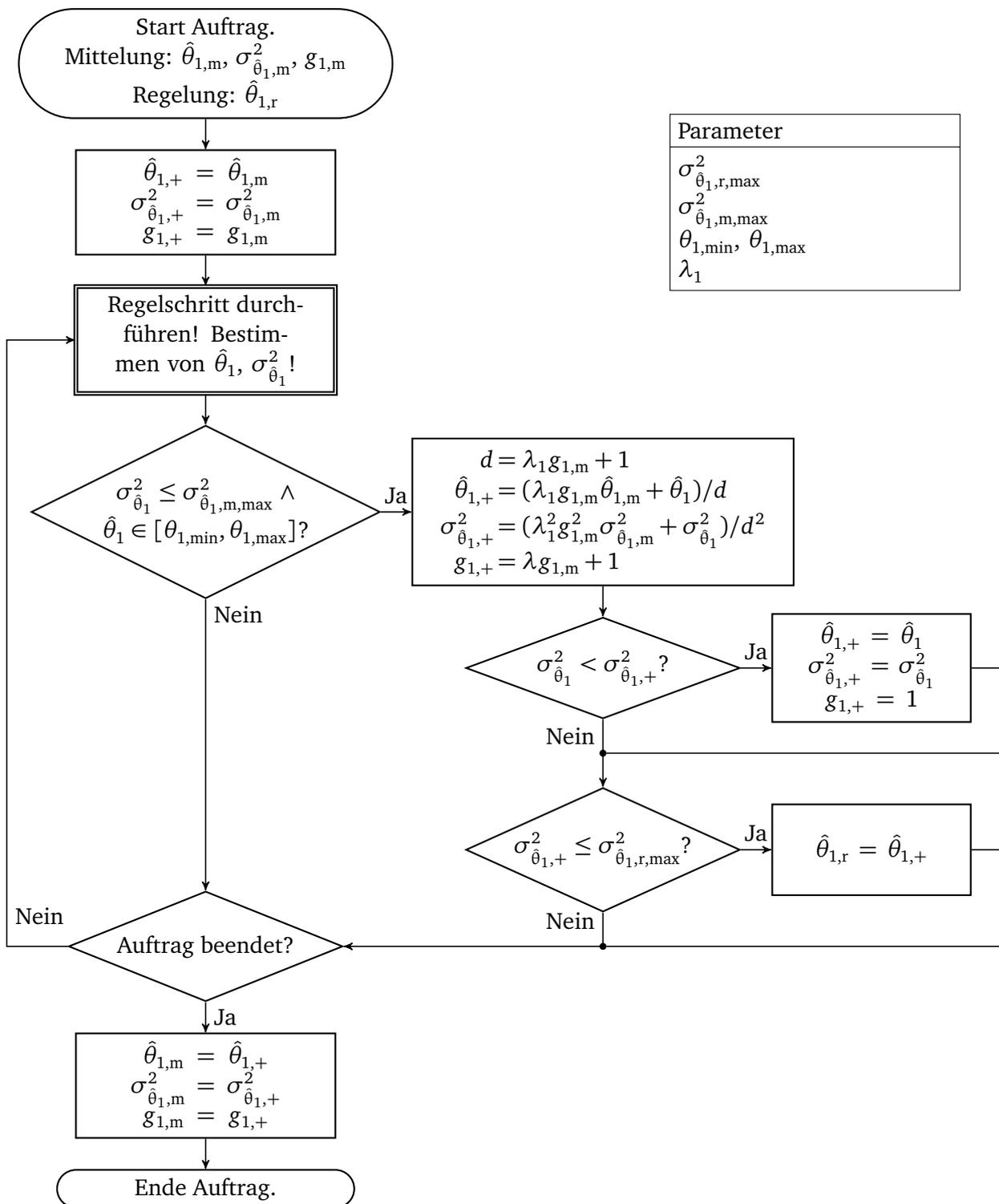


Abbildung 6.11: Vorgehen zur Mittelung mit Vergessensfaktor

Der Hauptteil beginnt mit der Durchführung eines Regelschritts und der Berechnung des Schätzwertes $\hat{\theta}_1$ und dessen Varianz $\sigma_{\hat{\theta}_1}^2$ mit den Methoden aus Abschnitt 6.4. Dann wird geprüft, ob die Varianz ausreicht, um $\hat{\theta}_1$ für die Mittelung zu verwenden. Diese Überprüfung sollte auch hier zumindest dann durchgeführt werden, wenn λ (deutlich) kleiner als eins ist, um zu schlechte Varianzen des Mittelwertes zu vermeiden.

Wenn dies nicht der Fall ist, dann geht es mit dem nächsten Regelschritt weiter, sofern noch einer durchgeführt wird. Wenn es die Varianz erlaubt, wird $\hat{\theta}_{1,+}$ zusammen mit $\sigma_{\hat{\theta}_{1,+}}$ nach den Gleichungen (6.36) und (6.38) bestimmt.

Als nächstes wird überprüft, ob der aktuelle Mittelwert eine höhere Varianz als der aktuelle Schätzwert hat. Ist dies der Fall, dann wird der aktuelle Schätzwert als neuer Startpunkt einer Mittelung verwendet und die entsprechenden Daten in die „+“-Variablen kopiert.

Wenn die Varianz des gemittelten $\hat{\theta}_{1,+}$ (bzw. des aktuellen Schätzwertes) kleiner als die maximal zulässige Varianz $\sigma_{\hat{\theta}_{1,r,\max}}^2$ zur Regelung ist, dann wird dieser Wert zur Regelung verwendet, $\hat{\theta}_{1,r} = \hat{\theta}_{1,+}$.

Wenn ein weiterer Regelschritt folgt, dann wird der Hauptteil des Algorithmus wieder durchlaufen. Falls der Druck beendet ist, werden die „+“-Werte in die „m“-Variablen kopiert und (zusammen mit dem aktuellen $\hat{\theta}_{1,r}$) für den nächsten Auftrag gespeichert.

Wahl von λ_1

Die Wahl von λ_1 hängt davon ab, wie sehr sich die Parameter von Auftrag zu Auftrag ändern können. Ist die mögliche Änderung sehr stark, so muss λ_1 eher klein gewählt werden. Die untere Grenze ist Null, und es kann durchaus sinnvoll sein, auch Werte unter 0,5 zu wählen. Es ist aber darauf zu achten, dass die Grenzen $\sigma_{\hat{\theta}_{1,m,\max}}$ und $\sigma_{\hat{\theta}_{1,r,\max}}$ aufeinander abgestimmt sind. Unterscheiden sich diese zu sehr, so kann es bei kleinen λ_1 möglich sein, dass die Grenze für die Regelung von den gemittelten Werten nie erreicht wird.

Die Grenze $\sigma_{\hat{\theta}_{1,m,\max}}$ kann sinnvollerweise nicht beliebig festgelegt werden, sondern sollte sich an den typisch auftretenden Varianzen der Einzelschätzwerte orientieren. Diese hängen davon ab, wie groß die Sprünge in θ_0 ausfallen, d. h. ob bei einem neuen Auftrag in der Regel eine größere Anfangsabweichung besteht oder nicht, oder ob typischerweise Änderungen der Sollgröße y_{soll} während eines Auftrags auftreten. Dies führt nämlich beides zu größeren Änderungen von u_1 und damit geringeren Varianzen der Schätzwerte. Ist $\sigma_{\hat{\theta}_{1,m,\max}}^2$ kleiner als die typisch auftretenden Varianzen gewählt, dann wäre die Mittelung de facto ausgeschaltet.

Um einen Anhaltswert für die Wahl von λ_1 und $\sigma_{\hat{\theta}_{1,m,\max}}$ zu erhalten, wird die Frage gestellt, welchen Wert $\sigma_{\hat{\theta}_{1,m,\max}}^2$ maximal haben darf, damit nach l gemittelten Werten die Varianz $\sigma_{\hat{\theta}_{1,m,l}}^2$ des Mittelwertes gerade der zur Regelung zulässigen Varianz $\sigma_{\hat{\theta}_{1,r,\max}}^2$ entspricht. Dazu wird vom ungünstigsten Fall ausgegangen, nämlich dass alle Einzelschätzwerte die maximal zulässige Varianz besitzen, die hier als

$$\sigma_{\hat{\theta}_{1,m,\max}}^2 \Big|_{\max,l}$$

bezeichnet wird. Die gerade genannte Forderung $\sigma_{\hat{\theta}_{1,m,l}}^2 \stackrel{!}{=} \sigma_{\hat{\theta}_{1,r,\max}}^2$ mit $\sigma_{\hat{\theta}_{1,l'}}^2 \equiv \sigma_{\hat{\theta}_{1,m,\max}}^2 \Big|_{\max,l}$ führt dann über eine ähnliche Rechnung, mit der (6.39) erhalten wurde, auf

$$\sigma_{\hat{\theta}_{1,m,\max}}^2 \Big|_{\max,l} = \sqrt{\frac{1 - \lambda_1^2}{1 - (\lambda_1^2)^l}} \cdot \frac{1 - \lambda_1^l}{1 - \lambda_1} \cdot \sigma_{\hat{\theta}_{1,r,\max}}^2,$$

wobei hier keine Varianzen, sondern die Standardabweichungen angegeben sind.

Die Verläufe von $\sigma_{\hat{\theta}_{1,m,\max}} \Big|_{\max,l}$, bezogen auf $\sigma_{\hat{\theta}_{1,r,\max}}$, für verschiedene l sind in Abbildung 6.12 dargestellt. Die Kurven beginnen für $\lambda_1 = 0$ alle bei eins, da bei $\lambda_1 = 0$ nur der aktuelle Einzelschätzwert berücksichtigt wird und dieser damit die geforderte Standardabweichung von $\sigma_{\hat{\theta}_{1,r,\max}}$ besitzen muss. Für $\lambda_1 = 1$ liegt die einfache Mittelung vor, und damit ergibt sich jeweils der Wert \sqrt{l} .

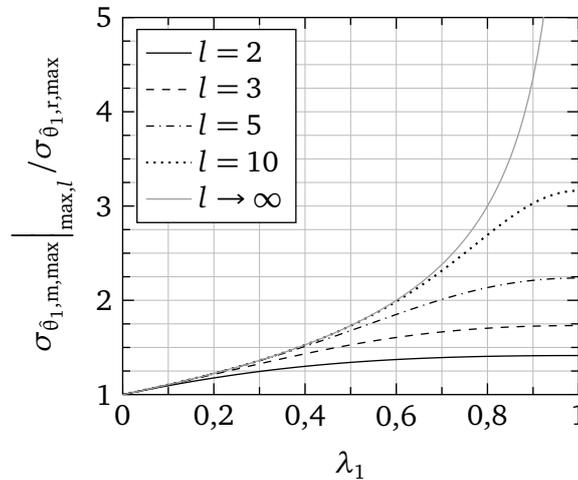


Abbildung 6.12: Zur Wahl von λ_1

Je größer λ_1 gewählt ist, desto größer kann auch die maximal zur Mittelung akzeptierte Varianz $\sigma_{\hat{\theta}_{1,m,\max}}^2$ eines einzelnen $\hat{\theta}_{1,l}$ sein. Es ist damit eine Abwägung zu treffen zwischen Reaktion auf veränderliche Parameter und Güte der Schätzung.

Kombination von varianz-optimaler Mittelung und Mittelung mit Vergessensfaktor

Die Mittelung mit Vergessensfaktor besitzt den Nachteil, dass auch sehr gute Schätzungen mit geringer Varianz von schlechteren Messungen dominiert werden (wenn $\lambda_1 \approx 1$, denn dann entspricht dieses Verfahren näherungsweise der einfachen Mittelung), bzw. schnell durch danach kommende, relativ schlechte Schätzwerte verdrängt werden können (wenn $\lambda_1 \ll 1$). Bei der varianz-optimalen Mittelung dagegen dominieren die Schätzwerte mit geringen Varianzen dauerhaft den Mittelwert, womit sich ändernden realen Parameterwerten nicht gefolgt werden kann.

Daher wird noch eine Kombination von varianz-optimaler Mittelung und der Mittelung mit Vergessensfaktor betrachtet, die sich dadurch auszeichnet, dass diese

- für $\lambda_1 \equiv 1$ der varianz-optimalen Mittelung und
- für $\sigma_{\hat{\theta}_{1,l}}^2 \equiv \sigma_{\hat{\theta}_1}^2$ der Mittelung mit Vergessensfaktor

entspricht.

Dazu wird zunächst die Mittelung mit Vergessensfaktor in der Form des rekursiven Least-Squares-Verfahrens

$$\gamma_{1,l} = \frac{p'_{1,m,l}}{p'_{1,m,l} + \lambda_1} \quad (6.40)$$

$$\hat{\theta}_{1,m,l+1} = (1 - \gamma_{1,l}) \cdot \hat{\theta}_{1,m,l} + \gamma_{1,l} \cdot \hat{\theta}_{1,l+1} \quad (6.41)$$

$$p'_{1,m,l+1} = (1 - \gamma_{1,l}) \cdot p'_{1,m,l} \cdot \frac{1}{\lambda_1} \quad (6.42)$$

mit den Anfangswerten

$$\hat{\theta}_{1,m,1} = \hat{\theta}_{1,1} \quad \text{und} \quad p'_{1,m,1} = 1$$

angegeben.¹⁶ Dieses ist äquivalent zu dem Verfahren nach den Gleichungen (6.36) und (6.37), wie im Folgenden kurz dargelegt wird. Dazu wird (6.40) in (6.41) eingesetzt, was auf

$$\hat{\theta}_{1,m,l+1} = \frac{\lambda_1 \cdot \hat{\theta}_{1,m,l} + p'_{1,m,l} \cdot \hat{\theta}_{1,l+1}}{\lambda_1 + p'_{1,m,l}} = \frac{\frac{\lambda_1}{p'_{1,m,l}} \cdot \hat{\theta}_{1,m,l} + \hat{\theta}_{1,l+1}}{\frac{\lambda_1}{p'_{1,m,l}} + 1}$$

führt. Für $g_{1,m,l} = \frac{1}{p'_{1,m,l}}$ entspricht diese Gleichung der Gleichung (6.36). Für den Kehrwert von $p'_{1,m,l+1}$ gilt mit Gl. (6.42) und (6.40)

$$\frac{1}{p'_{1,m,l+1}} = \left(\frac{\lambda_1}{\lambda_1 + p'_{1,m,l}} \cdot p'_{1,m,l} \cdot \frac{1}{\lambda_1} \right)^{-1} = \frac{\lambda_1}{p'_{1,m,l}} + 1.$$

Für $g_{1,m,l} = \frac{1}{p'_{1,m,l}}$ (und damit auch $g_{1,m,l+1} = \frac{1}{p'_{1,m,l+1}}$) entspricht dies Gl. (6.37), womit gezeigt ist, dass beide Verfahren das gleiche Ergebnis liefern.

Erweitert man den Bruch in Gl. (6.40) mit $\sigma_{\hat{\theta}_1}^2$, so erhält man mit $p_{1,m,l} = \sigma_{\hat{\theta}_1}^2 p'_{1,m,l}$ die Gleichung

$$\gamma_{1,l} = \frac{p_{1,m,l}}{p_{1,m,l} + \lambda_1 \cdot \sigma_{\hat{\theta}_1}^2} = \frac{p''_{1,m,l}}{p''_{1,m,l} + \sigma_{\hat{\theta}_1}^2}, \quad (6.43)$$

wobei im letzten Schritt die Abkürzung $p''_{1,m,l} = p_{1,m,l}/\lambda_1$ verwendet wurde. Die Gleichung (6.42) wird auf beiden Seiten mit $\sigma_{\hat{\theta}_1}^2$ multipliziert, was

$$p_{1,m,l+1} = (1 - \gamma_{1,l}) \cdot \frac{p_{1,m,l}}{\lambda_1} = (1 - \gamma_{1,l}) \cdot p''_{1,m,l} \quad (6.44)$$

ergibt. Damit ist auch der Anfangswert $p_{1,m,1} = \sigma_{\hat{\theta}_1}^2$ anstelle von $p'_{1,m,1} = 1$ zu verwenden.

Die Korrekturgleichung (6.41) für $\hat{\theta}_{1,l+1}$ bleibt dabei unverändert.

Zusammengefasst ergibt sich der Algorithmus zu

$$p''_{1,m,l} = \frac{1}{\lambda_1} p_{1,m,l} \quad (6.45)$$

$$\gamma_{1,l} = \frac{p''_{1,m,l}}{p''_{1,m,l} + \sigma_{\hat{\theta}_1}^2} \quad (6.46)$$

$$\hat{\theta}_{1,m,l+1} = (1 - \gamma_{1,l}) \cdot \hat{\theta}_{1,m,l} + \gamma_{1,l} \cdot \hat{\theta}_{1,l+1} \quad (6.47)$$

$$p_{1,m,l+1} = (1 - \gamma_{1,l}) \cdot p''_{1,m,l} \quad (6.48)$$

mit dem Anfangswert

$$p_{1,m,1} = \sigma_{\hat{\theta}_1}^2.$$

¹⁶ Die Messgleichung zu diesem LS-Problem lautet $\tilde{y}_l = 1 \cdot \theta_{1,m} + n_l$ mit den Messungen $\tilde{\mathbf{y}}^T = [\hat{\theta}_{1,1} \ \dots \ \hat{\theta}_{1,l}]$ und einem mittelwertfreien Rauschen n_l .

Dieser hat die eingangs geforderten Eigenschaften, d. h. für $\lambda_1 = 1$ ist dieser mit der optimalen Mittelung identisch, und für $\sigma_{\hat{\theta}_{1,l+1}}^2 \equiv \sigma_{\hat{\theta}_1}^2$ entspricht Gl. (6.45) bis (6.48) der Mittelung mit Vergessensfaktor λ_1 .

Durch Gl. (6.45) wird die „Varianz“ der bisherigen Schätzwerte vor der Bestimmung der optimalen Gewichtung in (6.46) durch die Multiplikation mit dem Kehrwert von λ etwas vergrößert, d. h. verschlechtert, womit sich die Gewichtung etwas zugunsten des neuen Schätzwertes $\hat{\theta}_{1,l+1}$ verschiebt. Berechnet man aber die Varianz des neuen gemittelten Wertes, so wäre eigentlich wieder die tatsächliche Varianz $\sigma_{\hat{\theta}_{1,m,l}}^2$ des bisherigen Mittelwertes zu verwenden, d. h.

$$\sigma_{\hat{\theta}_{1,m,l+1}}^2 = (1 - \gamma_{1,l})^2 \cdot \sigma_{\hat{\theta}_{1,m,l}}^2 + \gamma_{1,l}^2 \cdot \sigma_{\hat{\theta}_{1,l+1}}^2. \quad (6.49)$$

Um auf Gleichung (6.48) zu kommen, liegt aber auch die verfälschte Varianz zugrunde. Damit entspricht $p_{1,m,l}$ nicht mehr der tatsächlichen Varianz $\sigma_{\hat{\theta}_{1,l}}^2$. Dies ist notwendig, damit der Algorithmus für ein konstantes $\sigma_{\hat{\theta}_{1,l}}^2 \equiv \sigma_{\hat{\theta}_1}^2$ in die Mittelung mit Vergessensfaktor übergeht.

Soll auch die tatsächliche Varianz bestimmt werden, so wird der Algorithmus um die Gleichung (6.49) erweitert. Es kann jedoch auch $p''_{1,m,l+1}$ der Einfachheit halber als Varianz für den Vergleich mit $\sigma_{\hat{\theta}_{1,r,\max}}^2$ angesehen werden, um auf die Speicherung einer weiteren Variable zu verzichten.

Alternativ kann der Algorithmus auch in der „Kalmanfilter-Form“ formuliert werden. Dabei lautet die Propagationsgleichung für die „Varianz“ dann allgemein

$$p''_{1,m,l} = p_{1,m,l} + Q_{1,l}, \quad (6.50)$$

wobei $Q_{1,l}$ die Unsicherheit in der Parameterentwicklung zwischen zwei Mittelungen beschreibt. Wählt man

$$Q_{1,l} = p_{1,m,l} \cdot (\lambda_1^{-1} - 1),$$

so ergibt sich das gleiche Verhalten wie oben, was sich durch Einsetzen dieses $Q_{1,l}$ direkt aus (6.50) ergibt.

Um zu verhindern, dass ein einzelner Schätzwert mit sehr geringer Varianz das Lernen bei geringem Vergessen (λ_1 nahe eins) weitestgehend blockiert, kann die „Varianz“ über

$$p_{1,m,l} = \max \{ p_{1,m,l}, p_{1,\min} \} \quad (6.51)$$

nach unten begrenzt werden.

Der sich ergebende Algorithmus ist in Abbildung 6.13 dargestellt, wobei $p_{1,m}$ als „Näherung“ für $\sigma_{m,1}^2$ angesehen wird. Im Unterschied zu der Mittelung mit Vergessensfaktor (Abbildung 6.11) ist hier prinzipbedingt sichergestellt, dass die Varianz des gemittelten Wertes immer geringer ist als die des aktuellen Einzelschätzwertes, so dass auf die entsprechende Überprüfung verzichtet werden kann.

Für die Wahl von λ_1 gilt das bei der Mittelung mit Vergessensfaktor Gesagte.

Unterschiede bei der Mittelung der sekundären Parameter

Werden Schätzwerte der sekundären Parameter bei teilweise festgehaltenen Parametern ermittelt, werden zu diesen auch Korrekturvektoren bestimmt, die ebenfalls gemittelt werden müssen. Diese Mittelung wurde schon in Abschnitt 6.4.4 ab Seite 112 besprochen. Dort wurden die Gleichungen (6.26) und (6.28) angegeben.

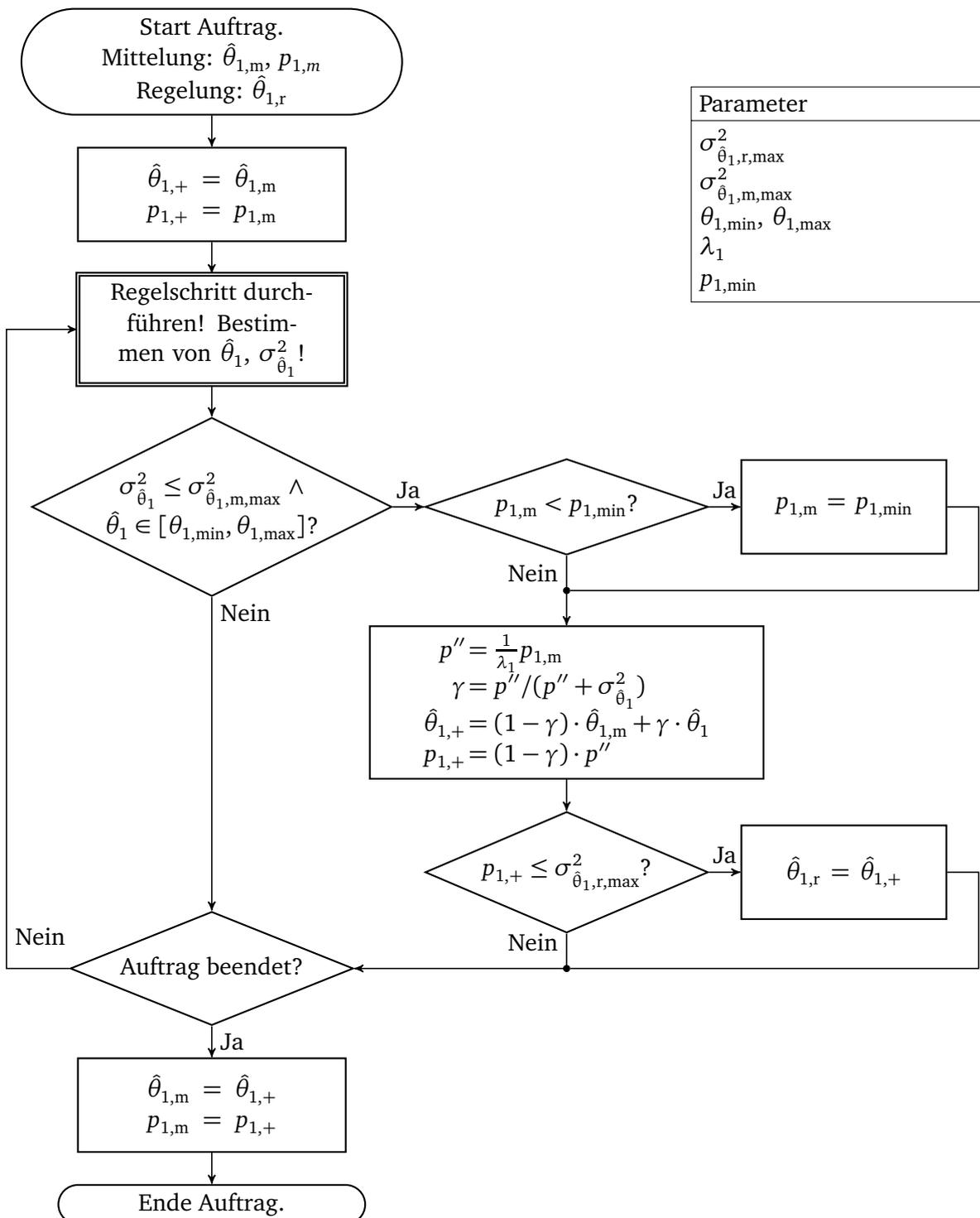


Abbildung 6.13: Vorgehen zur kombinierten Mittelung

Für den i -ten Parameter werden die Korrekturvektoren entsprechend dem Schätzwert nach

$$\mathbf{c}_{i|f,m}^T(l+1) = \frac{\lambda g_{i,m}(l) \cdot \mathbf{c}_{i|f,m}^T(l) + \mathbf{c}_{i|f}^T(l+1)}{\lambda g_{i,m}(l) + 1}$$

(Mittelung mit Vergessensfaktor) bzw.

$$\mathbf{c}_{i|f,m}^T(l+1) = (1 - \gamma_i(l)) \cdot \mathbf{c}_{i|f,m}^T(l) + \gamma_i(l) \cdot \mathbf{c}_{i|f}^T(l+1)$$

(Kombination der varianz-optimalen Mittelung und der Mittelung mit Vergessensfaktor) gemittelt.

Die Mittelung der eigentlichen Parameterschätzwerte geschieht nach den für den ersten Parameter angegebenen Gleichungen, nur ist darauf zu achten, dass sich der alte Mittelwert $\hat{\theta}_{i,m}$ und der neue Wert $\hat{\theta}_i(l+1)$ auf dieselben festen Parameterwerte beziehen.

6.5.3 Simulation

Testfall

Es ist wieder das in Tabelle 5.6 beschriebene System simuliert. Anders als in Abschnitt 5.5 werden aber mehrere Aufträge und auch Änderungen der sekundären Eingangsgröße u_2 simuliert. Da der Schätzer standardmäßig zu Beginn eines neuen Auftrages wie beschrieben eine neue Schätzung startet, können nur Änderungen von u_2 zur Schätzung von θ_2 genutzt werden, die innerhalb eines Auftrags stattfinden. Es sind die in Abbildung 6.14 dargestellten Aufträge simuliert. Dabei kennzeichnen die durchgezogenen schwarzen Linien neue Aufträge, die gestrichelten schwarzen Linien die Änderungen der sekundären Eingangsgröße innerhalb eines Auftrags.

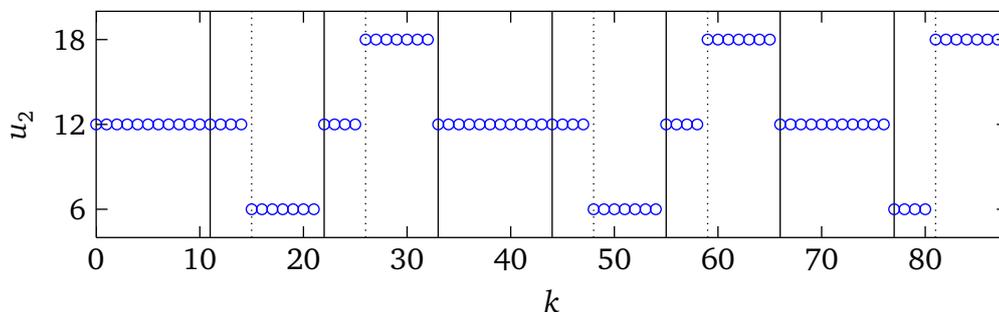


Abbildung 6.14: Verlauf der sekundären Eingangsgröße $u_{2,k}$ im Testfall

Zusätzlich verändert sich mit Beginn des vierten Auftrags der Parameter θ_0 des Systems. Dieser ist ab dem vierten Auftrag um 10 höher. Der Anfangswert für u_1 wird für jeden Auftrag anhand der Schätzwerte berechnet. Beim letzten Auftrag beträgt die Sollgröße 40 anstelle von 50, um die bis dahin identifizierten Parameter zur Regelung beurteilen zu können.

Ergebnisse

Es werden die folgenden Punkte anhand von Simulationen dargestellt und diskutiert:

- Vergleich Zwei- und Drei-Parameter-Schätzung für θ_2 ,
- Verhalten bei Sprung in θ_2 sowie
- Verhalten bei Sprung in θ_1 .

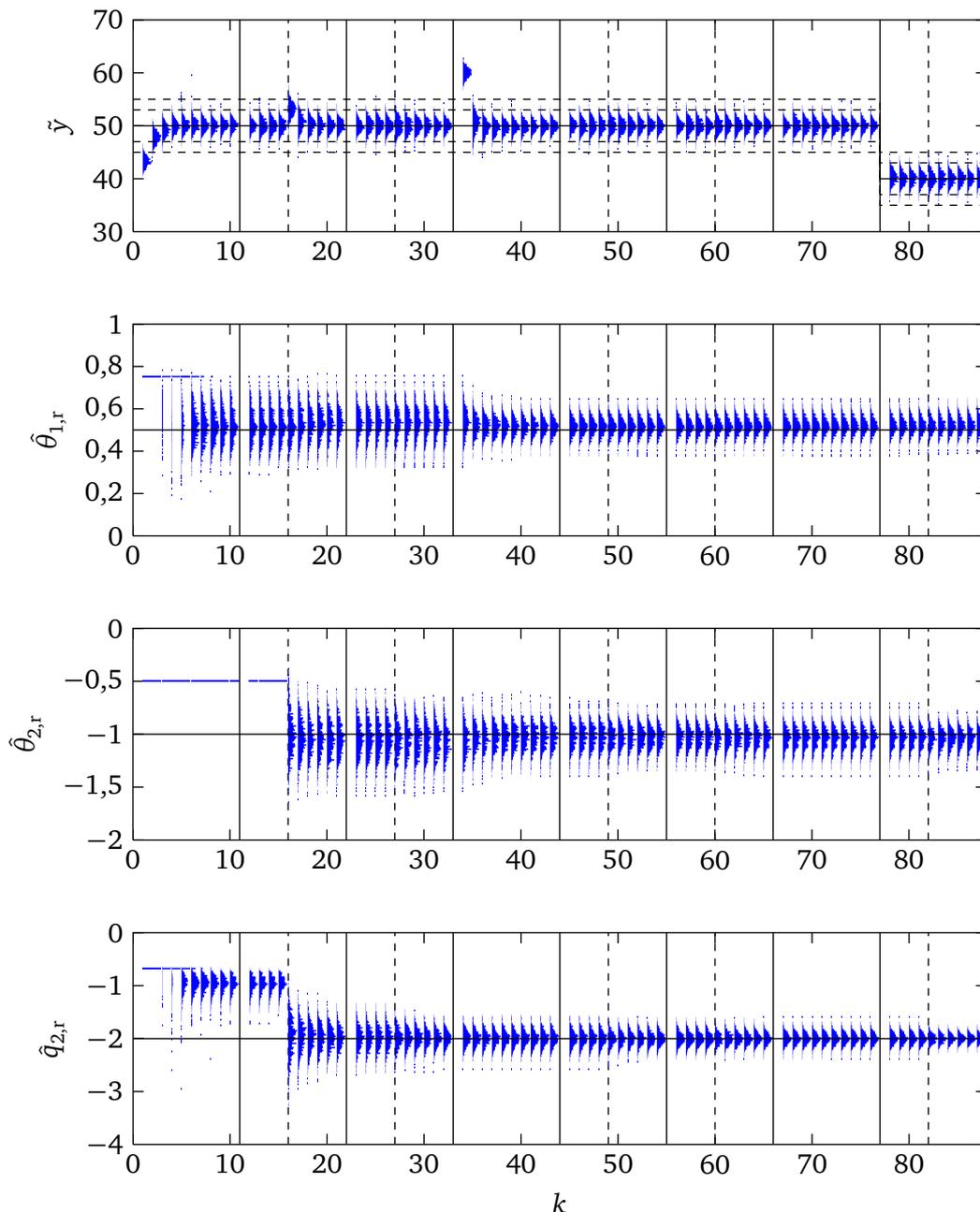


Abbildung 6.15: Standardverfahren, $\sigma_{\hat{\theta}_{2,r},\max} = \frac{1}{5} \cdot \theta_2$, $\sigma_{\hat{\theta}_{2,m},\max} = \frac{1}{2} \cdot \theta_2$ (Neue Schätzung bei jedem Auftrag) (1 000 Simulationen)

Vergleich Zwei- und Drei-Parameter-Schätzung für θ_2

In Abbildung 6.15 ist die Simulation des Standardtestfalls gezeigt, wobei der Parameter θ_2 über den Zwei-Parameter-Schätzer bei festgehaltenem $\hat{\theta}_1$ bestimmt wird. Die Grenzen $\sigma_{\hat{\theta}_{1,r},\max}$ und $\sigma_{\hat{\theta}_{2,r},\max}$ liegen jeweils auf einem Fünftel des jeweiligen Parameterwertes.

Im Vergleich zu Abbildung 6.2b liegt hier zu Beginn eine geringere Regelabweichung vor. Damit ist die Variation von u_1 geringer und entsprechend werden mehr Regelschritte als in Abbildung 6.2b benötigt, bis θ_1 in allen Fällen geschätzt werden kann.

Während des zweiten Auftrags ändert sich erstmals die sekundäre Eingangsgröße u_2 , was hier in allen Fällen direkt die Schätzung von θ_2 ermöglicht. Es zeigt sich in $\hat{\theta}_{1,r}$ ein kleiner Bias, der während des dritten Auftrags noch etwas größer wird. Dieser entsteht durch die Identifikation im geschlossenen Re-

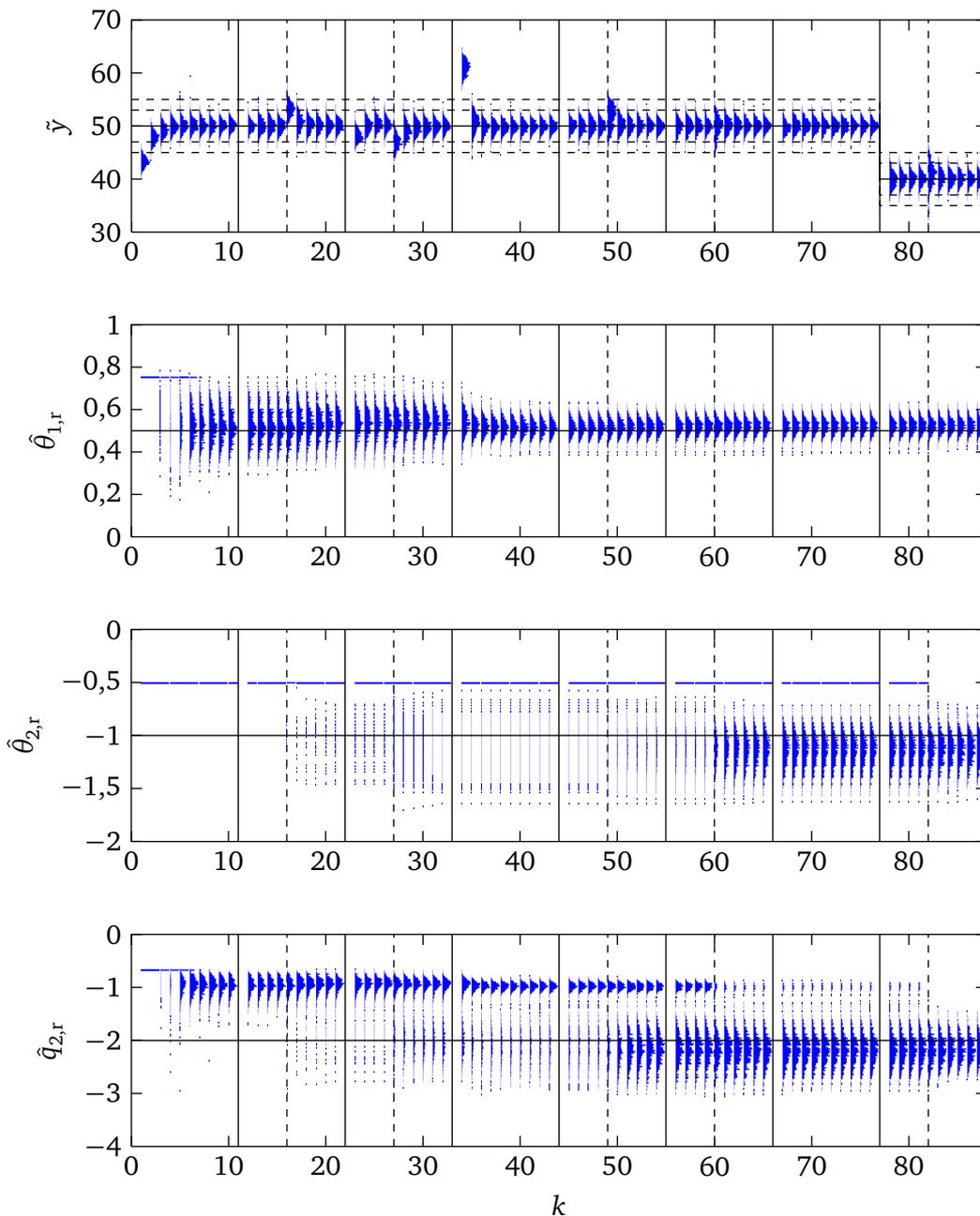


Abbildung 6.16: θ_2 über Drei-Parameter-Schätzung bestimmt, $\sigma_{\hat{\theta}_{2,r,\max}} = \frac{1}{5} \cdot \theta_2$, $\sigma_{\hat{\theta}_{2,m,\max}} = \frac{1}{2} \cdot \theta_2$ (1 000 Simulationen)

gelkreis und wird später in Abschnitt 6.6 noch diskutiert. Hier ist der Effekt zu beobachten, dass θ_2 ebenfalls betragsmäßig überschätzt wird, so dass $\{q_{2,r}\}_r$, das Verhältnis beider Größen, zumindest näherungsweise biasfrei ist. Auch darin liegt ein Vorteil der Zwei-Parameter-Schätzung für θ_2 . Da $\hat{\theta}_2$ immer auf das aktuelle $\hat{\theta}_{1,r}$ bezogen wird, wird das Verhältnis q_2 auch bei (nicht übertrieben großen) Fehlern des Schätzwertes für θ_1 noch gut geschätzt. Dies wird nochmal in Anhang D betrachtet. Hier werden beim Vergleich zunächst nur die in den Plots sichtbaren Effekte berücksichtigt.

Die gute Schätzung von q_2 zeigt sich schon im dritten Auftrag. Bei diesem ändert sich die sekundäre Eingangsgröße ebenfalls, aber die Kompensation funktioniert schon so gut, dass dies ohne die Markierungslinie kaum auszumachen wäre.

Durch die hohe Regelabweichung zu Beginn des vierten Auftrags verbessert sich die Schätzung von θ_1 nochmals deutlich, wovon durch die Korrektur auch die Schätzwerte für θ_2 etwas profitieren. Im weiteren Verlauf der Simulation können Änderungen der sekundären Eingangsgröße am Verlauf von $\{\tilde{y}\}_r$ gar nicht mehr ausgemacht werden.

In Abbildung 6.16 ist für denselben Testfall die Drei-Parameter-Schätzung für θ_2 verwendet. Mit derselben Varianzgrenze werden hier drei Änderungen der sekundären Eingangsgröße benötigt, bis zumindest in etwa der Hälfte der Fälle die Varianz des gemittelten Schätzwertes so gering ist, dass dieser zur Regelung verwendet werden kann. Auch kann bei den Schätzwerten für q_2 ein geringer Bias ausgemacht werden. Als Konsequenz zeigen sich auch bei den Änderungen der sekundären Eingangsgröße im dritten und fünften Auftrag noch deutliche Sprünge in $\{\tilde{y}\}_r$. Und auch bei den folgenden beiden Änderungen von u_2 ist die Streuung noch erhöht.

Die Schätzwerte für θ_1 sind vergleichbar mit denen aus Abbildung 6.15, da für diesen Parameter dasselbe Schätzverfahren verwendet wird. (Genau genommen sind die Schätzwerte im zweiten Fall minimal besser, da die Variation von u_1 etwas größer ist, da die schlechtere Kompensation von u_2 ausgeglichen werden muss.)

In Abbildung 6.17 ist die Varianzgrenze $\sigma_{\hat{\theta}_{2,m,\max}}$ auf ein Drittel des Parameterwertes gesenkt, um ein früheres Schätzen von θ_2 zu erreichen. Dies tritt auch ein, so liegt im Großteil der Fälle nach der zweiten Änderung der sekundären Eingangsgröße ein Schätzwert für θ_2 vor, der zur Regelung verwendet werden kann. Es zeigt sich in den Verläufen für $\{\tilde{y}\}_r$ aber, dass aufgrund der hohen Streuung und des Bias von $\{\hat{q}_{2,r}\}_r$ die Kompensation der Änderungen von u_2 insgesamt nicht so gut funktioniert wie in Abbildung 6.15, wo θ_2 durch die Zwei-Parameter-Schätzung bestimmt und immer an $\hat{\theta}_{1,r}$ angepasst wird.

Verhalten bei Sprung in θ_2

In Abbildung 6.18 und 6.19 ist der Testfall simuliert, bei dem sich zusätzlich zu Beginn des vierten Auftrags der sekundäre Parameter θ_2 von -1 auf $-0,75$ verändert.

In Abbildung 6.18 erfolgt die Mittelung der Parameter optimal bezüglich der Varianz, d. h. nach Abbildung 6.13 mit $\lambda_2 = 1$.

Man erkennt, wie mit jeder Änderung des sekundären Eingangs u_2 während der Aufträge fünf, sechs und acht der Schätzwert $\hat{\theta}_{2,r}$ etwas in Richtung des neuen Schätzwertes korrigiert wird, womit sich auch $\hat{q}_{2,r}$ dem neuen Wert von $-1,5$ langsam annähert. Der Fehler in $\hat{q}_{2,r}$ ist im Verlauf von \tilde{y} an den kleinen Fehlern bei den Änderungen von u_2 zu erkennen.

In Abbildung 6.19 ist $\lambda_2 = 0,5$ gewählt. Es ist deutlich zu erkennen, dass die Korrektur schneller stattfindet, was hier nur durch eine moderate Zunahme der Streuung von $\{\hat{\theta}_{2,r}\}_r$ bzw. $\{\hat{q}_{2,r}\}_r$ erkauft wird. Bei der Änderung der sekundären Eingangsgröße in Auftrag fünf zeigt sich noch ein Sprung in $\{\tilde{y}\}_r$, welcher unvermeidbar ist, da hier erst die Änderung des Parameters θ_2 erkannt werden kann. Bei der Änderung von u_2 im sechsten Auftrag und insbesondere im achten Auftrag fallen aber die Sprünge von $\{\tilde{y}\}_r$ deutlich geringer aus als bei der Wahl $\lambda_1 = 1$.

Verhalten bei Sprung in θ_1

In Abbildung 6.20 ist, analog zu Abbildung 6.18 ein Sprung des Parameters θ_1 zum fünften Auftrag gezeigt. Dabei ist zusätzlich zum bisher vorhandenen Sprung des Parameters θ_0 zum vierten Auftrag ein weiterer zum siebten Auftrag simuliert.

Aufgrund der sehr guten Anregung durch die anfänglich hohe Regelabweichung im vierten Auftrag ist die Varianz des Schätzwertes $\{\hat{\theta}_1\}_r$ im vierten Auftrag relativ gering. Dagegen ist die Anregung in den folgenden beiden Aufträgen sehr gering. Entsprechend dominiert der vierte Auftrag zunächst die Mittelwertbildung, und der Mittelwert nähert sich nur sehr langsam dem neuen Wert für θ_1 an. Mit der

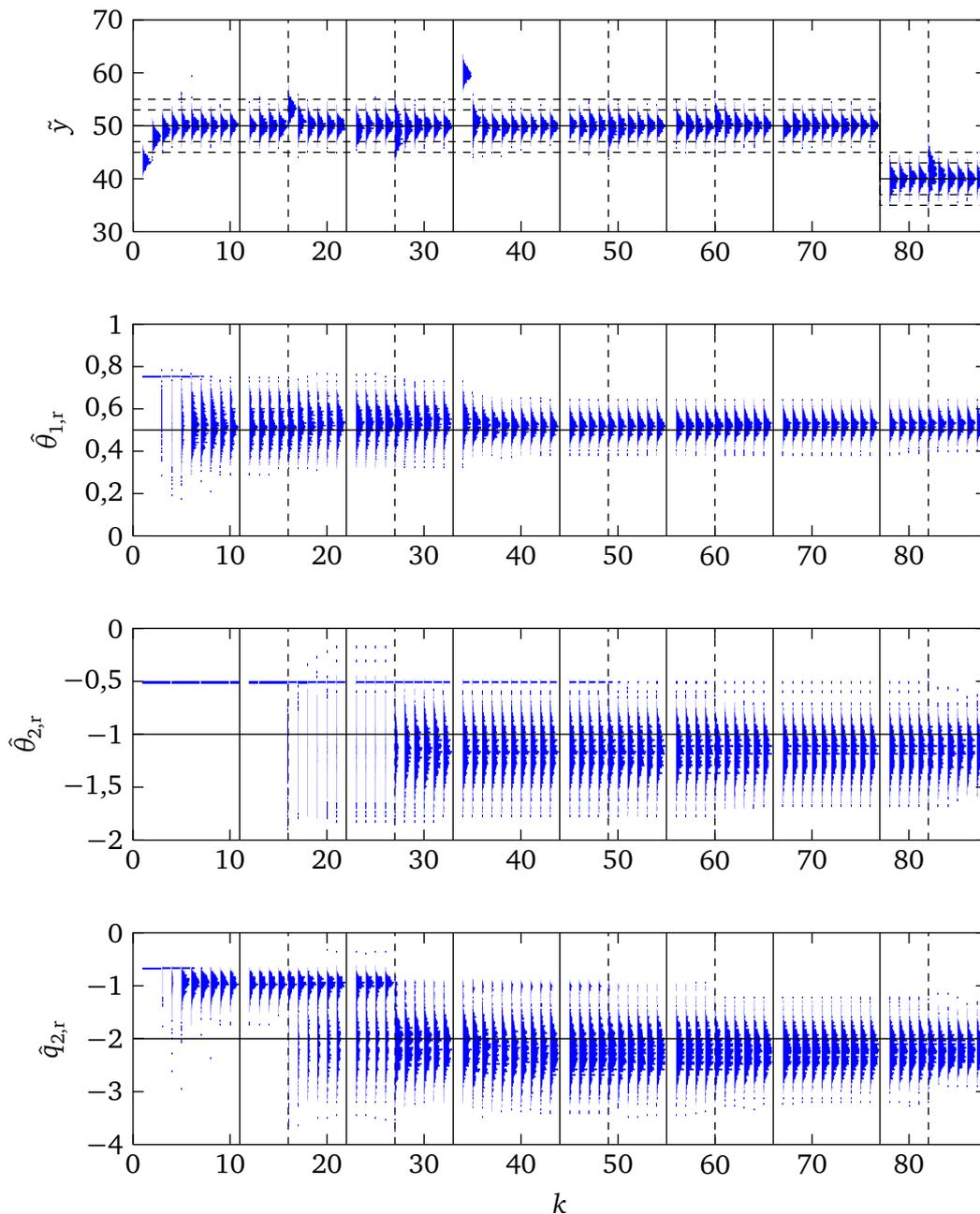


Abbildung 6.17: θ_2 über Drei-Parameter-Schätzung bestimmt, $\sigma_{\hat{\theta}_{2,r,\max}} = \frac{1}{3} \cdot \theta_2$, $\sigma_{\hat{\theta}_{2,m,\max}} = \frac{1}{2} \cdot \theta_2$ (1 000 Simulationen)

größeren Anregung durch den Sprung in θ_0 im siebten Auftrag ist die Korrektur des Schätzwertes etwas deutlicher, jedoch bei weitem noch nicht vollständig.

Ein weiterer wichtiger Punkt ist, dass durch den Sprung von θ_1 auch der Schätzwert für q_2 nicht mehr korrekt ist und das dieser bei jeder Änderung der sekundären Eingangsgröße nur langsam korrigiert wird. Dies liegt daran, dass θ_2 bezüglich eines falschen $\hat{\theta}_{1,r}$ geschätzt wird. Dies stellt aber kein prinzipielles Problem dar, es ist lediglich zu beachten, dass bei möglichen Sprüngen von θ_1 nicht nur die Mittelung der Schätzwerte von θ_1 , sondern auch die Mittelung der Schätzwerte von θ_2 mit einem Vergessensfaktor durchgeführt wird.

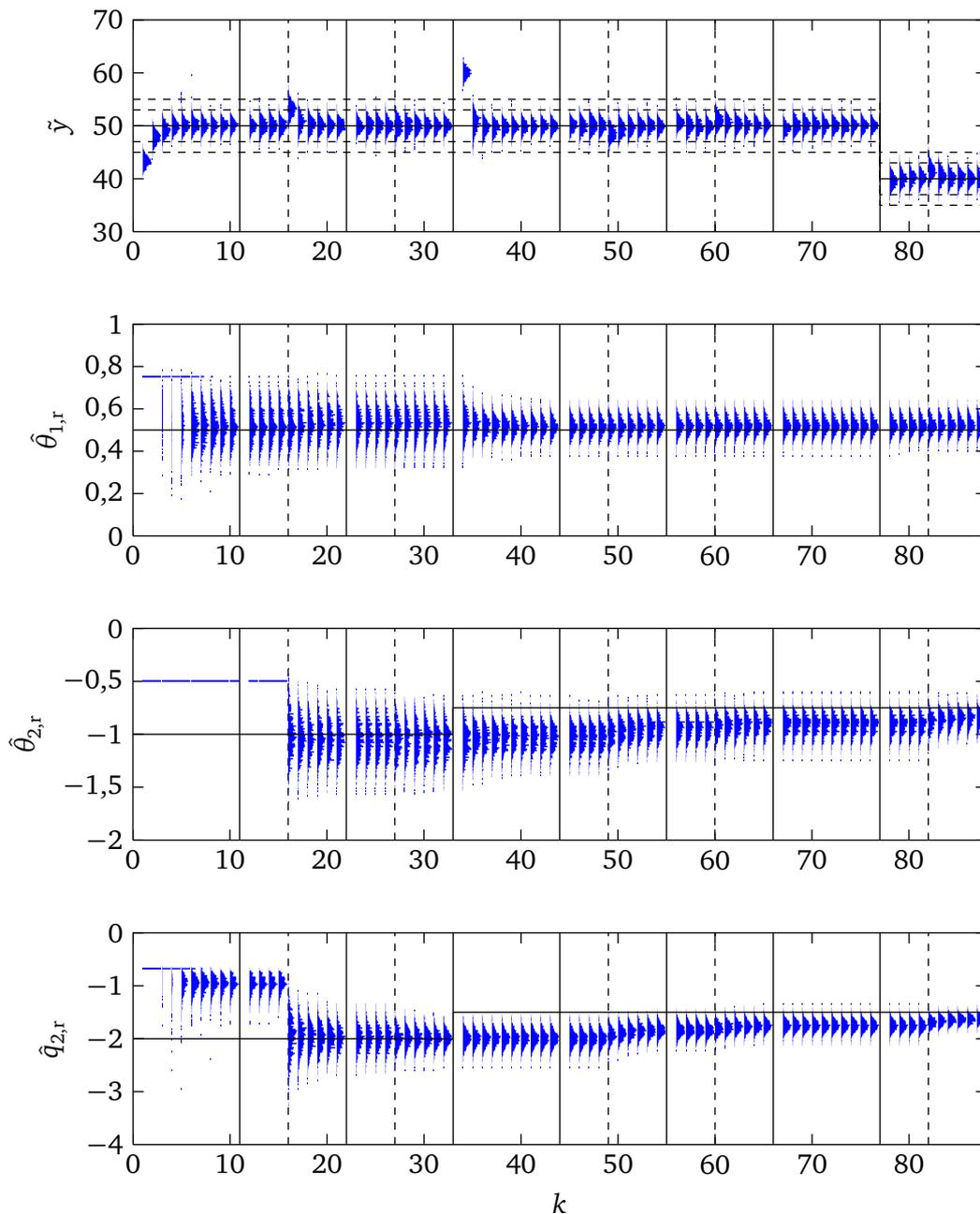


Abbildung 6.18: Sprung in θ_2 ab dem vierten Auftrag, $\lambda_2 = 1$ (1 000 Simulationen)

In Abbildung 6.21 ist $\hat{\theta}_1$ über eine Mittelung mit Vergessensfaktor und $\hat{\theta}_2$ über die Kombination von varianz-optimaler Mittelung und Mittelung mit Vergessensfaktor gemittelt. Die jeweiligen Vergessensfaktoren sind $\lambda_1 = 0,75$ und $\lambda_2 = 0,5$.

Es zeigt sich bei $\hat{\theta}_{1,r}$ und $\hat{q}_{2,r}$ ein deutlich besseres Anpassen an die neuen Parameterwerte. Dies wird aber mit einer Zunahme der Varianz der Schätzwerte erkaufte, was unvermeidbar ist.

Betrachtet man $\{\tilde{y}_k\}_r$, so zeigt sich, insbesondere im letzten Auftrag, ein deutlich besseres Regelverhalten.

In Abbildung 6.22 sind die Verteilungen von $\{\hat{\theta}_{1,r,k}\}_r$ für den Fall einer Kombination von varianz-optimaler Mittelung und Mittelung mit Vergessensfaktor dargestellt. Dabei sind die Parameter $\lambda_1 = 0,75$ und $p_{1,\min} = \frac{1}{2} \cdot \sigma_{\hat{\theta}_{1,r,\max}}^2$ verwendet. Aufgrund der beschriebenen Dominanz der niedrigen Varianz für $\{\hat{\theta}_{1,l'=4}\}_r$ des vierten Auftrags unterscheiden sich die Verläufe bis einschließlich zum Auftrag sechs in

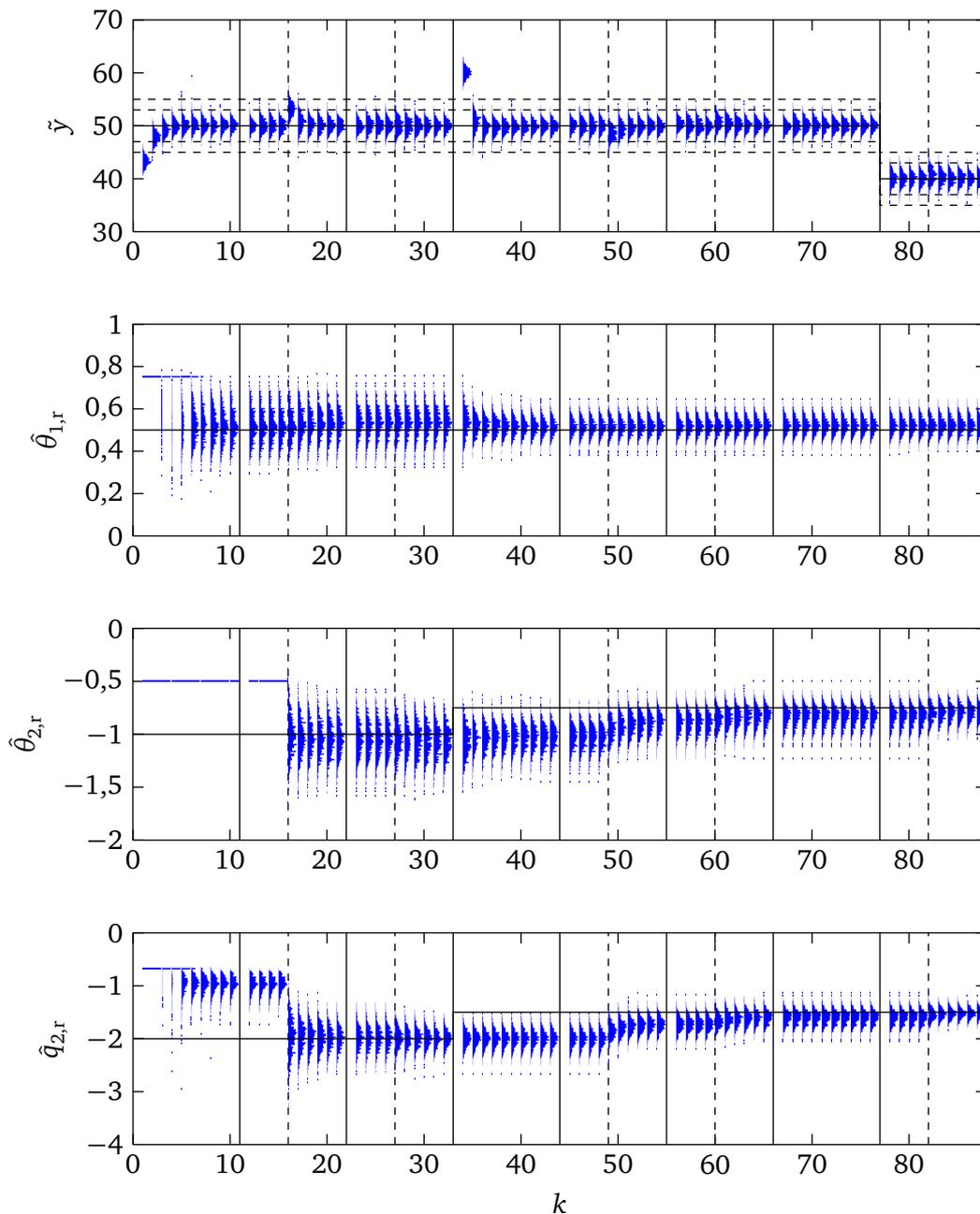


Abbildung 6.19: Sprung in θ_2 ab dem vierten Auftrag, $\lambda_2 = 0,5$ (1 000 Simulationen)

diesem Fall nicht von der reinen varianz-optimalen Mittelung (Abbildung 6.20). Im siebten Auftrag ergibt sich aber eine gute Korrektur, deren Varianz etwas geringer ist, als im Fall der reinen Mittelung mit Vergessensfaktor.

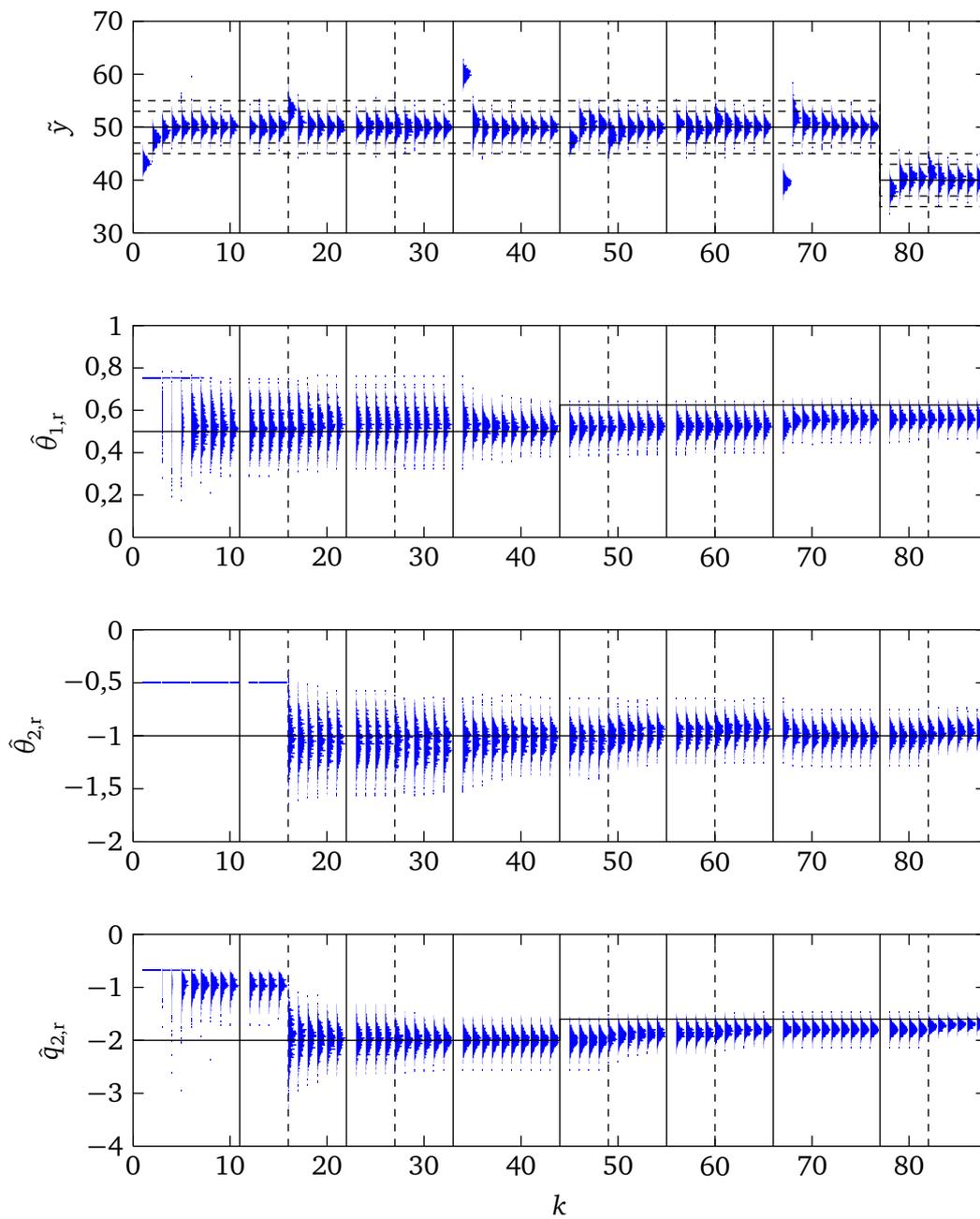


Abbildung 6.20: Sprung in θ_1 ab dem fünften Auftrag, $\lambda_1 = 1$, $\lambda_2 = 1$, varianz-optimale Mittelung (1 000 Simulationen)

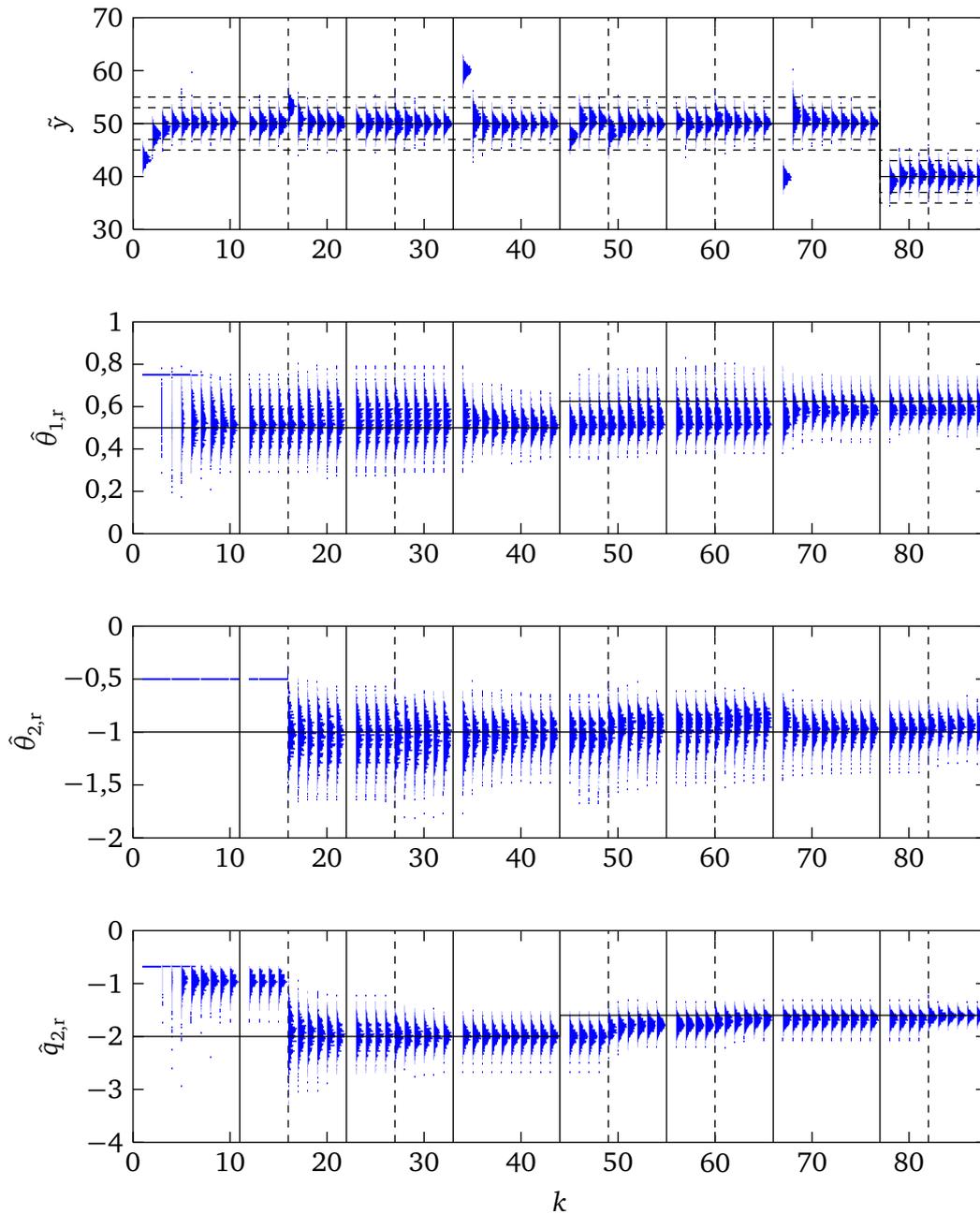


Abbildung 6.21: Sprung in θ_1 ab dem fünften Auftrag, $\lambda_1 = 0,75$, $\lambda_2 = 0,5$, Mittelung mit Vergessensfaktor (1 000 Simulationen)

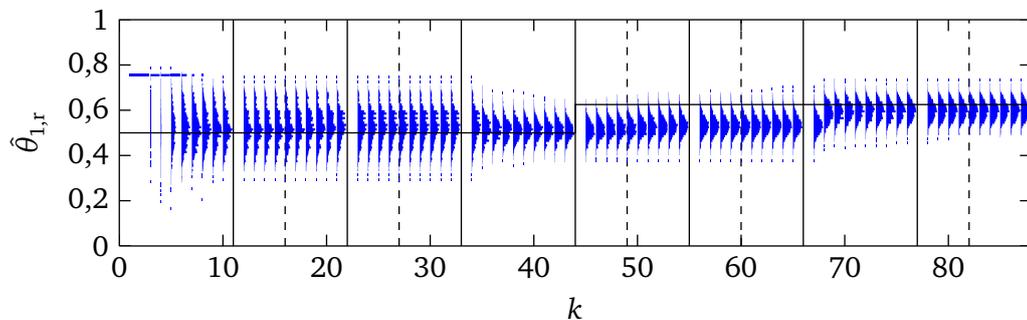


Abbildung 6.22: Sprung in θ_1 ab dem fünften Auftrag, $\lambda_1 = 0,75$, $\lambda_2 = 0,5$, Kombinierte Mittelung, $p_{1,\min} = \frac{1}{2} \cdot \sigma_{\hat{\theta}_{1,r,\max}}^2$ (1 000 Simulationen)

6.5.4 Auswahl eines Verfahrens

Als sinnvolle und robuste Lösung erweist sich die Mittelung nach Abschnitt 6.5.2. Zwar gehen dabei die Daten aus Aufträgen „verloren“, in denen die Maximalvarianz $\sigma_{\hat{\theta}_{1,m},\max}^2$ nicht erreicht wird, jedoch wurde begründet, dass sich Verfahren, die alle Messdaten verwenden, aufgrund der möglichen Sprünge des Parameters θ_0 nicht eignen.

Im Folgenden wird diskutiert, anhand welcher Überlegungen ein Verfahren zur Mittelung ausgewählt und parametrisiert werden kann.

Auswahlkriterien für ein Verfahren zur Mittelung

Eine wesentliche Entscheidung ist die, ob die Mittelung mit Vergessensfaktor oder die kombinierte Mittelung (varianz-optimal mit Vergessensfaktor) verwendet werden soll. Des Weiteren sind dann λ_1 und $\sigma_{\hat{\theta}_{1,m},\max}^2$ festzulegen.

Die Auswahl des Verfahrens zur Mittelung und die Wahl der dazugehörigen Parameter orientiert sich dabei an folgenden Punkten, die im Weiteren kurz diskutiert werden:

- Wie ist das allgemeine Niveau der Varianzen der Einzelschätzungen?
- Wie sind die Varianzen der Einzelschätzungen über die Aufträge verteilt?
- Sind häufige Parametersprünge zu erwarten, bzw. wie wesentlich ist eine schnelle Adaption an Parametersprünge?

Allgemeines Niveau der Varianzen der Einzelschätzungen

Besitzen die Einzelschätzungen (häufig) Varianzen im Bereich der für die Regelung zulässigen Varianz, so kann mit relativ niedrigen Vergessensfaktoren, d. h. hohem Vergessen gearbeitet werden. Die Varianzgrenze $\sigma_{\hat{\theta}_{1,m},\max}^2$ zur Mittelung wird dann nicht wesentlich höher als $\sigma_{\hat{\theta}_{1,r},\max}^2$ angesetzt (Faktor eins bis zwei).

Sind die Varianzen der Einzelschätzungen dagegen relativ hoch, so ist mit Vergessensfaktoren nahe eins zu arbeiten und $\sigma_{\hat{\theta}_{1,m},\max}^2$ entsprechend hoch zu wählen, um über eine große Anzahl an Einzelschätzungen zu mitteln. Auch sprechen hohe Varianzen für eine kombinierte Mittelung (die für λ nahe eins näherungsweise der varianz-optimalen Mittelung entspricht), um die schlechten Schätzwerte möglichst gut auszunutzen.

Verteilung der Varianzen der Einzelschätzungen über die Aufträge

Unterscheiden sich die Varianzen der Einzelschätzungen stark und kann auch eine große Anzahl an Aufträgen mit schlechter Anregung und hohen Varianzen hintereinander vorkommen, so spricht dies für eine kombinierte Mittelung und hohem λ . Alternativ kann auch die Grenze $\sigma_{\hat{\theta}_{1,m},\max}^2$ so niedrig angesetzt werden, dass die Aufträge mit schlechter Varianz einfach ignoriert werden, jedoch kann dann natürlich gar keine Adaption an Parameteränderungen stattfinden.

Bei gleichmäßigen Varianzen kann die Mittelung mit Vergessensfaktor verwendet werden. Diese hat dann den Vorteil, dass ein Ausreißer zum Guten, d. h. eine Einzelschätzung mit sehr geringer Varianz, die Adaption an veränderliche Parameterwerte nicht blockiert. Bei der kombinierten Mittelung ist in diesem Fall der Wert $p_{1,m}$ entsprechend nach unten zu begrenzen.

Häufigkeit von Parametersprüngen bzw. Notwendigkeit einer schnellen Adaption

Kann davon ausgegangen werden, dass der betreffende Parameter sich nicht verändert, so spricht dies natürlich für die kombinierte Mittelung mit einem hohen λ -Wert. Ganz auf eins sollte er jedoch nur gesetzt werden, wenn auch ausgeschlossen werden kann, dass eine größere Abweichung eines Schätzwertes unkritisch ist, bzw. die Wahrscheinlichkeit für das Auftreten eines kritischen Schätzfehlers durch niedrige Werte für $\sigma_{\hat{\theta}_{1,m},\max}$ und $\sigma_{\hat{\theta}_{1,r},\max}$ auf ein vertretbares Niveau sinkt.¹⁷

Treten häufig Parametersprünge auf bzw. wird auch bei selten vorkommenden Sprüngen eine sehr schnelle Adaption erwartet, so ist der Vergessensfaktor entsprechend gering zu wählen. Die Mittelung mit Vergessensfaktor ist dabei als robuster anzusehen, da bei dieser keine „Blockade“ der Anpassungen durch gute Einzelschätzungen vorkommen kann wie bei der kombinierten Mittelung.

Die Grenze $\sigma_{\hat{\theta}_{1,m},\max}$ ist so hoch anzusetzen, dass keine längeren Phasen auftreten, in denen die Einzelschätzwerte nicht zum Schätzen verwendet werden dürfen, da in diesen Phasen keine Adaption stattfinden kann.

Fazit

Geht man die drei oben beschriebenen Punkte der Reihe nach durch, so können sich durchaus widersprüchliche Empfehlungen ergeben. In diesen Fällen sind anhand von Simulationen und Versuchen am realen System Kompromisse zu finden.

6.6 Untersuchung des Bias

In den bisherigen Simulationsdaten lässt sich ein geringer Bias in der Schätzung von θ_1 erkennen, siehe z. B. Abbildung 6.15. Die Ursache für diesen Bias wird in diesem Abschnitt diskutiert. Dabei wird auch auf die Identifikation im geschlossenen Regelkreis eingegangen, die (eine) Ursache für den beobachteten Bias ist.

Zunächst wird der prinzipielle Einfluss einer Identifikation im geschlossenen Regelkreis auf den Bias beschrieben.

Die Untersuchung und Berechnung des Bias wird daraufhin für den ersten Schätzwert eines Auftrags, d. h. $k = 2$, behandelt. Im Anschluss wird der Bias für $k > 2$ besprochen. Es ist dabei wesentlich, die Schätzwerte zu begrenzen, um eine sinnvolle Definition des Bias zu erhalten.

Eine solche Begrenzung ist im Prinzip schon durch die Überprüfung der Varianz des Schätzwertes, $\sigma_{\hat{\theta}_1}^2 \leq \sigma_{\hat{\theta}_{1,m},\max}^2$ bzw. $\sigma_{\hat{\theta}_1}^2 \leq \sigma_{\hat{\theta}_{1,m},\max}^2$ gegeben, jedoch wird zunächst als einfachere Bedingung eine minimale anfängliche Regelabweichung, $|\tilde{\epsilon}_1| \geq \epsilon_{1,\min}$, verwendet. Diese lässt sich etwas einfacher interpretieren und ist für die numerische Berechnung des Bias vorteilhaft. Im Anschluss werden die Ergebnisse dann auf die „Varianz-Bedingung“ übertragen.

Diese Untersuchungen werden für den Fall durchgeführt, dass alle sekundären Eingänge konstant sind und somit nur eine Zwei-Parameter-Schätzung von θ_1 durchgeführt wird.

¹⁷ Wenn die Varianz einer Parameterschätzung gering ist, dann bedeutet das nicht, dass bei *einem* Auftrag mal ein schlechter Schätzwert, aber meistens gute Schätzwerte vorliegen, sondern dass bei den meisten Aufträgen gute Schätzwerte erreicht werden, bei manchen Aufträgen aber schlechte Schätzungen. Betrachtet man die gezeigten Verläufe, so sollte man sich auch bewusst machen, dass die Abweichungen bei $\{\tilde{y}_k\}_r$ relativ gering korreliert sind, d. h. die Werte für eine einzelne Simulation springen um den jeweiligen Mittelwert, während die anderen gezeigten Zufallsprozesse stark korreliert sind. D. h. ein großer Schätzfehler bei einem Parameter ist mit dem nächsten Schritt k' nicht ein geringer Schätzfehler oder einer mit anderem Vorzeichen, sondern meist bleibt dieser ein großer Schätzfehler.

6.6.1 Prinzipieller Einfluss des geschlossenen Regelkreises

Die Schätzung der Parameter erfolgt im geschlossenen Regelkreis. Dadurch ist die Biasfreiheit des LS-Schätzers nicht mehr gegeben, da die Annahme A2 (Seite 94), die Unkorreliertheit von Rauschen und Eingangswerten, nicht mehr gegeben ist.

Für den Schätzwert gilt

$$E\{\hat{\theta}\}_r = E\{(\Psi^T \Psi)^{-1} \Psi^T \cdot (y + n_y)\}_r = E\{(\Psi^T \Psi)^{-1} \Psi^T y\}_r + E\{(\Psi^T \Psi)^{-1} \Psi^T n_y\}_r = \theta + E\{(\Psi^T \Psi)^{-1} \Psi^T n_y\}_r,$$

wobei $n_y^T = [n_{y,1} \ n_{y,2} \ \dots \ n_{y,k}]$ der Vektor der Störgröße ist. Wenn das Eingangssignal und die Störung nicht korreliert sind, dann ergibt sich

$$E\{\Delta \hat{\theta}\}_r = E\{\hat{\theta}\}_r - \theta = E\{(\Psi^T \Psi)^{-1} \Psi^T\}_r \cdot E\{n_y\}_r = \mathbf{0},$$

wobei davon ausgegangen wird, dass die Störung mittelwertfrei ist. In diesem Fall ist der Schätzer also biasfrei.

Durch den Regler gilt aber

$$v_{k+1} = v_k + (y_{\text{soll}} - y_k - n_{y,k}) \cdot k_I.$$

Damit ist nicht nur $\{n_{y,k}\}_r$ ein Zufallsprozess, sondern auch $\{v_k\}_r$, der dazu noch mit $\{n_{y,k}\}_r$ korreliert ist. Betrachtet man die Zwei-Parameter-Schätzung, so lautet der Bias $\Delta \hat{\theta}$ nach k Messungen

$$E\{\Delta \hat{\theta}\}_r = E \left\{ \frac{1}{k \cdot \sum v_{k'}^2 - (\sum v_{k'})^2} \cdot \left[\sum v_{k'}^2 \cdot \sum n_{y,k'} - \sum v_{k'} \cdot \sum u_{k'} n_{y,k'} \right] \right\}_r.$$

Eine Störung $n_{y,k'}$ wirkt sich zwar nur auf spätere Werte von v aus, also $v_{k'+1}$, $v_{k'+2}$ usw., jedoch sind durch die Summen und den von $v_{k'}$ abhängigen Nenner alle $n_{y,k'}$ mit allen $v_{k'}$ multiplikativ verknüpft.

Behandlung in der Literatur

Bei der Identifikation dynamischer Systeme existieren verschiedene Ursachen für einen Bias, von denen die Korreliertheit von Eingangswerten und Messrauschen aufgrund eines geschlossenen Regelkreises nur eine ist. Identifiziert man dynamische Systeme durch Minimierung des Gleichungsfehlers, so darf das auf den Systemausgang wirkende Rauschen nicht weiß sein, sondern muss die Form $\frac{1}{A(z)} n(z)$ besitzen, wobei $A(z)$ das Nennerpolynom der Strecke und n weißes Rauschen ist [ISERMANN und MÜNCHHOF, 2011, S. 233]. Dann werden bezüglich des Gleichungsfehlers die Bedingungen (A2) und (A4) (Seite 94) erfüllt. Da diese Bedingung sehr einschränkend ist, wird dieser Thematik in der Literatur viel Platz eingeräumt. Eine Möglichkeit zur Behandlung dieser Problematik wird als „bias-correction“ oder „bias-compensation“ bezeichnet, siehe z. B. [SÖDERSTRÖM, 2013] für eine aktuelle Übersicht.

Die Identifikation im geschlossenen Regelkreis an sich ist auch Thema vieler Veröffentlichungen, jedoch liegt, insbesondere in der aktuelleren Literatur, der Schwerpunkt der Betrachtungen darauf, durch Identifikation im geschlossenen Regelkreis bessere – im Hinblick auf den Einsatz mit gegebenen Reglertypen oder zur robusten Regelung – Modelle zu erhalten [FORSELL und LJUNG, 1999, S. 1216]. Wenn dort von Bias die Rede ist, so ist meist der Bias gemeint, der durch einen falschen, zu einfachen Modellansatz entsteht („undermodeling“), siehe z. B. [VAN DEN HOF und SCHRAMA, 1995, S. 1751]. Einen ausführlichen Überblick über die etwas weiter zurückliegende Forschung auf dem Gebiet gibt der Übersichtsaufsatz

[GUSTAVSSON et al., 1977]. Dabei ist die Bestimmung der Identifizierbarkeit sowie der Genauigkeit ein wesentlicher Schwerpunkt.

Hier wird der Bias im Folgenden nicht anhand von Methoden aus der genannten Literatur besprochen. Es hat sich aufgrund des einfachen Systems sowie der speziellen Randbedingungen als zielführend erwiesen, den Bias direkt anhand der Verteilung des Schätzfehlers (numerisch) zu bestimmen. Aufgrund dessen, dass der Bias neben den Verteilungseigenschaften des Rauschens nur von der Reglerverstärkung α und der anfänglichen Regelabweichung ε_1 abhängt, ist dies mit einem vertretbaren Aufwand möglich.

Gemäß der allgemeinen Nomenklatur liegt hier eine direkte Identifikation vor, da die Identifikation anhand der Ein- und Ausgangsdaten des Prozesses erfolgt. Dagegen würde bei der indirekten Identifikation zunächst das Gesamtsystem, welches die Rückführung und den Regler umfasst, identifiziert werden, um dann aus diesem sowie dem bekannten Regler das Modell des eigentlichen Prozesses zu erhalten. [LJUNG, 1999, S. 435]

6.6.2 Bias für $k = 2$

In Abschnitt 4.5.2 wurde mit Gl. (4.22) der Schätzfehler für $k = 2$ zu

$$\Delta \hat{\vartheta}_{1,2} = \frac{n_2}{\kappa_1 \cdot (\varepsilon_1 - n_1)} - \frac{n_1}{\kappa_1 \cdot (\varepsilon_1 - n_1)} \quad (6.52)$$

bestimmt. Der Term mit n_2 hat für den Bias, den Erwartungswert des Schätzfehlers, keinen Einfluss, weshalb der Term

$$\Delta \hat{\vartheta}_{1,2}^* = -\frac{n_1}{\kappa_1 \cdot (\varepsilon_1 - n_1)} \quad (6.53)$$

alleine betrachtet wird.

Da $\{n_1\}_r$ eine Zufallsvariable ist, ist auch $\{\Delta \hat{\vartheta}_{1,2}^*\}_r$ eine Zufallsvariable. In Abbildung 6.23 sind die Verteilungsdichten von $\{\kappa_1 \cdot \Delta \hat{\vartheta}_{1,2}^*\}_r$ für verschiedene Werte von ε_1 gezeigt.

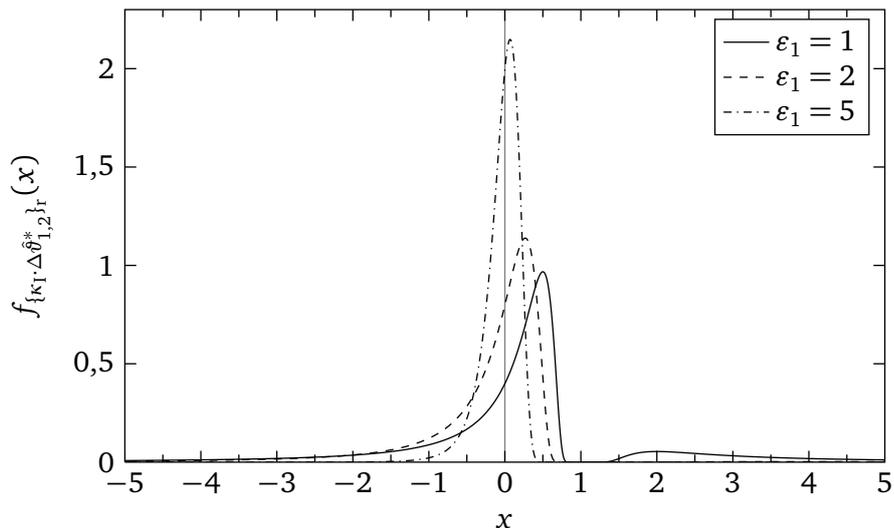


Abbildung 6.23: Verteilungsdichte von $\{\kappa_1 \cdot \Delta \hat{\vartheta}_{1,2}^*\}_r$ für verschiedene ε_1

Für $\varepsilon_1 \gg 1$ gilt

$$\Delta \hat{\vartheta}_{1,2}^* \approx \frac{-1}{\kappa_1 \varepsilon_1} \cdot n_1, \quad (6.54)$$

es ergibt sich mit wachsendem ε_1 eine immer engere Normalverteilung um null.

Es wurde in Abschnitt 4.5.2 schon darauf hingewiesen, dass der Erwartungswert von $\{\Delta\hat{\vartheta}_{1,2}^*\}_r$, der dem Bias $\Delta\hat{\vartheta}_{1,\text{mean},2}$ entspricht, nicht definiert ist. Um diesen sinnvoll bestimmen zu können, muss ausgeschlossen werden, dass der Nenner in (6.53) null wird.

Dies kann durch die Bedingung

$$|\tilde{\varepsilon}_1| = |n_1 - \varepsilon_1| \stackrel{!}{\geq} \varepsilon_{1,\text{min}}$$

geschehen. Für $k = 2$ entspricht diese auch der schon eingeführten Varianzgrenze $\sigma_{\hat{\vartheta}_{1,m,\text{max}}}^2$, denn für $k = 2$ gilt

$$\text{Var}\{\hat{\vartheta}_{1,2}\}_r = \frac{2}{(u_{1,2} - u_{1,1})^2} = \frac{2}{(\kappa_1 \tilde{\varepsilon}_1)^2}.$$

Drückt man $\sigma_{\hat{\vartheta}_{1,m,\text{max}}}$ anschaulich als Vielfaches des angenommenen Parameterwertes $\hat{\vartheta}_{1,r}$ aus, so ergibt sich

$$\text{Std}\{\hat{\vartheta}_{1,2}\}_r = \frac{\sqrt{2}}{\kappa_1 \cdot |\tilde{\varepsilon}_1|} \stackrel{!}{\leq} \gamma \hat{\vartheta}_{1,r}$$

als Bedingung an die Standardabweichung des ersten Schätzwertes. Umformen ergibt

$$|\tilde{\varepsilon}_1| \stackrel{!}{\geq} \frac{\sqrt{2}}{\gamma \hat{\vartheta}_{1,r} \kappa_1}$$

als Bedingung an die erste gemessene Regelabweichung¹⁸ und mit der Näherung $\hat{\vartheta}_{1,r} \kappa_1 \approx 1$ erhält man die Abschätzung

$$|\tilde{\varepsilon}_1| \stackrel{!}{\geq} \frac{\sqrt{2}}{\gamma}. \tag{6.55}$$

Im letzten Abschnitt wurde für $\sigma_{\hat{\vartheta}_{1,m,\text{max}}}$ meist der Wert $2,5 \cdot \sigma_{\hat{\vartheta}_{1,r,\text{max}}} = 2,5 \cdot \frac{1}{5} \cdot \vartheta_1$ gewählt, womit γ ungefähr bei 0,5 und damit $\varepsilon_{1,\text{min}} \approx 2,83$ ist.

In Abbildung 6.24 ist der Bias

$$\Delta\hat{\vartheta}_{1,\text{mean},2}(\varepsilon_1) = \text{E}\{\Delta\hat{\vartheta}_{1,2} \mid |\tilde{\varepsilon}_1| \geq \varepsilon_{1,\text{min}}\}_r,$$

der sich bei verschiedenen Begrenzungen von $|\tilde{\varepsilon}_1|$ ergibt, multipliziert mit $\kappa_1 = \alpha/\vartheta_1$ über ε_1 dargestellt.

Bei einer tatsächlichen Anfangsabweichung von null liegt unabhängig von $\varepsilon_{1,\text{min}}$ ein Bias von ϑ_1/α vor. Dieser fällt mit steigendem $|\varepsilon_1|$ umso schneller ab, je kleiner $\varepsilon_{1,\text{min}}$ gewählt ist. Für $\varepsilon_{1,\text{min}} = 3$ erreicht er bei ca. ± 4 negative Werte, um sich dann von unten der Null zu nähern.

Für kleinere $\varepsilon_{1,\text{min}}$ sind die beiden Minima im Negativen deutlich ausgeprägter und die Spitze schmäler. Die beiden Kurven für $\varepsilon_{1,\text{min}} = 0,1$ und $0,01$ liegen fast deckungsgleich übereinander. Allerdings kann daraus nicht geschlossen werden, dass sich ein Grenzverlauf des Bias für $\varepsilon_{1,\text{min}} \rightarrow 0$ ergibt. Es zeigt sich auch schon, dass die Kurve für $|\tilde{\varepsilon}_1| \geq 0,01$ in den Bereichen der Minima etwas zackiger verläuft, was

¹⁸ Der Wert der Variablen n_2 in der eingangs angegebenen Gleichung (6.52) für den Schätzfehler wird durch diese Bedingung nicht beschränkt. Dies rechtfertigt die Betrachtung des Termes $\Delta\hat{\vartheta}_{1,2}^*$ alleine.

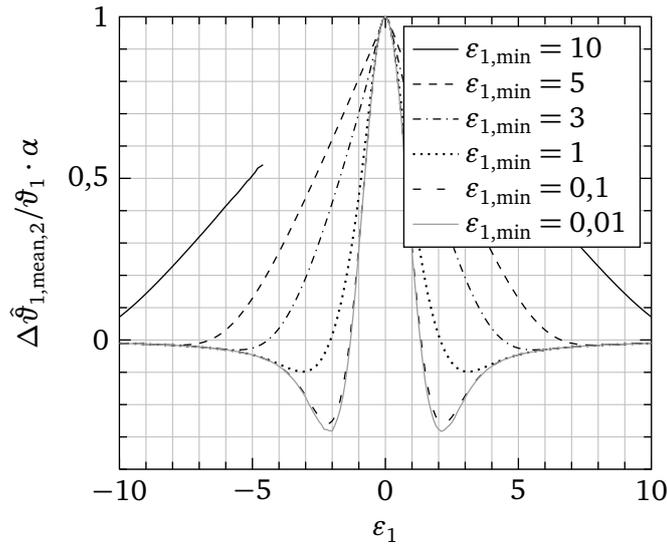


Abbildung 6.24: Bias für $\hat{\vartheta}_1$ bei $k = 2$ und verschiedenen Grenzen für ε_1 (Aus 10^7 Einzelwerten je Punkt bestimmt)

ein Zeichen dafür ist, dass viele extreme Einzelwerte auftreten und für eine genauere Berechnung die Anzahl der Simulationen noch erhöht werden müsste.

Erhöht man die Grenze $\varepsilon_{1, \text{min}}$, so erhöht sich auch der Bias für Werte von $|\varepsilon_1|$, die kleiner und ähnlich dem Grenzwert $\varepsilon_{1, \text{min}}$ sind. Besonders wird dies bei $\varepsilon_{1, \text{min}} = 10$ deutlich. Dies liegt daran, dass durch die gegebene Bedingung ein bestimmter Bereich von Werten, die n_1 annimmt, gar nicht weiter berücksichtigt wird. So sind bei $\varepsilon_{1, \text{min}} = 10$ und $\varepsilon_1 = -5$ nur Einzelwerte in der Mittelwertbildung berücksichtigt, bei denen n_1 einen größeren negativen Wert $n_1 \leq -5$ angenommen hat. Da der Erwartungswert von $\{n_2\}_r$ null ist, folgt daraus anschaulich, dass ϑ_1 unter dieser Bedingung im Mittel deutlich zu hoch geschätzt wird.

Entsprechend der äußerst geringen Wahrscheinlichkeit $P(\{n_1\}_r^\omega \leq -5)$ sind im Laufe der Simulation auch nur wenige Werte für die numerische Bestimmung des Mittelwertes an diesen Punkten aufgetreten, so dass die Kurve in diesem Bereich auch etwas unruhiger verläuft. Für $|\varepsilon_1| < 4,5$ wurde die Bedingung von keiner der 10^7 Einzelberechnungen erfüllt. Dies relativiert den hohen Bias bei $\varepsilon_{1, \text{min}} = 10$ und $|\varepsilon_1| = 5$ wieder, da dieser Fall praktisch nicht auftritt. Jedoch ist der Bias bei $|\varepsilon_1| \approx \varepsilon_{1, \text{min}}$ durchaus relevant, was später auch nochmal an einem Beispiel gezeigt wird.

Dieses „Beschneiden“ ist für die Bedingung $|\varepsilon_1| \geq 10$ auch in Abbildung 6.25 dargestellt. Dort sind die (auf den jeweiligen Maximalwert normierten, numerisch bestimmten) Verteilungsdichtefunktionen für verschiedene ε_1 dargestellt. Es ist deutlich zu erkennen, wie der Mittelwert mit sinkendem $|\varepsilon_1|$ nach rechts wandert.

Bei kleinen Werten von $\varepsilon_{1, \text{min}}$ (was einer hohen zulässigen Varianz $\sigma_{\hat{\vartheta}_{1, m, \text{max}}}^2$ entspricht) tritt also ein Bias infolge einer asymmetrischen „Grundverteilung“ auf, siehe Abbildung 6.23. Bei höheren Werten für $\varepsilon_{1, \text{min}}$ tritt ebenfalls ein Bias auf, der dadurch entsteht, dass die für größere ε_1 näherungsweise symmetrische und biasfreie Verteilung (6.54) durch die Bedingung $|\varepsilon_1| \geq \varepsilon_{1, \text{min}}$ für Werte von $|\varepsilon_1|$, die ungefähr im Bereich $|\varepsilon_1| \leq \varepsilon_{1, \text{min}} + 2$ liegen, deutlich beschnitten wird.

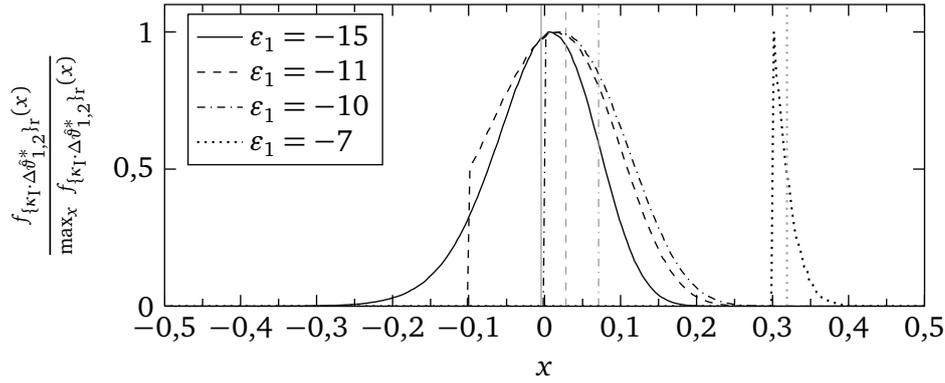


Abbildung 6.25: Verteilungsdichte von $\{\kappa_I \cdot \Delta\hat{\theta}_{1,2}^*\}_r$ (auf jeweiligen Maximalwert normiert) unter der Bedingung $|\varepsilon_1| \geq 10$ für verschiedene ε_1 (Die senkrechten grauen Striche markieren die jeweiligen Erwartungswerte)

6.6.3 Bias für $k > 2$

Für eine beliebige Anzahl an Messungen ergeben sich die Schätzwerte über $\hat{\theta} = \mathbf{L}^{-1}\mathbf{r}$. Bei der Zwei-Parameter-Schätzung lautet

$$\mathbf{L} = \begin{bmatrix} k & \sum v_{k'} \\ \sum v_{k'} & \sum v_{k'}^2 \end{bmatrix} \quad \text{und} \quad \mathbf{r} = \begin{bmatrix} \sum \tilde{y}_{k'} \\ \sum (\tilde{y}_{k'} v_{k'}) \end{bmatrix}.$$

Die Inverse \mathbf{L} ist damit

$$\mathbf{L}^{-1} = \frac{1}{k \cdot \sum v_{k'}^2 - (\sum v_{k'})^2} \cdot \begin{bmatrix} \sum v_{k'}^2 & -\sum v_{k'} \\ -\sum v_{k'} & k \end{bmatrix}.$$

Der betrachtete Schätzwert $\hat{\theta}_1$ ist dabei der zweite Eintrag in $\hat{\theta}$, so dass sich

$$\hat{\theta}_{1,k} = \frac{k \cdot \sum (\tilde{y}_{k'} v_{k'}) - \sum v_{k'} \cdot \sum \tilde{y}_{k'}}{k \cdot \sum v_{k'}^2 - (\sum v_{k'})^2}$$

ergibt. Verwendet man $\tilde{y}_{k'} = y_{k'} + n_{y,k'}$, so führt dies auf

$$\begin{aligned} \hat{\theta}_{1,k} &= \frac{k \cdot \sum ((y_{k'} + n_{y,k'}) \cdot v_{k'}) - \sum v_{k'} \cdot \sum (y_{k'} + n_{y,k'})}{k \cdot \sum v_{k'}^2 - (\sum v_{k'})^2} \\ &= \underbrace{\frac{k \cdot \sum (y_{k'} v_{k'}) - \sum v_{k'} \cdot \sum y_{k'}}{k \cdot \sum v_{k'}^2 - (\sum v_{k'})^2}}_{\theta_1} + \frac{k \cdot \sum (n_{y,k'} v_{k'}) - \sum v_{k'} \cdot \sum n_{y,k'}}{k \cdot \sum v_{k'}^2 - (\sum v_{k'})^2}. \end{aligned}$$

Der erste Summand entspricht dabei dem wahren Wert θ_1 , so dass sich der Schätzfehler in normierter Form zu

$$\Delta\hat{\theta}_{1,k} = \frac{k \cdot \sum (n_{k'} v_{k'}) - \sum v_{k'} \cdot \sum n_{k'}}{k \cdot \sum v_{k'}^2 - (\sum v_{k'})^2} \quad (6.56)$$

ergibt.

In Anhang C.5 ist gezeigt, dass für den Mittelwert von $\kappa_I \cdot \Delta \hat{\vartheta}_{1,k}$ bzw. $\Delta \hat{\vartheta}_{1,k} / \vartheta_1 \cdot \alpha$ ein Zusammenhang der Form

$$\kappa_I \cdot \Delta \hat{\vartheta}_{1,\text{mean},k} = \bar{E}\{\kappa_I \cdot \Delta \hat{\vartheta}_{1,k}\}_r = \kappa_I \cdot \Delta \hat{\vartheta}_{1,\text{mean},k}(\alpha, \varepsilon_1, \{n_1\}_r, \{n_2\}_r, \dots, \{n_{k-1}\}_r) \quad (6.57)$$

gilt, wobei $\bar{E}\{\kappa_I \cdot \Delta \hat{\vartheta}_{1,k}\}_r$ den Erwartungswert bezüglich einer geeigneten Bedingung meint. Eine wesentliche Aussage von (6.57) ist, dass der Streckenparameter ϑ_1 lediglich über α (und κ_I) einfließt. Damit kann ein Beispielsystem mit beliebigen ϑ_1 betrachtet werden und die Ergebnisse lassen sich dennoch einfach auf andere Strecken übertragen.

Zur Bestimmung des Erwartungswertes wird zunächst der Einfachheit halber wieder die Bedingung $|\varepsilon_1| \leq \varepsilon_{1,\text{min}}$ verwendet. Für $k > 2$ kann diese jedoch nicht mehr die eigentliche Varianzgrenze $\sigma_{\hat{\vartheta}_{1,m,\text{max}}}^2$ genau abbilden. Darauf wird im nächsten Abschnitt eingegangen.

In Abbildung 6.26 ist für den zweiten Schätzwert ($k = 3$) der Bias für verschiedene α -Werte dargestellt. Auch hier gilt wieder, dass bei $\varepsilon_{1,\text{min}} = 5$ der Wert für $\Delta \hat{\vartheta}_{1,\text{mean},k} / \vartheta_1 \cdot \alpha$ auch für kleine $|\varepsilon_1|$ existiert, diese jedoch numerisch nur mit größerem Aufwand bestimmt werden können.

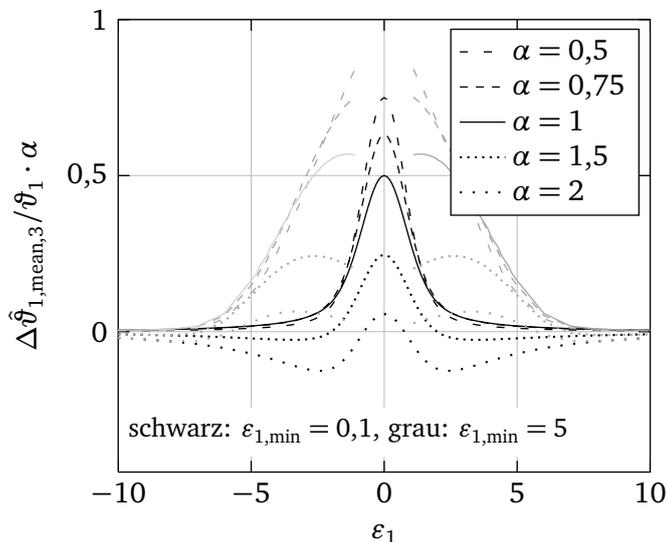


Abbildung 6.26: Bias des zweiten Schätzwertes ($k = 3$) für ϑ_1

Für $\alpha = 1$ können die dargestellten Werte $\Delta \hat{\vartheta}_{1,\text{mean},k} / \vartheta_1 \cdot \alpha$ direkt als relativer Bias interpretiert werden, für $\alpha \neq 1$ muss dafür noch durch α geteilt werden. Dies bedeutet, dass der relative Bias für $\alpha < 1$ größer und für $\alpha > 1$ betragsmäßig geringer ist, als die dargestellten Kurven.

Für $\alpha < 1$ ist der auf ϑ_1 / α normierte Bias in der Regel (etwas) höher, für $\alpha > 1$ geringer als bei $\alpha = 1$. Berechnet man den normierten Bias auch für extremere Werte für α , so zeigt sich (hier nicht dargestellt), dass für $\alpha \gg 1$ und $\varepsilon_{1,\text{min}} = 0,1$ der normierte Bias zunächst noch etwas mehr im Negativen verläuft als bei $\alpha = 2$, für weiter wachsende α jedoch ein Punkt erreicht wird, an dem der negative Bias (betragsmäßig) wieder abnimmt. Dies ist soweit auch einsichtig, da ein viel zu großes α zu einem instabilen, aufschwingenden Regelkreisverhalten führt und sich damit die Anregung für den Schätzer verbessert.

Für $\alpha \ll 1$ legen die numerischen Berechnungen die Vermutung nahe, dass sich der auf ϑ_1 / α normierte Bias einem Grenzverlauf annähert, der für kleine $|\varepsilon_1|$ oberhalb des für $\alpha = 0,5$ gezeigten Verlaufs liegt. Bestimmt man jedoch den relativen, auf ϑ_1 bezogenen Bias, indem man die gezeigten Kurven durch α teilt, so zeigt sich, dass der Bias für kleine α bei kleinen $|\varepsilon_1|$ stark zunimmt. Dies lässt sich dadurch begründen, dass die Anregung für kleine α sehr gering bis praktisch nicht mehr vorhanden ist. Auf die damit entstehende Frage nach der Stabilität des Bias wird weiter unten eingegangen.

Vergleicht man die Kurven der unterschiedlichen Grenzen $\varepsilon_{1,\min} = 0,1$ und 5 miteinander, so zeigt sich hier Gleiches wie schon bei $k = 2$. Der Bias ist ungefähr für $|\varepsilon_1| < 7$ bei der Schätzung mit $\varepsilon_{1,\min} = 5$ höher als bei $\varepsilon_{1,\min} = 0,1$, um dann in die Kurve der Schätzung mit $\varepsilon_{1,\min} = 0,1$ überzugehen.

In Abbildung 6.27 ist der Bias für die ersten vier Schätzwerte ($k = 2, \dots, 5$) dargestellt.

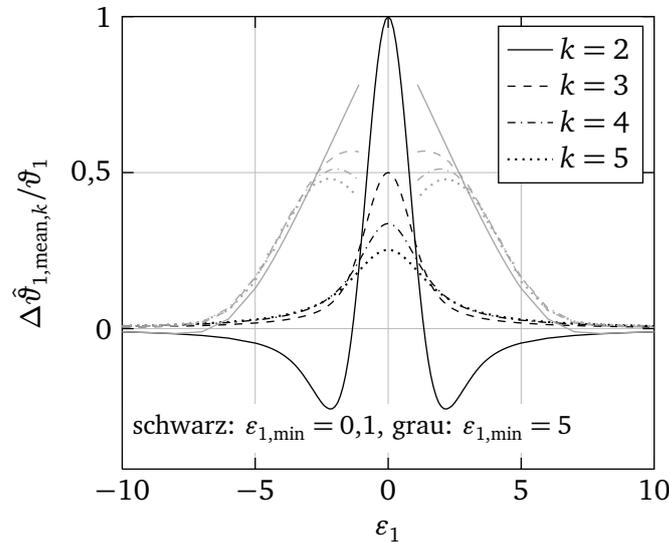


Abbildung 6.27: Bias der ersten vier Schätzwerte ($k = 2, \dots, 5$) für ϑ_1 bei $\alpha = 1$

Wieder zeigt sich, dass der Bias für $\varepsilon_{1,\min} = 5$ für kleinere $|\varepsilon_1|$ deutlich über dem bei $\varepsilon_{1,\min} = 0,1$ liegt. Im Bereich um $|\varepsilon_1| = 5$ ist auch eine leichte Vergrößerung des Bias in den Schritten $k = 3, \dots, 5$ bezüglich dem Bias bei $k = 2$ zu erkennen.

Für höhere k steigt der Aufwand zur numerischen Berechnung des Bias anhand der im Anhang angegebenen Gleichungen stark an. Um den Verlauf des Bias für höhere k zu bewerten, wurden daher einfache Simulationen des geregelten Systems durchgeführt und der Bias als Differenz des Mittelwertes von $\hat{\vartheta}_k$ aus jeweils 100 000 Simulationen mit dem bekannten wahren Parameterwert ϑ_1 bestimmt. Damit wurden die in Abbildung 6.28 gezeigten Verläufe ermittelt, wobei hier nur der Fall $\varepsilon_1 = 4$ dargestellt wird.

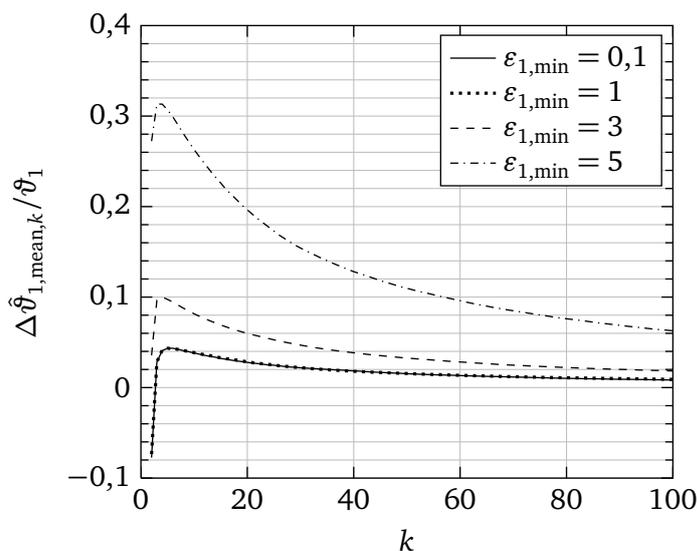


Abbildung 6.28: Verläufe des Bias über der Anzahl der Messungen, $\varepsilon_1 = 4$, $\alpha = 1$ (erstellt über jeweils 100 000 Simulationen)

Für $\varepsilon_{1,\min} = 0,1$ und $\varepsilon_{1,\min} = 1$ ergibt sich ein identischer Verlauf, der bei $k = 2$ mit einem negativen Bias beginnt, bei $k = 3$ einen kleineren positiven Wert hat und dann noch leicht zunimmt, um letztlich über eine steigende Anzahl an Messungen wieder abzunehmen. (Siehe dazu auch Abbildung 6.27.) Dass diese beiden Kurven hier identisch sind, liegt daran, dass mit $\varepsilon_1 = 4$ ein Wert gewählt wurde, der deutlich über $\varepsilon_{1,\min} = 1$ liegt.

Für $\varepsilon_{1,\min} = 3$ und $\varepsilon_{1,\min} = 5$ ergeben sich Verläufe, die deutlich über den beiden schon beschriebenen Verläufen liegen. Es zeigt sich, dass – insbesondere bei $\varepsilon_{1,\min} = 5$ – lange ein wesentlicher Bias vorliegt.

Zusammenfassung

Der Bias aufgrund der Korreliertheit der Eingangsdaten ist nur für kleine $\varepsilon_{1,\min}$ wesentlich. Für $|\tilde{\varepsilon}_1| \leq \varepsilon_{1,\min} + 2$ dominiert ein Bias, der aufgrund der Bedingung $|\tilde{\varepsilon}_1| \leq \varepsilon_{1,\min}$ entsteht, und nicht durch den durch die Identifikation im geschlossenen Regelkreis entstehenden Bias. Dieser dominiert wieder für $|\tilde{\varepsilon}_1| > \varepsilon_{1,\min} + 2$, jedoch ist dieser dann für vernünftige $\varepsilon_{1,\min}$ so gering, dass er nicht mehr wesentlich ist.

Allgemein hängt der Bias neben der Grenze $\varepsilon_{1,\min}$ auch von der anfänglichen Regelabweichung ε_1 , der Anzahl der Messungen k und der tatsächlichen Reglerverstärkung α ab, wobei die letzten drei Einflussgrößen die „Anregung“ des Schätzers beschreiben. Je besser der Schätzer angeregt wird, d. h. je höher die Variation der Eingangsgröße u_1 ist, umso geringer ist der Bias.

Dabei zeigt sich, dass bei der betrachteten Bedingung $|\tilde{\varepsilon}_1| \leq \varepsilon_{1,\min}$ der Bias stark von ε_1 abhängt, aber nur relativ langsam mit steigendem k fällt. Dies ist darauf zurückzuführen, dass der erste Regelschritt der dominante ist, d. h. die deutlich größte Änderung in u_1 erzeugt.

6.6.4 Berücksichtigung der „Varianz-Bedingung“

Bisher wurde als Bedingung, ob ein Schätzwert für den Bias berücksichtigt werden soll oder nicht, die gemessene Abweichung bei $k = 1$ verwendet. Tatsächlich wird jedoch, wie in Abschnitt 6.4.1 beschrieben, die Varianz $\sigma_{\hat{\vartheta}_1}^2$ des Schätzwertes herangezogen, um zu entscheiden, ob ein Schätzwert für die Mittelung verwendet wird. Damit ist es auch sinnvoll, diese Bedingung hier zu verwenden.

Um die Varianz der Schätzwerte zu bestimmen wird die Identifikation simuliert, womit der Bias auch nur noch durch Simulation bestimmt wird.

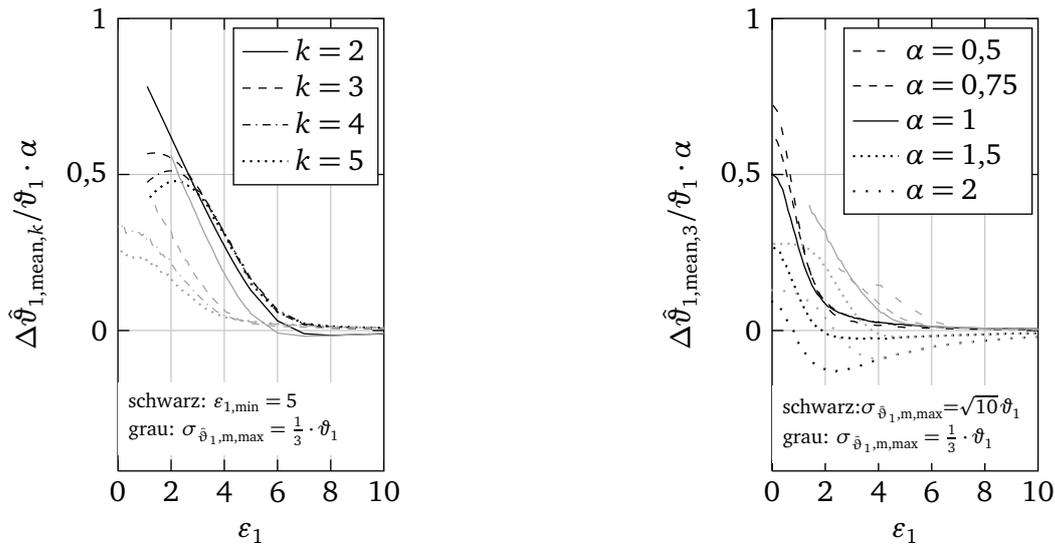
Für große Werte für $\sigma_{\hat{\vartheta}_1, m, \max}$ werden sehr viele, auch schlechte Schätzungen berücksichtigt und kaum Schätzwerte abgewiesen. Dieser Fall entspricht einem kleinen $\varepsilon_{1,\min}$, so dass sich beispielsweise mit $\sigma_{\hat{\vartheta}_1, m, \max} = \sqrt{10}$ auch sehr ähnliche Biasverläufe über ε_1 wie bei $\varepsilon_{1,\min} = 0,1$ ergeben. (Vergleiche Abbildung 6.29b und 6.26 für $k = 3$.)

In Abbildung 6.29a sind die Verläufe für den Bias gegenübergestellt, wobei für die bisher verwendete Bedingung die Grenze $\varepsilon_{1,\min} = 5$ und für die Alternative $\sigma_{\hat{\vartheta}_1, m, \max} = \frac{1}{3} \cdot \vartheta_1$ verwendet ist. Letztere ergibt nach Gl. (6.55) mit $|\tilde{\varepsilon}_1| \geq 4,2$ eine ähnliche Bedingung an $\tilde{\varepsilon}_1$, die jedoch nur für $\alpha = 1$ und $k = 2$ tatsächlich vergleichbar ist.

Entsprechend sind die beiden Verläufe für $k = 2$ auch sehr ähnlich. Bei höheren k zeigt sich jedoch, dass der Bias, der sich mit der Varianz-Bedingung $\sigma_{\hat{\vartheta}_1}^2 \leq \sigma_{\hat{\vartheta}_1, m, \max}^2$ ergibt, deutlich geringer wird.

In Abbildung 6.29b ist die Entsprechung zu Abbildung 6.26 für die Bedingung $\sigma_{\hat{\vartheta}_1}^2 \leq \sigma_{\hat{\vartheta}_1, m, \max}^2$ gezeigt.

Der schon bei Abbildung 6.29a beschriebene Effekt, dass der Bias für wachsende k schneller abfällt, lässt sich auch in Abbildung 6.30 erkennen. Die Grenzen $\sigma_{\hat{\vartheta}_1, m, \max}$ sind dort so gewählt, dass diese ungefähr den Grenzen $\varepsilon_{1,\min}$ in Abbildung 6.28 (für $\alpha = 1$ und $k = 2$) entsprechen.



(a) Vergleich von beiden Bedingungen für Schätzung, $\alpha = 1$

(b) Entsprechung zu Abbildung 6.26 ($k = 3$) bei Bedingung über $\sigma_{\hat{\vartheta}_{1,m,\max}}^2$

Abbildung 6.29: Bias bei Verwendung der Schätzbedingung über $\sigma_{\hat{\vartheta}_{1,m,\max}}^2$

Es ist kaum noch ein Unterschied zwischen den drei Kurven mit den hohen Varianzgrenzen zu erkennen. Für die niedrigste Varianzgrenze, die für $k = 2$ einem $\varepsilon_{1,\min}$ von etwa 7 entspricht, ist der Bias für $k = 2$ schon deutlich geringer als bei der Kurve für $\varepsilon_{1,\min} = 5$, und im weiteren Verlauf fällt der Bias auch deutlich schneller ab. Ab $k = 15$ ist kein Unterschied mehr zwischen den Kurven auszumachen.

In Abbildung 6.31a ist der Bias für verschiedene Anfangsabweichungen ε_1 verglichen. Wie eben schon bemerkt, ist der Bias für $\varepsilon_1 = 4$ schon ab $k = 2$ relativ gering. Für größere ε_1 sinkt dieser noch weiter. Für kleinere ε_1 erreicht der Bias aber noch sehr große Beträge. So ist der Bias bei $\varepsilon_1 = 0$ erst mit $k = 10$ auf 10 % (bei $\alpha = 1$) gefallen.

Je kleiner α ist, d. h. je geringer die Anregung, desto höher ist auch der Bias, was aus Abbildung 6.31b ersichtlich ist. Diese Beobachtung ist von Bedeutung, da die Absenkung der Reglerverstärkung für kleine ε_1 auch die Anregung vermindert.

6.6.5 Entwicklung des Bias über mehrere Aufträge

Stabilität des Bias

Wie sich gezeigt hat, hängt der Bias der Schätzung für ϑ_1 unter anderem vom Wert für α ab. Der Wert für α hängt wiederum wesentlich von dem Schätzwert $\hat{\vartheta}_{1,r}$ ab. Damit stellt sich die Frage nach der Stabilität des Bias, d. h. ob ein negativer Bias dazu führt, dass im Weiteren der Bias betragsmäßig noch größere negative Werte annimmt, oder umgekehrt, ob ein positiver Bias zu einer weiteren Vergrößerung des Bias führt.

Dazu wird davon ausgegangen, dass sich der Verlauf von $\Delta \hat{\vartheta}_{1,\text{mean},k} / \vartheta_1 \cdot \alpha$ für $\alpha \rightarrow 0$ und $\alpha \rightarrow \infty$ jeweils einem Grenzverlauf annähert, was die numerischen Berechnungen nahelegen.

Start mit negativem Bias

Startet man mit einem negativen Bias, dann bedeutet dies, dass das tatsächliche $\alpha > 1$ ist (eine „Wunschreglerverstärkung“ von $\hat{\alpha} = 1$ vorausgesetzt). In der Regel ist der Bias des neuen Schätzwertes damit po-

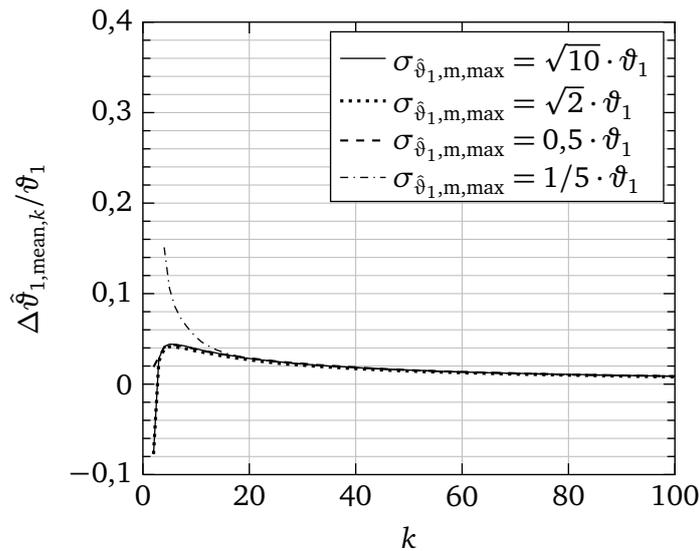


Abbildung 6.30: Verläufe des Bias über der Anzahl der Messungen bei Schätzbedingung über $\sigma_{\hat{\vartheta}_{1,m,\max}}^2$, $\varepsilon_1 = 4$, $\alpha = 1$ (erstellt über jeweils 100 000 Simulationen)

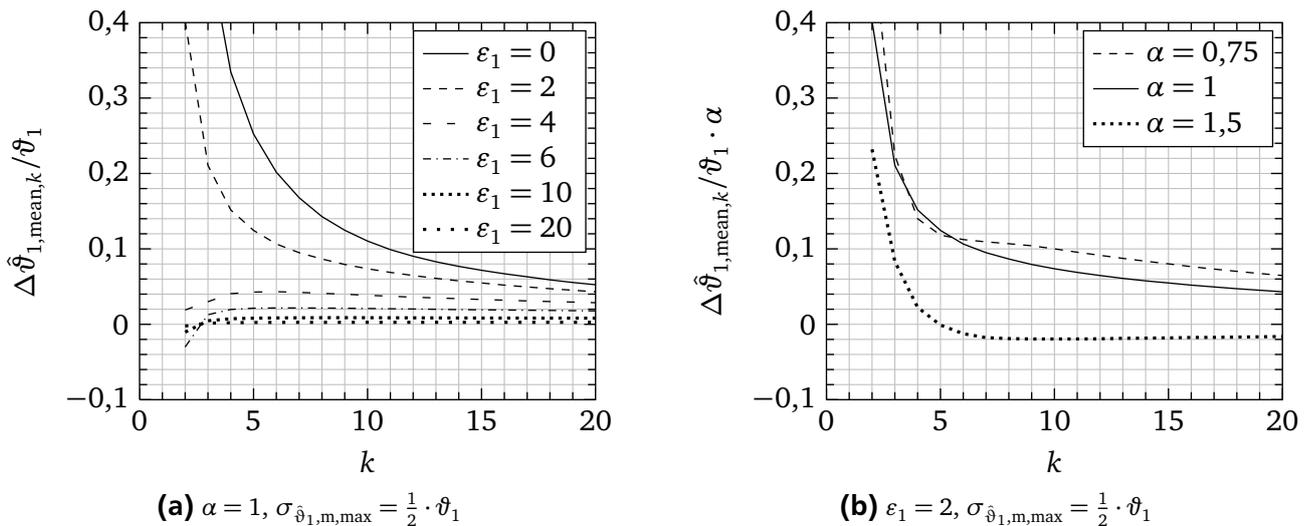


Abbildung 6.31: Verläufe des Bias über der Anzahl der Messungen bei Schätzbedingung über $\sigma_{\hat{\vartheta}_{1,m,\max}}^2$ für verschiedene Werte für ε_1 bzw. α (erstellt über jeweils 100 000 Simulationen)

sitiv (siehe Abbildung 6.26 bzw. 6.29b), d. h. der negative Bias des bisherigen Schätzwertes wird durch die Mittelung mit dem neuen Schätzwert in die richtige Richtung korrigiert.

Teilweise liegt bei $\alpha > 1$ aber auch ein negativer Bias vor. Da der Verlauf $\Delta\hat{\vartheta}_{1,\text{mean},k}/\vartheta_1 \cdot \alpha$ jedoch beschränkt ist, geht $\Delta\hat{\vartheta}_{1,\text{mean},k}/\vartheta_1$ für große α gegen Null. Damit wird ein negativer Bias über viele Aufträge nicht über alle Grenzen verstärkt.

Start mit positivem Bias

Beginnt man mit $\alpha < 1$, so ergibt sich in der Regel ein positiver Bias. Da $\Delta\hat{\vartheta}_{1,\text{mean},k}/\vartheta_1 \cdot \alpha$ für $\alpha \rightarrow 0$ nicht ebenfalls gegen Null strebt, geht der Bias $\Delta\hat{\vartheta}_{1,\text{mean},k}/\vartheta_1$ dabei gegen Unendlich.

Ein bisheriger, zu kleiner Schätzwert kann damit gerade in die falsche Richtung korrigiert werden und wird noch kleiner.

Diskussion

Dies ist natürlich nur eine grobe, qualitative Beschreibung der Entwicklung des Bias. Insbesondere ist damit noch keine mathematisch belastbare Aussage getroffen, ob es einen oder mehrere stabile Werte für den Bias gibt.

Dies liegt an der Komplexität einer mathematischen Beschreibung. Der Bias zum Zeitpunkt $k + 1$ ist im Allgemeinen eine nichtlineare Funktion der anfänglichen Regelabweichung und der Reglerverstärkung,

$$\frac{\Delta \hat{\vartheta}_{1,\text{mean},k+1}}{\hat{\vartheta}_1} = f(\varepsilon_1, \alpha_k),$$

die nicht analytisch gegeben ist, sondern nur numerisch bestimmt werden kann (Abbildung 6.29b). Aber auch wenn diese Funktion als bekannt angesehen wird, ist es nicht einfach möglich, vom Bias im k -ten Schritt auf den im $k + 1$ -ten Schritt zu schließen. Dies liegt daran, dass sowohl ε_1 als auch α_k Realisierungen von Zufallsvariablen sind. Anders als bei linearen Systemen können bei nichtlinearen Systemen die Zufallsvariablen nicht einfach durch die Erwartungswerte ersetzt werden. D. h. auch bei einem festen ε_1 und $\hat{\theta}_{1,r,k} = \hat{\theta}_{1,k}$, d. h. dass immer der aktuelle Schätzwert für θ_1 – ohne Überprüfung der Varianz und ohne Mittelung – verwendet werden würde, was bei $\hat{\alpha}_k = 1$ gleichbedeutend mit $\alpha_k = \theta_1 / \hat{\theta}_{1,k}$ ist, wäre es immer noch nicht einfach möglich, eine nichtlineare Differenzgleichung für die Entwicklung des Bias anzugeben.

Zudem treten „stabilisierende“ Effekte auf, die bisher noch nicht betrachtet wurden. So sind die Anforderungen an die Anregung desto schwieriger zu erfüllen, je kleiner α , d. h. je größer der Bias wird. Bei einem viel zu kleinen α ist eine große Anzahl an Messungen und/oder eine höhere Anfangsabweichung $|\varepsilon_1|$ notwendig, um den Schätzwert überhaupt zur Mittelung verwenden zu können. Wenn dieses notwendige k größer als die Anzahl der Messungen in einem Auftrag wird, findet keine Schätzung statt, und damit verschlechtert sich der Bias auch nicht mehr.

Daher wird die Entwicklung des Bias über viele Aufträge im Folgenden simulativ betrachtet.

Entwicklung des Bias über viele Aufträge

Nach den bisherigen Beobachtungen sind die wesentlichen Einflussfaktoren

- die Anzahl der Messungen k_{job} je Auftrag,
- die Anregung, welche hier über ein vorgebbares ε_1 simuliert wird,
- die Reglerverstärkung α sowie
- die Varianzgrenze $\sigma_{\hat{\vartheta}_{1,m},\text{max}}^2$.

Zunächst wird anhand von sehr ungünstigen Parametern die Stabilität des Bias betrachtet. Im Anschluss wird der Bias in Abhängigkeit der genannten Einflussgrößen verglichen.

Es werden dazu immer 250 Aufträge simuliert, wobei bei allen diesen Aufträgen die gleichen Parameter verwendet werden. In den gezeigten Plots werden immer die Verteilungen berücksichtigt, die am Ende jedes Auftrags vorliegen. Entsprechend ist an der Abszisse die Unabhängige l aufgetragen.

Stabilität

Um die Stabilität des Bias simulativ zu untersuchen, wird die Identifikation und Regelung mit sehr geringer Anregung simuliert, da dies nach den bisherigen Betrachtungen zu sehr hohen Bias-Werten führt.

D. h. der Anfangswert $u_{1,1}$ ist so gewählt, dass die Sollgröße im Mittel schon erreicht ist, $\varepsilon_1 = 0$. Damit sind keine größeren Korrekturen über u_1 nötig.

Die Anzahl der Messungen je Auftrag ist möglichst gering gewählt. Jeder Auftrag ist mit $k_{\text{Job}} = 3$ Messungen simuliert, da bei $k_{\text{Job}} = 2$ bei der bisher verwendeten Grenze $\sigma_{\hat{\vartheta}_{1,m,\max}} = 0,5 \cdot \vartheta_1$ keine nennenswerte Anzahl an „gültigen“ Schätzwerten entsteht.

Auf die Absenkung der Reglerverstärkung für kleine gemessene Regelabweichungen ist verzichtet, da dies dazu führen würde, dass die Anzahl der Regelschritte je Auftrag erhöht werden müsste, um genügend Anregung für die Schätzung zu erzeugen.

Für die Plots in Abbildung 6.32 sind jeweils 250 Aufträge simuliert. Die Einzelschätzwerte für ϑ_1 werden mittels einfacher Mittelung mit Vergessensfaktor gemittelt, wobei ein λ von 0,9 gewählt ist. Bei dieser Wahl ist berücksichtigt, dass zum einen der Verlauf des Schätzwertes aufgrund der biasbehafteten Schätzung der nahe zurückliegenden Werte betrachtet werden soll, und zum anderen, dass die Mittelung der Einzelwerte mit Varianzen nahe der gegebenen Grenze $\sigma_{\hat{\vartheta}_{1,m,\max}}^2$ auch einen für die Regelung gültigen Schätzwert ergibt. Aus diesem Grund ist hier die Varianzgrenze für die Regelung an die Grenze $\sigma_{\hat{\vartheta}_{1,m,\max}}$ angepasst. Neben dem bisher verwendeten Standardpaar $\sigma_{\hat{\vartheta}_{1,r,\max}} = 0,2 \cdot \vartheta_1$ und $\sigma_{\hat{\vartheta}_{1,m,\max}} = 0,5 \cdot \vartheta_1$ werden hier auch die Werte $\sigma_{\hat{\vartheta}_{1,r,\max}} = 0,4 \cdot \vartheta_1$ und $\sigma_{\hat{\vartheta}_{1,m,\max}} = 1 \cdot \vartheta_1$ verwendet. Damit erhält man in beiden Fällen ein Verhältnis von $\sigma_{\hat{\vartheta}_{1,m,\max}} / \sigma_{\hat{\vartheta}_{1,r,\max}} = 2,5$, welches bei $\lambda = 0,9$ mit zwischen fünf und zehn Mittelungen einen zur Regelung gültigen Schätzwert erzeugt. (Siehe dazu auch Abbildung 6.12 auf Seite 131.)

In Abbildung 6.32a wurde $\sigma_{\hat{\vartheta}_{1,m,\max}} = 0,5 \cdot \vartheta_1$ gesetzt. Es dauert teilweise bis zum 180. Auftrag, bis genügend Schätzwerte gemittelt wurden, um den Mittelwert zur Regelung zu verwenden. Dabei stellt sich ein Bias von ca. 0,25 bzw. 50 % ein.

Für die höhere, also weniger strenge Grenze $\sigma_{\hat{\vartheta}_{1,m,\max}} = \vartheta_1$ mit $\sigma_{\hat{\vartheta}_{1,r,\max}} = 0,4 \cdot \vartheta_1$ (Abbildung 6.32b) werden mehr Schätzwerte für die Mittelung akzeptiert, so dass ab ungefähr dem 15. Auftrag in den meisten Simulationen ein gemittelter Wert $\hat{\vartheta}_{1,r}$ zur Regelung zur Verfügung steht. Der Bias steigt bis zum 250. Auftrag auf einen Wert von knapp über 0,5 (100 %) an, wobei der Anstieg mit steigender Auftragszahl abnimmt.

In der Simulation zu Abbildung 6.32c ist ebenfalls $\sigma_{\hat{\vartheta}_{1,m,\max}} = \vartheta_1$ gesetzt, im Unterschied zu dem Beispiel vorher wird aber für die Regelung der Startwert $\hat{\vartheta}_{1,r} = 2,3 \cdot \vartheta_1 = 1,15$ verwendet, was einem α von 0,43 entspricht. Dadurch ist die Anregung noch weiter reduziert, so dass sich über die 250 Aufträge nur wenige zur Regelung gültige Schätzwerte ergeben. Diese nehmen im Mittel jedoch Werte von unter 1,15 an, was einem Bias von unter 0,65 entspricht. Dies legt nahe, dass der Bias auch bei einer weiteren Erhöhung der Auftragszahl nicht unbegrenzt steigt, sondern einen Endwert zwischen 0,5 und 0,65 annehmen wird.

Diese Simulation unterstützt damit die Annahme, dass die Varianz-Bedingung bei der Identifikation im adaptiven Regelkreis zu einem stabilen Bias führt.

Einflussfaktoren

Im Weiteren wird, etwas kompakter, nur noch der Mittelwert der Schätzwerte $\hat{\vartheta}_{1,r}$ gezeigt, und nicht mehr die Histogramme der zugrundeliegenden Verteilungen wie in Abbildung 6.32. Genauer wird der Mittelwert der Schätzwerte $\hat{\vartheta}_{1,r}$ noch in einen relativen Bias

$$\frac{\vartheta_{1,\text{mean},k}^*}{\vartheta_1} = \frac{E\{\hat{\vartheta}_{1,r,k}\}_r - \vartheta_1}{\vartheta_1}$$

umgerechnet. Dabei ist diese Größe mit einem „*“ gekennzeichnet, da diese nicht den Bias im bisherigen Sinne beschreibt, da zum einen nicht einzelne, sondern gemittelte Schätzwerte betrachtet werden,

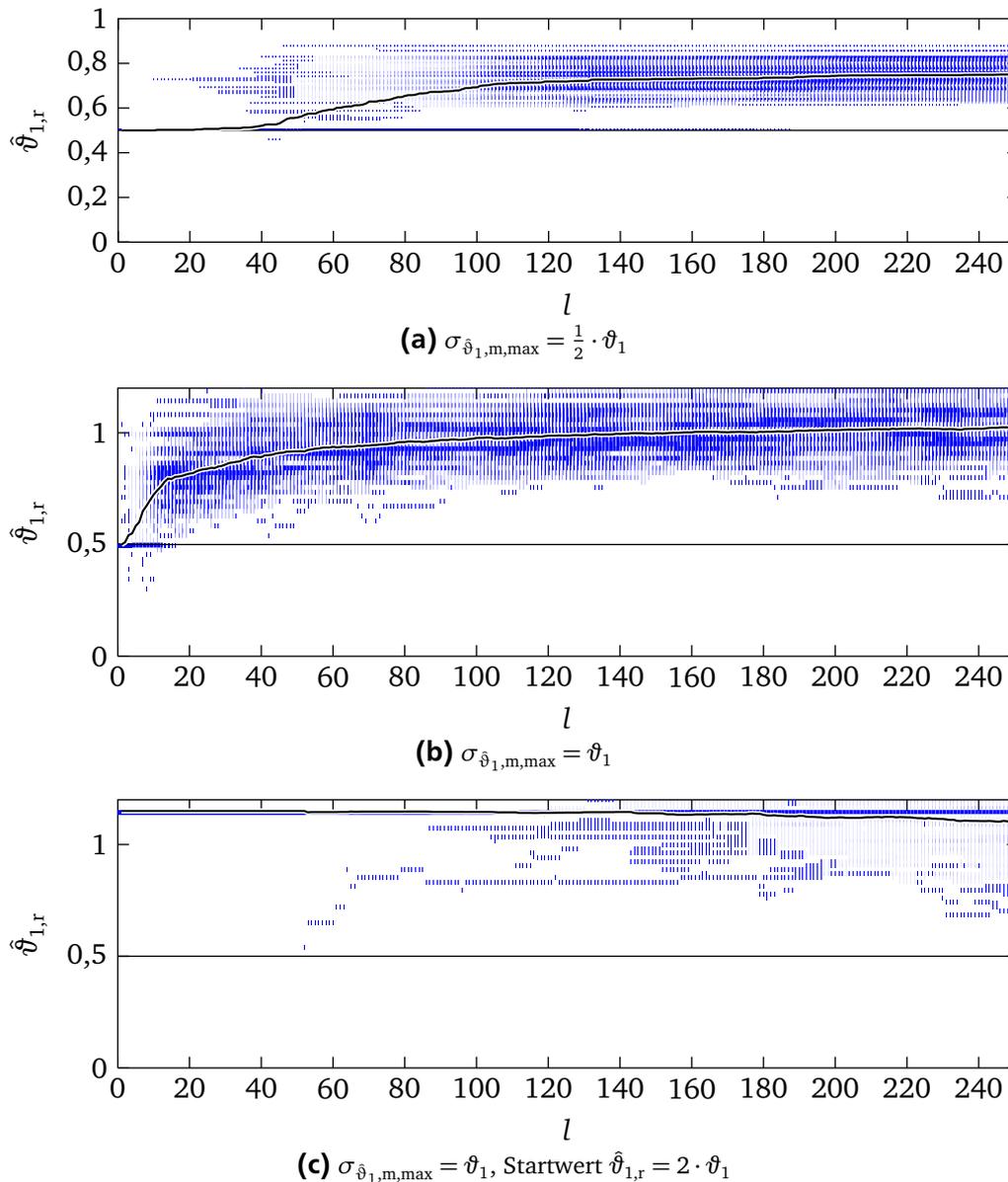


Abbildung 6.32: Schätzwert $\hat{\vartheta}_{1,r}$ über 250 Aufträge mit schwacher Anregung, jeweils der letzte Wert eines Auftrags ist angegeben (je 100 Simulationen)

und zum anderen auch die von außen vorgegebenen Anfangswerte für $\hat{\vartheta}_{1,r}$ in die Mittelung einbezogen werden.

Betrachtet man Abbildung 6.32b, so kann man grob zwei Phasen ausmachen. In der ersten Phase findet zunächst ein langsam zunehmender, dann relativ steiler Anstieg des „Bias“ $\vartheta_{1,\text{mean},k}^*$ statt, der sich dadurch erklärt, dass zunächst die Simulationen den Mittelwert von $\{\hat{\vartheta}_{1,r,k}\}_r$ dominieren, in denen noch der Anfangswert für $\hat{\vartheta}_{1,r}$ gültig ist. Daran schließt (in dem Beispiel bei etwa $k = 15$) die zweite Phase an, die die Entwicklung des Bias zeigt, der durch diesen selber hervorgerufen wird. Mit steigender Anzahl der Aufträge geht $\vartheta_{1,\text{mean},k}^*$ (näherungsweise) gegen einen Endwert.

Diese Phasen lassen sich relativ gut unterscheiden und so enthalten die für $\frac{\vartheta_{1,\text{mean},k}^*}{\vartheta_1}$ erstellten Graphen zwei wesentliche Informationen. Zum einen eine Aussage darüber, wie knapp die Varianzen der Einzelschätzwerte an der Varianzgrenze liegen und damit auch, ob überhaupt eine wesentliche Anzahl an

gültigen Einzelschätzwerten auftritt, und zum anderen natürlich eine Aussage über den sich letztlich einstellenden Bias.

In Abbildung 6.33a und 6.33b ist der beschriebene „Bias“ für verschiedene k_{Job} dargestellt, wobei bei 6.33a wie bei den Simulationen in Abbildung 6.32 nur eine minimale Anregung simuliert wird ($\varepsilon_1 = 0$) und bei 6.33b mit $\varepsilon_1 = 3$ eine mittlere Anregung vorliegt. Dabei liegt die Varianzgrenze in beiden Fällen bei $\sigma_{\hat{\vartheta}_{1,m,\max}}^2 = 0,5^2 \cdot \vartheta_1^2$.

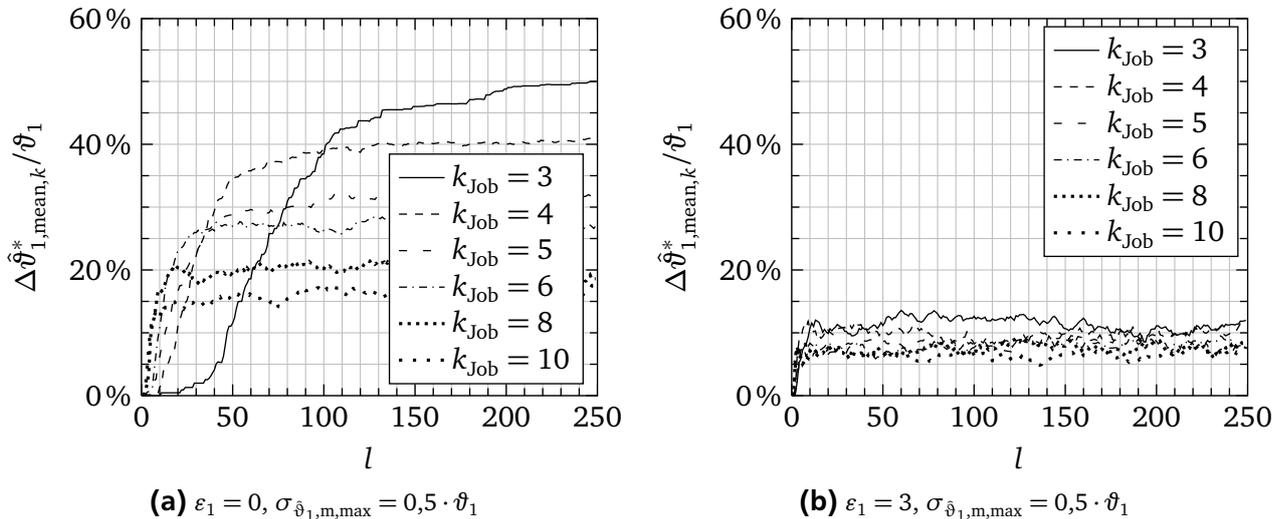


Abbildung 6.33: Entwicklung des Bias für $\hat{\vartheta}_{1,r}$ über 250 Aufträge bei verschiedenen Anregungen über ε_1 und bei verschiedenen Anzahlen k_{Job} an Messungen je Auftrag (je 100 Simulationen)

Mit steigender Anzahl an Messungen je Auftrag sinken die Varianzen der Einzelschätzwerte natürlich, was dazu führt, dass der Anteil an zulässigen Schätzwerten zunimmt und damit in Abbildung 6.33a die Kurven für den „Bias“ in Phase 1 schneller steigen. Daneben sinkt auch der sich einstellende Bias deutlich. So beträgt er bei $k_{\text{Job}} = 3$ nach dem 250-ten Auftrag 50 %, während er bei zehn Messungen je Auftrag knapp unter 20 % liegt.

Anhand Abbildung 6.33b ist ersichtlich, dass eine höhere Anregung über ε_1 die Varianz effektiv verringern kann. Auch zeigt sich, dass die Abhängigkeit des Bias von der Anzahl der Messungen je Auftrag mit einer steigenden anfänglichen Regelabweichung ε_1 stark zurückgeht. Hier liegen die erreichten Werte für den Bias innerhalb eines 5 %-Bandes.

Auch kann man an den Graphen in Abbildung 6.33 erkennen, dass der Bias umso schneller den Endwert erreicht, je höher die Anregung des Schätzers über k_{Job} bzw. ε_1 ist.

In Abbildung 6.34 ist der Einfluss der Varianzgrenze $\sigma_{\hat{\vartheta}_{1,m,\max}}^2$ verdeutlicht. Dazu sind drei verschiedene Werte für $\sigma_{\hat{\vartheta}_{1,m,\max}}$ verglichen. Die drei Plots unterscheiden sich wieder in der Anregung über ε_1 .

Bei minimaler Anregung, $\varepsilon_1 = 0$, ergibt $\sigma_{\hat{\vartheta}_{1,m,\max}} = 0,5 \cdot \vartheta_1$ einen Bias von 40 % (siehe auch Abbildung 6.33a). Erhöht man die Grenze auf $\sigma_{\hat{\vartheta}_{1,m,\max}} = \vartheta_1$, so werden deutlich mehr Einzelschätzwerte für die Mittelung berücksichtigt, was anhand des steileren Anstiegs des „Bias“ zu Beginn in Phase 1 erkennbar ist. Bei der schwachen Anregung dominiert der hohe Bias aufgrund der Identifikation im geschlossenen Regelkreis, so dass sich hier letztlich ein Bias von fast 80 % einstellt. Für die niedrige Varianzgrenze $\sigma_{\hat{\vartheta}_{1,m,\max}} = 0,25 \cdot \vartheta_1$ wurde kein einziger Einzelschätzwert für die Mittelung berücksichtigt.

In Abbildung 6.34b ist die Anregung mit $\varepsilon_1 = 4$ schon deutlich höher, was sich – wie schon oben gezeigt – in einem deutlich niedrigeren Bias niederschlägt. Es fällt hier auf, dass der Bias für die kleinste Grenze $\sigma_{\hat{\vartheta}_{1,m,\max}} = 0,25 \cdot \vartheta_1$ etwas über dem Bias liegt, der sich mit den anderen beiden Werten einstellt. Mit $\sigma_{\hat{\vartheta}_{1,m,\max}} = 0,5 \cdot \vartheta_1$ und $\sigma_{\hat{\vartheta}_{1,m,\max}} = \vartheta_1$ ergeben sich hier quasi die gleichen Werte, da die

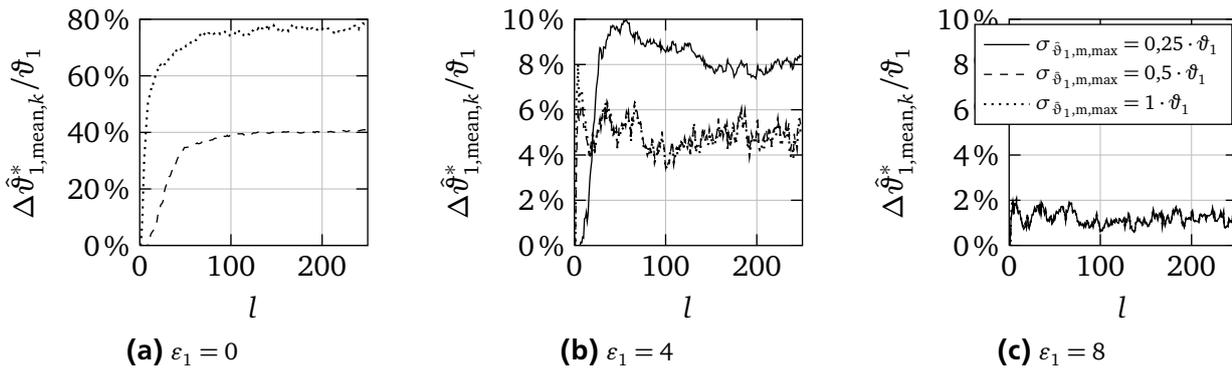


Abbildung 6.34: Entwicklung des Bias für $\hat{\vartheta}_{1,r}$ über 250 Aufträge bei verschiedenen Anregungen über ε_1 und bei verschiedenen Grenzen $\sigma_{\hat{\vartheta}_{1,m,max}}$, $k_{Job} = 4$ (je 100 Simulationen)

Anregung so groß ist, dass in beiden Fällen so gut wie jeder Einzelschätzwert berücksichtigt werden kann. Bei $\sigma_{\hat{\vartheta}_{1,m,max}} = 0,25 \cdot \vartheta_1$ hingegen werden nicht alle Einzelschätzwerte berücksichtigt. Der – im Vergleich zu den beiden anderen Kurven – zusätzliche Bias von ca. 1,5 % ist damit auf das Beschneiden der Verteilung $\{\hat{\vartheta}_1\}_r$ durch diese Bedingung zurückzuführen, während der Bias von etwa 5 % aufgrund der Identifikation im geschlossenen Regelkreis entsteht.

Erhöht man die Anregung weiter, so wird irgendwann der Punkt erreicht, an dem mit allen drei Varianzgrenzen das gleiche Ergebnis erzielt wird. So erhält man bei $\varepsilon_1 = 8$ in allen Fällen einen Bias von ca. 1,5 %, der durch die Identifikation im geschlossenen Regelkreis hervorgerufen wird.

Zusätzliche Bedingung über $\varepsilon_{1,min}$

Der sich einstellende Bias hängt also nicht nur von der Grenze $\sigma_{\hat{\vartheta}_{1,m,max}}$ ab, sondern auch signifikant von der Anregung, die im Wesentlichen durch die mittlere Regelabweichung ε_1 zu Beginn jedes Auftrags gegeben ist. Es ist allerdings nicht möglich, in dieser Hinsicht schlechte Aufträge von der Schätzung auszunehmen, indem die Bedingung $|\tilde{\varepsilon}_1| > \varepsilon_{1,min}$ *zusätzlich* überprüft wird.

Um dies zu verdeutlichen, sind in Abbildung 6.35 die Verläufe des Bias mit und ohne Berücksichtigung dieser zusätzlichen Bedingung für verschiedene Anregungen gegenübergestellt.

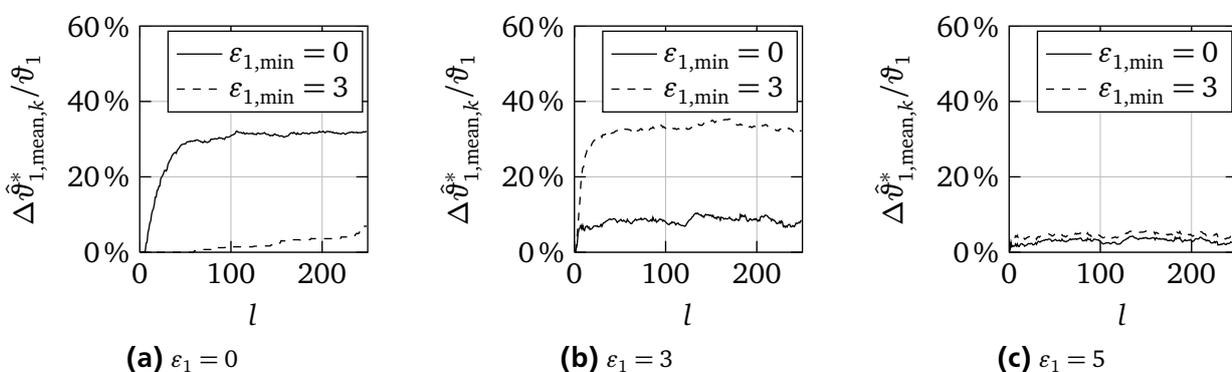


Abbildung 6.35: Entwicklung des Bias für $\hat{\vartheta}_{1,r}$ über 250 Aufträge bei verschiedenen Anregungen über ε_1 und bei verschiedenen zusätzlichen Bedingungen $|\tilde{\varepsilon}_1| > \varepsilon_{1,min}$, $\sigma_{\hat{\vartheta}_{1,m,max}} = 0,5 \cdot \vartheta_1$, $k_{Job} = 5$ (je 100 Simulationen)

In Abbildung 6.35a ist der Fall minimaler Anregung, d. h. $\varepsilon_1 = 0$, dargestellt. Der Bias, der sich unter Berücksichtigung von $|\tilde{\varepsilon}_1| > 3$ einstellt wird deutlich von dem gegebenen Anfangswert des Parameters dominiert.¹⁹ D. h. die Schätzung wird effektiv verhindert, was auch der Zweck dieser Bedingung ist.

Erhöht man die Anregung jedoch auf $\varepsilon_1 = 3$, so ergibt sich der in Abbildung 6.35b gezeigte Verlauf. Der Bias, der unter Berücksichtigung von $|\tilde{\varepsilon}_1| > 3$ bestimmt wird, ist deutlich höher als derjenige, der sich ohne diese Bedingung ergibt. Dies liegt daran, dass hier eine asymmetrische Beschneidung der Verteilung $\{\hat{\vartheta}_1\}_r$ vorgenommen wird.

Für höhere Anregungen nähert sich der Bias mit $|\tilde{\varepsilon}_1| > 3$ wieder dem Bias an, der sich ohne die zusätzliche Bedingung ergibt, wobei die Höhe des Bias insgesamt sinkt, siehe Abbildung 6.35c.

Fazit

Laut allen durchgeführten Simulationen ist der Bias stabil. Für kleine Anregungen und hohe Varianzgrenzen kann dieser jedoch große Werte annehmen. Es ist dabei aber keinesfalls empfehlenswert, die Anregung über die Bewertung der anfänglichen Regelabweichung zusätzlich zu überprüfen, da über eine solche Bedingung der Bias im Grenzbereich der zulässigen Anregung deutlich steigt.

D. h. die Grenze $\sigma_{\hat{\vartheta}_1, m, \max}$ muss entsprechend gewählt werden, so dass der auftretende Bias akzeptiert werden kann.

Wichtig ist auch zu bedenken, dass der Bias einzig durch die Grenze $\sigma_{\hat{\vartheta}_1, m, \max}$ und nicht $\sigma_{\hat{\vartheta}_1, r, \max}$ bestimmt ist. Dies liegt daran, dass die Mittelung der Einzelschätzwerte nur die Varianz, aber gerade nicht den Bias reduziert.

Wird die Absenkung der Reglerverstärkung bei kleinen Regelabweichungen verwendet, so sinkt die Anregung bei kleinen ε_1 nochmals deutlich, was den Bias entsprechend weiter erhöht.

Hier wurden immer gleichartige Aufträge simuliert. Tatsächlich werden aber die Anfangsabweichungen und Anzahl an Messungen variieren. Dann stellt sich ein entsprechend gemittelter Bias ein. Damit ist es aus Sicht der Biasvermeidung auch günstiger, eine eher varianz-optimale Mittelung zu verwenden. Denn dann werden Einzelschätzwerte mit guter Anregung und damit niedrigem Bias bevorzugt.

6.6.6 Kompensation

Bei der Behandlung des Bias stellt sich natürlich die Frage, ob bzw. in welchem Rahmen der hier beschriebene Bias in den Schätzwerten $\hat{\vartheta}_1$ korrigiert werden kann.

Durchführung der Bias-Kompensation

Als Einflussfaktoren auf den Bias wurden die Anzahl der Messungen k , die anfängliche Regelabweichung ε_1 , die Reglerverstärkung α sowie die Varianzgrenze $\sigma_{\hat{\vartheta}_1, m, \max}^2$ bestimmt. Dabei sind nur der erst- und der letztgenannte Faktor bekannt. Die anfängliche Regelabweichung ε_1 ist nicht messbar. Bekannt ist nur $\tilde{\varepsilon}_1$, das als Näherung für ε_1 verwendet werden kann. Für α wird der Wert eins angenommen. Damit wird in Kauf genommen, dass für $\alpha > 1$ die Kompensation zu hoch, und für $\alpha < 1$ zu niedrig ausfallen wird.

Zur Bias-Kompensation wird für $\alpha = 1$, das verwendete $\sigma_{\hat{\vartheta}_1, m, \max}$ (hier $\sigma_{\hat{\vartheta}_1, m, \max} = 0,5 \cdot \vartheta_1$) und eine gegebene Menge an Werten für ε_1 zwischen 0 und 10 der Bias in Abhängigkeit von k bestimmt, wie er auch in Abbildung 6.31a aufgetragen ist. Somit liegt ein zweidimensionales Kennfeld $\kappa_I \cdot \Delta \hat{\vartheta}_{1, \text{mean}}(k, \varepsilon_1)$ vor.

¹⁹ Würde man hier den Bias nur der Simulationen anzeigen, bei denen auch ein gültiger Schätzwert ermittelt wurde – also den Bias im engeren Sinne – wäre dieser deutlich höher.

Mit der Messung der ersten Regelabweichung $\tilde{\varepsilon}_1$ eines Auftrags wird aus diesem Kennfeld eine Kennlinie $\kappa_1 \cdot \Delta \hat{\vartheta}_{1,\text{mean}}(k)$ bestimmt, indem für ε_1 der gemessene Wert $\tilde{\varepsilon}_1$ eingesetzt wird.

Der aktuelle, theoretische Wert des Bias wird dann bestimmt, indem $\kappa_1 = \frac{1}{\hat{\vartheta}_{1,r}}$ gesetzt wird.²⁰ Der so bestimmte Wert wird von dem aktuellen Schätzwert abgezogen.

Wirkung der Bias-Kompensation für einen Auftrag

In Abbildung 6.36 ist der Bias für $\varepsilon_1 = 0$ und $\varepsilon_1 = 3$ über k aufgetragen, wobei $\alpha = 1$ und $\sigma_{\hat{\vartheta}_{1,m,\text{max}}} = 0,5 \cdot \vartheta_1$ gewählt sind. Für jeden der beiden Werte für ε_1 ist der Bias einmal ohne Kompensation, und einmal mit der beschriebenen Kompensation aufgetragen.

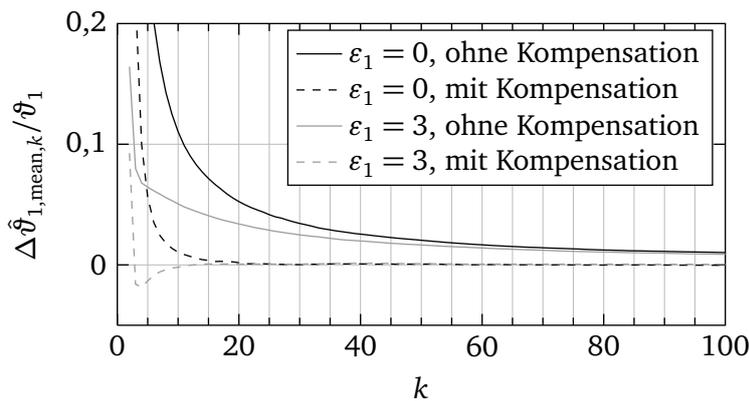


Abbildung 6.36: Verläufe des Bias über Anzahl der Messungen mit Bias-Kompensation $\alpha = 1$, $\sigma_{\hat{\vartheta}_{1,m,\text{max}}} = 0,5 \cdot \vartheta_1$ (erstellt über jeweils 100 000 Simulationen)

Es zeigt sich, dass in beiden Fällen ($\varepsilon_1 = 0$ und $\varepsilon_1 = 3$) der Bias deutlich verringert und für $k > 15$ näherungsweise vollständig kompensiert wird.

Für kleinere k ist die Kompensation jedoch nicht exakt, sondern sie ist unvollständig ($\varepsilon_1 = 0$) bzw. es wird leicht überkompensiert ($\varepsilon_1 = 3$). Dies ist darauf zurückzuführen, dass in diesem Bereich die Abhängigkeit des Bias von ε_1 noch relativ groß ist (siehe Abbildung 6.29a). Damit ist die Näherung $\tilde{\varepsilon}_1 \approx \varepsilon_1$ in diesem Bereich kritischer. Bei großen k ist der Bias über ε_1 relativ konstant, so dass die Unsicherheit in ε_1 keine Auswirkung mehr hat.

Wirkung der Bias-Kompensation über viele Aufträge

In Abbildung 6.37 ist die Identifikation ohne und mit Bias-Kompensation über 250 Aufträge verglichen. Dabei wurden immer $k_{\text{Job}} = 5$ Messungen je Auftrag simuliert, bei drei verschiedenen Werten für ε_1 .

Bei der minimalen Anregung $\varepsilon_1 = 0$ reduziert die Kompensation den Bias um zwei Drittel von ca. 30 % auf 10 %. Bei $\varepsilon_1 = 3$ ergibt sich ohne die Kompensation ein Bias von ungefähr 8 %, der mit der Bias-Kompensation leicht überkompensiert wird, so dass man einen Bias im Bereich von 0 bis -3 % erhält. Im letzten Beispiel mit $\varepsilon_1 = 5$ wird der Bias, der unkompensiert um 3 % liegt, auf 0 bis -1 % kompensiert.

²⁰ Führt man die Regelung mit einem anderen, konstanten Faktor $\hat{\alpha}$ durch, z.B. 0,95, so kann auch dieser verwendet werden. Jedoch darf keinesfalls bei Anwendung der Absenkung der Reglerverstärkung bei kleinen Regelabweichungen das im aktuellen Schritt verwendete $\hat{\alpha}_k$ verwendet werden, da dies bei dem hier beschriebenen Vorgehen zu viel zu hohen Biaswerten führen würde. Durch das Ignorieren der Absenkung der Reglerverstärkung wird der Bias durch das Kennfeld unterschätzt, so dass die Kompensation unvollständig ist. Dies muss in Kauf genommen werden.

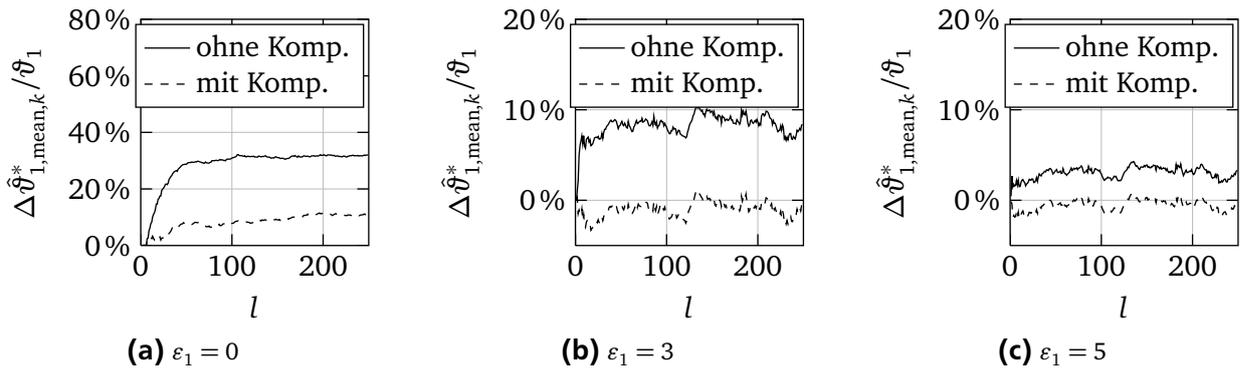


Abbildung 6.37: Entwicklung des Bias für $\hat{\vartheta}_{1,r}$ über 250 Aufträge bei verschiedenen Anregungen über ε_1 und Bias-Kompensation, $\sigma_{\hat{\vartheta}_{1,m},\max} = 0,5 \cdot \vartheta_1$, $k_{\text{Job}} = 5$ (je 100 Simulationen)

Wirkung der Bias-Kompensation bei Standard-Testfall

In Abbildung 6.38 ist die Wirkung der Kompensation im Standardtestfall aus Abbildung 6.15 gezeigt. Dort zeigt sich bei den Simulationen ohne Kompensation insbesondere Ende des zweiten und während des dritten Auftrags ein Bias, der mit der Bias-Kompensation deutlich geringer ausfällt. Auch ist der kleinere Bias, der sich ursprünglich in den letzten Aufträgen gezeigt hat ebenfalls verringert.

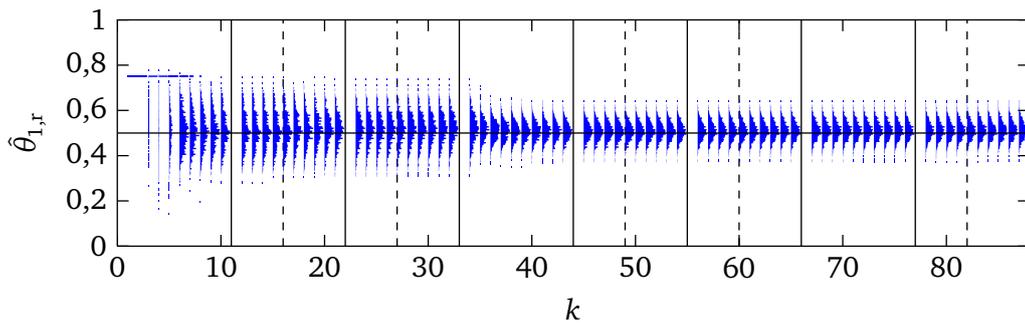


Abbildung 6.38: Verteilung von $\{\hat{\vartheta}_{1,r,k}\}_r$ mit Bias-Kompensation bei ansonsten gleichen Parametern wie bei Abbildung 6.15 (1 000 Simulationen)

Damit ist beispielhaft gezeigt, dass die Kompensation auch bei der Absenkung der Reglerverstärkung und bei Änderungen der sekundären Eingangsgrößen wirkungsvoll ist.

Allgemein sollte aber überprüft werden, ob die Regelabweichung nach einer Änderung der sekundären Eingangsgröße nicht signifikant gestiegen ist. Falls dies der Fall ist, erhöht dies die Anregung und es sollte auf die Kompensation verzichtet werden, da die tabellierten Bias-Werte nicht mehr gültig sind.

6.7 Erweiterungen des Schätzverfahrens

Das bisher beschriebene Verfahren arbeitet zufriedenstellend, solange

- die Anfangsschätzwerte der Parameter nicht übermäßig weit von den realen Werten entfernt liegen,
- das Prozessrauschen nicht deutlich über dem angenommenen Wert liegt, sowie
- keine Sprünge von θ_0 oder einzelne Ausreißer während eines Auftrags auftreten.

Sind diese Voraussetzungen nicht erfüllt, so kann ein aufschwingendes oder kriechendes Verhalten auftreten, oder es können zu schlechte Schätzwerte für die Regelung verwendet werden, was zu einem inakzeptablen Regelverhalten führt.

Dies wird im Folgenden kurz beschrieben und im Anschluss werden Erweiterungen des Schätzverfahrens angegeben, die diesen Problemen entgegenwirken.

Aufschwingendes Verhalten und Kriechen

Bei den in Abbildung 6.39a bzw. 6.39b gezeigten Beispielen ist $\hat{\theta}_{1,r}$ zu Beginn nur $0,45 \cdot \theta_1$. Wenn mit der Voreinstellung der Sollwert gut getroffen wurde, dann fällt dies zunächst nicht auf. Wenn sich aber die Regelabweichung langsam erhöht, so sind die Korrekturschritte des Reglers zu groß, was zu einem aufschwingenden Verhalten führt. Dieses dauert so lange an, bis die Variation von u_1 groß genug geworden ist, um einen neuen Schätzwert zur Regelung zu erhalten. Somit würde sich letztlich zwar ein stabiles Verhalten ergeben, allerdings kann dieses Verhalten in der Praxis schon nicht mehr akzeptabel sein. Besonders problematisch wird es, wenn es sich nicht um den ersten Auftrag für einen neuen Parametersatz handelt, sondern schon ein (für den aktuellen Auftrag falscher) Schätzwert existiert, der in die Mittelung mit einbezogen wird. Dies könnte auftreten, wenn ein falscher Parametersatz ausgewählt worden ist.

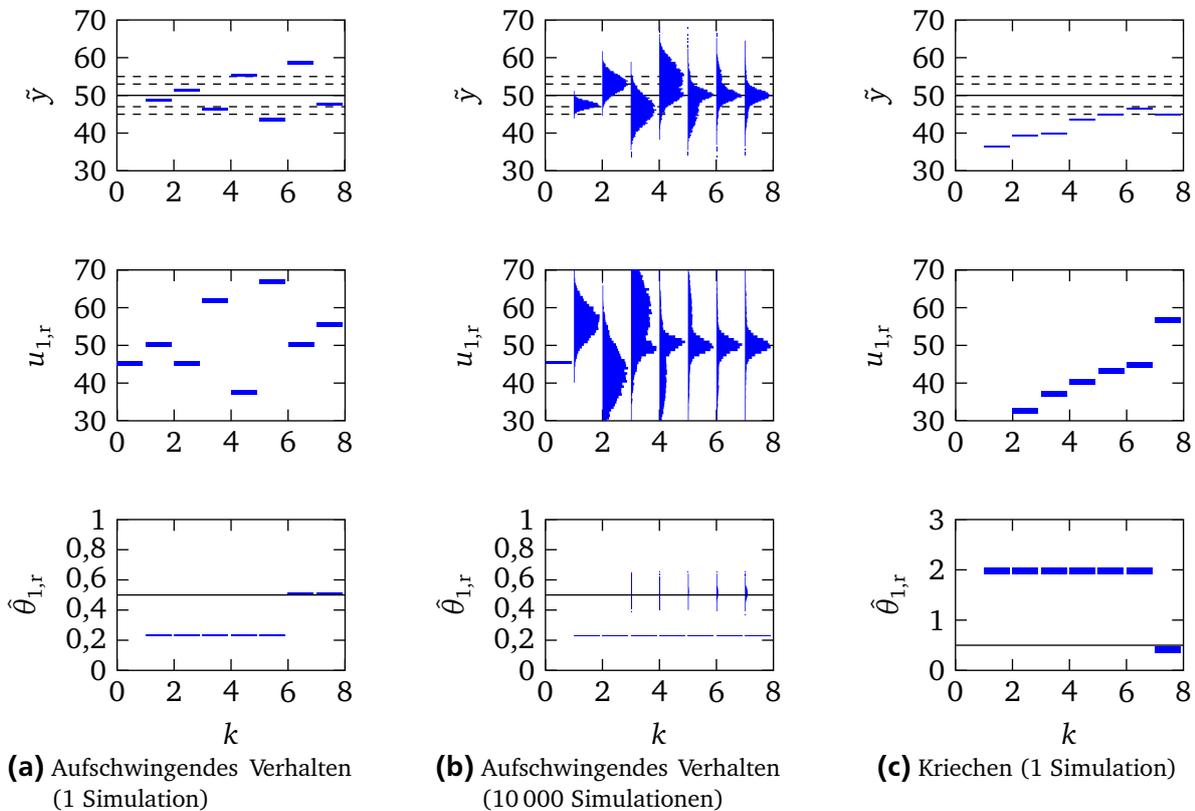


Abbildung 6.39: Beispiele für aufschwingendes und kriechendes Verhalten

In Abbildung 6.39b ist auch zu erkennen, dass selbst nach dem siebten Regelschritt bei weitem noch nicht in allen Simulationen der Parameterwert korrigiert wurde. Dies verdeutlicht, dass es bei kleinen anfänglichen Regelabweichungen lange dauern kann, bis sich das instabile Verhalten des geregelten Systems zeigt.

In Abbildung 6.39c ist das gegenteilige Verhalten an einem Beispiel gezeigt. Hier ist der Anfangswert von $\hat{\theta}_{1,r}$ deutlich zu hoch angesetzt, mit der Folge, dass die Korrekturen des Reglers zu schwach ausfallen. Dadurch bleibt aber auch die Variation der Eingangsgröße u_1 zunächst zu gering, so dass der für den Bediener offensichtlich falsche Wert erst nach einigen Schritten korrigiert wird.

In Abschnitt 6.7.1 wird ein Verfahren vorgestellt, welches die beschriebenen Probleme reduziert.

Ausreißer und hohes Messrauschen

Bei zu hohem Messrauschen stimmen die Varianzgrenzen nicht mehr und es werden zu schlechte Parameterschätzungen für die Regelung akzeptiert.

Ausreißer und Sprünge des Parameters θ_0 während eines Auftrags können zu grob falschen Schätzwerten der Parameter führen, die nicht durch einen hohen Wert für die Varianz erkannt werden können. Ein Beispiel wird weiter unten in Abbildung 6.44a auf Seite 172 gegeben.

Eine einfache Möglichkeit, solche Fehler abzufangen, ist die Berücksichtigung der geschätzten Varianz des Messrauschens. Dies wird in Abschnitt 6.7.2 besprochen. Eine weitere Verbesserung kann durch die gezielte Prüfung auf Ausreißer oder Sprünge erreicht werden. Dies ist in dem angegebenen Abschnitt jedoch nur kurz skizziert.

Auch wenn die hier erwähnten und im Folgenden vorgestellten Methoden die Ergebnisse effizient verbessern, können immer noch vereinzelt Fälle auftreten, in denen kein zufriedenstellendes Verhalten vorliegt. Dies wird in Abschnitt 6.7.3 diskutiert.

6.7.1 Test des Schätzwertes $\hat{\theta}_{1,r}$ auf Plausibilität

Eine Möglichkeit zum Verhindern des unter den Stichworten „Aufschwingendes Verhalten“ und „Kriechen“ beschriebenen Verhaltens, die in diesem Abschnitt behandelt wird, besteht darin, die Aussage „Es mag sein, dass es mit den gegebenen Messwerten nicht möglich ist, den Parameter θ_1 ausreichend sicher zu schätzen. Aber der aktuell verwendete Wert von $\hat{\theta}_{1,r}$ stimmt mit Sicherheit nicht!“, die jeder Betrachter der Versuchsdaten als richtig erkennen würde, mathematisch auszudrücken. Zudem muss dann der Wert für $\hat{\theta}_{1,r}$ entsprechend korrigiert werden.

Um zu prüfen, wie der zur Regelung verwendete Schätzwert $\hat{\theta}_{1,r}$ und der aus den aktuellen Messungen geschätzte Wert $\hat{\theta}_1$ zusammenpassen, wird angenommen, dass der aktuell bestimmte Schätzwert $\hat{\theta}_1$ normalverteilt mit der Varianz $\sigma_{\hat{\theta}_1}^2$ ist,

$$\{\hat{\theta}_1\}_r \sim N(\mu, \sigma_{\hat{\theta}_1}^2).$$

Es wird nun die Hypothese

$$H_0: \mu = \hat{\theta}_{1,r}$$

geprüft. Es handelt sich hierbei um einen Test des Mittelwerts einer Normalverteilung bei bekannter Streuung. Dieser ist beispielsweise in [STORM, 2001, S. 174ff] beschrieben und diskutiert.²¹

Die Wahrscheinlichkeit p_1 dafür, dass ein Schätzwert $\{x\}_r^\omega$ für θ_1 einen Wert annimmt, der mindestens so weit von $\mu = \hat{\theta}_{1,r}$ wie $\hat{\theta}_1$ entfernt ist lautet

$$\begin{aligned} p_1 &= P(|\{x\}_r^\omega - \hat{\theta}_{1,r}| \geq |\hat{\theta}_1 - \hat{\theta}_{1,r}|) \\ &= 2 - 2 \cdot \Phi\left(\frac{|\hat{\theta}_1 - \hat{\theta}_{1,r}|}{\sigma_{\hat{\theta}_1}}\right). \end{aligned}$$

Die Hypothese H_0 wird abgelehnt wenn die Wahrscheinlichkeit p_1 eine festzulegende Grenze $p_{H_0,\min}$ unterschreitet,

$$p_1 \stackrel{!}{\leq} p_{H_0,\min}.$$

²¹ Geht man davon aus, dass σ_y^2 nicht bekannt ist, und ersetzt man dieses durch den Schätzwert $\hat{\sigma}_y^2$, so erhält man eine t -Verteilung für $\{\hat{\theta}_1\}_r$, die ebenfalls als Grundlage eines statistischen Tests verwendet werden kann [STORM, 2001, S. 185ff]. Allerdings wäre dieser hier aufgrund der sehr wenigen Messwerte nicht aussagekräftig genug.

Es muss aber nicht jedes Mal p_1 aus den aktuellen Daten neu bestimmt werden, was aufgrund der Verteilungsfunktion Φ der Standardnormalverteilung auch aufwendig wäre. Durch Umformen ergibt sich

$$\frac{|\hat{\theta}_1 - \hat{\theta}_{1,r}|}{\sigma_{\hat{\theta}_1}} \geq \Phi^{-1} \left(\frac{2 - p_{H_0,\min}}{2} \right) \quad (6.58)$$

als Bedingung für das Ablehnen der Hypothese, wobei Φ^{-1} die Umkehrfunktion von Φ ist. Der Wert von $\Phi^{-1}((2 - p_{H_0,\min})/2)$ kann dabei für ein festes $p_{H_0,\min}$ vorher bestimmt und in der Steuerung hinterlegt werden. In Tabelle 6.3 sind die Werte dieser Funktion für ein paar $p_{H_0,\min}$ angegeben.

Tabelle 6.3: Werte für $\Phi^{-1} \left(\frac{2 - p_{H_0,\min}}{2} \right)$

$p_{H_0,\min}$ [%]	10	5	1	0,5	0,1
$\Phi^{-1} \left(\frac{2 - p_{H_0,\min}}{2} \right)$	1,64	1,96	2,58	2,81	3,29

Wenn die Bedingung (6.58) erfüllt ist, dann ist der aktuelle Wert für $\hat{\theta}_{1,r}$ zu unwahrscheinlich und sollte korrigiert werden. Dies ist nochmal in Abbildung 6.40a veranschaulicht. Mit dieser Bedingung wird überprüft, ob der Wert für $\hat{\theta}_{1,r}$ innerhalb der grau markierten Bereiche liegt.

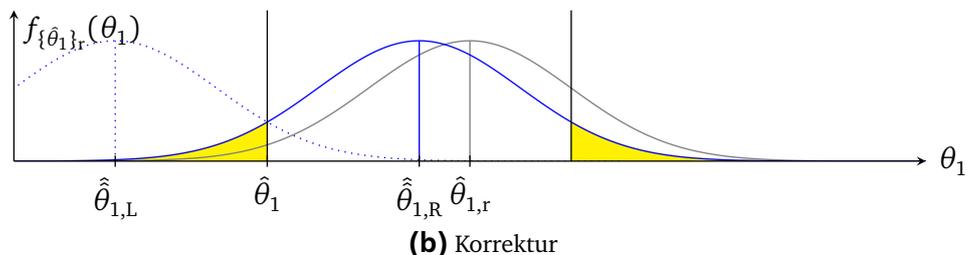
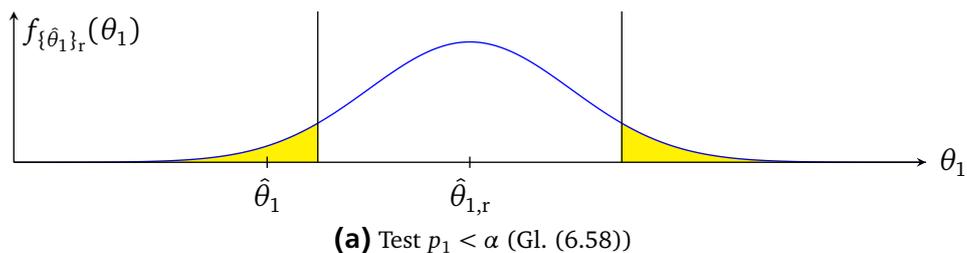


Abbildung 6.40: Veranschaulichung des Test auf Plausibilität und die Korrektur von $\hat{\theta}_1$

Korrektur

Hat sich bei dem Test der Bedingung (6.58) gezeigt, dass $\hat{\theta}_{1,r}$ (wahrscheinlich) nicht dem wahren θ_1 entspricht, dann sollte dieser, von der Regelung verwendete Wert, entsprechend angepasst werden.

Dazu werden die Werte für $\hat{\theta}_{1,r}$ bestimmt, mit denen die Bedingung (6.58) gerade erfüllt würde. Es ergibt sich also

$$\hat{\theta}_{1,L} = \hat{\theta}_1 - \sigma_{\hat{\theta}_1} \cdot \Phi^{-1} \left(\frac{2 - p_{H_0,\min}}{2} \right) \quad (6.59)$$

bzw.

$$\hat{\theta}_{1,R} = \hat{\theta}_1 + \sigma_{\hat{\theta}_1} \cdot \Phi^{-1} \left(\frac{2 - p_{H_0,\min}}{2} \right), \quad (6.60)$$

für die linke und rechte Grenze des Wertebereichs $(\hat{\theta}_{1,L}, \hat{\theta}_{1,R})$, welcher Bedingung (6.58) nicht erfüllt, und für den damit die Hypothese H_0 nicht abgelehnt wird. Ist Bedingung (6.58) erfüllt, so liegt $\hat{\theta}_{1,r}$ außerhalb dieses Bereichs, und $\hat{\theta}_{1,r}$ wird auf die nächstgelegene Grenze, also $\hat{\theta}_{1,L}$ oder $\hat{\theta}_{1,R}$ gesetzt. Liegt er innerhalb des Bereichs, so ist die Bedingung (6.58) erfüllt gewesen und keine Korrektur nötig. Damit muss (6.58) auch nicht mehr gesondert geprüft werden, und es ergibt sich das in Abbildung 6.41 dargestellte Vorgehen.

Das Prinzip der Korrektur ist in Abbildung 6.40b nochmals grafisch dargestellt.

Wenn man $p_{H_0,\min} = 1$ setzen würde, dann wäre $\Phi^{-1}\left(\frac{2-p_{H_0,\min}}{2}\right) = \Phi^{-1}(0,5) = 0$ und damit ergäbe sich mit Gl. (6.59) oder (6.60)

$$\hat{\theta}_1 = \hat{\theta}_{1,L} = \hat{\theta}_{1,R} = \hat{\theta}_1.$$

Dieses $\hat{\theta}_1$ wäre damit dasselbe, welches sich durch den normalen Schätzalgorithmus ergeben würde. Aber wenn man den Test auf Plausibilität durchführt, dann wird der normale Schätzer gerade nicht verwendet, da die Varianz der Schätzung zu groß wäre. Somit ist klar, dass hier $p_{H_0,\min}$ nicht zu groß und insbesondere nicht gleich eins gewählt werden darf. Auf der anderen Seite darf das $p_{H_0,\min}$ nicht zu klein gewählt werden, da ansonsten der Bereich, für den $\hat{\theta}_{1,r}$ akzeptiert wird, zu groß wird. Es geht hier im Prinzip auch nicht mehr darum, einen „guten“ Schätzwert zu bestimmen, da dies mit den gegebenen Daten einfach nicht möglich ist. Ziel ist es lediglich, „offensichtlich“ unrealistisch große Werte für $\hat{\theta}_{1,r}$ zu vermeiden, da dies zu einer sehr langsamen Korrektur von u_1 und damit sehr vielen notwendigen Regelschritten führt. Dazu wird $\hat{\theta}_{1,r}$ „behutsam“ korrigiert.

Beispiel – Kriechen

In Abbildung 6.42 ist die Wirkung dieses Verfahrens an einem Beispiel gezeigt. Dabei wird für die Regelung zunächst von einem $\hat{\theta}_{1,r}$ ausgegangen, welches dem Vierfachen des wahren Parameters θ_1 entspricht. Damit wird mit dem ersten Regelschritt (im Mittel) nur ein Viertel der Regelabweichung korrigiert. Die damit verbundene Änderung der primären Eingangsgröße u_1 reicht noch nicht aus, um den Schätzwert von θ_1 zur Regelung verwenden zu können.

In Abbildung 6.42a ist der hier beschriebene Test nicht verwendet und damit wird auch im nächsten Regelschritt eine deutlich zu geringe Korrektur durchgeführt. Auch mit diesem Schritt ist der geschätzte Parameter noch zu unsicher. Erst nach dem fünften Regelschritt ($k = 6$) ist der zur Regelung verwendete Wert $\hat{\theta}_{1,r}$ aktualisiert, so dass bei $k = 8$ (hier nicht mehr dargestellt) der Sollwert im Mittel erreicht ist.

Abbildung 6.42b zeigt das Verhalten, wenn $p_{H_0,\min} = 10\%$ gesetzt wird. Hierbei wird nach dem ersten Schritt festgestellt, dass das vorliegende $\hat{\theta}_{1,r} = 4 \cdot \theta_1$ wahrscheinlich nicht zutreffen kann und $\hat{\theta}_{1,r}$ entsprechend korrigiert. Der Mittelwert von $\hat{\theta}_{1,r}$ liegt nach dieser Korrektur bei ca. 0,8, so dass die Korrektur im nächsten Regelschritt auch noch nicht ideal in dem Sinne ist, dass im Mittel der Sollwert erreicht wird. Mit dem nächsten Schritt wird der Wert für $\hat{\theta}_{1,r}$ nochmals weiter korrigiert, und mit dem folgenden Schritt reicht die Anregung über u_1 aus, den aktuellen Schätzwert für θ_1 zur Regelung zu verwenden.

Mit einer weitere Erhöhung von $p_{H_0,\min}$ auf 30% (Abbildung 6.42c) wird $\hat{\theta}_{1,r}$ bei $k = 2$ etwas offensiver korrigiert, so dass der Mittelwert bei etwa 0,7 liegt, womit auch der Mittelwert der Regelgröße bei $k = 3$ etwas näher am Sollwert liegt. Jedoch wird dies durch mehr Ausreißer (hohe Überschwinger) erkauft.

Beispiel – Aufschwingen

Für den Fall des aufschwingenden Verhaltens ist das Verfahren an dem Beispiel Abbildung 6.39b demonstriert. Dazu sind in Abbildung 6.43 die Ergebnisse für die zwei Grenzen $p_{H_0,\min} = 10\%$ und 30% gezeigt, wobei jeweils einmal eine einzelne Simulation und einmal 10 000 Simulationen dargestellt sind.

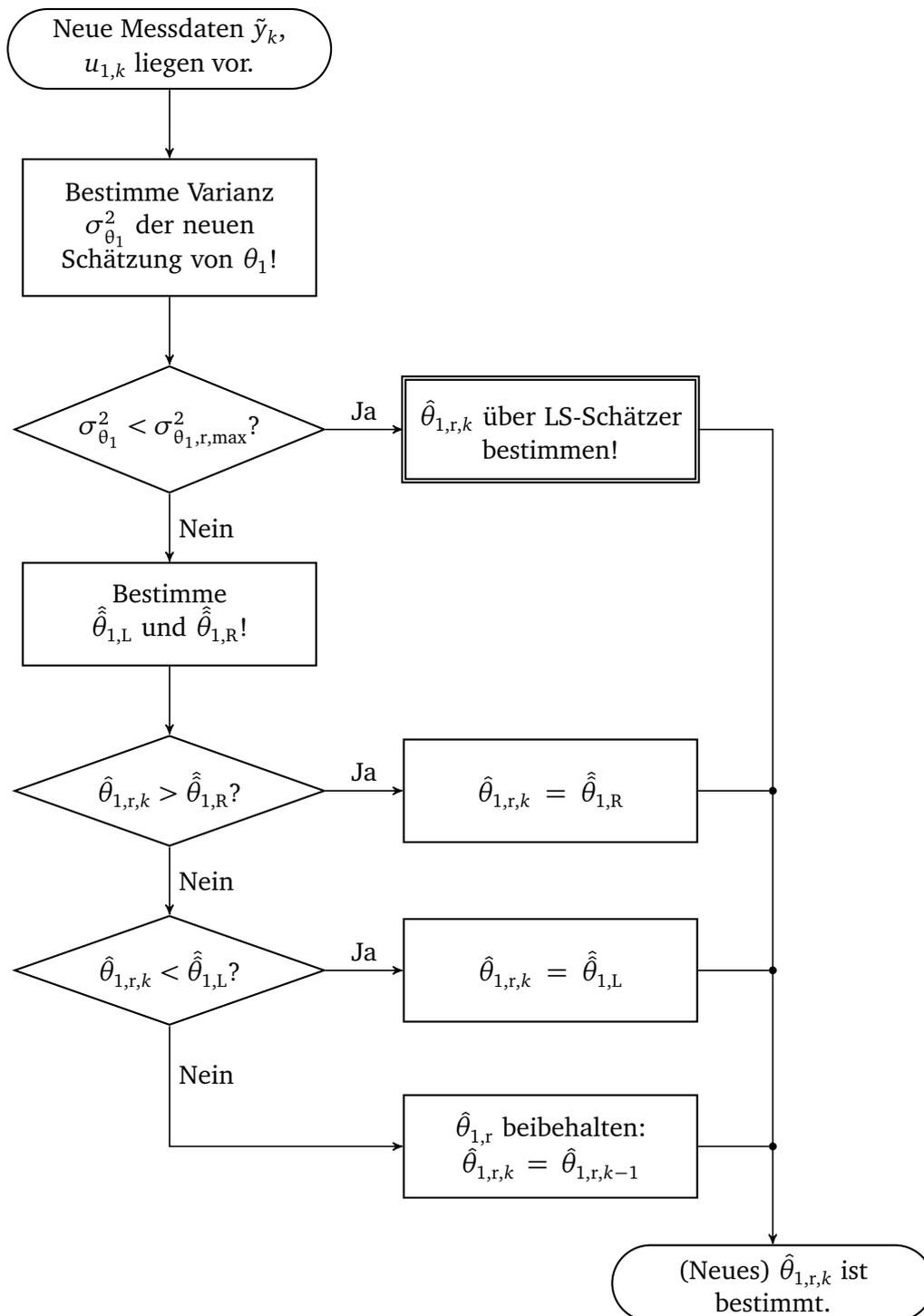


Abbildung 6.41: Vorgehen zur Parameterschätzung von θ_1 mit Plausibilitätsprüfung

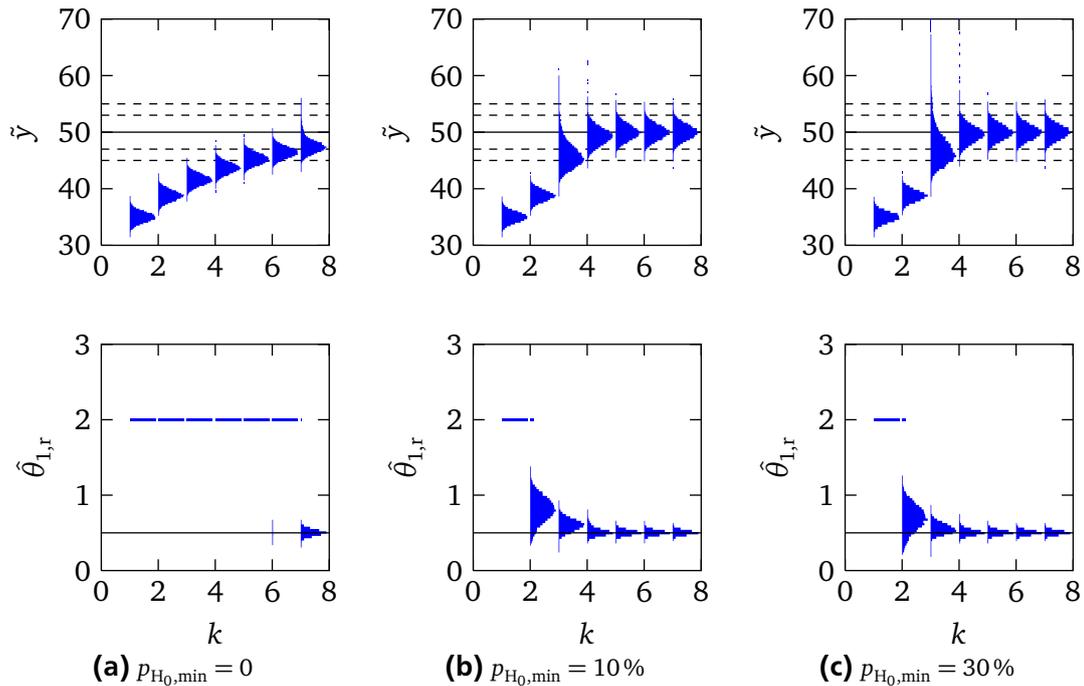


Abbildung 6.42: Beispiele für verschiedene Grenzen des Plausibilitätstests ($\sigma_{\hat{\theta}_{1,r,\max}} = \frac{1}{10} \cdot \theta_1$, $\hat{\theta}_{1,r,0} = 4 \cdot \theta_1$)
(Daten aus je 10 000 Simulationen)

Die Reglerverstärkung ist hierbei konstant $\alpha = 1$, um die stabilisierende Wirkung der Absenkung der Reglerverstärkung bzw. Filterung auszuschalten.

Mit beiden Werten für $p_{H_0,\min}$ wird ein deutlich besseres Verhalten als in Abbildung 6.39b erreicht.

Insgesamt, d. h. bei den 10 000 Simulationen, zeigt sich nur eine geringe Verbesserung bei Verwendung der Grenze $p_{H_0,\min} = 30\%$ im Vergleich zu $p_{H_0,\min} = 10\%$. Im gezeigten Einzelfall wird bei $p_{H_0,\min} = 30\%$ der Schätzwert $\hat{\theta}_{1,r}$ jedoch schon einen Schritt früher korrigiert.

Fazit

Würde man den anfangs angenommenen Wert für $\hat{\theta}_{r,1}$ in dem gezeigten Beispiel zum Kriechen weiter erhöhen, so würde die Korrektur zu Beginn noch schwächer ausfallen. Im Grenzfall $\hat{\theta}_{r,1} \rightarrow \infty$ würde gar keine Korrektur mehr stattfinden.

Die in diesem Abschnitt vorgestellte Methode zum Erkennen und Korrigieren von aufschwingendem und kriechendem Verhalten beruht auf der Bewertung des angenommenen Wertes $\hat{\theta}_{1,r}$ anhand des aus den aktuellen Daten geschätzten $\hat{\theta}_1$. Dazu muss die Varianz $\sigma_{\hat{\theta}_1}^2$ zwar formal keinen maximalen Wert einhalten, aber praktisch würde das Intervall $[\hat{\theta}_{1,L}, \hat{\theta}_{1,R}]$ bei einem zu kleinen $\sigma_{\hat{\theta}_1}^2$ zu groß, um den Wert für $\hat{\theta}_{1,r}$ sinnvoll einzuschränken.

Damit würde in einem solchen extremen Fall ein sehr langsames Kriechen auftreten, bevor der Regelparameter automatisch angepasst würde, wobei davon auszugehen ist, dass der Bediener in solchen Fällen selber eingreifen würde. Um eine möglichst gute Akzeptanz der Regelung zu erreichen, könnte es gewünscht sein, dass auch dieser Fall automatisch festgestellt werden sollte. Dazu kann der Verlauf der gemessenen Ausgangsgröße, ohne Berücksichtigung der Eingangsgröße, betrachtet werden.

Ähnliches gilt, wenn ein aufschwingendes Verhalten beginnt. Es kann einige Messungen dauern, bis die Anregung ausreicht, um $\hat{\theta}_{1,r}$ zu korrigieren, obwohl für den Bediener schon ein instabiles Verhal-

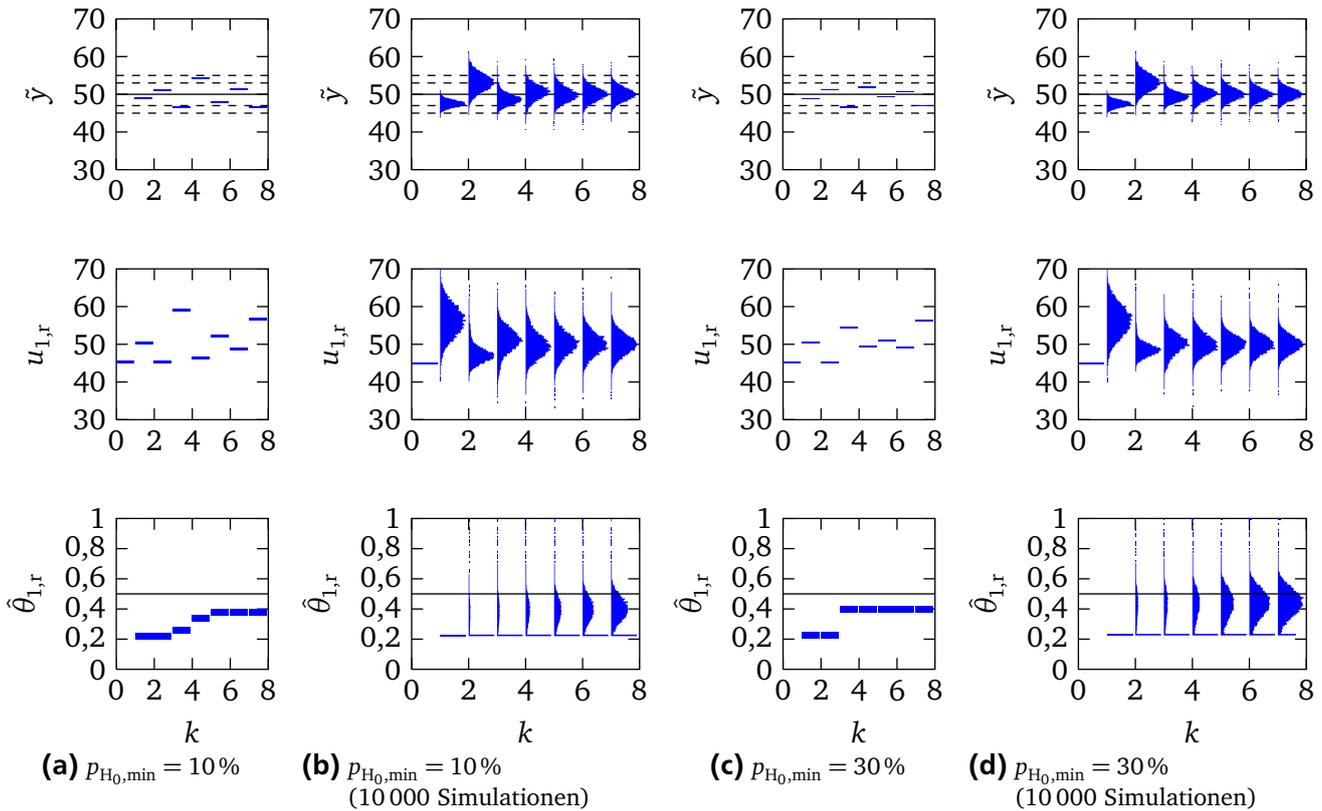


Abbildung 6.43: Beispiele für verschiedene Grenzen des Plausibilitätstests ($\sigma_{\hat{\theta}_{1,r, \max}} = \frac{1}{10} \cdot \theta_1$, $\hat{\theta}_{1,r,0} = 4 \cdot \theta_1$)
(Daten aus je 10 000 Simulationen)

ten zu erkennen ist. Hier wäre eine Möglichkeit, dass bei wechselnden, betragsmäßig anwachsenden Regelabweichungen die Reglerverstärkung „testweise“ reduziert wird.

6.7.2 Berücksichtigen der geschätzten Varianz des Messrauschens/Ausreißer

Überprüfen der Varianz

Über Gl. (6.30) ist eine Möglichkeit gegeben, einen Schätzwert $\hat{\sigma}_y^2$ für die Varianz zu erhalten. Dieser kann genutzt werden, zu prüfen, ob das angenommene σ_y^2 tatsächlich der Varianz der Messung entspricht. Ist dies nicht der Fall, d. h. ist $\hat{\sigma}_y^2$ deutlich größer als σ_y^2 , so ist dies ein Hinweis darauf, dass entweder die Messung stärker verrauscht ist als angenommen und damit die Varianzgrenzen der Parameter ($\sigma_{\hat{\theta}_{1,r, \max}}^2, \dots$) nicht mehr stimmen, oder dass ein Ausreißer oder Sprung in θ_0 aufgetreten ist. In beiden Fällen führt eine Schätzung mit diesen Messdaten nicht zu den gewünschten Ergebnissen.

Um zu verhindern, dass die Schätzung mit ungünstigen Messdaten durchgeführt wird, sollte überprüft werden, ob

$$\hat{\sigma}_y^2 \stackrel{!}{<} (f_{\sigma_y, \max} \cdot \sigma_y)^2,$$

mit einem geeigneten $f_{\sigma_y, \max}$, erfüllt ist. Ist dies nicht der Fall, dann ist es die einfachste Maßnahme, alle bisherigen Daten des aktuellen Auftrags zu verwerfen, d. h. es wird eine Reinitialisierung des Schätzalgorithmus durchgeführt, wodurch auch die zur Regelung verwendeten Werte $\hat{\theta}_{1,r}$ und $\hat{\theta}_{2,r}$ auf die zu Beginn des Auftrags gültigen zurückgesetzt werden.

Ungünstig wäre bei diesem Vorgehen, dass bei einem Wegdriften des Parameters θ_0 während eines längeren Auftrags zu einem späten Zeitpunkt alle Messdaten, und damit auch gute Daten vom Beginn des Auftrags, verworfen werden. Es sollte für die Anwendung dazu überprüft werden, ob solche Situationen häufig auftreten. Falls ja, wäre es die einfachste Lösung, nach einer gewissen Zeit nach Auftragsbeginn die Werte grundsätzlich nicht mehr zur Schätzung zu verwenden, sondern nur noch eine Regelung durchzuführen.

Beispiel

In Abbildung 6.44 ist die Wirkung der Überprüfung der Varianz des Prozessrauschens an einem Beispiel dargestellt. Es sind jeweils vier Aufträge (ohne Änderung der sekundären Eingänge) simuliert. Nach dem dritten Regelschritt im dritten Auftrag – diese Stelle ist in den Plots mit einer senkrechten gestrichelten Linie markiert – ändert sich θ_0 sprunghaft um fünf auf einen Wert von 55. Nach dem dritten Auftrag gilt wieder der alte Wert von 50.

In Abbildung 6.44a ist diese Überprüfung der Varianz des Prozessrauschens nicht aktiv. Man erkennt deutlich, wie sich die Schätzung von θ_1 ab der Stelle, an der der Sprung von θ_0 aufgetreten ist, verschlechtert.

Im unteren Plot ist die geschätzte Standardabweichung dargestellt. Die tatsächliche Standardabweichung von $\sigma_y = 1$ ist mit einer horizontalen schwarzen Linie eingezeichnet. Der mit einer gestrichelten Linie markierte Wert von 1,5 wird als Grenze verwendet werden. Man kann an diesem Plot auch gut erkennen, dass ein biasbehafteter Schätzer für die Varianz verwendet wird, da der Mittelwert der Verteilungen $\{\hat{\sigma}_{y,k}\}_r$ bei jedem Auftrag zu Beginn deutlich unter eins liegt und sich dieser erst mit steigender Anzahl an Schritten nähert. Es ist auch gut zu erkennen, dass die Varianz der Schätzung der Varianz, also $\text{Var}\{\hat{\sigma}_{y,k}\}_r$ zu Beginn jedes Auftrags zunächst sehr hoch ist. In Summe liegen die „Oberkanten“ der Verteilungen $\text{Var}\{\hat{\sigma}_{y,k}\}_r$ aber auf einem ähnlichen Niveau. Damit ist auch das in Abschnitt 6.4.6 beschriebene Vorgehen gerechtfertigt, den biasbehafteten Varianzschätzer zu verwenden.

In Abbildung 6.44b ist für $f_{\sigma_y, \max}$ der Wert $1,5 \cdot \sigma_y$ verwendet. Man erkennt, dass die Schätzung falscher Parameterwerte in Auftrag drei effektiv verhindert wird. Jedoch werden dazu auch Nachteile in Kauf genommen.

So werden bei einem geringen, aber nicht vernachlässigbaren Anteil der Simulationen zu keinem Zeitpunkt oder erst sehr spät überhaupt Parameter geschätzt, wogegen zuvor in 100 % der Fälle der Parameter θ_1 im Laufe des ersten Auftrags geschätzt wurde. Eine Folge davon sind die extremeren Ausreißer nach unten von $\hat{\theta}_{1,r}$ im dritten Auftrag, die wiederum die vereinzelt ($< 0,1\%$) Ausreißer in \tilde{y}_k zur Folge haben. Die extrem kleinen Werte für $\hat{\theta}_{1,r}$ treten dann auf, wenn die geschätzte Varianz im dritten Auftrag trotz der Störung unter der Grenze bleibt, und bisher kein Wert für θ_1 geschätzt werden konnte, weil diese Grenze im ersten Auftrag überschritten wurde. Damit besteht $\hat{\theta}_{1,r}$ hier ausschließlich aus dem falschen Wert des dritten Auftrags.

Fazit

Die Methode, die geschätzte Varianz zu überprüfen, ist sehr einfach umzusetzen. Es muss zur Schätzung von σ_y^2 nur ein einzelner Skalar zusätzlich gespeichert werden, unabhängig von der Anzahl der Messungen.

Nachteilig ist eine geringe „Trennschärfe“ bei kleineren Störungen. Bei der in diesem Beispiel gezeigten Störung handelt es sich schon um einen Grenzfall in dem Sinne, dass diese (ohne Gegenmaßnahmen) schon einen deutlichen Fehler bei der Schätzung des Parameters θ_1 hervorruft, wodurch man schon nahe an die Stabilitätsgrenze von $\hat{\theta}_{1,r} = 0,5 \cdot \theta_1$ kommt. Anhand des unteren Plots in Abbildung 6.44a erkennt man auch, dass die hier gezeigte Störung die untere Grenze darstellt, was die Unterscheidung von „guten“ und „schlechten“ Aufträgen angeht.

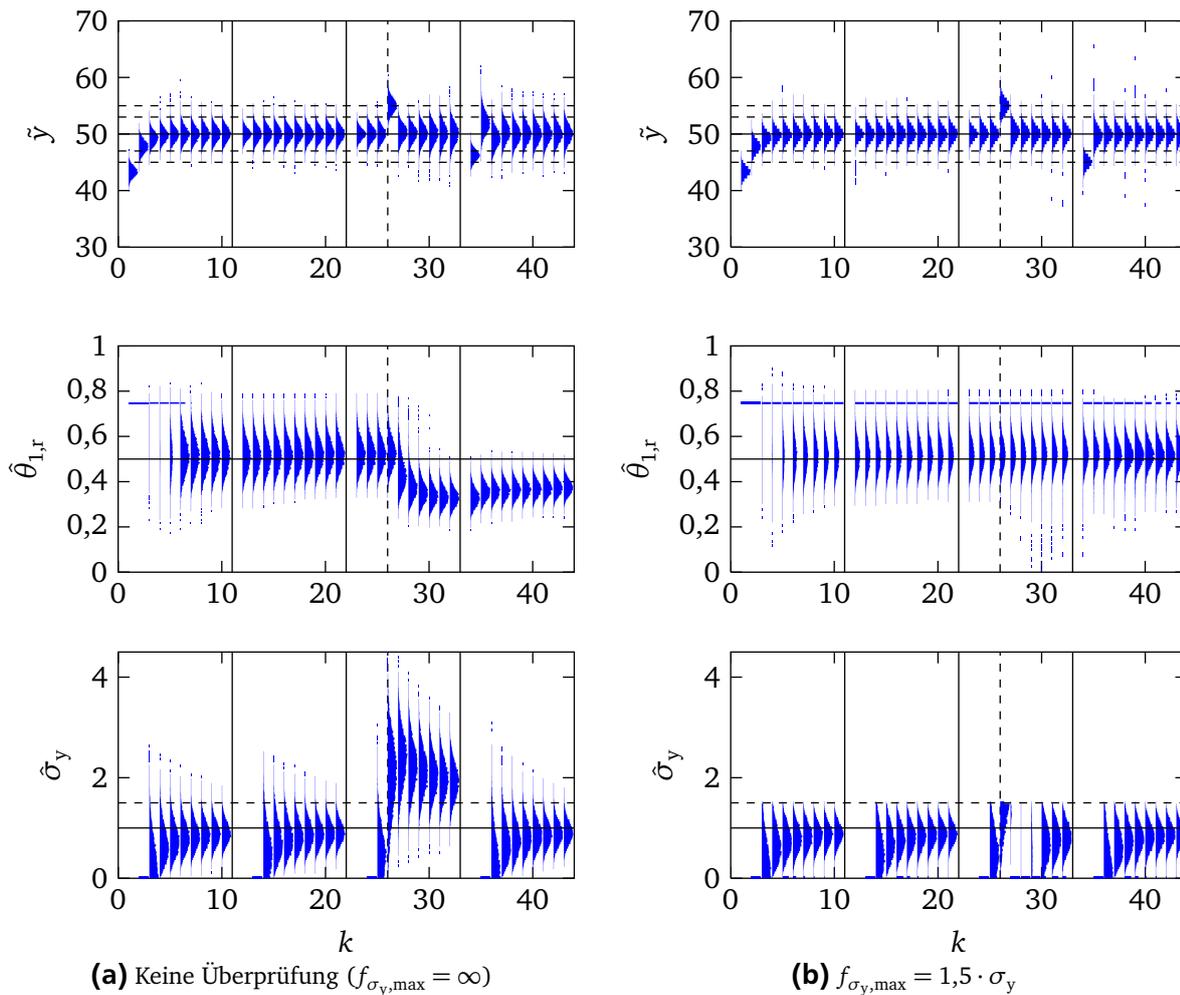


Abbildung 6.44: Überprüfung der Varianz der Prozessrauschens (je 10 000 Simulationen)

Eine weitere Absenkung der Grenze $f_{\sigma_y, \max}$ würde zu einer drastischen Reduzierung der gültigen Schätzwerte führen, bei einer Erhöhung würden zu viele schlechte Werte akzeptiert werden.

Inwieweit dies für den vorliegenden Prozess problematisch ist, hängt auch davon ab, wie die Ausreißer und Störungen verteilt sind. Wenn häufig kleinere Störungen auftreten, dann reicht die Überprüfung der geschätzten Varianz alleine nicht aus. Es kann dann, wie weiter unten skizziert, versucht werden, die Messdaten punktweise auf Ausreißer oder Sprünge zu testen. Sind in der Regel auch immer eine höhere Anzahl an „guten“ Aufträgen mit ausreichender Anregung zwischen „schlechten“ Aufträgen zu erwarten, kann auch mit höheren Vergessensfaktoren (weniger Vergessen) in der Mittelung gearbeitet werden, wodurch der Einfluss der Werte aus den „schlechten“ Aufträgen durch die Mittelung über viele Messungen reduziert wird.

Wenn die Anzahl der Aufträge mit kleineren Störungen geringer ist, bzw. deren Einfluss durch die angesprochenen Maßnahmen reduziert ist, so wäre auch zu prüfen, ob die Erhöhung der Grenze $f_{\sigma_y, \max}$ möglich ist. Könnte diese z. B. auf einen Wert zwischen 2 und 2,5 eingestellt werden, so würde dies die Abweisung von „guten“ Aufträgen deutlich reduzieren. Größere Störungen würden dadurch aber immer noch sicher erkannt.

Bei der Wahl von $f_{\sigma_y, \max}$ ist auch zu beachten, in welchem Bereich die Varianz σ_y^2 des Messrauschens bei dem gegebenen Prozess schwanken kann bzw. wie dieser Wert ggf. von den Prozessparametern abhängen kann.

Der Nachteil der Methode, alle Messdaten zu verwerfen, wenn die Varianz zu hoch ist, liegt auf der Hand. Auch wenn nur ein einziger oder zwei Messwerte aufgrund zeitlich begrenzter Störungen, wie z. B. ein Fehler bei der Durchführung einer einzelnen Messung, nicht zu den anderen passen, so werden auch die Informationen, die in den „guten“ Messungen stecken, verworfen.

Bestimmt man die Ausreißer in den bisherigen Messdaten, so könnten nur diese entfernt werden und mit den übrigen Messdaten eine normale Schätzung durchgeführt werden. Der Nachteil dabei liegt aber darin, dass zurückliegende Messungen gespeichert werden müssen und vor Allem der Berechnungsaufwand und die Komplexität des Programmes stark ansteigen.

Bestimmen einzelner Ausreißer

Zur Detektion der Ausreißer müssten auch zurückliegende Messpunkte im Speicher vorgehalten werden.

In der Literatur findet man Methoden, bei denen die Residuen jedes Messpunktes bestimmt werden und für den Messpunkt mit dem höchsten Residuum ein stochastischer Test durchgeführt wird. [FREUND et al., 2006, S. 120ff]

Da nur sehr wenige Messungen vorliegen, können bei solchen Tests die Intervalle, für die ein Messpunkt noch akzeptiert wird, nur sehr konservativ ausgelegt sein. Alternativ kann man die Schätzung der Varianz wie beschrieben durchführen, wobei jeweils ein Messwert nicht verwendet wird. Sinkt die Varianz einer solchen Schätzung signifikant (im Vergleich zu den Veränderungen beim Weglassen anderer Punkte), so kann der nicht verwendete Messwert als Ausreißer betrachtet und dauerhaft verworfen werden. Dies kann so durchgeführt werden, solange es sich nicht um den letzten Messwert handelt. Sollte dies der Fall sein, dann müsste noch abgewartet werden, ob nicht vielleicht ein Sprung anstelle eines Ausreißers vorliegt.

Bestimmen von Sprüngen

Der Test auf einen Sprung ist nochmals aufwendiger, da ein weiterer Parameter hinzukommt. Tritt der Sprung in y_0 zur Messung $k = k_s$ auf, so kann die zu identifizierende Funktion in der Form

$$y_k = \theta_0 + \theta_1 \cdot (u_{1,k} - u_{1,N}) + \theta_2 \cdot (u_{2,k} - u_{2,N}) + \theta_{0,s} \sigma(k - k_s)$$

dargestellt werden, wobei es sich bei $\sigma(k)$ um den Einheitssprung handelt und $\theta_{0,s}$ die Änderung des Parameters θ_0 ist. Zu dieser Funktion können die Parameter

$$\boldsymbol{\theta}^T = [\theta_0 \quad \theta_1 \quad \theta_2 \quad \theta_{0,s}]$$

über eine Least-Squares-Schätzung bestimmt werden. Der Eingangsgrößenvektor lautet in diesem Fall

$$\boldsymbol{\psi}^T = [1 \quad u_{1,k} - u_{1,N} \quad u_{2,k} - u_{2,N} \quad \sigma(k - k_s)] .$$

Für dieses Schätzproblem ist der Schätzwert $\hat{\sigma}_y^2$ für die Varianz des Messrauschens für jedes mögliche (und sinnvolle) k_s zu berechnen. Reduziert sich der Wert von $\hat{\sigma}_y^2$ für ein k_s deutlich, so ist dies ein Hinweis darauf, dass zu diesem Zeitpunkt ein Sprung in θ_0 aufgetreten ist.

Ist ein Sprung aufgetreten, so ist noch zu entscheiden, wie mit den Messdaten zu verfahren ist. Die Maßnahme, die den wenigsten „Verlust“ an Informationen beinhaltet, wäre

- eine Schätzung mit den Daten bis zum Sprung durchzuführen und sich zu merken und

- die aktuelle Schätzung mit den Daten ab dem Sprung fortzuführen.

Solange die Daten vor dem Sprung die bessere Schätzung liefern, werden diese Parameter verwendet. Wenn die Daten nach dem Sprung besser geworden sind, können die Daten vor dem Sprung endgültig verworfen werden. Nur der Parameter θ_0 sollte immer aus den jüngsten Messdaten ermittelt werden.

Diese Beschreibung zeigt schon die hohe Komplexität dieser Maßnahmen. Ob diese sich lohnen, sollte im Einzelfall geprüft werden.

6.7.3 Diskussion der Erweiterungen

Verwendet man ausschließlich die bisher vorgestellten Verfahren zur Identifikation und Überprüfung der Werte, so ist bei realistischen Anforderungen im Mittel eine zufriedenstellende Robustheit der Gesamtregelung gegen anfänglich deutlich falsch angenommener Parameterwerte sowie Ausreißern bzw. Sprüngen von θ_0 während eines Auftrags zu erreichen.

Jedoch zeigen sich bei den Simulationen immer vereinzelt Ausreißer, bei denen schlechtes oder inakzeptables Regelverhalten auftritt. Teilweise wurden auch schon weitere mögliche Maßnahmen angesprochen, ohne diese jedoch weiter auszuführen.

Die Plausibilitätsprüfung sowie die Überprüfung der geschätzten Varianz sind noch relativ „elegante“ Maßnahmen, die einfach zu berechnen und mit wenigen, anschaulichen Größen zu parametrieren sind. Soll der Verlauf der gemessenen Ausgangsgröße ebenfalls bewertet werden, bzw. gezielt nach Ausreißern und Sprüngen gesucht werden, so sind zum einen mehr Daten zu speichern und zum anderen steigt die Anzahl der zu parametrierenden Einflussgrößen stark an.

Eine weitere Stufe wäre die Hinzunahme von Regeln, die aufgrund der „Geschichte“ der Messdaten und den Erfahrungen mit dem Prozess festlegen, welche Daten zur Schätzung verwendet werden, oder wann die anderen Maßnahmen aktiviert werden sollen.

Je spezieller die Maßnahmen werden, desto kritischer ist jedoch auch das angenommene Rauschmodell zu betrachten. Das einfache Modell des unkorrelierten, gaußverteilten Rauschens leistet schon gute Dienste zum Einstellen der Parameter, sofern es näherungsweise dem tatsächlichen Rauschen entspricht. Wenn aber auch Fälle berücksichtigt werden sollen, die im Bereich von unter einem Prozent oder Promille auftreten, stellt sich die Frage, ob dann eine einfache Näherung des Rauschprozesses ausreicht.

Auch sind die Anforderungen an die Regelung zu überprüfen. So wurden bisher keinerlei Annahmen über den gerade vorliegenden Prozess gemacht, abgesehen davon, dass $\theta_1 > 0$ und $\theta_2 < 0$ ist. Kann unter diesen Umständen beim ersten Einlernen eines Parametersatzes ein unterstützender Eingriff des Bedieners in die Regelung akzeptiert werden, relativiert sich die Problematik des Kriechens oder Aufschwingens wieder. (Ideal zum Einlernen wäre es ohnehin, die Eckpunkte des möglichen Parameterbereichs abzufahren. Die damit erreichbaren Varianzen der Schätzwerte liegen deutlich unter denen, die sich während normaler Regelvorgänge ergeben.)

Auch wurde bisher nicht berücksichtigt, ob die Anfangsparameter von außen vorgegeben waren, oder ob diese auf Schätzungen bei vorangegangenen Aufträgen beruhen. Ändert sich ein Parameter bei einem neuen Auftrag signifikant, so ist dies ein Hinweis darauf, dass entweder eine Störung vorliegt oder ggf. auch der falsche Parametersatz geladen wurde. In diesen Fällen ist auch eine Rückfrage beim Bediener sinnvoll. Beschränkt man die Änderung von Parametern bei Parametersätzen, die auf Schätzungen vorangegangener Aufträge beruhen, auf einen sinnvollen Wertebereich, so wären z. B. die extremen Unterschätzungen von θ_1 , wie in Abbildung 6.44b gezeigt, abzufangen.

6.8 Zusammenfassung des Algorithmus

Es wird der vorgestellte Least-Squares-Algorithmus verwendet. Die Parameter werden für jeden Auftrag einzeln geschätzt und gemittelt.

Die Daten eines „Parametersatzes“, d. h. die Daten, die auch nach Beenden eines Auftrags zu speichern sind, sind in Tabelle 6.4 zusammengefasst. Die nur über einen Auftrag zu speichernde Daten sind in Tabelle 6.5 aufgelistet.

Die Parameter zum Konfigurieren des Algorithmus sind in Tabelle 6.6 zusammengefasst.

Tabelle 6.4: Über mehrere Aufträge zu speichernde Daten (Materialdaten)

Größe	Variable
$\hat{\theta}_{0,r'}, \hat{\theta}_{1,r'}, \hat{\theta}_{2,r'}, c_{2 1,r'}$	Parameter für Regelung
$\hat{\theta}_{1,m}, \sigma_{\hat{\theta}_{1,m}}^2, g_{1,m}$	Mittelung $\hat{\theta}_1$ (Alternativ: $\hat{\theta}_{1,m}, p_{1,m}$)
$\hat{\theta}_{2,m}, \sigma_{\hat{\theta}_{2,m}}^2, g_{2,m}, c_{2 1,m}$	Mittelung $\hat{\theta}_2$ (Alternativ: $\hat{\theta}_{2,m}, p_{2,m}, c_{2 1,m}$)

Tabelle 6.5: Während eines Auftrags zu speichernde Daten

Größe	Variable
L, r, Y, k	Variablen für Schätzalgorithmus
$\hat{\theta}_{0,r}, \hat{\theta}_{1,r}, \hat{\theta}_{2,r}, c_{2 1,r}$	Aktuelle Parameter für Regelung
$\hat{\theta}_{1,+}, \sigma_{\hat{\theta}_{1,+}}^2, g_{1,+}$	Aktuelle Mittelung $\hat{\theta}_1$ (Alternativ: $\hat{\theta}_{1,+}, p_{1,+}$)
$\hat{\theta}_{2,+}, \sigma_{\hat{\theta}_{2,+}}^2, g_{2,+}, c_{2 1,+}$	Aktuelle Mittelung $\hat{\theta}_2$ (Alternativ: $\hat{\theta}_{2,+}, p_{2,+}, c_{2 1,+}$)

Tabelle 6.6: Parameter

Größe	Variable
Parameter Least-Squares-Verfahren mit Mittelung	
σ_y^2	Varianz des Messrauschens
$\sigma_{\hat{\theta}_{1,r,\max}}^2$	Maximale Varianz von $\hat{\theta}_1$ für Regelung
$\sigma_{\hat{\theta}_{2,r,\max}}^2$	Maximale Varianz von $\hat{\theta}_2$ für Regelung
$\sigma_{\hat{\theta}_{1,m,\max}}^2$	Maximale Varianz von $\hat{\theta}_1$ für Mittelung
$\sigma_{\hat{\theta}_{2,m,\max}}^2$	Maximale Varianz von $\hat{\theta}_2$ für Mittelung
$\lambda_1 (, p_{1,\min})$	Parameter zur Mittelung von θ_1
$\lambda_2 (, p_{2,\min})$	Parameter zur Mittelung von θ_2
$[\theta_{1,\min}, \hat{\theta}_{1,\max}]$	Grenzen für θ_1
$[\theta_{2,\min}, \hat{\theta}_{2,\max}]$	Grenzen für θ_2
Parameter Erweiterungen	
$p_{H_0,\min}$	Wahrscheinlichkeitsgrenze zur Ablehnung von H_0
$f_{\sigma_y,\max}$	Maximal zulässiger Faktor für $\hat{\sigma}_y$

Aktionen beim Vorliegen neuer Messdaten

In Abbildung 6.45 ist der Ablauf beschrieben, der während eines Auftrags beim Vorliegen neuer Messdaten durchgeführt wird. Im Folgenden werden die doppelt umrandeten „Subsysteme“ kurz erläutert.

Schätzer reseten

Beim Reset werden alle in Tabelle 6.5 aufgezählten Größen auf ihren Initialisierungswert gesetzt. D. h. dass die Variablen L , r und Y „genullt“ werden. Die „+“-Werte werden auf die „m“-Werte der Materialdaten gesetzt und die Listen mit den Messpunkten geleert. Die aktuellen Reglerparameter („r“) werden wieder mit den „r“-Werten der Materialdaten überschrieben.

$\hat{\theta}_1$ bestimmen und mitteln

$\hat{\theta}_1$ wird immer aus der Drei-Parameter-Schätzung bestimmt. Der Schätzwert wird wie in Abschnitt 6.5.2 beschrieben gemittelt.

$\hat{\theta}_2$ bestimmen und mitteln

$\hat{\theta}_2$ wird immer aus der Zwei-Parameter-Schätzung mit festgehaltenem $\hat{\theta}_{1,r}$ bestimmt. Der Schätzwert wird wie in Abschnitt 6.5.2 beschrieben gemittelt.

Plausibilität $\hat{\theta}_{1,r}$ prüfen

Dieser Test wird wie oben beschrieben durchgeführt und $\hat{\theta}_{1,r}$ gegebenenfalls korrigiert.

Korrektur der $\hat{\theta}_2$ -Werte

Da die Werte für $\hat{\theta}_{2,r}$, $\hat{\theta}_{2,m}$ und $\hat{\theta}_{2,+}$ über die Zwei-Parameter-Schätzung bestimmt werden, müssen diese immer, wenn sich $\hat{\theta}_{1,r}$ ändert, auch korrigiert werden. Dies geschieht über

$$\begin{aligned}\hat{\theta}_{2,r,k} &= \hat{\theta}_{2,r,k-1} + c_{2|1,r} \cdot (\hat{\theta}_{1,r,k} - \hat{\theta}_{1,r,k-1}) \\ \hat{\theta}_{2,m,k} &= \hat{\theta}_{2,m,k-1} + c_{2|1,m} \cdot (\hat{\theta}_{1,r,k} - \hat{\theta}_{1,r,k-1}) \\ \hat{\theta}_{2,+,k} &= \hat{\theta}_{2,+,k-1} + c_{2|1,+} \cdot (\hat{\theta}_{1,r,k} - \hat{\theta}_{1,r,k-1}).\end{aligned}$$

(Dabei ist die Korrektur für die „m“-Werte an einer temporären, für den Auftrag angelegten Kopie der Materialparameter durchzuführen, oder aber es muss in den Materialparametern zusätzlich noch der Referenzwert für $\hat{\theta}_1$ abgelegt werden.)

$\hat{\theta}_{0,r}$ schätzen

Der Parameter θ_0 wird nach Gl. (6.10) bezüglich der aktuell verwendeten Reglerparameter geschätzt, wie eingangs in Abschnitt 6.4.1 begründet.

Aktionen beim Starten und Beenden eines Auftrags

Starten eines Auftrags

Beim Starten eines neuen Auftrags werden dieselben Aktionen wie im Block „Schätzer reseten“ beschrieben durchgeführt.

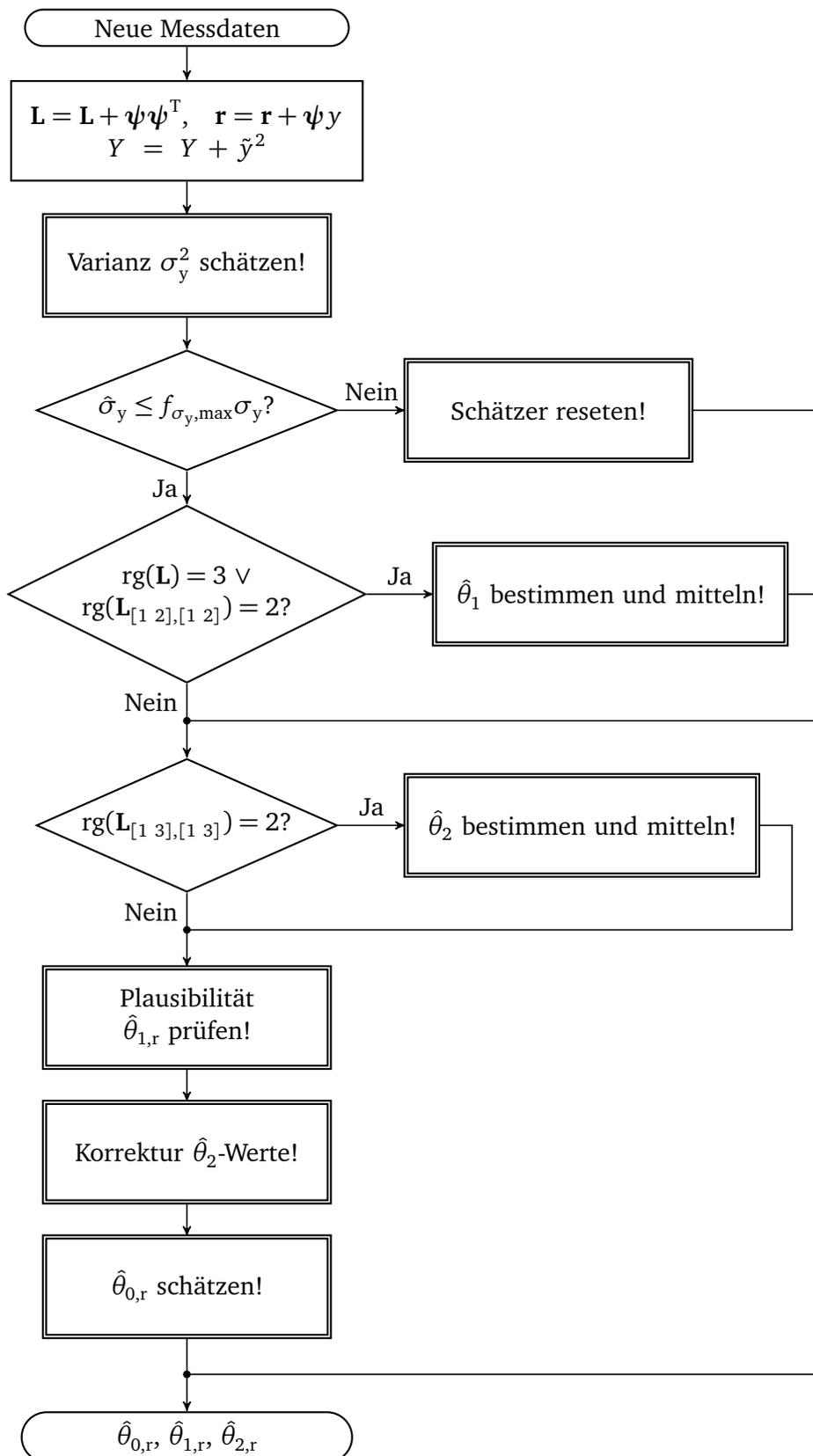


Abbildung 6.45: Ablaufplan Messung komplett

Beenden eines Auftrags

Die „+“-Werte des aktuellen Auftrags werden als „m“-Werte der Materialdaten gespeichert. Die „r“-Werte des aktuellen Auftrags werden als „r“-Werte der Materialdaten gespeichert.

6.9 Adaptive Regelung

Zuletzt soll die verwendete, in Abbildung 6.1 auf Seite 90 gezeigte Struktur des Regelkreises mit Adaption in den Kontext der Adaptiven Regelung eingeordnet werden: Es handelt sich hierbei um eine selbst-einstellende Regelung (self-tuning regulator) [CHALAM, 1987, S. 189ff] bzw. indirekte adaptive Regelung [TAO, 2003, S. 295], da der Regler zunächst anhand der als bekannt vorausgesetzter Streckenparameter entworfen wurde. Im Betrieb werden dann die Streckenparameter unabhängig vom Regler geschätzt und die Reglerparameter entsprechend angepasst.

Eine wesentliche Eigenschaft eines adaptiven Regelkreises, wie allgemein auch jedes anderen Regelkreises, ist dessen Stabilität. Es wird im Rahmen dieser Arbeit jedoch verzichtet, tiefer auf die Stabilität des geschlossenen Regelkreises einzugehen. Anschaulich verfügt der Regelkreis schon über eine stabilisierende Tendenz, da hohe Regelabweichungen zu guten Parameterschätzungen führen, die wiederum den Regelkreis gut dämpfen. Da darüber hinaus die Dynamik der Regelstrecke – ein einzelnes Verzögerungsglied – sehr einfach ist, hängt das Verhalten im nächsten Schritt lediglich von der aktuellen Regelabweichung sowie der Reglerverstärkung ab. Eine Verbesserung der Schätzung des Prozessparameters führt daher immer zu einer Verbesserung der Regelung.

Auch ist schon des Öfteren in den Diskussionen der Verfahren angeklungen, dass die Anforderungen, die ein Bediener an die Regelung stellt, deutlich strenger sind, als es mit (einfachen) Stabilitätskriterien zu beschreiben wäre.

6.10 Messungen

Abschließend soll eine Messung an dem Beispielsystem der Druckmaschine vorgestellt werden. Bei diesen Versuchen wurde die in Tabelle 6.8 angegebene Folge an „Aufträgen“ abgearbeitet. Dabei wurde jeweils ins OK geregelt und dann mit dem nächsten Auftrag fortgefahren, wobei der Schätzer während diesen Blöcken nicht resetet wurde.

Entsprechung der Größen

In Tabelle 6.7 sind die in der Messung verwendeten Größen den bisher hier verwendeten Bezeichnungen gegenübergestellt.

Tabelle 6.7: Entsprechung der Größen

Ausgangsgröße	y	Färbung	F
Primäre Eingangsgröße	u_1	Temperatur Rasterwalze	T
Sekundäre Eingangsgröße	u_2	Druckgeschwindigkeit	v
Primärer Parameter	θ_1		k_T
Sekundärer Parameter	θ_2		k_v

Die Messgröße wird als F bezeichnet und ist in IST angegeben. Dabei bedeutet standardmäßig ein IST-Wert von 200, dass der Farbort (bestmöglich) getroffen wurde, die Farbdichte also ideal ist. Ein IST-Wert von 180 sagt aus, dass die Farbdichte um 10% erhöht werden muss, um den gewünschten Farbort zu erreichen. Es kann jedoch auch während einer Regelung der IST-Sollwert verändert werden, was hier in den Aufträgen drei und vier gemacht wurde.

Der Soll-Farbort bezieht sich immer auf einen Referenzbogen. Im Versuch wird dieser zweimal gewechselt (von Referenzbogen A auf B beim vierten Auftrag und wieder zurück beim fünften Auftrag). Dies stellt aus Sicht der Regelung jeweils einen Sprung des Parameters θ_0 dar.

Tabelle 6.8: „Aufträge“ im Versuch

Auftrag	Testform	Ref.	Wechsel Druckgeschwindigkeiten und Sollwert
1	T1	A	9 000 Dr/h → 12 000 Dr/h → 15 000 Dr/h
2	T2	A	9 000 Dr/h → 15 000 Dr/h
3	T2	A	9 000 Dr/h → 15 000 Dr/h → $F_{\text{soll}} = 180$ IST
4	T1	B	9 000 Dr/h → 15 000 Dr/h → $F_{\text{soll}} = 220$ IST
5	T1	A	9 000 Dr/h

Referenzbogen: A: 12 000 Dr/h, 32,5 °C, B: 9 000 Dr/h, 20 °C

Darstellung in den Plots

Die senkrechten gestrichelten Linien markieren Wechsel der Druckgeschwindigkeit und Änderungen der Sollgröße, die durchgezogenen Linien markieren neue „Aufträge“.

In den Plots für \tilde{F} stellt die waagrechte durchgezogene Linie den Sollwert und die gepunkteten Linien die 3%- und 6%-Grenzen dar.

Bei den Plots für \hat{k}_T und \hat{k}_V markiert die waagrechte durchgezogene Linie den Parameterwert, der (nachträglich) aus allen Messdaten der jeweiligen Versuchswoche ermittelt wurde.

Ergebnisse

Aufgrund der guten Startwerte hat es bei allen Druckwerken etwas gedauert, bis k_T geschätzt werden konnte. Das Regelverhalten entspricht den Anforderungen. So verbleibt man nach dem Start immer innerhalb der 6%-Toleranzgrenze, meistens auch in der 3%-Toleranzgrenze.

Die hohen Abweichungen zu Beginn der Aufträge vier und fünf bei Schritt 12 und 18 liegen daran, dass bei dem vierten Auftrag ein anderer Referenzbogen mit niedrigerer Flächendeckung verwendet wurde, ohne dies dem Regler „mitzuteilen“.

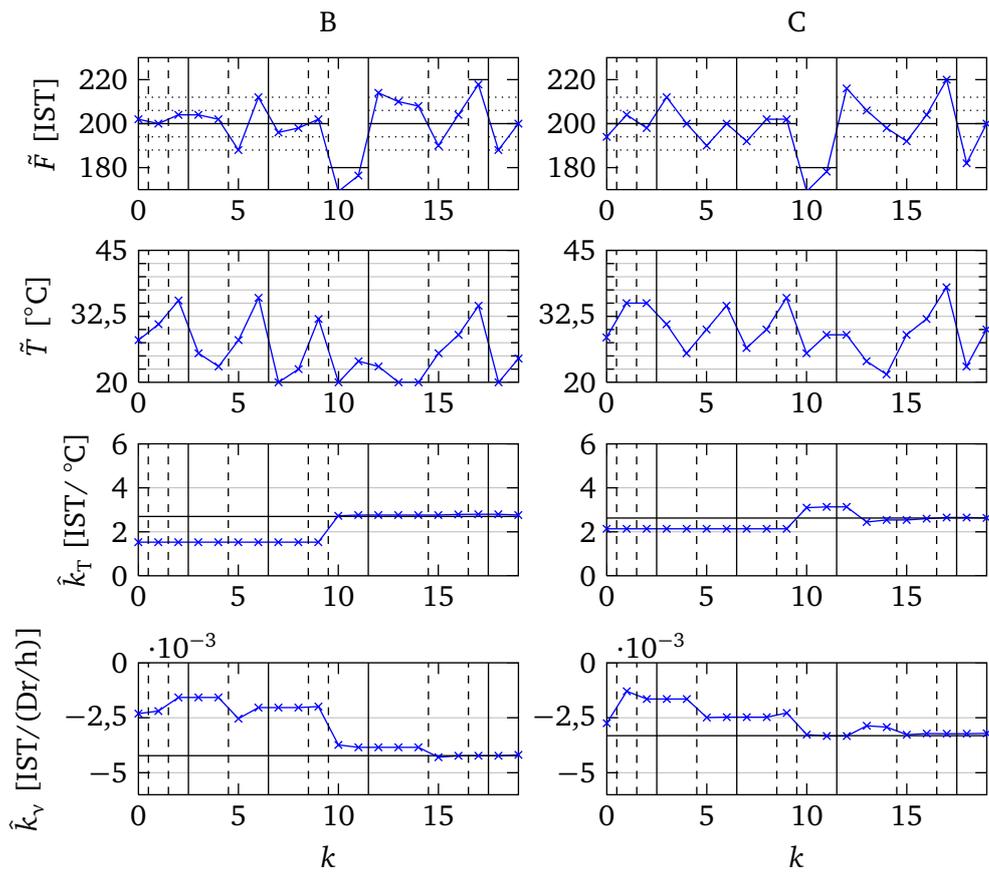
Bei Druckwerk B war der Parameter \hat{k}_T anfangs stark unterschätzt, was die große Schwankung der gemessenen Ausgangsgröße von $k = 5$ auf $k = 6$ erklärt. Mit der ersten Sollwertänderung bei $k = 10$ konnte der Schätzwert dann aber korrigiert werden.

6.11 Zusammenfassung

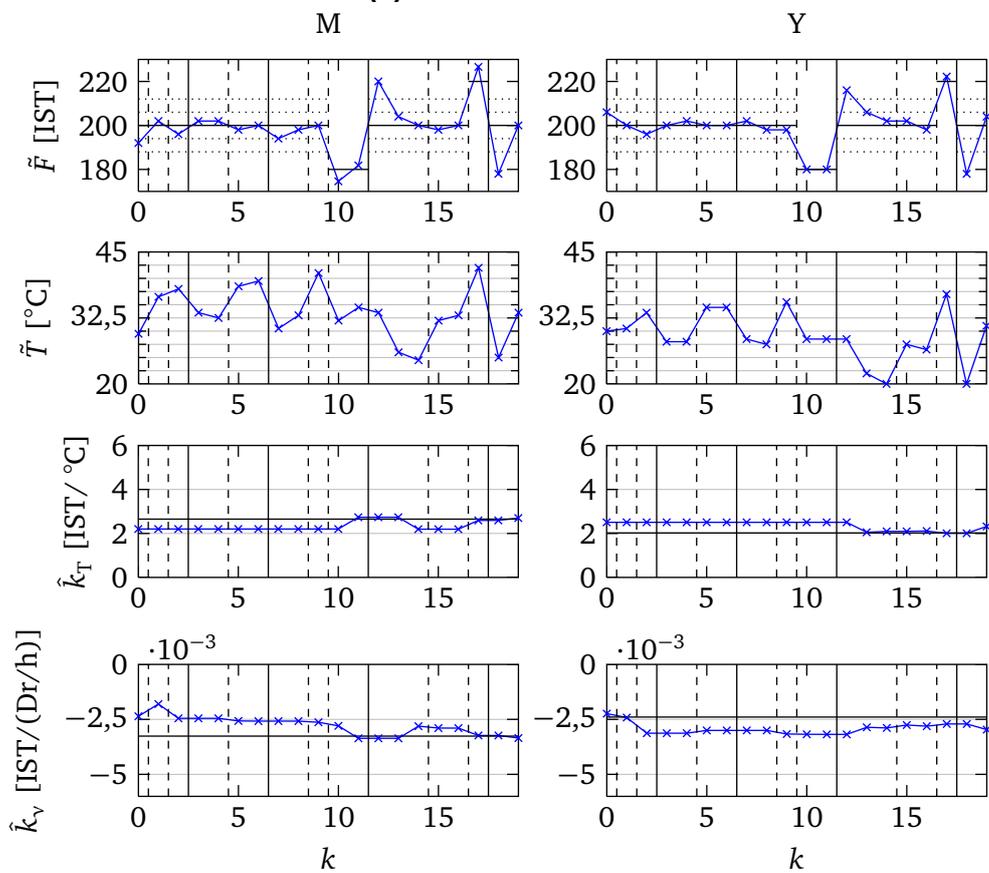
In diesem Kapitel wurde die Online-Identifikation der Prozessparameter θ behandelt. Dazu wurde nach der Präzisierung der Identifikationsaufgabe die Literatur aufgeführt und kurz diskutiert, die sich mit der gegebenen Thematik befasst.

Es wurde das bekannte Least-Squares-Verfahren, welches die Basis des hier vorgestellten Verfahrens bildet, eingeführt und dessen Eigenschaften sowie insbesondere die Einflussfaktoren auf die Varianz besprochen.

Daraufhin wurde die Anwendung des LS-Verfahrens auf die Identifikation während eines Auftrags besprochen. Hierbei ist besonders das neu entwickelte Verfahren zur Bestimmung der sekundären Parameter hervorzuheben. Aufgrund der Tatsache, dass diese für die Regelung immer im Verhältnis zum primären Parameter verwendet werden, konnte dargelegt werden, dass es vorteilhaft ist, die sekundären



(a) Druckwerke B und C



(b) Druckwerke M und Y

Abbildung 6.46: Messdaten aus Versuchen an der realen Druckmaschine

Parameter bezüglich eines festen primären Parameters zu schätzen. Dabei konnte ebenfalls gezeigt werden, dass die so ermittelten Werte später immer exakt an sich ändernde Referenzwerte des primären Parameters angepasst werden können. Dabei bedeutet „exakt“, dass sich nach der Korrektur für die sekundären Parameter die Werte ergeben, die sich auch ergeben hätten, wenn direkt mit den neuen Referenzwerten gerechnet worden wäre.

Danach wurde die Identifikation über mehrere Aufträge behandelt, wobei gezeigt wurde, dass eine Mittelung der in einzelnen Aufträgen ermittelten Schätzwerte die günstigste Variante darstellt. Es wurden verschiedene Möglichkeiten zur Mittelung besprochen und Anhaltspunkte zur Auswahl und Parametrierung abgeleitet.

Im anschließenden Abschnitt wurde auf den Bias eingegangen, der durch die Identifikation im geschlossenen Regelkreis sowie der eingeführten Varianzgrenzen entsteht.

Es wurden verschiedene „kritische“ Fälle besprochen, die bei der Online-Identifikation auftreten können, und Erweiterungen vorgestellt, um diese zu vermeiden bzw. deren Einfluss zu reduzieren.

Nach der Zusammenfassung des Algorithmus wurde dieser noch kurz in den Kontext der Adaptiven Regelung eingeordnet sowie ein Beispiel für die Messungen am realen System gezeigt.

7 Zusammenfassung

Ausgangspunkt dieser Arbeit ist das Bestreben, Einrichtvorgänge bei Druckmaschinen mit minimaler Makulatur, d. h. minimalem Ausschuss durchzuführen. Als Qualitätsmerkmal wird dabei die Färbung betrachtet. Dies bedeutet, dass eine gegebene Sollfärbung mit einer möglichst geringen Anzahl an Regelschritten erreicht werden soll. Dabei sollen unbekannte Streckenparameter anhand der Messungen, die im Rahmen der normalen Regelschritte durchgeführt werden, identifiziert werden.

Dieser Aufgabenstellung liegt ein einfaches, statisches Prozessmodell zugrunde. Die wesentlichen Herausforderungen liegen hier darin, die Identifikation mit einer äußerst geringen Anzahl an Messungen durchzuführen, sowie dem Wunsch, bezogen auf das Rauschen sehr enge zulässige Toleranzgrenzen für die Regelabweichung erreichen zu können.

Im Kapitel *Festlegung des Untersuchungsgegenstandes* werden ausgehend von dem motivierenden Anwendungsbeispiel das Prozessmodell sowie die Randbedingungen definiert, welche in dieser Arbeit behandelt werden.

Regelung

Nach der kurzen Besprechung der Regelung des ungestörten Systems im Kapitel *Regelung des ungestörten Systems* sowie der Einführung der verwendeten, stochastischen Größen und Zusammenhänge in *Modellierung der Störungen* befasst sich das Kapitel *Regelung des gestörten Systems* mit der Regelung des vorliegenden stochastischen Systems.

Dazu werden die Regelziele in eine der stochastischen Natur des Systems angebrachte Form gebracht und Größen eingeführt, die eine Bewertung und den Vergleich verschiedener Regler zulassen. Dabei ist ein besonderer Wert darauf gelegt, die grundsätzlichen Grenzen bezüglich der erreichbaren Güte herauszustellen.

Es wird gezeigt, dass die Minimierung des Mean-Squared-Errors ein geeignetes globales Ziel ist, welches aus den anderen Zielen folgt. Basierend auf dieser Erkenntnis werden ein MSE-optimaler Regler sowie weitere Varianten eines I-Reglers entworfen und verglichen.

Identifikation/Adaption

Den zweiten, hauptsächlichen Schwerpunkt dieser Arbeit stellt die Identifikation der Prozessparameter dar, welche im Kapitel *Online-Identifikation und Adaption* behandelt wird.

Es zeigt sich, dass aufgrund der am Beispielsystem beobachteten Parameterunsicherheiten eine Aufteilung der Identifikation in zwei Schritte die günstigste Herangehensweise darstellt. D. h. es werden anhand der Messdaten jeweils eines Auftrags die Parameter geschätzt und diese Einzelschätzwerte dann über mehrere Aufträge gemittelt. Bezüglich der Mittelung der Einzelschätzwerte werden zwei Methoden verglichen.

Bei der Identifikation der sekundären Parameter, d. h. die Identifikation der Abhängigkeit der Färbung von der Druckgeschwindigkeit im Fall des betrachteten Druckprozesses, tritt durch die gegebenen Randbedingungen besonders das Problem der Multikollinearität auf. Dies bedeutet, dass aufgrund der starken linearen Abhängigkeit die Varianzen der geschätzten Parameter sehr hoch sind. Zur Lösung dieser Problematik wird eine dem Verwendungszweck der Parameter angepasste Lösung vorgeschlagen, die auf einer Schätzung der sekundären Parameter bezüglich des aktuell verwendeten primären Parameters beruht.

Die so gewonnenen und gemittelten Schätzwerte können nachträglich an einen sich ändernden Wert des primären Parameters angepasst werden, womit eine maximale Ausnutzung der zur Verfügung stehenden Messdaten erreicht wird.

Eine weitere Besonderheit bezüglich der Identifikation entsteht daraus, dass zum einen in einem geschlossenen Regelkreis identifiziert wird, und zum anderen nur wenige Messungen für jeden Auftrag anfallen. Durch die Identifikation im geschlossenen Regelkreis sind die Schätzwerte biasbehaftet. Dieser Bias kann aufgrund der nur sehr geringen Anzahl an Messungen kritische Größenordnungen annehmen. Damit spielt der Bias hier eine große Rolle und wird entsprechend ausführlich untersucht. Neben der Darstellung der Einflussfaktoren auf den Bias wird in dieser Arbeit auch eine kennfeldbasierte Methode zur Kompensation, d. h. Verminderung des Bias vorgeschlagen.

Zuletzt werden Erweiterungen vorgeschlagen, die auch bei extremen Abweichungen der Anfangswerte der Parameter sowie einem Fehler in der angenommenen Varianz des Messrauschens zu einem guten Regelverhalten führen.

Ausblick

In dieser Arbeit wurden Prozesse betrachtet, die linear in den Parametern und affin in den Eingängen sind. Um eine größere Klasse von Prozessen behandeln zu können, könnten auch weitere Abhängigkeiten von den Eingangsgrößen zugelassen werden. So könnte der Prozessausgang auch von höheren Potenzen der Eingangsgrößen abhängen und es könnten auch Produkte verschiedener Eingangsgrößen in der Prozessgleichung zugelassen werden.

Solange der Prozess linear in den Parametern ist, ist eine lineare Least-Squares-Schätzung prinzipiell genauso möglich, was eingangs auch erwähnt wurde. Die Auswirkungen der Schätzung mit festgehaltenem primären Parameter, welche hier den wesentlichen Vorteil erbracht hat, wären jedoch für diesen Fall neu zu untersuchen und zu bewerten.

In den Untersuchungen wurde ein normalverteiltes weißes Rauschen angenommen. Es hat sich gezeigt, dass die Verteilung des Rauschens nur eine untergeordnete Rolle spielt. Verwendet man beispielsweise gleichverteiltes weißes Rauschen mit gleicher Varianz, so ist keine signifikante Verschlechterung des Regelverhaltens zu beobachten. Liegt jedoch eine Korrelation der Rauschwerte zu verschiedenen Abtastpunkten vor, so kann dies einen negativen Einfluss auf die Güte der Adaption und Regelung haben. Wird ein Prozess betrachtet, bei dem die Annahme eines weißen Rauschens nicht (näherungsweise) erfüllt ist, wäre dies im Entwurf zu berücksichtigen.

A Stochastik

A.1 Verteilungen

In diesem Abschnitt sind die wesentlichen Gleichungen und Eigenschaften der Normal- sowie der Gleichverteilung angegeben. Diese finden sich beispielsweise in [STORM, 2001, S. 62ff] (Normalverteilung) und [STORM, 2001, S. 61f] (Gleichverteilung).

Normalverteilung

Die Verteilungsdichte der Normalverteilung $N(\mu, \sigma^2)$ ist als

$$\varphi(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

definiert. Damit ergibt sich für die Verteilungsfunktion

$$\Phi(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt .$$

Mit der Abkürzung

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-\tau^2} d\tau$$

erhält man auch

$$\Phi(x; \mu, \sigma^2) = \frac{1}{2} \left(1 + \operatorname{erf} \left(\frac{x - \mu}{\sqrt{2}\sigma} \right) \right) .$$

Standardnormalverteilung

Die Standardnormalverteilung ist eine Normalverteilung mit dem Mittelwert $\mu = 0$ und der Varianz $\sigma^2 = 1$,

$$\begin{aligned} \varphi(x) &= \varphi(x; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} , \\ \Phi(x) &= \Phi(x; 0, 1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt . \end{aligned}$$

Die Verteilungsdichtefunktion und die Verteilungsfunktion sind in Abbildung A.1 dargestellt.

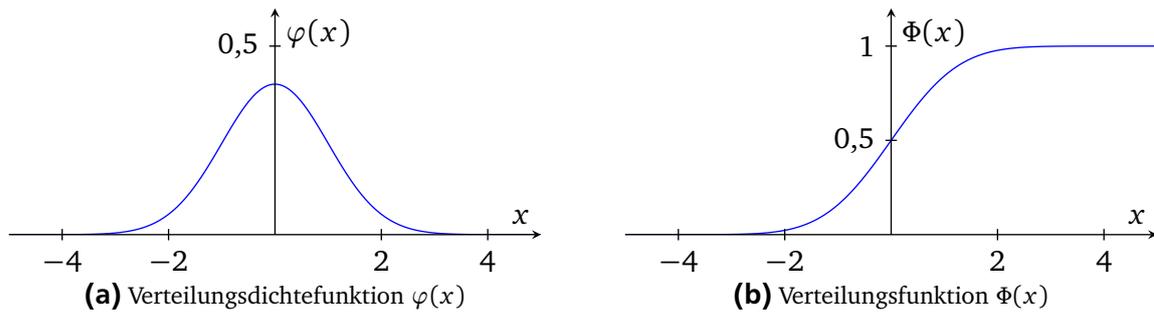


Abbildung A.1: Standardnormalverteilung $N(0, 1)$

Normalverteilung

Die Verteilungsdichtefunktion und Verteilungsfunktion einer beliebigen Normalverteilung $\{x\}_p \sim N(\mu, \sigma)$ kann über

$$\varphi(x; \mu, \sigma^2) = \frac{1}{\sigma} \varphi\left(\frac{x - \mu}{\sigma}\right)$$

und

$$\Phi(x; \mu, \sigma^2) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

aus den jeweiligen Funktionen der Standardnormalverteilung berechnet werden.¹

Die Wahrscheinlichkeit, dass eine Realisierung x einer normalverteilten Zufallsvariable $\{x\}_r \sim N(\mu, \sigma^2)$ einen Wert im Intervall (a, b) annimmt, ist

$$P(a < \{x\}_r^\omega < b) = \Phi(b; \mu, \sigma^2) - \Phi(a; \mu, \sigma^2) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right). \quad (\text{A.1})$$

Für den Sonderfall $a = \mu - c$ und $b = \mu + c$, d. h. dass eine Realisierung x innerhalb des Intervalls $[\mu - c, \mu + c]$ liegt, erhält man

$$P(|\{x\}_r^\omega - \mu| < c) = \Phi\left(\frac{c}{\sigma}\right) - \Phi\left(\frac{-c}{\sigma}\right) = \Phi\left(\frac{c}{\sigma}\right) - \left(1 - \Phi\left(\frac{c}{\sigma}\right)\right) = -1 + 2\Phi\left(\frac{c}{\sigma}\right), \quad (\text{A.2})$$

wobei von der Symmetrieeigenschaft $\Phi(-z) = 1 - \Phi(z)$ der Verteilungsfunktion der Standardnormalverteilung Gebrauch gemacht wurde.

Die Wahrscheinlichkeit, dass eine Realisierung x außerhalb des Intervalls $[\mu - c, \mu + c]$ liegt, beträgt damit

$$P(|\{x\}_r^\omega - \mu| \geq c) = 1 - P(|\{x\}_r^\omega - \mu| < c) = 2 - 2\Phi\left(\frac{c}{\sigma}\right). \quad (\text{A.3})$$

Für folgende speziellen Intervalle sind die Wahrscheinlichkeiten $P(|\{x\}_r^\omega - \mu| < c)$ hier angegeben:

- 68,27 % aller Werte liegen in einer $\pm\sigma$ -Umgebung von μ
- 95,45 % aller Werte liegen in einer $\pm 2\sigma$ -Umgebung von μ
- 99,73 % aller Werte liegen in einer $\pm 3\sigma$ -Umgebung von μ

¹ $\Phi\left(\frac{x - \mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x - \mu}{\sigma}} e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x e^{-\frac{(u - \mu)^2}{2\sigma^2}} du$ mit der Substitution $u = \sigma \cdot t + \mu$.

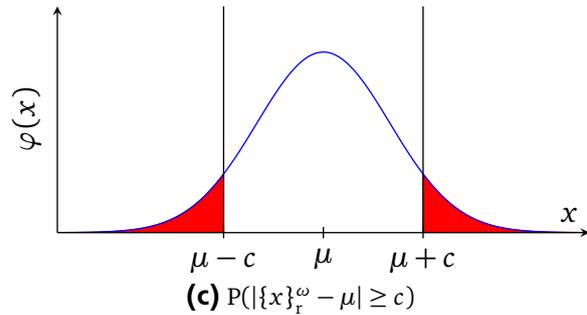
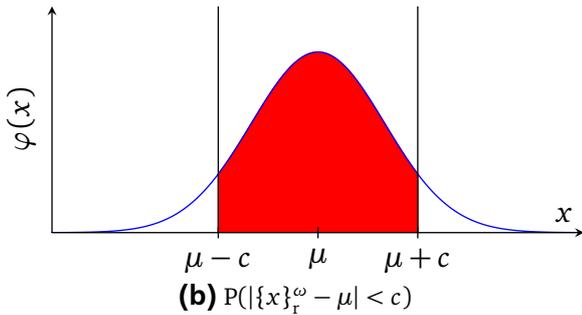
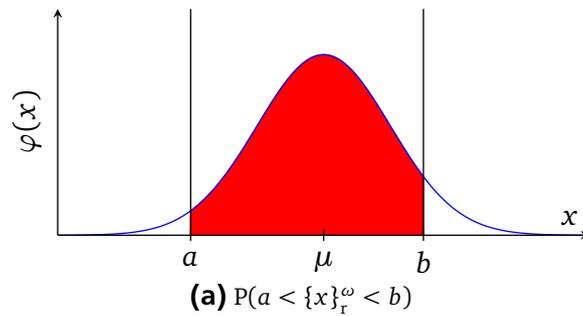


Abbildung A.2: Darstellung der Wahrscheinlichkeiten dafür, dass $\{x\}_r^\omega$ in bestimmten Intervallen liegt

Gleichverteilung

Schreibweise der Gleichverteilung im Intervall $[\mu - a, \mu + a]$:

$$\{x\}_r \sim G(\mu, a).$$

Es gilt für die Verteilungsfunktion

$$F_{\{x\}_r}(x) = \begin{cases} 0 & \text{für } x < \mu - a \\ \frac{1}{2a} \cdot (x + a) & \text{für } \mu - a \leq x \leq \mu + a \\ 1 & \text{für } x > \mu + a \end{cases}$$

und für die Verteilungsdichte

$$f_{\{x\}_r}(x) = \begin{cases} 0 & \text{für } x < \mu - a \vee x > \mu + a \\ \frac{1}{2a} & \text{für } \mu - a \leq x \leq \mu + a. \end{cases}$$

Damit ergibt sich auch der Mittelwert und die Varianz von $\{x\}_r \sim G(\mu, a)$ zu

$$E\{x\}_r = \mu \quad \text{und} \\ \text{Var}\{x\}_r = \int_{-a}^a x^2 \frac{1}{2a} dx = \frac{1}{3} a^2.$$

A.2 Rechnungen zur bedingten Verteilung

Hier sind die Rechnungen aufgeführt, die in Abschnitt 4.5.1 benutzt wurden, um die Abbildungen 4.2 und 4.3 zu erstellen.

Bestimmung der Verteilung $\{\tilde{\varepsilon}_1\}_r$

Die gesuchte Zufallsvariable ist über den Zusammenhang

$$\tilde{\varepsilon}_1 = \varepsilon_1 - n_1$$

gegeben.

Zur Bestimmung der Verteilung von $\{\tilde{\varepsilon}_1\}_r$ wird die Dichte $f_{\{n_1\}_r, \{\varepsilon_1\}_r}$ benötigt. Da $\{n_1\}_r$ und $\{\varepsilon_1\}_r$ unkorreliert sind, ergibt sich diese Dichtefunktion einfach über die Multiplikation beider einzelner Dichtefunktionen, und man erhält mit $D = d_{\max} - d_{\min}$

$$f_{\{n_1\}_r, \{\varepsilon_1\}_r}(n_1, \varepsilon_1) = \varphi(n_1) \cdot \begin{cases} \frac{1}{D} & d_{\min} < d < d_{\max} \\ 0 & \text{sonst} \end{cases}$$

Diese Funktion ist für $\{\varepsilon_1\}_r \sim G(2,5, 5)$ auch in Abbildung A.3a dargestellt.

Nach [RINNE, 2008, S. 221] ergibt sich die gesuchte Dichtefunktion zu

$$\begin{aligned} f_{\{\tilde{\varepsilon}_1\}_r}(\tilde{\varepsilon}_1) &= \int_{-\infty}^{\infty} f_{\{n_1\}_r, \{\varepsilon_1\}_r}(\varepsilon_1 - \tilde{\varepsilon}_1, \varepsilon_1) \cdot d\varepsilon_1 \\ &= \int_{-\infty}^{\infty} \varphi(\varepsilon_1 - \tilde{\varepsilon}_1) \cdot \begin{cases} \frac{1}{D} & d_{\min} < d < d_{\max} \\ 0 & \text{sonst} \end{cases} \cdot d\varepsilon_1 \\ &= \frac{1}{D} \cdot \int_{d_{\min}}^{d_{\max}} \varphi(\varepsilon_1 - \tilde{\varepsilon}_1) \cdot d\varepsilon_1 \\ &= \frac{1}{D} \cdot \int_{d_{\min}}^{d_{\max}} \varphi(\varepsilon_1; \tilde{\varepsilon}_1, 1) \cdot d\varepsilon_1 \\ &= \frac{1}{D} \cdot (\Phi(d_{\max}; \tilde{\varepsilon}_1, 1) - \Phi(d_{\min}; \tilde{\varepsilon}_1, 1)) \end{aligned}$$

Bestimmung der bedingten Verteilung $\{n_1|\tilde{\varepsilon}_1\}_r$

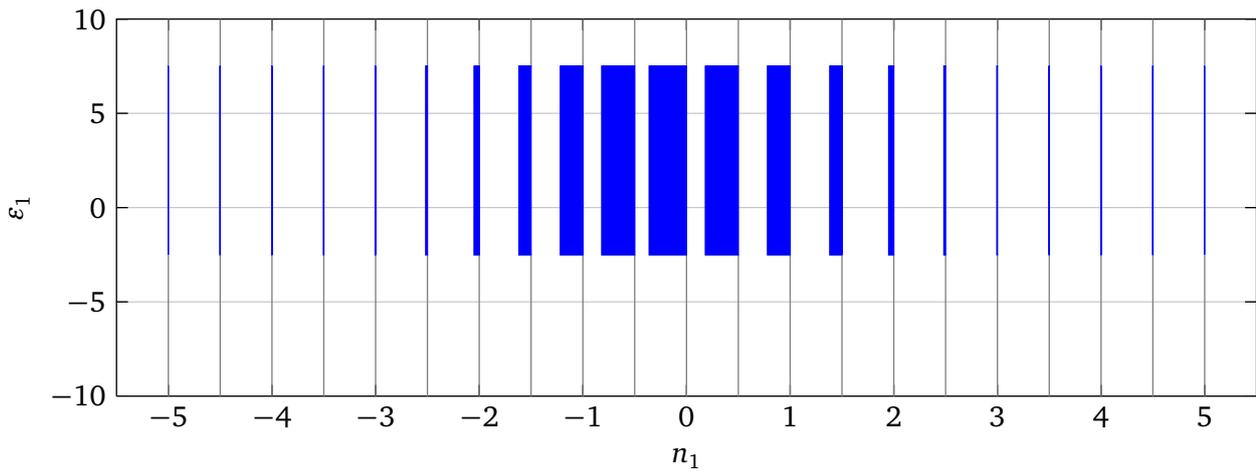
Zur Bestimmung der bedingten Verteilung $\{n_1|\tilde{\varepsilon}_1\}_r$ nach [RINNE, 2008, S. 199f] muss die zweidimensionale Verteilungsdichte $f_{\{n_1\}_r, \{\tilde{\varepsilon}_1\}_r}(n_1, \tilde{\varepsilon}_1)$ bekannt sein. Hier sind $\{n_1\}_r$ und $\{\tilde{\varepsilon}_1\}_r$ nicht mehr unkorreliert.

Es gilt

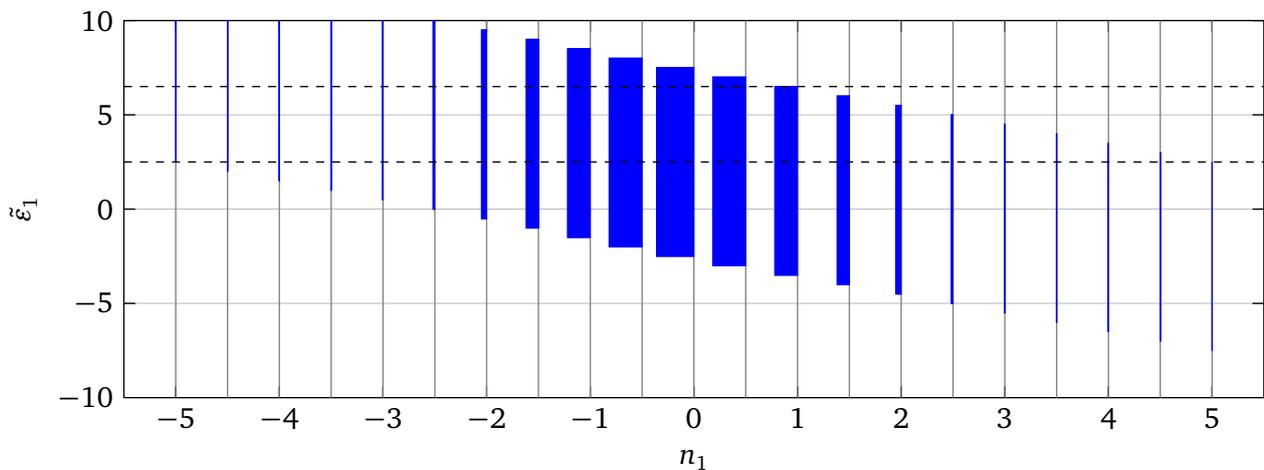
$$\tilde{\varepsilon}_1 = \varepsilon_1 - n_1,$$

wobei $\{\varepsilon_1\}_r$ und $\{n_1\}_r$ unkorreliert sind. Somit erhält man $f_{\{n_1\}_r, \{\tilde{\varepsilon}_1\}_r}(n_1, \tilde{\varepsilon}_1)$ anschaulich aus Abbildung A.3a, indem man die Werte um n_1 entlang der ε_1 -Achse verschiebt. Es ergibt sich die in Abbildung A.3b dargestellte Funktion, die

$$\begin{aligned} f_{\{n_1\}_r, \{\tilde{\varepsilon}_1\}_r}(n_1, \tilde{\varepsilon}_1) &= \varphi(n_1) \cdot \begin{cases} \frac{1}{D} & \tilde{\varepsilon}_1 + n_1 > d_{\min} \quad \wedge \quad \tilde{\varepsilon}_1 + n_1 < d_{\max} \\ 0 & \text{sonst} \end{cases} \\ &= \varphi(n_1) \cdot \begin{cases} \frac{1}{D} & \tilde{\varepsilon}_1 - d_{\max} < n_1 < \tilde{\varepsilon}_1 - d_{\min} \\ 0 & \text{sonst} \end{cases} \end{aligned}$$



(a) Zweidimensionale Dichte $f_{\{n_1\}_r, \{\varepsilon_1\}_r}(n_1, \varepsilon_1)$



(b) Zweidimensionale Dichte $f_{\{n_1\}_r, \{\tilde{\varepsilon}_1\}_r}(n_1, \tilde{\varepsilon}_1)$

Abbildung A.3: Zweidimensionale Dichten (Der Wert der Dichtefunktionen ist jeweils durch die Dicke der senkrechten Linien dargestellt.)

lautet. Daraus kann die Randdichte

$$\begin{aligned}
 f_{\{\tilde{\varepsilon}_1\}_r}(\tilde{\varepsilon}_1) &= \int_{-\infty}^{\infty} f_{\{n_1\}_r, \{\tilde{\varepsilon}_1\}_r}(n_1, \tilde{\varepsilon}_1) \cdot dn_1 \\
 &= \frac{1}{D} \cdot \int_{\tilde{\varepsilon}_1 - d_{\max}}^{\tilde{\varepsilon}_1 - d_{\min}} \varphi(n_1) \cdot dn_1 \\
 &= \frac{1}{D} \cdot (\Phi(\tilde{\varepsilon}_1 - d_{\min}) - \Phi(\tilde{\varepsilon}_1 - d_{\max}))
 \end{aligned}$$

bestimmt werden. Zuletzt ergibt sich die gesuchte Verteilung zu

$$\begin{aligned}
 f_{\{n_1 | \tilde{\varepsilon}_1\}_r}(n_1) &= \frac{f_{\{n_1\}_r, \{\tilde{\varepsilon}_1\}_r}(n_1, \tilde{\varepsilon}_1)}{f_{\{\tilde{\varepsilon}_1\}_r}(\tilde{\varepsilon}_1)} \\
 &= \begin{cases} \frac{\varphi(n_1)}{\Phi(\tilde{\varepsilon}_1 - d_{\min}) - \Phi(\tilde{\varepsilon}_1 - d_{\max})} & \tilde{\varepsilon}_1 - d_{\max} < n_1 < \tilde{\varepsilon}_1 - d_{\min} \\ 0 & \text{sonst.} \end{cases}
 \end{aligned}$$

Dies entspricht der auf das Intervall $[\tilde{\varepsilon}_1 - d_{\max}, \tilde{\varepsilon}_1 - d_{\min}]$ gestutzten Normalverteilung [RINNE, 2008, S. 220].

A.3 Optimale gewichtete Mittelung von Zufallszahlen

Zwei Zufallszahlen

Es wird mit $\tau \in [0, 1]$ der gewichtete Mittelwert der zwei Zufallszahlen $\{x_a\}_r$ und $\{x_b\}_r$ berechnet,

$$x_{ab} = \tau x_a + (1 - \tau) \cdot x_b.$$

Die Frage ist, wie τ gewählt werden muss, damit $\{x_{ab}\}_r$ die geringstmögliche Varianz besitzt. Dabei sind die Varianzen $\text{Var}\{x_a\}_r = \sigma_a^2$ und $\text{Var}\{x_b\}_r = \sigma_b^2$ bekannt.

Für die Varianz von $\{x_{ab}\}_r$ gilt

$$\sigma_{x_{ab}}^2 = \tau^2 \sigma_a^2 + (1 - \tau)^2 \cdot \sigma_b^2$$

und damit

$$\frac{d\sigma_{x_{ab}}^2}{d\tau} = 2\sigma_a^2 \tau - 2 \cdot (1 - \tau) \cdot \sigma_b^2.$$

Mit der Forderung $\frac{d\sigma_{x_{ab}}^2}{d\tau} \stackrel{!}{=} 0$ ergibt sich

$$\tau_{\text{opt}} = \frac{\sigma_b^2}{\sigma_a^2 + \sigma_b^2}$$

als kritischer Punkt. Da dies der einzige kritische Punkt ist und die Varianz für $\tau \rightarrow \pm\infty$ gegen unendlich geht, beschreibt dieser das Minimum und damit das optimale τ .

Für $\sigma_a^2 \gg \sigma_b^2$ ergibt sich $\tau_{\text{opt}} \approx 0$, d. h. x_a wird quasi verworfen ($x_{ab} \approx x_b$). Entsprechend gilt bei $\sigma_a^2 \ll \sigma_b^2$, dass $\tau_{\text{opt}} \approx 1$ ist, und damit $x_{ab} \approx x_a$.

Für die minimale Varianz ergibt sich durch Einsetzen

$$\sigma_{x_{ab}, \text{min}}^2 = \left(\frac{\sigma_b^2}{\sigma_a^2 + \sigma_b^2} \right)^2 \cdot \sigma_a^2 + \left(1 - \frac{\sigma_b^2}{\sigma_a^2 + \sigma_b^2} \right)^2 \cdot \sigma_b^2 = \frac{\sigma_b^4 \sigma_a^2 + \sigma_a^4 \sigma_b^2}{(\sigma_a^2 + \sigma_b^2)^2} = \frac{\sigma_a^2 \sigma_b^2}{\sigma_a^2 + \sigma_b^2}$$

oder

$$\sigma_{x_{ab}, \text{min}}^2 = \left(\frac{1}{\sigma_a^2} + \frac{1}{\sigma_b^2} \right)^{-1}.$$

Gewichtetes Mittel

Allgemeiner gilt für die Mittelung von N Zufallszahlen $\{x_i\}_r$ mit dem Mittelwert $E\{x_i\}_r = x$ und den Varianzen $\text{Var}\{x_i\}_r = \sigma_i^2$, dass sich der Mittelwert x_m mit der geringsten Varianz über

$$x_m = \frac{\sum \frac{1}{\sigma_i^2} x_i}{\sum \frac{1}{\sigma_i^2}}$$

ergibt. Die Varianz lautet in diesem Fall

$$\text{Var}\{x_m\}_r = \frac{1}{\sum \frac{1}{\sigma_i^2}}.$$

Das dieses gewichtete Mittel optimal bezüglich der Varianz ist ergibt sich beispielsweise aus den Eigenschaften des gewichteten Least-Squares-Verfahren ([RAWLINGS et al., 1998, S. 414ff] oder [ISERMANN und MÜNCHHOF, 2011, S. 279f], siehe auch Anhang B.2.1) und dessen Anwendung auf die Systemgleichung $y_k = x$ mit dem einzigen zu schätzenden Parameter x .

B Das Least-Squares-Verfahren

B.1 Einführung

Die Herleitungen finden sich z. B. in [ISERMANN und MÜNCHHOF, 2011, S. 205ff]. Alternative Literatur ist im Haupttext angegeben.

Es wird angenommen, dass sich der Prozess durch die Gleichung

$$y_{k'} = \theta_{00} + \theta_1 u_{1,k'} + \dots + \theta_p u_{p,k'}$$

beschreiben lässt. Dabei sind u_i die Eingangsgrößen, y die Ausgangsgröße und θ_{00} und θ_i die Parameter des Prozesses. Ziel ist es nun, die Parameter durch Messungen zu bestimmen. Die Messungen werden durch einen Messfehler n_y gestört, so dass die gemessene Größe

$$\tilde{y}_{k'} = \theta_{00} + \theta_1 u_{1,k'} + \dots + \theta_p u_{p,k'} + n_{y,k'}$$

im Allgemeinen nicht mit der realen Größe $y_{k'}$ übereinstimmt. Die geschätzten Parameter werden mit einem „Dach“ gekennzeichnet, d. h. $\hat{\theta}_{00}$ und $\hat{\theta}_i$. Berechnet man mit diesen Werten den Prozessausgang, so erhält man den Schätzwert

$$\hat{y}_{k'} = \hat{\theta}_{00} + \hat{\theta}_1 u_{1,k'} + \dots + \hat{\theta}_p u_{p,k'}$$

Der Schätzfehler

$$\hat{e}_{k'} = \tilde{y}_{k'} - \hat{y}_{k'}$$

gibt die Abweichung zwischen gemessenem und geschätztem Ausgangswert an. Ziel ist es nun, anhand von k Messungen die Schätzwerte $\hat{\theta}_{00}$ und $\hat{\theta}_i$ so zu bestimmen, dass die quadratische Summe der Schätzfehler minimal wird, es soll also das Gütemaß

$$V = \sum_{k'=1}^k \hat{e}_{k'}^2$$

minimiert werden.

Vektorielle Notation

Für die Behandlung solcher Probleme ist es günstig, die Gleichungen vektoriell zu schreiben. Der Prozess wird dabei über

$$y_{k'} = \boldsymbol{\psi}_{k'}^T \boldsymbol{\theta}$$

beschrieben, wobei $\boldsymbol{\psi}_{k'}^T$ der Messvektor und $\boldsymbol{\theta}$ der Parametervektor ist. Im vorliegenden Fall gilt also

$$\boldsymbol{\psi}_{k'}^T = \begin{bmatrix} 1 & u_{1,k'} & \dots & u_{p,k'} \end{bmatrix} \quad \text{und} \\ \boldsymbol{\theta} = \begin{bmatrix} \theta_{00} & \theta_1 & \dots & \theta_p \end{bmatrix}^T.$$

Fasst man die k Gleichungen

$$\begin{aligned} y_1 &= \theta_{00} + \theta_1 u_{1,1} + \cdots + \theta_p u_{p,1} \\ y_2 &= \theta_{00} + \theta_1 u_{1,2} + \cdots + \theta_p u_{p,2} \\ &\vdots \\ y_k &= \theta_{00} + \theta_1 u_{1,k} + \cdots + \theta_p u_{p,k} \end{aligned}$$

zusammen, so kann dies auch als

$$\mathbf{y} = \Psi \boldsymbol{\theta}$$

mit

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_k \end{bmatrix} \quad \text{und} \quad \Psi = \begin{bmatrix} \boldsymbol{\psi}_1^T \\ \vdots \\ \boldsymbol{\psi}_k^T \end{bmatrix} = \begin{bmatrix} 1 & u_{1,1} & \cdots & u_{p,1} \\ \vdots & \vdots & & \vdots \\ 1 & u_{1,k} & \cdots & u_{p,k} \end{bmatrix}$$

geschrieben werden. Mit dem Vektor der Schätzfehler $\hat{\mathbf{e}} = \tilde{\mathbf{y}} - \hat{\mathbf{y}}$ kann das Gütemaß auch

$$V = \sum_{k'=1}^k \hat{e}_{k'}^2 = \hat{\mathbf{e}}^T \hat{\mathbf{e}}$$

geschrieben werden. Für den gesuchten Parametervektor $\hat{\boldsymbol{\theta}}$ gilt, dass dieser das Gütemaß V minimieren soll,

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} V(\boldsymbol{\theta}).$$

Setzt man \mathbf{e} in V ein, so erhält man

$$\begin{aligned} V &= (\tilde{\mathbf{y}} - \Psi \hat{\boldsymbol{\theta}})^T \cdot (\tilde{\mathbf{y}} - \Psi \hat{\boldsymbol{\theta}}) \\ &= \tilde{\mathbf{y}}^T \tilde{\mathbf{y}} - \tilde{\mathbf{y}}^T \Psi \hat{\boldsymbol{\theta}} - (\Psi \hat{\boldsymbol{\theta}})^T \tilde{\mathbf{y}} + (\Psi \hat{\boldsymbol{\theta}})^T \Psi \hat{\boldsymbol{\theta}} \\ &= \tilde{\mathbf{y}}^T \tilde{\mathbf{y}} - \tilde{\mathbf{y}}^T \Psi \hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^T \Psi^T \tilde{\mathbf{y}} + \hat{\boldsymbol{\theta}}^T \Psi^T \Psi \hat{\boldsymbol{\theta}}. \end{aligned}$$

Die erste Ableitung nach dem geschätzten Parametervektor $\hat{\boldsymbol{\theta}}$ ist¹

$$\begin{aligned} \frac{dV}{d\hat{\boldsymbol{\theta}}} &= -\frac{d\tilde{\mathbf{y}}^T \Psi \hat{\boldsymbol{\theta}}}{d\hat{\boldsymbol{\theta}}} - \frac{d\hat{\boldsymbol{\theta}}^T \Psi^T \tilde{\mathbf{y}}}{d\hat{\boldsymbol{\theta}}} + \frac{d\hat{\boldsymbol{\theta}}^T \Psi^T \Psi \hat{\boldsymbol{\theta}}}{d\hat{\boldsymbol{\theta}}} \\ &= -\Psi^T \tilde{\mathbf{y}} - \Psi^T \tilde{\mathbf{y}} + 2\Psi^T \Psi \hat{\boldsymbol{\theta}} \\ &= -2\Psi^T \tilde{\mathbf{y}} + 2\Psi^T \Psi \hat{\boldsymbol{\theta}}. \end{aligned}$$

Eine notwendige Bedingung für das Minimum ist

$$\frac{dV}{d\hat{\boldsymbol{\theta}}} \stackrel{!}{=} \mathbf{0}$$

und damit erhält man

$$\hat{\boldsymbol{\theta}} = (\Psi^T \Psi)^{-1} \Psi^T \tilde{\mathbf{y}}. \tag{B.1}$$

(Da V nur ein Extremum hat und V anschaulich nach unten beschränkt und nach oben unbeschränkt ist, handelt es sich bei dem Extremum um das Minimum von V .)

¹ Dabei wird von folgenden Regeln gebrauch gemacht: $\frac{dx^T \mathbf{b}}{dx} = \mathbf{b}$, $\frac{d\mathbf{b}^T x}{dx} = \mathbf{b}$ und $\frac{dx^T A x}{dx} = 2A x$.

Zur Herleitung der Schätzgleichung $\hat{\boldsymbol{\theta}} = (\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T \tilde{\mathbf{y}}$ wurden keine Anforderungen an den Messfehler $n_{y,k'}$ gestellt.

Um folgende Eigenschaften des Schätzers zu bestimmen, müssen aber Annahmen über $n_{y,k'}$ getroffen werden.

Erwartungswert

[WOOLDRIDGE, 2003, S. 790f]

Für den Erwartungswert gilt allgemein

$$\begin{aligned} E\{\hat{\boldsymbol{\theta}}\}_r &= E\{(\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T \tilde{\mathbf{y}}\}_r \\ &= E\{(\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T (\mathbf{y} + \mathbf{n}_y)\}_r \\ &= E\{(\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T \mathbf{y} + (\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T \mathbf{n}_y\}_r. \end{aligned}$$

Der erste Summand $(\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T \mathbf{y}$ ist genau $\boldsymbol{\theta}$ und keine Zufallsvariable, womit man

$$E\{\hat{\boldsymbol{\theta}}\}_r = \boldsymbol{\theta} + E\{(\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T \mathbf{n}_y\}_r$$

erhält. Unter der Annahme, dass die Einträge in $\boldsymbol{\Psi}$ und das Messrauschen \mathbf{n}_y nicht korreliert sind, kann der Erwartungswert für das Produkt $(\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T \mathbf{n}_y$ aufgeteilt werden, so dass sich

$$E\{\hat{\boldsymbol{\theta}}\}_r = \boldsymbol{\theta} + E\{(\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T\}_r \cdot E\{\mathbf{n}_y\}_r$$

ergibt. Ist weiter das Messrauschen mittelwertfrei, d. h. $E\{\mathbf{n}_y\}_r = \mathbf{0}$, so gilt

$$E\{\hat{\boldsymbol{\theta}}\}_r = \boldsymbol{\theta}.$$

Unter den Voraussetzungen, dass

- der reale Prozess durch den angenommenen Zusammenhang beschrieben wird ($\mathbf{y} = \boldsymbol{\Psi} \boldsymbol{\theta}$),
- die Eingangsgrößen und das Rauschen nicht korreliert sind ($E\{(\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T \mathbf{n}_y\}_r = E\{(\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T\}_r \cdot E\{\mathbf{n}_y\}_r$) und
- das Rauschen mittelwertfrei ist ($E\{\mathbf{n}_y\}_r = \mathbf{0}$),

ist der Least-Squares-Schätzer demnach biasfrei.

Varianz

[WOOLDRIDGE, 2003, S. 791]

Die Kovarianzmatrix der geschätzten Parameter ist definiert als

$$\mathbf{P} = E\left\{ \left(\hat{\boldsymbol{\theta}} - E\{\hat{\boldsymbol{\theta}}\}_r \right) \cdot \left(\hat{\boldsymbol{\theta}} - E\{\hat{\boldsymbol{\theta}}\}_r \right)^T \right\}_r.$$

Unter den oben gemachten Voraussetzungen gilt $E\{\hat{\boldsymbol{\theta}}\}_r = \boldsymbol{\theta}$, und damit

$$\begin{aligned} \mathbf{P} &= E\left\{ \left((\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T \tilde{\mathbf{y}} - \boldsymbol{\theta} \right) \cdot \left((\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T \tilde{\mathbf{y}} - \boldsymbol{\theta} \right)^T \right\}_r \\ &= E\left\{ \left((\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T \mathbf{y} + (\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T \mathbf{n}_y - \boldsymbol{\theta} \right) \cdot \left((\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T \mathbf{y} + (\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T \mathbf{n}_y - \boldsymbol{\theta} \right)^T \right\}_r \\ &= E\left\{ \left((\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T \mathbf{n}_y \right) \cdot \left((\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T \mathbf{n}_y \right)^T \right\}_r \\ &= E\left\{ \left(\boldsymbol{\Psi}^T \boldsymbol{\Psi} \right)^{-1} \boldsymbol{\Psi}^T \mathbf{n}_y \mathbf{n}_y^T \boldsymbol{\Psi} \cdot \left(\boldsymbol{\Psi}^T \boldsymbol{\Psi} \right)^{-1} \right\}_r. \end{aligned} \tag{B.2}$$

Die Eingangsmatrix Ψ wird als deterministisch angenommen, so dass diese aus dem Erwartungswert herausgezogen werden kann. Außerdem gilt $((\Psi^T \Psi)^{-1})^T = (\Psi^T \Psi)^{-1}$, da $\Psi^T \Psi$ symmetrisch ist. Somit erhält man

$$\mathbf{P} = (\Psi^T \Psi)^{-1} \Psi^T \cdot E\{\mathbf{n}_y \mathbf{n}_y^T\}_r \cdot \Psi \cdot (\Psi^T \Psi)^{-1}.$$

Nimmt man weiter an, dass $\{n_{y,k'}\}_r$ unkorreliert ist und für alle k' dieselbe Varianz σ_y^2 besitzt, so ist $E\{\mathbf{n}_y \mathbf{n}_y^T\}_r = \sigma_y^2 \cdot \mathbf{I}$. Damit ergibt sich die Varianz der geschätzten Parameter zu

$$\mathbf{P} = \sigma_y^2 \cdot (\Psi^T \Psi)^{-1}. \quad (\text{B.3})$$

Varianz, wenn Eingänge auch Zufallsvariablen sind

Mit der Abkürzung $\mathbf{M} = \Psi \cdot (\Psi^T \Psi)^{-1}$ kann Gl. (B.2) auch als

$$(\mathbf{P})_{qp} = E\left\{ \sum_{i=1}^k (\mathbf{M}^T)_{qi} \sum_{j=1}^k (\mathbf{n}_y \mathbf{n}_y^T)_{ij} (\mathbf{M})_{jp} \right\}_r = E\left\{ \sum_{i=1}^k \sum_{j=1}^k (\mathbf{M}^T)_{qi} (\mathbf{n}_y \mathbf{n}_y^T)_{ij} (\mathbf{M})_{jp} \right\}_r$$

geschrieben werden. Die Erwartungswertbildung kann immer in die Summe gezogen werden, und nimmt man weiter die Unkorreliertheit von $\{\mathbf{M}\}_r$ und $\{\mathbf{n}_y\}_r$ an, so folgt

$$(\mathbf{P})_{qp} = \sum_{i=1}^k \sum_{j=1}^k E\{(\mathbf{M}^T)_{qi} (\mathbf{M})_{jp}\}_r \cdot E\{(\mathbf{n}_y \mathbf{n}_y^T)_{ij}\}_r.$$

Unter der Annahme, dass $\{n_{y,k'}\}_r$ unkorreliert ist und für alle k' dieselbe Varianz σ_y^2 besitzt,

$$E\{(\mathbf{n}_y \mathbf{n}_y^T)_{ij}\}_r = \begin{cases} \sigma_y^2 & i = j \\ 0 & i \neq j, \end{cases}$$

vereinfacht sich dies weiter zu

$$(\mathbf{P})_{qp} = \sigma_y^2 \cdot \sum_{i=1}^k E\{(\mathbf{M}^T)_{qi} (\mathbf{M})_{ip}\}_r,$$

bzw.

$$\mathbf{P} = \sigma_y^2 \cdot E\{\mathbf{M}^T \mathbf{M}\}_r = \sigma_y^2 \cdot E\{(\Psi^T \Psi)^{-1}\}_r.$$

B.1.1 Schätzung der Varianz des Messrauschens

Die Herleitung folgt hier [WOOLDRIDGE, 2003, S. 793].

Für den geschätzten Messfehler $\hat{\mathbf{n}}_y$ gilt

$$\hat{\mathbf{n}}_y = \tilde{\mathbf{y}} - \hat{\mathbf{y}} = \tilde{\mathbf{y}} - \Psi \hat{\boldsymbol{\theta}} = \tilde{\mathbf{y}} - \Psi (\Psi^T \Psi)^{-1} \Psi^T \tilde{\mathbf{y}} = (\mathbf{I} - \Psi (\Psi^T \Psi)^{-1} \Psi^T) \tilde{\mathbf{y}} = \Pi_{\Psi^\perp} \tilde{\mathbf{y}},$$

wobei $\Pi_{\Psi^\perp} = (\mathbf{I} - \Psi \cdot (\Psi^T \Psi)^{-1} \Psi^T)$ die Projektionsmatrix auf das senkrechte Komplement zum Spaltenraum von Ψ darstellt [HOFFMANN et al., 2005, S. 289].

Mit $\tilde{\mathbf{y}} = \mathbf{y} + \mathbf{n}_y = \mathbf{\Psi}\boldsymbol{\theta} + \mathbf{n}_y$ und $\mathbf{\Pi}_{\Psi^\perp}\mathbf{\Psi} = \mathbf{0}$ folgt

$$\hat{\mathbf{n}}_y = \mathbf{\Pi}_{\Psi^\perp} \cdot (\mathbf{\Psi}\boldsymbol{\theta} + \mathbf{n}_y) = \mathbf{\Pi}_{\Psi^\perp} \mathbf{n}_y .$$

Die Projektionsmatrix $\mathbf{\Pi}_{\Psi^\perp}$ ist symmetrisch, so dass $\mathbf{\Pi}_{\Psi^\perp}^T \mathbf{\Pi}_{\Psi^\perp} = \mathbf{\Pi}_{\Psi^\perp} \mathbf{\Pi}_{\Psi^\perp} = \mathbf{\Pi}_{\Psi^\perp}$ gilt. Damit ergibt sich

$$\hat{\mathbf{n}}_y^T \hat{\mathbf{n}}_y = \mathbf{n}_y^T \mathbf{\Pi}_{\Psi^\perp}^T \mathbf{\Pi}_{\Psi^\perp} \mathbf{n}_y = \mathbf{n}_y^T \mathbf{\Pi}_{\Psi^\perp} \mathbf{n}_y .$$

Weiter gilt²

$$\hat{\mathbf{n}}_y^T \hat{\mathbf{n}}_y = \text{tr}(\mathbf{\Pi}_{\Psi^\perp} \mathbf{n}_y \mathbf{n}_y^T) ,$$

und damit erhält man für den Erwartungswert

$$E\{\hat{\mathbf{n}}_y^T \hat{\mathbf{n}}_y\}_r = \text{tr}(\mathbf{\Pi}_{\Psi^\perp} \cdot E\{\mathbf{n}_y \mathbf{n}_y^T\}_r) = \sigma_y^2 \cdot \text{tr}(\mathbf{\Pi}_{\Psi^\perp}) .$$

Für die Spur von $\mathbf{\Pi}_{\Psi^\perp}$ gilt

$$\text{tr}(\mathbf{\Pi}_{\Psi^\perp}) = \text{tr}(\mathbf{I}_k) - \text{tr}(\mathbf{\Psi} (\mathbf{\Psi}^T \mathbf{\Psi})^{-1} \mathbf{\Psi}^T)$$

und mit der Umformung³ $\text{tr}(\mathbf{\Psi} (\mathbf{\Psi}^T \mathbf{\Psi})^{-1} \mathbf{\Psi}^T) = \text{tr}((\mathbf{\Psi}^T \mathbf{\Psi})^{-1} \mathbf{\Psi}^T \mathbf{\Psi}) = \text{tr}(\mathbf{I}_{p+1})$ ergibt sich letztlich

$$E\{\hat{\mathbf{n}}_y^T \hat{\mathbf{n}}_y\}_r = \sigma_y^2 \cdot (k - p - 1) .$$

Damit ist

$$\hat{\sigma}_y^2 = \frac{1}{k - p - 1} \cdot \hat{\mathbf{n}}_y^T \hat{\mathbf{n}}_y = \frac{1}{k - p - 1} \cdot \tilde{\mathbf{y}}^T \cdot \left(\mathbf{I} - \mathbf{\Psi} (\mathbf{\Psi}^T \mathbf{\Psi})^{-1} \mathbf{\Psi}^T \right) \cdot \tilde{\mathbf{y}} \quad (\text{B.4})$$

ein biasfreier Schätzer für die Varianz des Messrauschens.

B.2 Erweiterungen

B.2.1 Weighted Least Squares

[ISERMANN und MÜNCHHOF, 2011, S. 279ff]

Bei dem normalen Least-Squares-Verfahren werden die k Schätzfehler $e_{k'}$ im Gütemaß gleich gewichtet. Es können aber auch Gewichtsfunktionen $w_{k'} > 0$ eingeführt werden, mit denen eine unterschiedliche Gewichtung der einzelnen Schätzfehler erhalten wird. So lautet V in diesem Fall

$$V = w_1 \cdot e_1^2 + w_2 \cdot e_2^2 + \dots + w_k \cdot e_k^2 = \mathbf{e}^T \mathbf{W} \mathbf{e} ,$$

mit der diagonalen Gewichtungsmatrix

$$\mathbf{W} = \text{diag}([w_1 \ w_2 \ \dots \ w_k])$$

Aus diesem Gütemaß ergibt sich bei gleichem Vorgehen wie oben die Schätzgleichung

$$\hat{\boldsymbol{\theta}} = (\mathbf{\Psi}^T \mathbf{W} \mathbf{\Psi})^{-1} \mathbf{\Psi}^T \mathbf{W} \mathbf{y} .$$

Es können auch nicht-diagonale Gewichtungsmatrizen \mathbf{W} zugelassen werden. Ist die Kovarianzmatrix $E\{\mathbf{e}\mathbf{e}^T\}_r$ des Schätzfehlers $\{e_k\}_r$ bekannt, so kann

$$\mathbf{W} = \left(E\{\mathbf{e}\mathbf{e}^T\}_r \right)^{-1} .$$

gewählt werden, um die Varianz der Schätzwerte $\{\hat{\boldsymbol{\theta}}\}_r$ zu minimieren.

Das gleiche Vorgehen wie beim normalen Least-Squares-Verfahren führt darauf, dass auch das Weighted-Least-Squares-Verfahren biasfrei ist (unter gleichen Voraussetzungen) und die Kovarianzmatrix

$$\mathbf{P} = (\mathbf{\Psi}^T \mathbf{W} \mathbf{\Psi})^{-1} \cdot \mathbf{\Psi}^T \mathbf{W} \cdot E\{\mathbf{e}\mathbf{e}^T\}_r \cdot \mathbf{W}^T \mathbf{\Psi} \cdot (\mathbf{\Psi}^T \mathbf{W} \mathbf{\Psi})^{-1} .$$

besitzt.

² $\mathbf{x}^T \mathbf{A} \mathbf{x} = \sum_m \sum_l x_m a_{ml} x_l = \sum_m \sum_l a_{ml} x_l x_m = \sum_m \sum_l a_{ml} \cdot (\mathbf{x} \mathbf{x}^T)_{lm} = \sum_m (\mathbf{A} \cdot \mathbf{x} \mathbf{x}^T)_{mm} = \text{tr}(\mathbf{A} \cdot \mathbf{x} \mathbf{x}^T)$

³ Für $\mathbf{A} \in \mathbb{R}^{n \times m}$ und $\mathbf{B} \in \mathbb{R}^{m \times n}$ gilt $\text{tr}(\mathbf{A}\mathbf{B}) = \sum_{k=1}^n (\mathbf{A}\mathbf{B})_{kk} = \sum_{k=1}^n \sum_{i=1}^m a_{ki} b_{ik} = \sum_{i=1}^m \sum_{k=1}^n b_{ik} a_{ki} = \text{tr}(\mathbf{B}\mathbf{A})$.

B.2.2 Exponentielles Vergessen

[ISERMANN und MÜNCHHOF, 2011, S. 281ff]

Ein Sonderfall des Weighted-Least-Squares ist das Least-Squares mit exponentiellem Vergessen. Bei diesem hat die Gewichtungsmatrix die Form

$$\mathbf{W} = \text{diag}([\lambda^k, \lambda^{k-1}, \dots, \lambda, 1]) ,$$

wobei $0 < \lambda < 1$ ist. Damit werden die Schätzfehler der weiter zurückliegenden Messungen geringer gewichtet als die der neueren Messungen.

B.3 Iterative Berechnung

Wenn die Schätzung online erfolgen soll, dann kommt jeden Abtastschritt ein neuer Datensatz hinzu. Es müssen aber nicht in jedem Abtastschritt Produkte immer größer werdender Matrizen berechnet werden, sondern man kann die Berechnung iterativ ausführen.

Für die Informationsmatrix $\mathbf{L}_{k+1} = \Psi_{k+1}^T \Psi_{k+1}$ gilt

$$\Psi_{k+1}^T \Psi_{k+1} = \begin{bmatrix} \Psi_k^T & \psi_{k+1} \end{bmatrix} \begin{bmatrix} \Psi_k \\ \psi_{k+1}^T \end{bmatrix} = \Psi_k^T \Psi_k + \psi_{k+1} \psi_{k+1}^T$$

und damit

$$\mathbf{L}_{k+1} = \mathbf{L}_k + \psi_{k+1} \psi_{k+1}^T .$$

Analog ergibt sich

$$\mathbf{r}_{k+1} = \Psi_{k+1}^T \mathbf{y}_{k+1} = \begin{bmatrix} \Psi_k^T & \psi_{k+1} \end{bmatrix} \begin{bmatrix} \mathbf{y}_k \\ y_{k+1} \end{bmatrix} = \Psi_k^T \mathbf{y}_k + \psi_{k+1} y_{k+1} = \mathbf{r}_k + \psi_{k+1} y_{k+1} .$$

Zusammengefasst kann das LS-Verfahren damit in der Form

$$\hat{\boldsymbol{\theta}}_k = \mathbf{L}_k^{-1} \mathbf{r}_k \tag{B.5}$$

mit

$$\mathbf{L}_k = \begin{cases} \mathbf{0} & k = 0 \\ \mathbf{L}_{k-1} + \psi_k \psi_k^T & k > 0 \end{cases} \tag{B.6}$$

und

$$\mathbf{r}_k = \begin{cases} \mathbf{0} & k = 0 \\ \mathbf{r}_{k-1} + y_k \psi_k & k > 0 \end{cases} \tag{B.7}$$

geschrieben werden.

B.4 Rekursives Least-Squares-Verfahren

Eine weitere Möglichkeit den Least-Squares-Schätzwert online zu bestimmen ist die Verwendung des rekursiven Least-Squares-Algorithmus, der

$$\begin{aligned}\gamma_k &= \frac{\mathbf{P}'_k \boldsymbol{\psi}_{k+1}}{\boldsymbol{\psi}_{k+1}^T \mathbf{P}'_k \boldsymbol{\psi}_{k+1} + 1} \\ \hat{\boldsymbol{\theta}}_{k+1} &= \hat{\boldsymbol{\theta}}_k + \gamma_k \cdot (y_{k+1} - \boldsymbol{\psi}_{k+1}^T \hat{\boldsymbol{\theta}}_k) \\ \mathbf{P}'_{k+1} &= (\mathbf{I} - \gamma_k \boldsymbol{\psi}_{k+1}^T) \cdot \mathbf{P}'_k\end{aligned}$$

lautet. Hier fällt auch die Matrixinversion weg. Die Matrix \mathbf{P}' ist die durch σ_y^2 geteilte Kovarianzmatrix der geschätzten Parameter ($\mathbf{P} = \sigma_y^2 \mathbf{P}'$) und muss für den ersten Schritt vorgegeben werden. [ISERMANN und MÜNCHHOF, 2011, S. 270ff]

Alternativ kann der Algorithmus auch mit der tatsächlichen Kovarianzmatrix \mathbf{P} angegeben werden,

$$\begin{aligned}\gamma_k &= \frac{\mathbf{P}_k \boldsymbol{\psi}_{k+1}}{\boldsymbol{\psi}_{k+1}^T \mathbf{P}_k \boldsymbol{\psi}_{k+1} + \sigma_y^2} \\ \hat{\boldsymbol{\theta}}_{k+1} &= \hat{\boldsymbol{\theta}}_k + \gamma_k \cdot (y_{k+1} - \boldsymbol{\psi}_{k+1}^T \hat{\boldsymbol{\theta}}_k) \\ \mathbf{P}_{k+1} &= (\mathbf{I} - \gamma_k \boldsymbol{\psi}_{k+1}^T) \cdot \mathbf{P}_k\end{aligned}$$

Weighted-Least-Squares

$$\begin{aligned}\gamma_k &= \frac{\mathbf{P}'_k \boldsymbol{\psi}_{k+1}}{\boldsymbol{\psi}_{k+1}^T \mathbf{P}'_k \boldsymbol{\psi}_{k+1} + \frac{1}{w_{k+1}}} \\ \hat{\boldsymbol{\theta}}_{k+1} &= \hat{\boldsymbol{\theta}}_k + \gamma_k \cdot (y_{k+1} - \boldsymbol{\psi}_{k+1}^T \hat{\boldsymbol{\theta}}_k) \\ \mathbf{P}'_{k+1} &= (\mathbf{I} - \gamma_k \boldsymbol{\psi}_{k+1}^T) \cdot \mathbf{P}'_k\end{aligned}$$

[ISERMANN und MÜNCHHOF, 2011, S. 281]

Vergessensfaktor

$$\begin{aligned}\gamma_k &= \frac{\mathbf{P}'_k \boldsymbol{\psi}_{k+1}}{\boldsymbol{\psi}_{k+1}^T \mathbf{P}'_k \boldsymbol{\psi}_{k+1} + \lambda} \\ \hat{\boldsymbol{\theta}}_{k+1} &= \hat{\boldsymbol{\theta}}_k + \gamma_k \cdot (y_{k+1} - \boldsymbol{\psi}_{k+1}^T \hat{\boldsymbol{\theta}}_k) \\ \mathbf{P}'_{k+1} &= (\mathbf{I} - \gamma_k \boldsymbol{\psi}_{k+1}^T) \cdot \mathbf{P}'_k \frac{1}{\lambda}\end{aligned}$$

[ISERMANN und MÜNCHHOF, 2011, S. 283]

B.5 Varianz bei teilweise festgehaltenen Parametern

Der folgende Ausgangspunkt des Beweises ist auch in [RAWLINGS et al., 1998, S. 233f] als Aufgabe formuliert.

Aus der Aufteilung $\Psi = [\Psi_e \quad \Psi_f]$ der Eingangswerte folgt

$$\mathbf{L} = \Psi^T \Psi = \begin{bmatrix} \Psi_e^T \Psi_e & \Psi_e^T \Psi_f \\ \Psi_f^T \Psi_e & \Psi_f^T \Psi_f \end{bmatrix}.$$

Mit der entsprechenden Aufteilung

$$\mathbf{L}^{-1} = \begin{bmatrix} (\mathbf{L}^{-1})_e & (\mathbf{L}^{-1})_{ef} \\ (\mathbf{L}^{-1})_{fe} & (\mathbf{L}^{-1})_f \end{bmatrix}$$

für die Inverse gilt mit der in der Fußnote auf Seite 110 angegebenen Formel zur Inversion von Blockmatrizen

$$(\mathbf{L}^{-1})_e = (\Psi_e^T \Psi_e)^{-1} + (\Psi_e^T \Psi_e)^{-1} \cdot \Psi_e^T \Psi_f \cdot (\Psi_f^T \Psi_f - \Psi_f^T \Psi_e \cdot (\Psi_e^T \Psi_e)^{-1} \cdot \Psi_e^T \Psi_f)^{-1} \cdot \Psi_f^T \Psi_e \cdot (\Psi_e^T \Psi_e)^{-1}.$$

Die Diagonaleinträge von $(\mathbf{L}^{-1})_e$ entsprechen (bis auf den Faktor σ_y^2) den Varianzen der Schätzwerte $\hat{\theta}_e$ im Fall, dass alle Parameter geschätzt werden, während die Diagonaleinträge der Matrix $(\Psi_e^T \Psi_e)^{-1}$ den Varianzen von $\hat{\theta}_e$ entsprechen, wenn nur diese Parameter geschätzt und θ_f festgehalten werden.

Um zu zeigen, dass die Varianzen der zweitgenannten Schätzvariante geringer (oder gleich) sind, kann gezeigt werden, dass alle Diagonaleinträge der Matrix

$$\begin{aligned} \Delta &= (\mathbf{L}^{-1})_e - (\Psi_e^T \Psi_e)^{-1} \\ &= (\Psi_e^T \Psi_e)^{-1} \cdot \underbrace{\Psi_e^T \Psi_f \cdot (\Psi_f^T \Psi_f - \Psi_f^T \Psi_e \cdot (\Psi_e^T \Psi_e)^{-1} \cdot \Psi_e^T \Psi_f)^{-1} \cdot \Psi_f^T \Psi_e}_{\mathbf{M}} \cdot \underbrace{(\Psi_e^T \Psi_e)^{-1}}_{\mathbf{N}} \cdot \underbrace{(\Psi_e^T \Psi_e)^{-1}}_{\mathbf{S}} \end{aligned}$$

größer oder gleich Null sind, bzw. dass Δ positiv semidefinit ist.

Mit den angegebenen Abkürzungen hat Δ die Struktur

$$\Delta = \mathbf{S}^T \mathbf{N}^T \mathbf{M}^{-1} \mathbf{N} \mathbf{S},$$

wobei $\mathbf{S} = (\Psi_e^T \Psi_e)^{-1}$ regulär ist. Damit hat auch Δ dieselbe Definitheit wie $\mathbf{N}^T \mathbf{M}^{-1} \mathbf{N}$. Eine Matrix der Form $\mathbf{N}^T \mathbf{M}^{-1} \mathbf{N}$ mit positiv definiten \mathbf{M}^{-1} ist zumindest positiv-semidefinit. Somit ist nur noch zu zeigen, dass

$$\mathbf{M} = \Psi_f^T \cdot (\mathbf{I} - \Psi_e \cdot (\Psi_e^T \Psi_e)^{-1} \cdot \Psi_e^T) \cdot \Psi_f = \Psi_f^T \cdot \Pi_{\Psi_e \perp} \cdot \Psi_f$$

positiv definit ist. Die Matrix

$$\Pi_{\Psi_e \perp} = \mathbf{I} - \Psi_e \cdot (\Psi_e^T \Psi_e)^{-1} \cdot \Psi_e^T$$

beschreibt die orthogonale Projektion auf den Unterraum von \mathbb{R}^k , der senkrecht auf den Spalten von Ψ_e steht [HOFFMANN et al., 2005, S. 289]. Die Matrix Ψ_f kann in zwei Summanden $\Psi_{f,e}$ und $\Psi_{f,e \perp}$ zerlegt werden, $\Psi_f = \Psi_{f,e} + \Psi_{f,e \perp}$, wobei $\Psi_{f,e}$ im Spaltenraum von Ψ_e liegt und die Spalten von $\Psi_{f,e \perp}$ senkrecht zu den Spalten von Ψ_e stehen. Es gilt dann $\Pi_{\Psi_e \perp} \cdot \Psi_f = \Psi_{f,e \perp}$ und damit ergibt sich

$$\mathbf{M} = \Psi_f^T \cdot \Pi_{\Psi_e \perp} \cdot \Psi_f = (\Psi_{f,e}^T + \Psi_{f,e \perp}^T) \cdot \Psi_{f,e \perp} = \Psi_{f,e \perp}^T \cdot \Psi_{f,e \perp}.$$

Da $\Psi_{f,e \perp}$ maximalen Rang hat, was aus der Voraussetzung folgt, dass $\Psi = [\Psi_e \quad \Psi_f] = [\Psi_e \quad \Psi_{f,e} + \Psi_{f,e \perp}]$ maximalen Rang hat und $\text{rg}(\Psi_e) = \text{rg}([\Psi_e \quad \Psi_{f,e}])$ ist, ist \mathbf{M} positiv definit.

Weil Δ demnach positiv semidefinit ist, sind deren Diagonaleinträge Null oder positiv. Damit ist gezeigt, dass die Varianz der kleinen Schätzung immer besser ist als die der Gesamtschätzung.

B.6 Weiterführen einer Schätzung bei bekannten Parameteränderungen

Es sind \mathbf{L}_k und \mathbf{r}_k gegeben, die bei Messungen an einem System mit den (realen) Parametern $\boldsymbol{\theta}$ bestimmt wurden. Es wird nun davon ausgegangen, dass sich die Systemparameter *nach* diesen Messungen geändert haben. Die neuen Parameter sind

$$\boldsymbol{\theta}^* = \boldsymbol{\theta} + \boldsymbol{\delta} .$$

Gesucht ist nun eine Korrektur von \mathbf{L}_k und \mathbf{r}_k , so dass die aus den zurückliegenden Messungen erhaltenen Daten weiter genutzt werden können.

Hierzu werden die zurückliegenden Messwerte korrigiert, als ob diese mit den neuen Parametern gemessen worden wären. Für einen einzelnen, korrigierten Messwert \tilde{y}^* bedeutet das

$$\tilde{y}_{k'}^* = \tilde{y}_{k'} + \boldsymbol{\psi}_{k'}^T \boldsymbol{\delta} ,$$

bzw.

$$\tilde{\mathbf{y}}_k^* = \tilde{\mathbf{y}}_k + \boldsymbol{\Psi}_k \boldsymbol{\delta}$$

für alle Messwerte. Da nur die Messwerte, aber nicht die Eingangswerte korrigiert sind, bleibt \mathbf{L}_k unverändert, $\mathbf{L}_k^* = \mathbf{L}_k$. Für den korrigierten Vektor \mathbf{r}_k^* ergibt sich

$$\mathbf{r}_k^* = \boldsymbol{\Psi}_k^T \tilde{\mathbf{y}}_k^* = \boldsymbol{\Psi}_k^T \cdot (\tilde{\mathbf{y}}_k + \boldsymbol{\Psi}_k \boldsymbol{\delta}) = \boldsymbol{\Psi}_k^T \tilde{\mathbf{y}}_k + \boldsymbol{\Psi}_k^T \boldsymbol{\Psi}_k \boldsymbol{\delta} = \mathbf{r}_k + \mathbf{L}_k \boldsymbol{\delta} .$$

B.7 Extended-Kalmanfilter

Die Herleitung der Gleichungen des Extended-Kalmanfilters ist beispielsweise in [KAMEN und SU, 1999, S. 269ff] und [ISERMANN und MÜNCHHOF, 2011, S. 547f] zu finden. Die Anwendung zur Identifikation ist in [ISERMANN und MÜNCHHOF, 2011, S. 548f] beschrieben.

Der Extended-Kalmanfilter ist ein Verfahren zur Zustandsbeobachtung nichtlinearer Systeme. D. h. dieser bestimmt für ein System

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k, k) + \mathbf{n}_{x,k} \\ \mathbf{y}_k &= \mathbf{h}(\mathbf{x}_k, \mathbf{u}_k, k) + \mathbf{n}_{y,k} \end{aligned}$$

aus bekannten Eingangswerten \mathbf{u}_k und Messungen \mathbf{y}_k der Ausgangsgröße einen Schätzwert $\hat{\mathbf{x}}$ des Systemzustands. Darüber hinaus liefert er auch eine Schätzung \mathbf{P} der Kovarianzmatrix des geschätzten Zustands. \mathbf{n}_x ist das Zustandsrauschen und \mathbf{n}_y das Messrauschen, die die Kovarianzmatrizen

$$\begin{aligned} \mathbf{Q}_k &= E\{\mathbf{n}_{x,k} \mathbf{n}_{x,k}^T\}_r \quad \text{und} \\ \mathbf{R}_k &= E\{\mathbf{n}_{y,k} \mathbf{n}_{y,k}^T\}_r \end{aligned}$$

besitzen.

Die Gleichungen des Extended-Kalmanfilters (für diskrete Systeme) lauten

$$\text{Prädiktion: } \hat{\mathbf{x}}_{k+1,-} = \mathbf{f}(\hat{\mathbf{x}}_k, \mathbf{u}_k) \tag{B.8}$$

$$\mathbf{P}_{k+1,-} = \mathbf{A}_k \mathbf{P}_{k,+} \mathbf{A}_k^T + \mathbf{Q}_k \tag{B.9}$$

$$\text{Korrektur: } \mathbf{K}_k = \mathbf{P}_{k,-} \mathbf{C}_k^T \cdot (\mathbf{C}_k \mathbf{P}_{k,-} \mathbf{C}_k^T + \mathbf{R}_k)^{-1} \tag{B.10}$$

$$\hat{\mathbf{x}}_{k,+} = \hat{\mathbf{x}}_{k,-} + \mathbf{K}_k \cdot (\mathbf{y}_k - \mathbf{h}(\hat{\mathbf{x}}_{k,-}, \mathbf{u}_k)) \tag{B.11}$$

$$\mathbf{P}_{k,+} = (\mathbf{I} - \mathbf{K}_k \mathbf{C}_k) \cdot \mathbf{P}_{k,-} , \tag{B.12}$$

wobei $\mathbf{A}_k = \left. \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \right|_{\hat{\mathbf{x}}_k, \mathbf{u}_k}$ und $\mathbf{C}_k = \left. \frac{\partial \mathbf{h}}{\partial \mathbf{x}} \right|_{\hat{\mathbf{x}}_k, \mathbf{u}_k}$ ist.

Anwendung auf das vorliegende System

Im vorliegenden Fall vereinfachen sich die Gleichungen. Das betrachtete „System“ stellt einen statischen Zusammenhang dar und besitzt keine eigenen Zustände. Um die Parameter zu identifizieren, werden diese als Zustände $\mathbf{x} = \boldsymbol{\theta}$ interpretiert. Damit ergeben sich die System- und Ausgangsgleichungen

$$\begin{aligned}\mathbf{x}_{k+1} &= \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k, k) + \mathbf{n}_{x,k} = \mathbf{x}_k + \mathbf{n}_{x,k} \\ y_k &= h(\mathbf{x}_k, \mathbf{u}_k) + n_{y,k} = x_1 + x_2 \cdot u_1 + \cdots + x_{p+1}u_p + k_T u_k + n_{y,k}.\end{aligned}$$

Für \mathbf{A} und \mathbf{c}^T ergibt sich damit

$$\begin{aligned}\mathbf{A}_k &= \left. \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \right|_{\hat{\mathbf{x}}_k} = \mathbf{I}_{p+1} \\ \mathbf{c}_k^T &= \left. \frac{\partial h}{\partial \mathbf{x}} \right|_{\hat{\mathbf{x}}_k, \mathbf{u}_k} = [1 \quad u_{1,k} \quad u_{2,k} \quad \cdots \quad u_{p,k}].\end{aligned}$$

Damit können die Gleichungen (B.8) bis (B.12) in die Form

$$\mathbf{k}_k = \frac{\mathbf{P}_k \mathbf{c}_k}{\mathbf{c}_k^T \mathbf{P}_k \mathbf{c}_k + r_k} \quad (\text{B.13})$$

$$\hat{\mathbf{x}}_{k+1} = \hat{\mathbf{x}}_k + \mathbf{k}_k \cdot (y_k - \mathbf{c}_k^T \hat{\mathbf{x}}_k) \quad (\text{B.14})$$

$$\mathbf{P}_{k+1} = (\mathbf{I} - \mathbf{k}_k \mathbf{c}_k^T) \cdot \mathbf{P}_k + \mathbf{Q}_k \quad (\text{B.15})$$

gebracht werden.

Aufgrund der besonderen Struktur des Systems gilt hier auch $\mathbf{c}_k^T \hat{\mathbf{x}}_k = h(\mathbf{x}_k, \mathbf{u}_k)$ exakt. Man hätte \mathbf{c}_k^T auch direkt als zeitvariante Ausgangsmatrix des linearen Systems $\mathbf{x}_{k+1} = \mathbf{A}_k \mathbf{x}_k$, $y_k = \mathbf{c}_k^T \mathbf{x}_k$ auffassen können, und die Gleichungen des Kalmanfilters für lineare, diskrete Systeme anwenden können. Diese entsprechen den Gleichungen (B.8) bis (B.12), wenn Gl. (B.8) durch

$$\hat{\mathbf{x}}_{k+1|-} = \mathbf{A}_k \hat{\mathbf{x}}_k + \mathbf{B}_k \hat{\mathbf{u}}_k$$

und Gl. (B.11) durch

$$\hat{\mathbf{x}}_{k,+} = \hat{\mathbf{x}}_{k,-} + \mathbf{K}_k \cdot (y_k - (\mathbf{C}_k \hat{\mathbf{x}}_{k,-} + \mathbf{D} \mathbf{u}_k))$$

ersetzt wird. Es ergibt sich also hier das gleiche Ergebnis.

Der Kalmanfilter (B.13 – B.15) entspricht bis auf den Term \mathbf{Q}_k dem rekursiven Weighted-Least-Squares-Verfahren mit einem Gewicht $w_k = 1/r_k$. (Die weiteren Größen entsprechen sich wie folgt: $\mathbf{k} = \boldsymbol{\gamma}$, $\mathbf{c} = \boldsymbol{\psi}$, $\hat{\mathbf{x}} = \hat{\boldsymbol{\theta}}$.) Für den Term \mathbf{Q}_k gibt es keine Entsprechung, da das Least-Squares-Verfahren nur von Störungen am Ausgang ausgeht, d. h. für das Least-Squares-Verfahren gilt $\mathbf{Q}_k \equiv \mathbf{0}$.

C Herleitungen

C.1 Beweis zur Mittelung sekundärer Parameter

Es ist etwas übersichtlicher, nicht den Zusammenhang $\hat{\theta}''_{\text{elf},m}(l, l+1) = \hat{\theta}'_{\text{elf},m}(l)$ zu beweisen, sondern den etwas allgemeineren Zusammenhang

$$\hat{\theta}''_{\text{elf},m}(l, x) = \hat{\theta}'_{\text{elf},m}(l) - \mathbf{C}_{\text{elf},m}(l) \cdot (\hat{\theta}_f(x) - \hat{\theta}_f(l+1)), \quad (\text{C.1})$$

wobei $\hat{\theta}_f(x)$ ein beliebiger Bezugspunkt ist. Für $\hat{\theta}_f(x) = \hat{\theta}_f(l+1)$ geht dies wieder in den gesuchten Zusammenhang über.

Die Behauptung (C.1) wird mittels der Methode der vollständigen Induktion bewiesen (siehe z.B. [VON FINCKENSTEIN et al., 2000, S. 24]).

Induktionsanfang: Für $l = 2$ ergibt sich aus (6.22) direkt

$$\hat{\theta}''_{\text{elf},m}(2, x) = \frac{g(1) \cdot (\hat{\theta}_{\text{elf}}(1) - \mathbf{C}_{\text{elf}}(1) \cdot (\hat{\theta}_f(x) - \hat{\theta}_f(1))) + \hat{\theta}_{\text{elf}}(2) - \mathbf{C}_{\text{elf}}(2) \cdot (\hat{\theta}_f(x) - \hat{\theta}_f(2))}{g(1) + 1}. \quad (\text{C.2})$$

Auf der anderen Seite lautet Gl. (6.25) für $l = 2$

$$\hat{\theta}'_{\text{elf},m}(2) = \hat{\theta}_{\text{elf},m}(2) - \mathbf{C}_{\text{elf},m}(2) \cdot (\hat{\theta}_f(x) - \hat{\theta}_f(2)). \quad (\text{C.3})$$

Weiter gilt mit (6.24)

$$\hat{\theta}_{\text{elf},m}(2) = \frac{g(1) \cdot \hat{\theta}'_{\text{elf},m}(1) + \hat{\theta}_{\text{elf}}(2)}{g(1) + 1}, \quad (\text{C.4})$$

wobei sich $\hat{\theta}'_{\text{elf},m}(1)$ wiederum aus Gl. (6.25) mit $l = 1$ und den Anfangswerten $\hat{\theta}_{\text{elf},m}(1) = \hat{\theta}_{\text{elf}}(1)$ und $\mathbf{C}_{\text{elf},m}(1) = \mathbf{C}_{\text{elf}}(1)$ zu

$$\hat{\theta}'_{\text{elf},m}(1) = \hat{\theta}_{\text{elf}}(1) - \mathbf{C}_{\text{elf}}(1) \cdot (\hat{\theta}_f(2) - \hat{\theta}_f(1)) \quad (\text{C.5})$$

ergibt. Aus Gl. (6.26) folgt

$$\mathbf{C}_{\text{elf},m}(2) = \frac{g(1) \cdot \mathbf{C}_{\text{elf}}(1) + \mathbf{C}_{\text{elf}}(2)}{g(1) + 1}. \quad (\text{C.6})$$

Setzt man (C.4), (C.5) und (C.6) in (C.3) ein, so erhält man für $\hat{\theta}'_{\text{elf},m}(2)$ denselben Ausdruck wie für $\hat{\theta}''_{\text{elf},m}(2, x)$ in Gl. (C.2), womit der Induktionsanfang abgeschlossen ist

Induktionsschritt: Es verbleibt zu zeigen, dass Gl. (C.1) für $l = l^* + 1$,

$$\hat{\theta}''_{\text{elf},m}(l^* + 1, x) = \hat{\theta}'_{\text{elf},m}(l^* + 1) - \mathbf{C}_{\text{elf},m}(l^* + 1) \cdot (\hat{\theta}_f(x) - \hat{\theta}_f(l^* + 2)), \quad (\text{C.7})$$

gilt, unter der Voraussetzung das dieselbe Gleichung (C.1) für $l = l^*$,

$$\hat{\theta}_{\text{elf},m}''(l^*, x) = \hat{\theta}_{\text{elf},m}'(l^*) - \mathbf{C}_{\text{elf},m}(l^*) \cdot (\hat{\theta}_f(x) - \hat{\theta}_f(l^* + 1)), \quad (\text{C.8})$$

erfüllt ist.

Es gilt nach Gl. (6.22) für $l' = l^*$

$$\hat{\theta}_{\text{elf},m}''(l^* + 1, x) = \frac{g(l^*) \cdot \hat{\theta}_{\text{elf},m}''(l^*, x) + \hat{\theta}_{\text{elf}}(l^* + 1) - \mathbf{C}_{\text{elf}}(l^* + 1) \cdot (\hat{\theta}_f(x) - \hat{\theta}_f(l^* + 1))}{g(l^*) + 1}.$$

Ersetzt man $\hat{\theta}_{\text{elf},m}''(l^*, x)$ durch Gl. (C.8), so ergibt sich nach Umstellen

$$\begin{aligned} \hat{\theta}_{\text{elf},m}''(l^* + 1, x) &= \frac{g(l^*) \cdot \hat{\theta}_{\text{elf},m}'(l^*) + \hat{\theta}_{\text{elf}}(l^* + 1)}{g(l^*) + 1} \\ &\quad - \frac{g(l^*) \cdot \mathbf{C}_{\text{elf},m}(l^*) + \mathbf{C}_{\text{elf}}(l^* + 1)}{g(l^*) + 1} \cdot (\hat{\theta}_f(x) - \hat{\theta}_f(l^* + 1)). \end{aligned}$$

Der erste Summand der rechten Seite entspricht nach Gl. (6.24) gerade $\hat{\theta}_{\text{elf},m}(l^* + 1)$, und der Bruch des zweiten Summanden ist nach Gl. (6.26) $\mathbf{C}_{\text{elf},m}(l^* + 1)$, so dass sich weiter

$$\begin{aligned} \hat{\theta}_{\text{elf},m}''(l^* + 1, x) &= \hat{\theta}_{\text{elf},m}(l^* + 1) - \mathbf{C}_{\text{elf},m}(l^* + 1) \cdot (\hat{\theta}_f(x) - \hat{\theta}_f(l^* + 1)) \\ &= \hat{\theta}_{\text{elf},m}'(l^* + 1) + \mathbf{C}_{\text{elf},m}(l^* + 1) \cdot (\hat{\theta}_f(l^* + 2) - \hat{\theta}_f(l^* + 1)) \\ &\quad - \mathbf{C}_{\text{elf},m}(l^* + 1) \cdot (\hat{\theta}_f(x) - \hat{\theta}_f(l^* + 1)) \end{aligned}$$

ergibt, wobei $\hat{\theta}_{\text{elf},m}(l^* + 1)$ über Gl. (6.25) ersetzt wurde. Durch Zusammenfassen der $\mathbf{C}_{\text{elf},m}(l^* + 1)$ -Terme ergibt sich der Zusammenhang (C.7) und die Behauptung ist bewiesen.

C.2 Empirische Varianz und Kovarianz

Die im Folgenden vorgestellten Zusammenhänge können auch [WOOLDRIDGE, 2003, S. 676f] entnommen werden.

Der Mittelwert einer Anzahl k von Werten $v_{k'}$ wird im Folgenden mit

$$\bar{v} = \frac{1}{k} \sum_{k'} v_{k'}$$

abgekürzt.

Es gilt

$$\sum_{k'} \left(v_{k'} - \frac{1}{k} \sum_{k''=1}^k v_{k''} \right) = \sum_{k'} (v_{k'} - \bar{v}) = \sum_{k'} v_{k'} - \bar{v} \cdot \sum_{k'} 1 = 0,$$

was nichts weiter aussagt, dass ein mittelwertbefreites Signal den Mittelwert Null hat.

Für die gemessene quadratische Abweichung gilt

$$\begin{aligned} \sum_{k'} (v_{k'} - \bar{v})^2 &= \sum_{k'} v_{k'}^2 - 2 \cdot \bar{v} \cdot \sum_{k'} v_{k'} + \bar{v}^2 \cdot k \\ &= \sum_{k'} v_{k'}^2 - 2 \cdot k \cdot \bar{v}^2 + k \cdot \bar{v}^2 \\ &= \sum_{k'} v_{k'}^2 - k \cdot \bar{v}^2 \end{aligned} \quad (\text{C.9})$$

bzw.

$$\sum^k \left(v_{k'} - \frac{1}{k} \sum_{k''=1}^k v_{k''} \right)^2 = \sum^k v_{k'}^2 - \frac{1}{k} \left(\sum^k v_{k'} \right)^2. \quad (\text{C.10})$$

Für die gemessene „Kovarianz“ gilt

$$\begin{aligned} \sum^k (v_{k'} - \bar{v})(w_{k'} - \bar{w}) &= \sum^k v_{k'} \cdot (w_{k'} - \bar{w}) - \bar{v} \cdot \sum^k (w_{k'} - \bar{w}) \\ &= \sum^k v_{k'} w_{k'} - \bar{w} \sum^k v_{k'} \\ &= \sum^k v_{k'} w_{k'} - k \cdot \bar{v} \bar{w} \end{aligned} \quad (\text{C.11})$$

$$= \sum^k v_{k'} w_{k'} - \frac{1}{k} \cdot \sum^k v_{k'} \cdot \sum^k w_{k'}. \quad (\text{C.12})$$

C.3 Varianz bei Drei-Parameter-Schätzung

Mit dem Messvektor

$$\boldsymbol{\psi}_{k'}^T = [1 \quad v_{k'} \quad w_{k'}]$$

($w_{k'} = u_{2,k'} - u_{2,N}$) erhält man

$$\boldsymbol{\psi}_{k'} \boldsymbol{\psi}_{k'}^T = \begin{bmatrix} 1 & v_{k'} & w_{k'} \\ v_{k'} & v_{k'}^2 & v_{k'} w_{k'} \\ w_{k'} & v_{k'} w_{k'} & w_{k'}^2 \end{bmatrix}$$

und damit

$$\boldsymbol{\Psi}_k^T \boldsymbol{\Psi}_k = \sum^k (\boldsymbol{\psi}_{k'} \boldsymbol{\psi}_{k'}^T) = \begin{bmatrix} k & \sum^k v_{k'} & \sum^k w_{k'} \\ \sum^k v_{k'} & \sum^k v_{k'}^2 & \sum^k (v_{k'} w_{k'}) \\ \sum^k w_{k'} & \sum^k (v_{k'} w_{k'}) & \sum^k w_{k'}^2 \end{bmatrix}.$$

Für die auf σ_y^2 normierten Varianzen $1/\sigma_y^2 \cdot [\sigma_{\hat{\theta}_0}^2 \quad \sigma_{\hat{\theta}_1}^2 \quad \sigma_{\hat{\theta}_2}^2]^T = \text{diag}((\boldsymbol{\Psi}_k^T \boldsymbol{\Psi}_k)^{-1})$ der Schätzwerte ergibt sich daraus

$$\begin{aligned} \frac{1}{\sigma_y^2} \cdot \begin{bmatrix} \sigma_{\hat{\theta}_0}^2 \\ \sigma_{\hat{\theta}_1}^2 \\ \sigma_{\hat{\theta}_2}^2 \end{bmatrix} &= \frac{1}{\det(\boldsymbol{\Psi}_k^T \boldsymbol{\Psi}_k)} \begin{bmatrix} \sum^k v_{k'}^2 \cdot \sum^k w_{k'}^2 - \left(\sum^k (v_{k'} w_{k'}) \right)^2 \\ k \cdot \sum^k w_{k'}^2 - \left(\sum^k w_{k'} \right)^2 \\ k \cdot \sum^k v_{k'}^2 - \left(\sum^k v_{k'} \right)^2 \end{bmatrix} \\ &= \frac{k}{\det(\boldsymbol{\Psi}_k^T \boldsymbol{\Psi}_k)} \begin{bmatrix} \frac{1}{k} \sum^k v_{k'}^2 \cdot \sum^k w_{k'}^2 - \frac{1}{k} \left(\sum^k (v_{k'} w_{k'}) \right)^2 \\ \sum_{k''=1}^k (w_{k''} - \bar{w})^2 \\ \sum_{k''=1}^k (v_{k''} - \bar{v})^2 \end{bmatrix} \end{aligned} \quad (\text{C.13})$$

und für die Kovarianzen

$$\begin{aligned}
\frac{1}{\sigma_y^2} \cdot \begin{bmatrix} \text{Cov}(\{\hat{\theta}_0\}_r, \{\hat{\theta}_1\}_r) \\ \text{Cov}(\{\hat{\theta}_0\}_r, \{\hat{\theta}_2\}_r) \\ \text{Cov}(\{\hat{\theta}_1\}_r, \{\hat{\theta}_2\}_r) \end{bmatrix} &= \begin{bmatrix} ((\Psi_k^T \Psi_k)^{-1})_{2,1} \\ ((\Psi_k^T \Psi_k)^{-1})_{3,1} \\ ((\Psi_k^T \Psi_k)^{-1})_{3,2} \end{bmatrix} \\
&= \frac{1}{\det(\Psi_k^T \Psi_k)} \cdot \begin{bmatrix} \sum^k (v_{k'} w_{k'}) \cdot \sum^k w_{k'} - \sum^k v_{k'} \cdot \sum^k w_{k'}^2 \\ \sum^k (v_{k'} w_{k'}) \cdot \sum^k v_{k'} - \sum^k v_{k'}^2 \cdot \sum^k w_{k'} \\ -N \sum^k (v_{k'} w_{k'}) + \sum^k v_{k'} \cdot \sum^k w_{k'} \end{bmatrix} \\
&= \frac{k}{\det(\Psi_k^T \Psi_k)} \cdot \begin{bmatrix} \frac{1}{k} \sum^k (v_{k'} w_{k'}) \cdot \sum^k w_{k'} - \frac{1}{k} \sum^k v_{k'} \cdot \sum^k w_{k'}^2 \\ \frac{1}{k} \sum^k (v_{k'} w_{k'}) \cdot \sum^k v_{k'} - \frac{1}{k} \sum^k v_{k'}^2 \cdot \sum^k w_{k'} \\ - \sum^k (v_{k'} - \bar{v})(w_{k'} - \bar{w}) \end{bmatrix}, \quad (\text{C.14})
\end{aligned}$$

mit $\bar{v} = \frac{1}{k} \sum^k v_{k'}$, $\bar{w} = \frac{1}{k} \sum^k w_{k'}$ und

$$\begin{aligned}
\det(\Psi_k^T \Psi_k) &= - \sum^k w_{k'}^2 \cdot \left(\sum^k v_{k'} \right)^2 + 2 \sum^k v_{k'} \cdot \sum^k (v_{k'} w_{k'}) \cdot \sum^k w_{k'} - \sum^k v_{k'}^2 \cdot \left(\sum^k w_{k'} \right)^2 \\
&\quad - k \cdot \left(\sum^k v_{k'} w_{k'} \right)^2 + k \cdot \sum^k v_{k'}^2 \cdot \sum^k w_{k'}^2. \quad (\text{C.15})
\end{aligned}$$

C.4 Äquivalenz von Gl. (6.35) und (6.36)

Es soll gezeigt werden, dass Gl. (6.35),

$$\hat{\theta}_{1,m,l+1} = \frac{\sum_{l'=1}^{l+1} \lambda^{l+1-l'} \hat{\theta}_{1,l'}}{\sum_{l'=1}^{l+1} \lambda^{l+1-l'}}, \quad (\text{C.16})$$

und Gl. (6.36),

$$\hat{\theta}_{1,m,l+1} = \frac{\lambda g_{1,m,l} \cdot \hat{\theta}_{1,m,l} + \hat{\theta}_{1,l+1}}{\lambda g_{1,m,l} + 1}, \quad \text{mit } g_{1,m,l} = \sum_{l'=1}^l \lambda^{l-l'}, \quad (\text{C.17})$$

äquivalent sind.

Für $l = 1$ erhält man mit beiden Gleichungen und $\hat{\theta}_{1,m,1} = \hat{\theta}_{1,1}$

$$\hat{\theta}_{1,m,2} = \frac{\lambda \hat{\theta}_{1,1} + \hat{\theta}_{1,2}}{\lambda + 1}.$$

Jetzt ist noch zu zeigen, dass Gl. (C.17) ausgehend von

$$\hat{\theta}_{1,m,l} = \frac{\sum_{l'=1}^l \lambda^{l-l'} \hat{\theta}_{1,l'}}{\sum_{l'=1}^l \lambda^{l-l'}},$$

gleich (C.16) ist. Setzt man $g_{1,m,l}$ in (C.17) ein,

$$\hat{\theta}_{1,m,l+1} = \frac{\lambda \cdot \sum_{l'=1}^l \lambda^{l-l'} \cdot \hat{\theta}_{1,m,l} + \hat{\theta}_{1,l+1}}{\lambda \cdot \sum_{l'=1}^l \lambda^{l-l'} + 1},$$

und drückt $\hat{\theta}_{1,m,l}$ entsprechend (C.16) aus, so ergibt sich

$$\begin{aligned} \hat{\theta}_{1,m,l+1} &= \frac{\lambda \cdot \sum_{l'=1}^l \lambda^{l-l'} \cdot \frac{\sum_{l''=1}^{l'} \lambda^{l-l''} \hat{\theta}_{1,m,l''}}{\sum_{l''=1}^{l'} \lambda^{l-l''}} + \hat{\theta}_{1,l+1}}{\lambda \cdot \sum_{l'=1}^l \lambda^{l-l'} + 1} \\ &= \frac{\sum_{l'=1}^l \lambda^{l-l'+1} \hat{\theta}_{1,l'} + \lambda^0 \hat{\theta}_{1,l+1}}{\sum_{l'=1}^l \lambda^{l-l'+1} + \lambda^0} \\ &= \frac{\sum_{l'=1}^{l+1} \lambda^{l+1-l'} \hat{\theta}_{1,l'}}{\sum_{l'=1}^{l+1} \lambda^{l+1-l'}}. \end{aligned}$$

Die letzte Zeile entspricht Gl. (C.16), und damit ist die behauptete Äquivalenz gezeigt.

Für $g_{1,m,l}$ gilt

$$g_{1,m,l} = \sum_{l'=1}^l \lambda^{l-l'} = \sum_{n=l-1}^0 \lambda^n = \sum_{n=0}^{l-1} \lambda^n = \frac{1 - \lambda^l}{1 - \lambda}$$

wobei im letzten Schritt genutzt wurde, dass für die abgebrochene geometrische Reihe $\sum_{n=0}^k x^n = \frac{1-x^{k+1}}{1-x}$ für $|x| < 1$ gilt [MERZIGER et al., 2001].

C.5 Bias im Fall $k > 2$

In diesem Abschnitt wird Gl. (6.56) von Seite 150, die hier nochmals als (C.18) aufgeführt ist,

$$\Delta \hat{\theta}_{1,k} = \frac{k \cdot \sum (n_{k'} v_{k'}) - \sum v_{k'} \cdot \sum n_{k'}}{k \cdot \sum v_{k'}^2 - (\sum v_{k'})^2}, \quad (\text{C.18})$$

genauer betrachtet.

Im weiteren Verlauf wird dazu auch auf die Gl. (5.23) von Seite 62 zurückgegriffen,

$$\eta_k = -(1 - \alpha)^{k-1} \cdot (\eta_{\text{soll}} - \eta_1) + \eta_{\text{soll}} - \sum_{k'=1}^{k-1} (1 - \alpha)^{k-k'-1} \alpha \cdot n_{k'}. \quad (\text{C.19})$$

Nenner

Nach Abschnitt C.2 gilt für den Nenner von (C.18) auch¹

$$k \cdot \sum v_{k'}^2 - \left(\sum v_{k'} \right)^2 = k \cdot \sum \left(v_{k'} - \frac{1}{k} \sum_{k''} v_{k''} \right)^2.$$

¹ Wenn nicht anders angegeben, ist k' die Variable, über die die Summe gebildet wird. Die Summenbildung erfolgt immer von 1 bis k .

Weitere Umformungen ergeben

$$k \cdot \sum \left(v_{k'} - \frac{1}{k} \sum_{k''} v_{k''} \right)^2 = k \cdot \sum \left(\frac{1}{k} \sum_{k''} v_{k'} - \frac{1}{k} \sum_{k''} v_{k''} \right)^2 = k \cdot \sum \left(\frac{1}{k} \sum_{k''} (v_{k'} - v_{k''}) \right)^2,$$

in der $v_{k'} - v_{k''}$ durch

$$v_{k'} - v_{k''} = \frac{1}{\vartheta_1} \cdot (\eta_{k'} - \vartheta_{0S}) - \frac{1}{\vartheta_1} \cdot (\eta_{k''} - \vartheta_{0S}) = \frac{1}{\vartheta_1} \cdot (\eta_{k'} - \eta_{k''}) = \frac{\kappa_I}{\alpha} \cdot (\eta_{k'} - \eta_{k''})$$

ersetzt werden können. Damit ergibt sich für den Nenner von (C.18)

$$k \cdot \sum \left(v_{k'} - \frac{1}{k} \sum_{k''} v_{k''} \right)^2 = k \cdot \left(\frac{\kappa_I}{\alpha} \right)^2 \cdot \sum \left(\frac{1}{k} \sum_{k''} (\eta_{k'} - \eta_{k''}) \right)^2 = k \cdot \left(\frac{\kappa_I}{\alpha} \right)^2 \cdot \sum \left(\eta_{k'} - \frac{1}{k} \sum_{k''} \eta_{k''} \right)^2.$$

Zähler

Zunächst wird der Zähler von (C.18) zu

$$\begin{aligned} k \cdot \sum (n_{k'} v_{k'}) - \sum v_{k'} \cdot \sum n_{k'} &= \sum \left(n_{k'} \cdot \left(k \cdot v_{k'} - \sum_{k''} v_{k''} \right) \right) \\ &= k \cdot \sum \left(n_{k'} \cdot \left(v_{k'} - \frac{1}{k} \cdot \sum_{k''} v_{k''} \right) \right) \end{aligned}$$

umgeschrieben und dann weiter zu

$$k \cdot \sum \left(n_{k'} \cdot \left(v_{k'} - \frac{1}{k} \cdot \sum_{k''} v_{k''} \right) \right) = k \cdot \sum \left(n_{k'} \cdot \left(\frac{1}{k} \cdot \sum_{k''} (v_{k'} - v_{k''}) \right) \right)$$

umgeformt. Analog zu dem Vorgehen bei der Umformung des Nenners ergibt sich durch die Substitution von $v_{k'} - v_{k''}$ der Term

$$\begin{aligned} k \cdot \sum \left(n_{k'} \cdot \left(\frac{1}{k} \cdot \sum_{k''} (v_{k'} - v_{k''}) \right) \right) &= \frac{\kappa_I}{\alpha} \cdot k \cdot \sum \left(n_{k'} \cdot \left(\frac{1}{k} \cdot \sum_{k''} (\eta_{k'} - \eta_{k''}) \right) \right) \\ &= \frac{\kappa_I}{\alpha} \cdot k \cdot \sum \left(n_{k'} \cdot \left(\eta_{k'} - \frac{1}{k} \cdot \sum_{k''} \eta_{k''} \right) \right). \end{aligned}$$

Gesamter Bruch

Verwendet man die jeweils letzten Ausdrücke aus den Abschnitten „Nenner“ und „Zähler“, so kann für Gl. (C.18) auch

$$\Delta \hat{\vartheta}_{1,k} = \frac{\frac{\kappa_I}{\alpha} \cdot k \cdot \sum \left(n_{k'} \cdot \left(\eta_{k'} - \frac{1}{k} \cdot \sum_{k''} \eta_{k''} \right) \right)}{\left(\frac{\kappa_I}{\alpha} \right)^2 \cdot k \cdot \sum \left(\eta_{k'} - \frac{1}{k} \sum_{k''} \eta_{k''} \right)^2}$$

geschrieben werden. Nun wird auf beiden Seiten mit κ_1 multipliziert, so dass sich folgende (äquivalente) Möglichkeiten ergeben, $\kappa_1 \cdot \Delta \hat{\vartheta}_{1,k}$ auszudrücken:

$$\kappa_1 \cdot \Delta \hat{\vartheta}_{1,k} = \alpha \cdot \frac{\sum (n_{k'} \cdot (\eta_{k'} - \frac{1}{k} \cdot \sum_{k''} \eta_{k''}))}{\sum (\eta_{k'} - \frac{1}{k} \sum_{k''} \eta_{k''})^2} \quad (\text{C.20})$$

$$= \alpha \cdot \frac{\sum (n_{k'} \cdot \frac{1}{k} \cdot \sum_{k''} (\eta_{k'} - \eta_{k''}))}{\sum (\frac{1}{k} \cdot \sum_{k''} (\eta_{k'} - \eta_{k''}))^2} \quad (\text{C.21})$$

$$= \alpha \cdot \frac{\sum (n_{k'} \eta_{k'}) - \sum n_{k'} \cdot \sum \eta_{k'}}{k \cdot \sum \eta_{k'}^2 - (\sum \eta_{k'})^2}, \quad (\text{C.22})$$

mit $\eta_{k'}$ nach Gl. (C.19).

Diskussion

Aus Gl. (C.19) ist zu erkennen, dass η_k in der Form

$$\eta_k = \eta_{\text{soll}} + f(\alpha, \varepsilon_1, k, n_1, \dots, n_{k-1})$$

gegeben ist, wobei $f(\cdot)$ für einen aus Gl. (C.19) ablesbaren Zusammenhang der angegebenen Argumente steht. Daraus lässt sich zweierlei folgern.

Betrachtet man den Ausdruck (C.21), so erkennt man, dass nur die Differenzen von jeweils zwei Werten von η_k relevant sind. Daher hängt $k_1 \cdot \Delta \hat{\vartheta}_{1,k}$ nicht von η_{soll} ab.

Zweitens kommen im Nenner Werte von $n_{k'}$ nur über $\eta_{k'}$ vor, wobei k' maximal k ist. D.h. im Nenner kommen die Werte n_1 bis n_{k-1} , aber nicht n_k vor. Betrachtet man den Zähler, so wird ersichtlich, dass n_k dort niemals als Quadrat oder höhere Potenz auftreten kann. Damit kann bei der Bildung des Erwartungswerts n_k aus allen Summanden, in denen es vorkommt, ausgeklammert werden. Da davon ausgegangen wird, dass das Rauschen unkorreliert ist und der Erwartungswert Null ist, fallen damit alle Summanden weg, in denen der Faktor n_k vorkommt. Für die Erwartungswertbildung kann also $n_k = 0$ gesetzt werden.

Fasst man dies zusammen, so ergibt sich als Fazit

$$E\{\kappa_1 \cdot \Delta \hat{\vartheta}_{1,k}\}_r = E\{\kappa_1 \cdot \Delta \hat{\vartheta}_{1,k} | n_k = 0\}_r = f(\alpha, \varepsilon_1, \{n_1\}_r, \dots, \{n_{k-1}\}_r),$$

d. h. der Erwartungswert von $\kappa_1 \cdot \Delta \hat{\vartheta}_{1,k}$ wird neben den Verteilungen $\{n_{k'}\}_r$ des Rauschens nur von den Faktoren α und ε_1 beeinflusst.

D Schätzung von q_i

Der primäre Parameter θ_1 wird direkt für die Regelung verwendet. Die sekundären Parameter treten im Regel- bzw. Steuerungsgesetz (3.14) jedoch immer nur in der Form θ_i/θ_1 , also geteilt durch den primären Parameter auf. Daher werden hier die Eigenschaften der Schätzung des Quotienten

$$q_i = \frac{\theta_i}{\theta_1}$$

genauer betrachtet, wobei sich auf ein System mit nur einem sekundären Parameter θ_2 beschränkt wird.

D.1 Möglichkeiten zur Schätzung von q_i

Prinzipiell kann die Schätzung von $q_2 = \theta_2/\theta_1$ auf drei Arten erfolgen.

Bei der in dieser Arbeit eingeführten und verwendeten Variante werden θ_1 und θ_2 unabhängig voneinander geschätzt und gemittelt, wobei $\hat{\theta}_2$ über eine Zwei-Parameter-Schätzung mit festgehaltenem $\hat{\theta}_1 = \hat{\theta}_{1,r}$ bestimmt wird. Der Schätzwert für θ_2 wird dabei laufend an den aktuell verwendeten Wert von $\hat{\theta}_{1,r}$ angepasst. Der Wert für \hat{q}_2 wird bei Bedarf über $\hat{q}_2 = \hat{\theta}_2/\hat{\theta}_{1,r}$ berechnet. Diese Möglichkeit wird im Weiteren als Variante „2P“ bezeichnet.

Alternativ können θ_1 und θ_2 ebenfalls unabhängig voneinander geschätzt und gemittelt werden, wobei jedoch θ_2 wie θ_1 über die Drei-Parameter-Schätzung geschätzt wird. Dies bedeutet, dass bei Aufträgen mit einem Sprung in u_2 beide Parameter über eine Drei-Parameter-Schätzung geschätzt werden, während ansonsten nur θ_1 über eine Zwei-Parameter-Schätzung ermittelt wird. Die Werte für $\hat{\theta}_1$ und $\hat{\theta}_2$ werden jeweils für sich über die Aufträge gemittelt und $\hat{q}_2 = \hat{\theta}_2/\hat{\theta}_1$ bei Bedarf berechnet. Dies wird hier als Variante „3P“ bezeichnet.

Zuletzt können für jeden Auftrag mit entsprechender Anregung θ_1 und θ_2 über eine Drei-Parameter-Schätzung geschätzt werden. Aus den Werten $\hat{\theta}_1$ und $\hat{\theta}_2$ wird direkt \hat{q}_2 bestimmt. Dieses \hat{q}_2 wird dann – wie schon $\hat{\theta}_1$ (und $\hat{\theta}_2$) – über mehrere Aufträge gemittelt. In diesem Fall beruht die Schätzung für \hat{q}_2 nur auf Aufträgen, bei denen ein Sprung in u_2 aufgetreten ist. Dies ist Variante „D“.

In Tabelle D.1 sind die drei beschriebenen Varianten nochmals zusammengestellt. Zudem sind die Größen angegeben, mit denen in diesem Abschnitt die jeweiligen Schätzwerte bezeichnet werden. Dabei steht die Größe ohne „+“ im Index für den Schätzwert, der sich aus den Messdaten des aktuellen Auftrags ergibt. Die „+“-Größe ist der Wert, der sich aus der Mittelung der zurückliegenden Aufträge sowie dem Wert aus dem aktuellen Auftrag ergibt.

Für die letztgenannte Variante wird nur der gemittelte Wert $\hat{q}_{2,3P,+}$ angegeben, da der Wert aus dem aktuellen Auftrag gleich dem Wert $\hat{q}_{2,D}$ wäre, und dieser darüber hinaus für den Schätzer „3P“ nicht so aussagekräftig ist.

Der Wert von \hat{q}_2 ergibt sich in allen Fällen als Quotient von Zufallsvariablen, weshalb zunächst die Eigenschaften solcher Quotienten betrachtet werden. Insbesondere interessiert der Erwartungswert $E\{\hat{q}_2 - q_2\}_r$ sowie die Varianz $\text{Var}\{\hat{q}_2 - q_2\}_r = \text{Var}\{\hat{q}_2\}_r$.

Tabelle D.1: Varianten zur Schätzung von q_2

Variante	Größen	Kurzbeschreibung
2P	$\hat{q}_{2,2P}, \hat{q}_{2,2P,+}$	\hat{q}_2 wird bei Bedarf aus $\hat{\theta}_1$ und $\hat{\theta}_2$ bestimmt, wobei $\hat{\theta}_2$ aus der Zwei-Parameter-Schätzung mit festem $\hat{\theta}_1$ stammt. (In dieser Arbeit eingeführte und verwendete Schätzmethode.)
D	$\hat{q}_{2,D}, \hat{q}_{2,D,+}$	Schätzwert \hat{q}_2 wird immer direkt als Quotient von $\hat{\theta}_2$ und $\hat{\theta}_1$ berechnet, wobei diese aus einer Drei-Parameter-Schätzung bestimmt werden. Gemittelt wird \hat{q}_2 .
3P	$\hat{q}_{2,3P,+}$	Schätzwert aus Quotient von $\hat{\theta}_2$ und $\hat{\theta}_1$, wobei $\hat{\theta}_2$ aus der Drei-Parameter-Schätzung und $\hat{\theta}_1$ je nach Anregung aus der Drei- oder Zwei-Parameter-Schätzung bestimmt wird.

D.2 Quotienten von Zufallsvariablen

Literatur

Für zwei Zufallsvariablen $\{x_1\}_r$ und $\{x_2\}_r$ mit $E\{x_i\}_r = \mu_i$, $\text{Var}\{x_i\}_r = \sigma_i^2$ und dem Korrelationskoeffizienten $\rho = \text{Cov}(\{x_1\}_r, \{x_2\}_r) / (\sigma_1 \sigma_2)$ sind für den Quotienten $\{y\}_r = \{x_1/x_2\}_r$, unter Voraussetzung der Existenz der jeweiligen Größe, in [RINNE, 2008, S. 222] folgende Näherungen für den Erwartungswert und die Varianz angegeben. Es gilt

$$E\{y\}_r \approx \frac{\mu_1}{\mu_2} \cdot \left(1 + \left(\frac{\sigma_2}{\mu_2} \right)^2 - \rho \frac{\sigma_1 \sigma_2}{\mu_1 \mu_2} \right) \quad (\text{D.1})$$

und

$$\text{Var}\{y\}_r \approx \left(\frac{\mu_1}{\mu_2} \right)^2 \cdot \left(\left(\frac{\sigma_1}{\mu_1} \right)^2 + \left(\frac{\sigma_2}{\mu_2} \right)^2 - 2\rho \frac{\sigma_1 \sigma_2}{\mu_1 \mu_2} \right). \quad (\text{D.2})$$

Der Fall, dass es sich bei x_1 und x_2 um normalverteilte Variablen handelt, wird in [MARSAGLIA, 1965] und [MARSAGLIA, 2006] behandelt, wobei es sich bei der zweiten Quelle im Prinzip um eine Erläuterung der ersten handelt. Der Ausgangspunkt ist mit dem hier vorliegenden gleich, in dem Sinne dass der Quotient von zwei Parametern einer Schätzung betrachtet wird. Eine alternative Vorgehensweise zur Berechnung der Verteilungsdichte ist in [HINKLEY, 1969] (und der Korrektur [HINKLEY, 1970]) gegeben.

Anwendung

Wendet man Gl. (D.1) auf das vorliegende Problem an, so lautet der Näherungswert für den Erwartungswert¹ des Schätzfehlers,

$$E\{\hat{q}_2 - q_2\}_r \approx q_2 \cdot \left(\frac{\sigma_{\hat{\theta}_1}^2}{\theta_1^2} - \rho_{\hat{\theta}_1 \hat{\theta}_2} \frac{\sigma_{\hat{\theta}_2} \sigma_{\hat{\theta}_1}}{\theta_2 \theta_1} \right) = q_2 \cdot \left(\frac{\sigma_{\hat{\theta}_1}^2}{\theta_1^2} - \frac{\text{Cov}(\{\hat{\theta}_1\}_r, \{\hat{\theta}_2\}_r)}{\theta_1 \theta_2} \right), \quad (\text{D.3})$$

woraus sich ablesen lässt, dass der Erwartungswert von \hat{q}_2 im Allgemeinen nicht dem wahren q_2 entspricht, auch wenn die Schätzwerte $\hat{\theta}_1$ und $\hat{\theta}_2$ biasfrei sind. So sorgt die Varianz der Zufallsvariablen im Nenner, $\sigma_{\hat{\theta}_1}^2$, für einen Bias. Hat der Korrelationskoeffizient $\rho_{\hat{\theta}_1 \hat{\theta}_2}$ bzw. die Kovarianz $\text{Cov}(\{\hat{\theta}_1\}_r, \{\hat{\theta}_2\}_r)$ das gleiche Vorzeichen wie $\theta_1 \cdot \theta_2$, so verringert dies den Bias.² Haben die genannten Terme unterschied-

¹ Es wird hierzu davon ausgegangen, dass $\theta_1 \gg \sigma_{\hat{\theta}_1}$ ist bzw. dass der Erwartungswert über eine geeignete Bedingung ermittelt wird, damit dieser sinnvoll definiert ist. Siehe dazu auch [MARSAGLIA, 2006].

² Sofern $\rho_{\hat{\theta}_1 \hat{\theta}_2} \frac{\sigma_{\hat{\theta}_2} \sigma_{\hat{\theta}_1}}{\theta_2 \theta_1}$ nicht so groß wird, dass ein betragsmäßig großer negativer Bias entsteht. Dies wäre aber auch nur für $\left| \frac{\sigma_{\hat{\theta}_2}}{\theta_2} \right| > \left| \frac{\sigma_{\hat{\theta}_1}}{\theta_1} \right|$ möglich.

liche Vorzeichen, so wird der Bias größer. Aber auch bei perfekter Korrelation $\rho_{\hat{\theta}_1\hat{\theta}_2} = 1$ bzw. -1 , je nach $\text{sign}(\theta_1\theta_2)$, und damit auch $\sigma_{\hat{\theta}_1} = \sigma_{\hat{\theta}_2}$ wird \hat{q}_2 nur dann erwartungstreu, wenn $\theta_1 = \theta_2$ wäre.

In Gl. (C.14) ist die Kovarianz der Schätzwerte $\hat{\theta}_1$ und $\hat{\theta}_2$ mit

$$\text{Cov}(\{\hat{\theta}_1\}_r, \{\hat{\theta}_2\}_r) = -\sigma_y^2 \cdot \frac{k}{\det(\Psi^T\Psi)} \cdot \sum (u_{1,k'} - \bar{u}_1)(u_{2,k'} - \bar{u}_2)$$

angegeben. Die Determinante von $\Psi^T\Psi$ ist dabei immer positiv, da diese Matrix symmetrisch und positiv definit ist. Der Faktor $\sum (u_{1,k'} - \bar{u}_1)(u_{2,k'} - \bar{u}_2)$ stellt die mit $k - 1$ multiplizierte empirische Kovarianz der Eingangsgrößen u_1 und u_2 dar. Im geregelten (bzw. gesteuerten) Betrieb ist dessen Vorzeichen positiv, wenn θ_1 und θ_2 ein unterschiedliches Vorzeichen besitzen, und negativ, wenn die Parameter gleiches Vorzeichen haben, da für die u_2 -Kompensation $\Delta u_1 = -\frac{\hat{\theta}_{2,r}}{\hat{\theta}_{1,r}} \Delta u_2$ gilt. Damit ist das Vorzeichen des Quotienten $\text{Cov}(\{\hat{\theta}_1\}_r, \{\hat{\theta}_2\}_r)/(\theta_1\theta_2)$ immer positiv und der Bias wird verringert.

Für die Varianz ergibt sich nach Gl. (D.2) der Näherungswert

$$\text{Var}\{\hat{q}_2\}_r \approx \left(\frac{\theta_2}{\theta_1}\right)^2 \cdot \left(\left(\frac{\sigma_{\hat{\theta}_2}}{\theta_2}\right)^2 + \left(\frac{\sigma_{\hat{\theta}_1}}{\theta_1}\right)^2 - 2\rho_{\hat{\theta}_1\hat{\theta}_2} \frac{\sigma_{\hat{\theta}_2}\sigma_{\hat{\theta}_1}}{\theta_2\theta_1} \right). \quad (\text{D.4})$$

Je kleiner θ_1 ist, desto größer wird auch die Varianz. Analog zum Erwartungswert zeigt sich auch wieder, dass die Korrelation der Eingangsgrößen durch die Regelung bzw. Steuerung die Varianz verringert.

In Abbildung D.1 sind die nach den in [MARSAGLIA, 1965] und [MARSAGLIA, 2006] bzw. [HINKLEY, 1969] beschriebenen Methoden die Verteilungsdichtefunktionen der Zufallsvariable

$$\{\hat{q}_2\}_r = \left\{ \begin{array}{c} \hat{\theta}_2 \\ \hat{\theta}_1 \end{array} \right\}_r$$

für verschiedene Werte des Korrelationskoeffizienten $\rho_{\hat{\theta}_1\hat{\theta}_2}$ dargestellt. Dabei ist $\{\hat{\theta}_1\}_r \sim N(0,5, 0,25^2)$ und $\{\hat{\theta}_2\}_r \sim N(-1, 0,5^2)$. Dies entspricht dem bisher meist gewählten Verhältnis $\sigma_{\hat{\theta}_{i,m,\max}}/\theta_i$ von 0,5. Somit stellen die Kurven die Verteilungsdichten für den Fall dar, wenn die Schätzwerte $\hat{\theta}_1$ und $\hat{\theta}_2$ gerade noch zur Mittelung verwendet werden dürfen.

Zusätzlich ist auch die Verteilungsdichte der Normalverteilung $N(-2, 1)$ gezeigt. Dieses stellt die idealisierte Verteilung des Schätzwertes \hat{q}_2 dar, wenn dieser gerade noch zur Mittelung verwendet werden darf.

Aus Abbildung D.1 kann abgelesen werden, dass eine hohe (hier negative) Korrelation die Streuung drastisch reduziert. Für $\rho_{\hat{\theta}_1\hat{\theta}_2} = -0,95$ ist das Maximum um -2 auch deutlich ausgeprägter als das der Normalverteilung. Es ist aber auch zu beachten, dass die Normalverteilung schneller abfällt, wenn man sich vom Mittelwert weiter entfernt. Für $\rho_{\hat{\theta}_1\hat{\theta}_2} = -0,95$ ist dies in dieser Darstellung nicht zu erkennen, jedoch für $\rho_{\hat{\theta}_1\hat{\theta}_2} = 0$ und $-0,5$. Dies bedeutet, dass die Wahrscheinlichkeit sehr großer Abweichungen von $\{\hat{q}_2\}_r = \{\hat{\theta}_2/\hat{\theta}_1\}_r$ deutlich größer ist als bei einer Normalverteilung.³

Zusammenfassend können die folgenden beiden Punkte aus der Diskussion entnommen werden:

- Bei gegebenen Verhältnissen $\sigma_{\hat{\theta}_1}/\theta_1$ und $\sigma_{\hat{\theta}_2}/\theta_2$ ist das Maximum der Verteilungsdichte deutlich besser ausgeprägt und liegt näher am wahren Wert, wenn die Korrelation hoch ist.
- Je kleiner θ_1 bzw. je größer $\sigma_{\hat{\theta}_1}$ umso höher ist die Wahrscheinlichkeit für extreme Werte von \hat{q}_2 .

³ Dies ist auch der Grund weshalb (ohne Bedingungen) für $\{\hat{\theta}_2/\hat{\theta}_1\}_r$ weder der Erwartungswert noch die Varianz definiert ist.

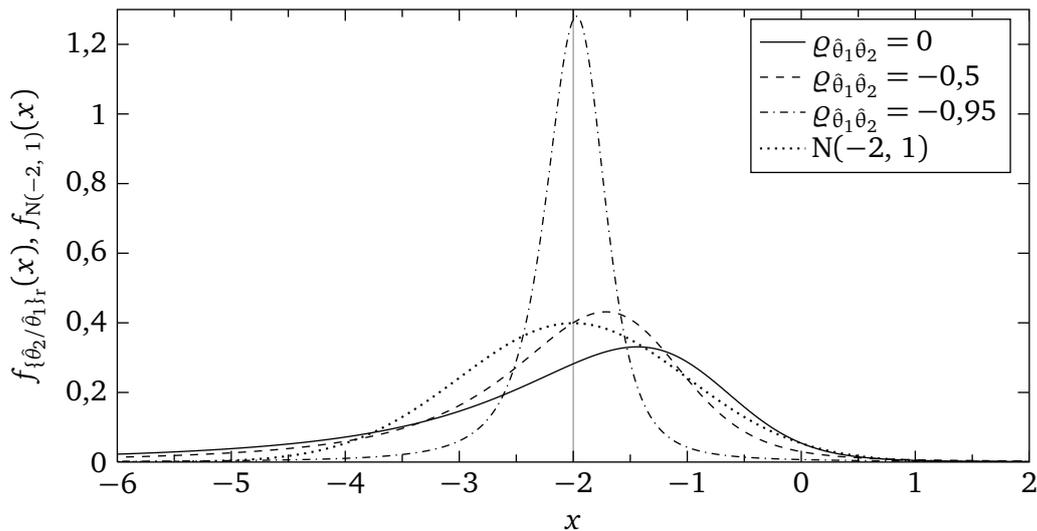


Abbildung D.1: Verteilungsdichtefunktionen für $\hat{q}_2 = \hat{\theta}_2 / \hat{\theta}_1$

D.3 Vergleich der Schätzer für q_2

Es wird wieder der Testfall aus Abschnitt 6.5.3 verwendet, wobei hier aber zunächst in jedem Auftrag ein Sprung von u_2 auftritt. Bei den Simulationen ist zur Regelung immer der Schätzer „2P“ verwendet worden. Damit beruhen die Schätzwerte aller Schätzer hier auf den gleichen Eingangsdaten.

In Abbildung D.2 sind die Verteilungen der Ausgangsgröße \tilde{y}_k , der Standardabweichungen $\sigma_{\hat{\theta}_1, k}$ und $\sigma_{\hat{\theta}_2, k}$ sowie die des Korrelationskoeffizienten $\rho_{\hat{\theta}_1, \hat{\theta}_2, k}$ aus der Drei-Parameter-Schätzung dargestellt. Daher sind für die Standardabweichungen immer erst ab den Änderungen von u_2 , die durch die senkrechten gestrichelten Linien markiert sind, Werte angegeben.

Die Verläufe variieren zwar etwas in Abhängigkeit des zur Regelung verwendeten Schätzers, jedoch sind die gezeigten Verläufe für alle im Weiteren besprochenen Fälle repräsentativ.

Im ersten und vierten Auftrag ist die Anregung aufgrund der hohen Anfangsabweichung (beim ersten Auftrag wegen schlechter Startwerte, beim vierten Auftrag wegen eines Sprungs des Parameters θ_0) und der Änderung von u_2 relativ gut, was sich in den niedrigen Werten von $\sigma_{\hat{\theta}_1}$ und $\sigma_{\hat{\theta}_2}$ zeigt. Dafür ist die Korrelation zwischen den beiden Schätzwerten nur mäßig ausgeprägt. Diese erreicht bei beiden Aufträgen jeweils einen Wert knapp unter $-0,5$.

In den anderen Aufträgen ist die Anregung nicht besonders gut. Dafür ist die Korrelation zwischen $\hat{\theta}_1$ und $\hat{\theta}_2$ sehr gut ausgeprägt und erreicht Werte nahe -1 .

Änderung von u_2 in jedem Auftrag

In Abbildung D.3 sind die Ergebnisse der verschiedenen Schätzer dargestellt. Dabei ist noch zu bemerken, dass bei den „+“-Größen, solange noch kein Wert geschätzt werden kann oder die Varianz für die Mittelung zu groß ist, der Wert $-0,667$ angegeben wird. Dies entspricht dem Quotienten der falschen Anfangswerte $\hat{\theta}_{2,r}$ und $\hat{\theta}_{1,r}$ und erleichtert das Erkennen, in welchem Anteil der Simulationen ein zur Mittelung gültiger Schätzwert erreicht wurde.

Es zeigt sich, dass alle Schätzer näherungsweise biasfrei sind. Die Varianz bzw. Streuung der Schätzung „2P“ ist die geringste und die von „D“ und „3P“ sind ähnlich, wobei „2P“ Ausreißer zeigt.

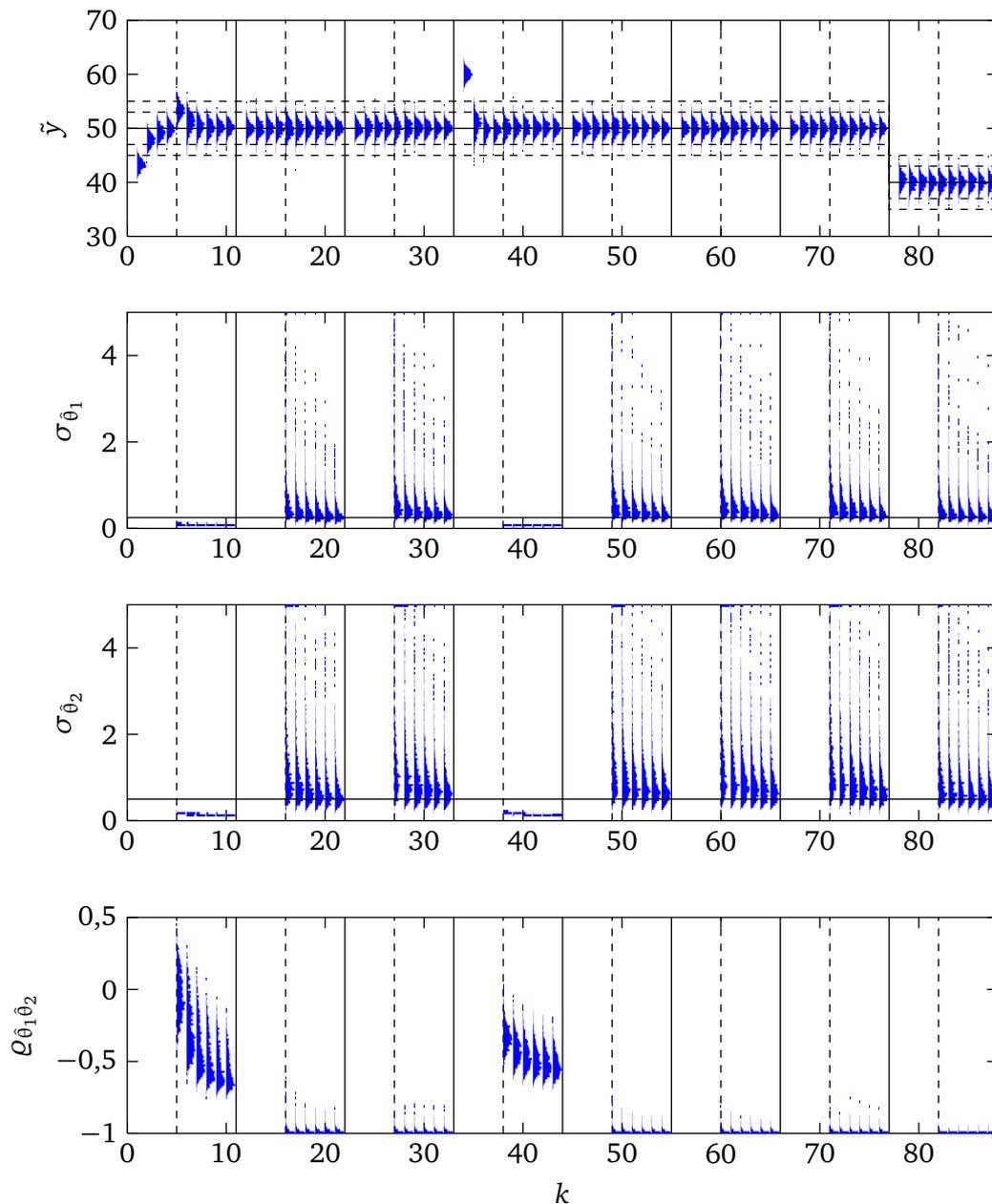


Abbildung D.2: Verteilungen der Ausgangsgröße und der Varianzen der Schätzwerte bei Sprüngen von u_2 in jedem Auftrag (Die waagrechten Linien bei $\sigma_{\hat{\theta}_1}$ und $\sigma_{\hat{\theta}_2}$ markieren die Werte für $\sigma_{\hat{\theta}_i, m, \max}$, 1 000 Simulationen)

Verdoppelt man die Grenzen der Standardabweichungen zur Mittelung (Vervierfachung der Varianzgrenzen) für den Schätzer „D“, so ergeben sich die in Abbildung D.4 dargestellten Schätzwerte. (Die Einzelschätzwerte $\hat{q}_{2,D}$ sind die selben wie in Abbildung D.3.)

Es zeigt sich eine deutlich geringere Streuung um den wahren Wert, die jetzt vergleichbar mit der des Schätzers „2P“ ist, jedoch nehmen auch die Ausreißer zu.

Änderung von u_2 in einem Teil der Aufträge

In Abbildung D.5 sind die (gemittelten) Schätzwerte der drei Schätzer für den Fall dargestellt, dass nicht mehr in allen Aufträgen eine Änderung von u_2 auftritt. Für den Schätzer „2P“ ist dies identisch zu dem

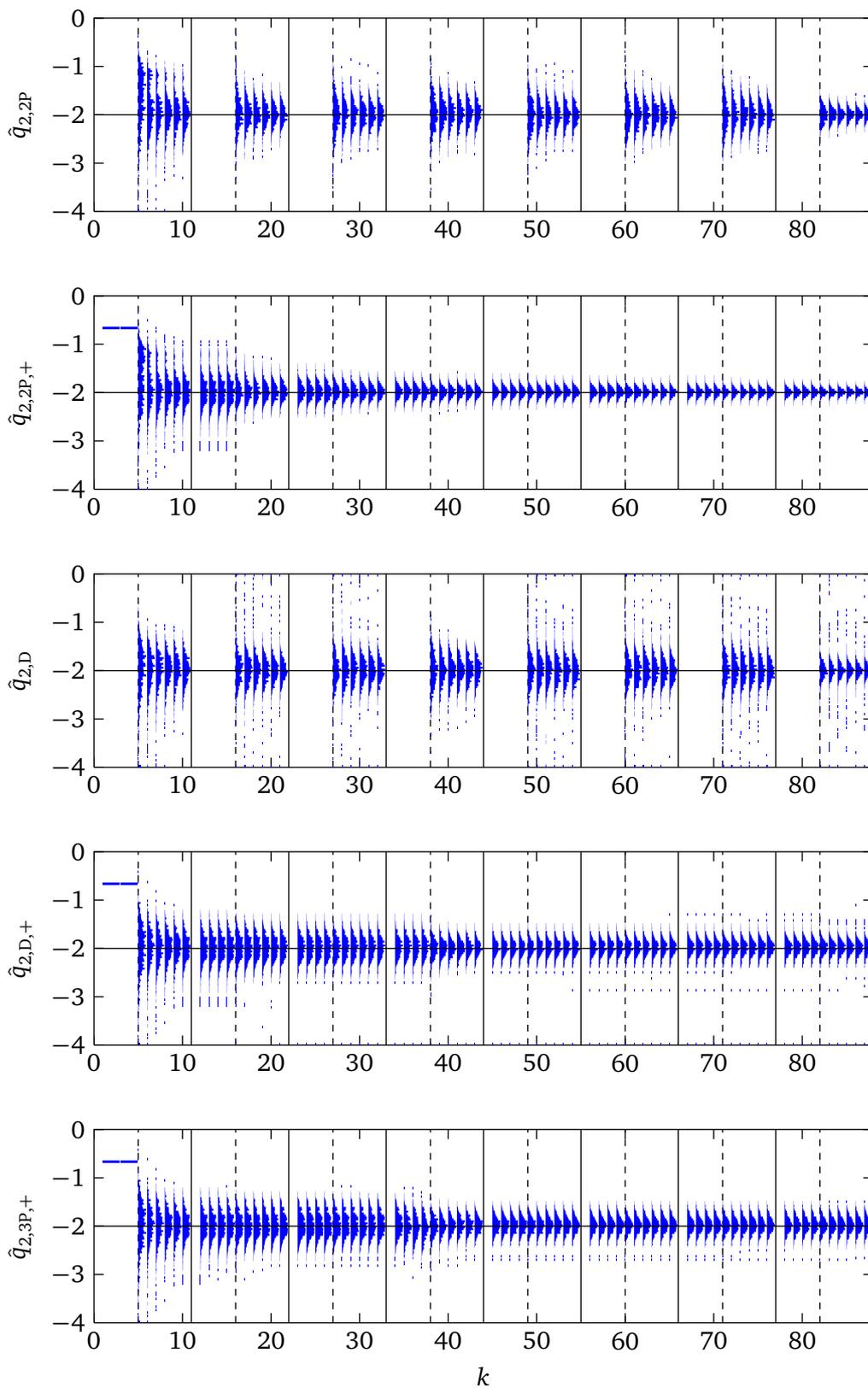


Abbildung D.3: Verteilungen der Schätzwerte \hat{q}_2 der verschiedenen Schätzer aus Tabelle D.1 (1 000 Simulationen)

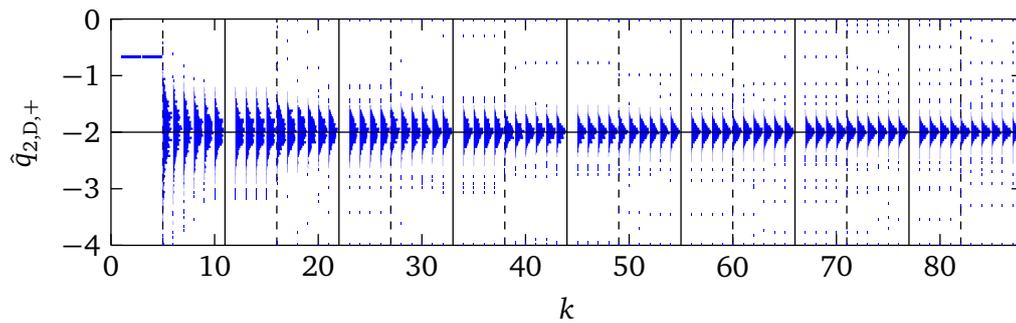


Abbildung D.4: Verteilungen der Schätzwerte \hat{q}_2 des Schätzers „D“ bei höheren Grenzen für Mittelung:
 $\sigma_{\hat{\theta}_{1,m,\max}} = \theta_1, \sigma_{\hat{\theta}_{2,m,\max}} = \theta_2$ (1 000 Simulationen)

in Abbildung 6.15 gezeigten Verlauf, für den Schätzer „3P“ wurde dieser Fall schon in Abbildung 6.16 gezeigt.⁴

Mit dem Schätzer „D“ dauert es ähnlich lange wie bei „3P“ bis die Varianz der Schätzung gut genug für eine Mittelung ist. Die Streuung ist deutlich geringer als bei „3P“, aber – besonders zum letzten Auftrag hin – noch höher als bei „2P“. Zudem treten auch vereinzelt Ausreißer auf.

Die Schätzung von „3P“ ist biasbehaftet. Dies kann hier unter anderem darauf zurückgeführt werden, dass durch die getrennte Mittelung von $\hat{\theta}_1$ und $\hat{\theta}_2$, wobei nicht in allen Aufträgen beide Parameter geschätzt werden konnten, die Korrelation zwischen diesen beiden Schätzwerten verloren geht, welche sich günstig auf die Verteilungseigenschaften von \hat{q}_2 auswirkt.

Schätzung bei dauerhaft falschem Wert für $\hat{\theta}_{1,r}$

Zuletzt ist in Abbildung D.6 noch die Schätzung über „2P“ und „3P“ für den Fall betrachtet, dass $\hat{\theta}_{1,r}$ über die ganze Simulation einen falschen Wert besitzt. (Dies wurde hier erreicht, indem die Wahl $\sigma_{\hat{\theta}_{1,m,\max}} = 0$ getroffen wurde und der falsche Anfangswert $\hat{\theta}_{1,r} = 1,5 \cdot \theta_1$ damit nie korrigiert worden ist.)

Es zeigt sich, dass der Schätzer „2P“ dabei einen leichten Bias entwickelt. Dieser entsteht in den Aufträgen eins und vier, in denen zu Beginn eine größere Anfangsabweichung vorliegt und wird im folgenden Abschnitt noch genauer betrachtet. Der Schätzer „D“ verwendet den falschen Wert $\hat{\theta}_{1,r}$ nicht und würde daher dasselbe Verhalten wie in Abbildung D.3 zeigen. Der Schätzer „3P“ zeigt einen größeren Bias. Dabei gilt ungefähr, dass $E\{\hat{q}_{2,3P,+}\}_r/q_2 = \theta_1/\hat{\theta}_{1,r} = 2/3$ ist.

D.3.1 Bias von \hat{q}_2 bei der Zwei-Parameter-Schätzung

Beim Schätzer „2P“ wird q_2 durch

$$\hat{q}_2 = \frac{\hat{\theta}_2}{\hat{\theta}_{1,r}}$$

⁴ Die Plots sind jedoch nicht identisch. Dies liegt daran, dass zum einen hier die aktuell gemittelten „+“-Werte und nicht die zur Regelung verwendeten „r“-Werte dargestellt sind. Zum anderen ist hier in allen Fällen der Schätzer „2P“ für die Regelung verwendet, so dass sich beim Schätzer „3P“ auch dadurch Unterschiede ergeben. Da der Schätzer „2P“ schneller bessere Schätzwerte als „3P“ liefert, ist die Anregung hier geringer als in Abbildung 6.16.

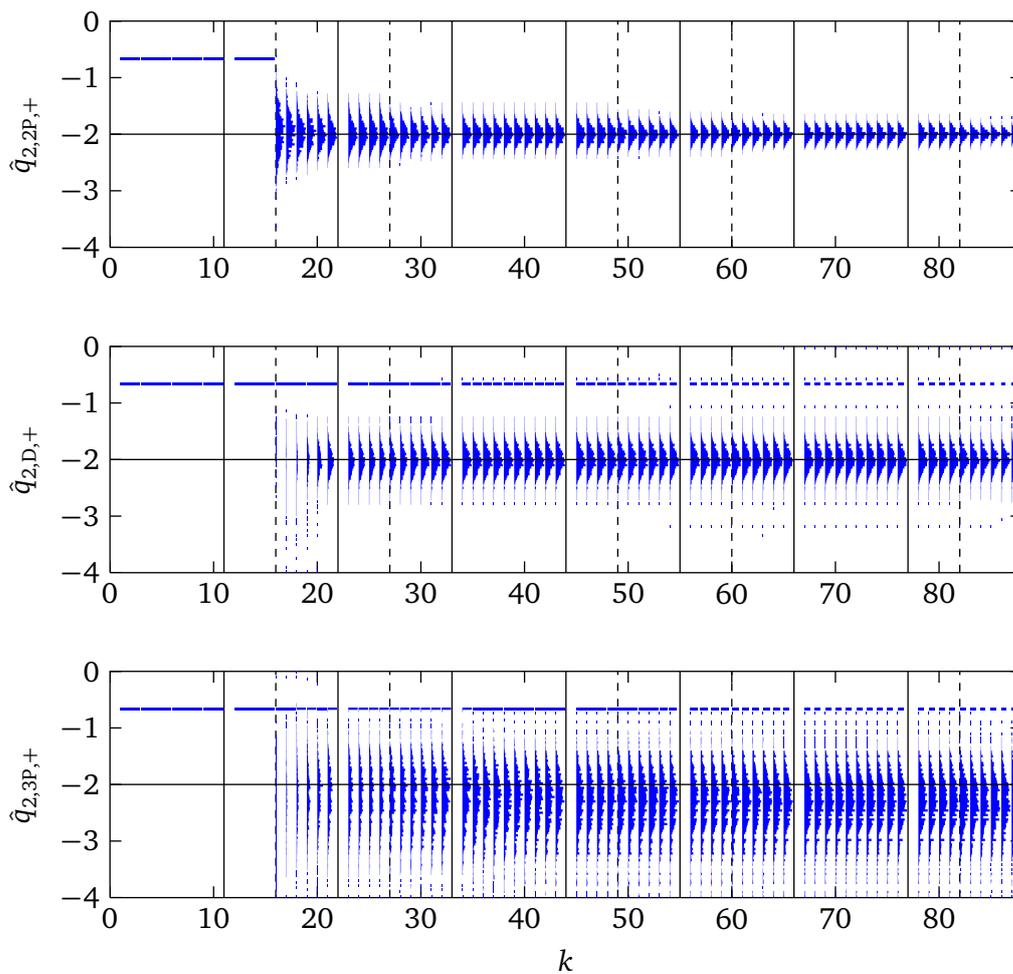


Abbildung D.5: Verteilungen der Schätzwerte \hat{q}_2 der verschiedenen Schätzer aus Tabelle D.1 bei Änderungen von u_2 nur in den Aufträgen 2, 3, 5, 6 und 7 (1 000 Simulationen)

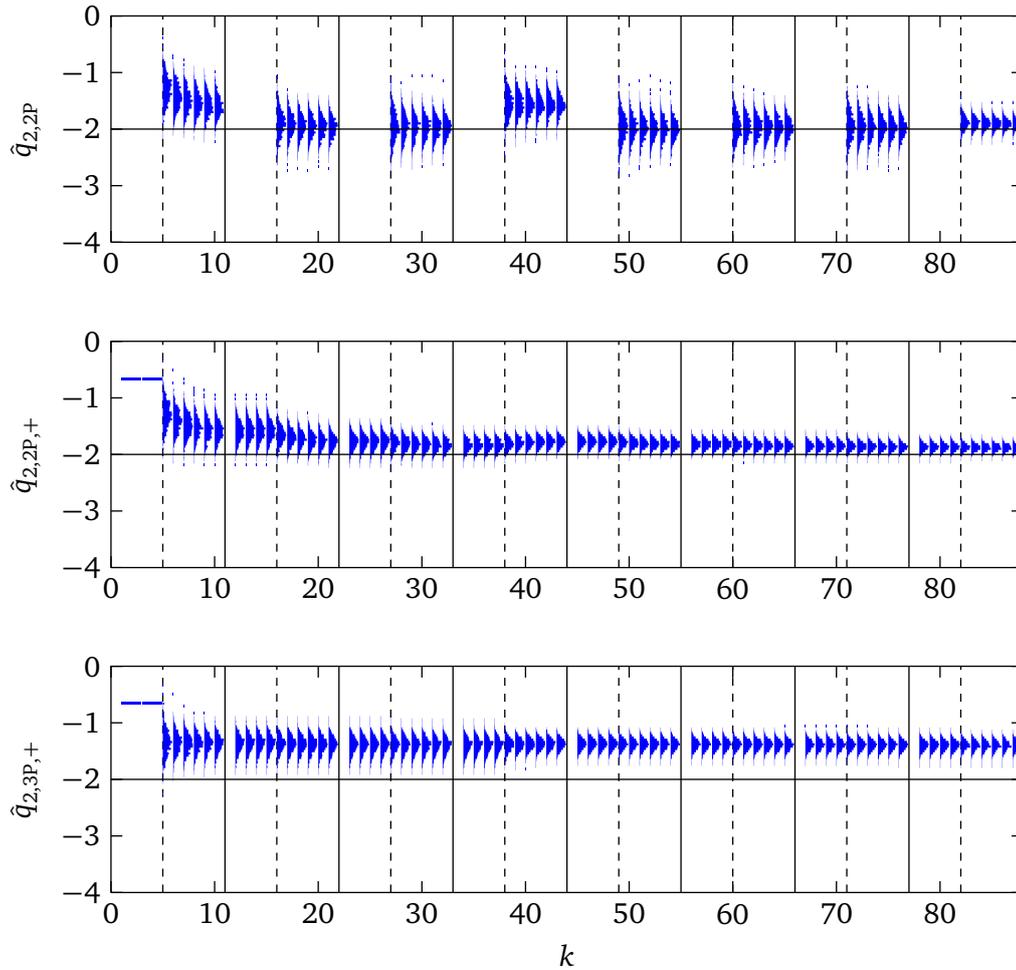


Abbildung D.6: Verteilungen der Schätzwerte \hat{q}_2 der Schätzer „2P“ und „3P“ bei $\hat{\theta}_{1,r} \equiv 1,5 \cdot \theta_1$ (1 000 Simulationen)

bestimmt, wobei $\hat{\theta}_2$ durch eine Schätzung mit festgehaltenem $\hat{\theta}_{1,r}$ bestimmt wurde. Für die Differenz des so bestimmten Schätzwertes \hat{q}_2 und dem wahren q_2 gilt

$$\begin{aligned} \hat{q}_2 - q_2 &= \begin{bmatrix} 0 & 1 \end{bmatrix} \cdot (\Psi_2^T \Psi_2)^{-1} \Psi_2^T \cdot \left(\frac{\mathbf{y} - \hat{\theta}_{1,r} \mathbf{u}_1 + \mathbf{n}}{\hat{\theta}_{1,r}} - \frac{\mathbf{y} - \theta_1 \mathbf{u}_1}{\theta_1} \right) \\ &= \begin{bmatrix} 0 & 1 \end{bmatrix} \cdot (\Psi_2^T \Psi_2)^{-1} \Psi_2^T \cdot \left(\frac{\mathbf{y}}{\hat{\theta}_{1,r}} + \frac{\mathbf{n}}{\hat{\theta}_{1,r}} - \frac{\mathbf{y}}{\theta_1} \right). \end{aligned}$$

Es wird davon ausgegangen, dass $\hat{\theta}_{1,r}$ hauptsächlich durch andere Aufträge bestimmt ist, so dass $\{n\}_r$ und $\{\hat{\theta}_{1,r}\}_r$ in guter Näherung unkorreliert sind, womit $E\{\mathbf{n}/\hat{\theta}_{1,r}\}_r = E\{\mathbf{n}\}_r \cdot E\{1/\hat{\theta}_{1,r}\}_r = 0$ gilt. Damit erhält man

$$E\{\hat{q}_2 - q_2\}_r = \underbrace{\begin{bmatrix} 0 & 1 \end{bmatrix} \cdot (\Psi_2^T \Psi_2)^{-1} \Psi_2^T \mathbf{y}}_K \cdot \left(E\left\{ \frac{1}{\hat{\theta}_{1,r}} \right\}_r - \frac{1}{\theta_1} \right).$$

Der Ausdruck K entspricht der Schätzung des Parameters θ_2 der Ansatzfunktion $y = \theta_0 + \theta_2 u_2$, also ohne den Einfluss von u_1 zu berücksichtigen,

$$K = \frac{\sum (y_{k'} - \bar{y}) \cdot (u_{2,k'} - \bar{u}_2)}{\sum (u_{2,k'} - \bar{u}_2)^2} = \frac{s_{y u_2}}{s_{u_2 u_2}}.$$

D. h. der Bias ist umso geringer

- je geringer die Abweichung von $E \left\{ \frac{1}{\hat{\theta}_{1,r}} \right\}_r - \frac{1}{\theta_1}$ und
- je geringer die Variation des (wahren) Prozessausgangs $y_{k'}$ ist.

Damit erklärt sich die Beobachtung in Abbildung D.6, dass der Bias bei gleichem, falschem $\hat{\theta}_{1,r}$ dann größer ist, wenn die Anfangsabweichung und damit die Variation von y_k größer ist.

D.3.2 Fazit

Auch hier hat sich gezeigt, dass der Schätzer „3P“ gegenüber dem Schätzer „2P“ deutlich schlechter ist. Dies liegt daran, dass

- die Korrelation der Schätzwerte nicht ausgenutzt wird,
- sich der Bias von $\hat{\theta}_1$ aufgrund der Identifikation im geschlossenen Regelkreis direkt auf \hat{q}_2 auswirkt und
- große Fehler auftreten können, solange nur einer der beiden Parameter geschätzt werden kann und für den anderen noch ein falscher Anfangswert $\hat{\theta}_{1,r}$ bzw. $\hat{\theta}_{2,r}$ verwendet wird.

Bezüglich des Schätzers „2P“ hat der Schätzer „D“ den kleinen Vorteil, dass dieser (näherungsweise) biasfrei ist, während der Schätzer „2P“ bei einem größeren falschen Wert von $\hat{\theta}_{1,r}$ einen Bias zeigt. Dies ist aber keine wesentliche Einschränkung des Schätzers „2P“. Stärker sind die Nachteile des Schätzers „D“ zu bewerten:

- Es können nur Aufträge genutzt werden, in denen sich beide Eingänge ändern.
- Aufgrund der zugrundeliegenden Verteilung treten systematisch häufiger große Ausreißer auf.

Insbesondere der letzte Punkt stellt eine wesentliche Einschränkung des Verfahrens „D“ dar.

Daher ist der Schätzer „2P“ als die für das vorliegende Problem beste Lösung identifiziert. Dieser zeigt die geringste Streuung aller Schätzer, ohne das größere Ausreißer auftreten.

Zwar kann bei falschem $\hat{\theta}_{1,r}$ ein Bias auftreten, jedoch fällt dieser geringer aus als der Fehler in $\hat{\theta}_{1,r}$, womit sich auch eine, schon in den Simulationen in Abschnitt 6.5.3 bemerkte, „Robustheit“ gegenüber einem Bias des Schätzers für θ_1 erklärt.

Die geringe Streuung wird bei dem Schätzer „2P“ dadurch erreicht, dass durch die vorgestellte Möglichkeit der Korrektur bzw. Anpassung von $\hat{\theta}_2$ an den aktuellen Wert von $\hat{\theta}_{1,r}$ auch die Aufträge einen effektiven Beitrag zur Schätzung von q_2 leisten, bei denen keine Änderung von u_2 auftritt.

Literaturverzeichnis

- [ÅSTRÖM, 1970] ÅSTRÖM, K. J. (1970). *Introduction to Stochastic Control Theory*. Academic Press, New York.
- [AULL, 2008] AULL, M. (2008). *Lehr- und Arbeitsbuch Druck*. Verlag Beruf + Schule, Itzehoe, 8. Aufl.
- [BERTI et al., 2010] BERTI, C., M. JURKEWITZ und J. RAUTERT (2010). Patentanmeldung, DE 10 2010 046 957 A1.
- [BRAMMER und SIFFLING, 1975] BRAMMER, K. und G. SIFFLING (1975). *Stochastische Grundlagen des Kalman-Bucy-Filters Wahrscheinlichkeitsrechnung und Zufallsprozesse*. Methoden der Regelungstechnik. R. Oldenbourg Verlag, München, Wien.
- [CHALAM, 1987] CHALAM, V. V. (1987). *Adaptive Control Systems*. Dekker, New York.
- [DAHMEN und REUSKEN, 2008] DAHMEN, W. und A. REUSKEN (2008). *Numerik für Ingenieure und Naturwissenschaftler*. Springer, Berlin, Heidelberg, 2. Aufl.
- [DIECKHOFF, 1999] DIECKHOFF, F. (1999). *Ursachenuntersuchung für prozeßbedingte Qualitätsmängel im Anilox-Offset*. Diplomarbeit, Bergische Universität Gesamthochschule Wuppertal.
- [DIETRICH und SCHULZE, 2003] DIETRICH, E. und A. SCHULZE (2003). *Statistische Verfahren zur Maschinen- und Prozessqualifikation*. Hanser, München, etc., 4. Aufl.
- [ELTER und PFEIFFER, 2008] ELTER, P. und N. PFEIFFER (2008). *Verbessertes Kennlinienlernen in Druckmaschinen*. Patentanmeldung, DE 10 2008 013 744 A1.
- [VON FINCKENSTEIN et al., 2000] FINCKENSTEIN, K. G. F. VON, J. LEHN, H. SCHELLHAAS und H. WEGMANN (2000). *Arbeitsbuch Mathematik für Ingenieure I. Analysis*. Teubner, Stuttgart, etc.
- [VON FINCKENSTEIN et al., 2002] FINCKENSTEIN, K. G. F. VON, J. LEHN, H. SCHELLHAAS und H. WEGMANN (2002). *Arbeitsbuch Mathematik für Ingenieure II. Differentialgleichungen, Funktionentheorie, Numerik, Statistik*. Teubner, Stuttgart, etc.
- [FÖLLINGER, 1993] FÖLLINGER, O. (1993). *Lineare Abtastsysteme*. Oldenbourg, München, etc., 5. Aufl.
- [FORSSELL und LJUNG, 1999] FORSELL, U. und L. LJUNG (1999). *Closed-loop identification revisited*. Automatica, 35(7):1215–1241.
- [FREUND et al., 2006] FREUND, R. J., W. J. WILSON und P. SA (2006). *Regression Analysis*. Elsevier, Burlington, MA, etc., 2. Aufl.
- [GUSTAVSSON et al., 1977] GUSTAVSSON, I., L. LJUNG und T. SÖDERSTRÖM (1977). *Identification of processes in closed loop – identifiability and accuracy aspects*. Automatica, 13(1):59–75.
- [HARTUNG, 2009] HARTUNG, J. (2009). *Statistik*. Oldenbourg, München, 15. Aufl.
- [HINKELMANN und KEMPTHORNE, 2008] HINKELMANN, K. und O. KEMPTHORNE (2008). *Design and Analysis of Experiments Volume 1 – Introduction to Experimental Design*. Wiley Interscience, Hoboken, NJ, 2. Aufl.
- [HINKLEY, 1969] HINKLEY, D. V. (1969). *On the ratio of two correlated normal random variables*. Biometrika, 56(3):635–639.
- [HINKLEY, 1970] HINKLEY, D. V. (1970). *Correction: On the ratio of two correlated normal random variables*. Biometrika, 57(3):683.

- [HIPPE, 2006] HIPPE, P. (2006). *Windup in Control*. Springer, London.
- [HÄNSLER, 2001] HÄNSLER, E. (2001). *Statistische Signale*. Springer, Berlin, etc., 3. Aufl.
- [HOFFMANN et al., 2005] HOFFMANN, A., B. MARX und W. VOGT (2005). *Mathematik für Ingenieure 1*. Pearson Studium, München.
- [HOFFMANN et al., 2006] HOFFMANN, A., B. MARX und W. VOGT (2006). *Mathematik für Ingenieure 2*. Pearson Studium, München.
- [HOLZHÜTER, 2009] HOLZHÜTER, T. (2009). *Zustandsregelung*. Skript, Fachhochschule Hamburg.
- [HORN und JOHNSON, 1985] HORN, R. A. und C. R. JOHNSON (1985). *Matrix Analysis*. Cambridge University Press, New York.
- [ISERMANN, 1989] ISERMANN, R. (1989). *Digital Control Systems*, Bd. I. Springer, Berlin, Heidelberg, 2. Aufl.
- [ISERMANN, 1991] ISERMANN, R. (1991). *Digital Control Systems*, Bd. II. Springer, Berlin, Heidelberg, 2. Aufl.
- [ISERMANN und MÜNCHHOF, 2011] ISERMANN, R. und M. MÜNCHHOF (2011). *Identification of Dynamic Systems*. Springer, Berlin, Heidelberg.
- [KAMEN und SU, 1999] KAMEN, E. W. und J. K. SU (1999). *Introduction to Optimal Estimation*. Springer, London, etc.
- [KAMISKE und BRAUER, 2011] KAMISKE, G. F. und J.-P. BRAUER (2011). *Qualitätsmanagement von A bis Z*. Carl Hanser Verlag, München, Wien, 7. Aufl.
- [KIPPHAN, 2000] KIPPHAN, H., Hrsg. (2000). *Handbuch der Printmedien*. Springer, Berlin, etc.
- [KOCZY, 1998] KOCZY, A. (1998). *Statistische Prozessregelung für die Kleinserienfertigung*. Shaker, Aachen. Dissertation.
- [KRAUS, 1986] KRAUS, J. F. (1986). *Das Vergessen in rekursiven Parameterschätzverfahren*. Dissertation, ETH Zürich.
- [LENZ, 2012] LENZ, E. (2012). *Farbregelung beim Anicolor-Farbwerk (Abschlussbericht)*. Technischer Bericht, Heidelberger Druckmaschinen AG.
- [LENZ, 2013] LENZ, E. (2013). *Stochastische adaptive Farbregelung bei Bogendruckmaschinen mit Kurzfarbwerk*. In: *Workshop KRK, 17. Juni 2013, Garching*.
- [LJUNG, 1999] LJUNG, L. (1999). *System Identification Theory for the User*. Prentice Hall, Upper Saddle River, NJ, 2. Aufl.
- [LOVE, 2007] LOVE, J. (2007). *Process Automation Handbook*. Springer, London.
- [LUNZE, 2007] LUNZE, J. (2007). *Regelungstechnik 1*. Springer, Berlin etc., 6 Aufl.
- [LUNZE, 2008] LUNZE, J. (2008). *Regelungstechnik 2*. Springer, Berlin, Heidelberg, 5. Aufl.
- [MARSAGLIA, 1965] MARSAGLIA, G. (1965). *Ratios of Normal Variables and Ratios of Sums of Uniform Variables*. *Journal of the American Statistical Association*, 60:193–204.
- [MARSAGLIA, 2006] MARSAGLIA, G. (2006). *Ratios of Normal Variables*. *Journal of Statistical Software*, 16(4).
- [MAYER und PFEIFFER, 2003] MAYER, M. und N. PFEIFFER (2003). *Lernende Farbführung*. Patentanmeldung, DE 103 12 998 A 1.
- [MEDITCH, 1968] MEDITCH, J. S. (1968). *Minimum mean-square error stochastic linear control*. *International Journal of Control*, 8(5):441–456.

-
- [MEDITCH, 1969] MEDITCH, J. S. (1969). *Stochastic Optimal Linear Estimation and Control*. McGraw-Hill, New York, etc.
- [MERZIGER et al., 2001] MERZIGER, G., G. MÜHLBACH, D. WILLE und T. WIRTH (2001). *Formeln und Hilfen zur höheren Mathematik*. Binomi, 4. Aufl.
- [MICHELS et al., 2011] MICHELS, J., D. SCHAFFRATH und W. SCHÖNBERGER (2011). Patentanmeldung, DE 10 2011 008 592 A1.
- [NELLES, 1999] NELLES, O. (1999). *Nonlinear System Identification with Local Linear Neuro-Fuzzy Models*. Shaker, Aachen. Dissertation.
- [PFEIFFER, 1988] PFEIFFER, D. (1988). *Kompensation thermisch bedingter Bearbeitungsfehler durch prozessnahe Qualitätsregelung*. Springer, Berlin, Heidelberg. Dissertation.
- [PFEIFFER, 1995] PFEIFFER, N. (1995). *Verfahren zur Steuerung der Farbführung einer Druckmaschine*. Patent, DE 39 13 382 C2.
- [PFEIFFER und LENZ, 2013] PFEIFFER, N. und E. LENZ (2013). *Farbregelungskonzept für Druckmaschinen mit Kurzfarbwerk*. Patentanmeldung (eingereicht am 08.05.2013), DE 10 2013 007 907.1.
- [RAWLINGS et al., 1998] RAWLINGS, J. O., S. G. PANTULA und D. A. DICKEY (1998). *Applied Regression Analysis : A Research Tool*. Springer, New York, 2. Aufl.
- [RINNE, 2008] RINNE, H. (2008). *Taschenbuch der Statistik*. Harri Deutsch Verlag, Frankfurt am Main, 3. Aufl.
- [SAWODNY, 1996] SAWODNY, O. (1996). *Modellbasierte Qualitätsregelung in der spanenden Fertigung am Beispiel der Drehbearbeitung*. VDI-Verlag, Düsseldorf. Dissertation.
- [SÖDERSTRÖM, 2013] SÖDERSTRÖM, T. (2013). *Comparing some classes of bias-compensating least squares methods*. *Automatica*, 49(3):840–845.
- [STIEBNER, 1992] STIEBNER, E. D. (1992). *Bruckmann's Handbuch der Drucktechnik*. Bruckmann, München, 5. Aufl.
- [STORM, 2001] STORM, R. (2001). *Wahrscheinlichkeitsrechnung, mathematische Statistik und statistische Qualitätskontrolle*. Fachbuchverlag Leipzig im Carl Hanser Verlag, München, Wien, 11. Aufl.
- [TAO, 2003] TAO, G. (2003). *Adaptive Control Design and Analysis*. Wiley, Hoboken, NJ.
- [VAN DEN HOF und SCHRAMA, 1995] VAN DEN HOF, P. M. und R. J. SCHRAMA (1995). *Identification and control – Closed-loop issues*. *Automatica*, 31(12):1751–1770.
- [WESTKÄMPER, 1991] WESTKÄMPER, E., Hrsg. (1991). *Integrationspfad Qualität*. Springer, Berlin, etc.
- [WOOLDRIDGE, 2003] WOOLDRIDGE, J. M. (2003). *Introductory Econometrics*. Thomson South-Western, Mason, OH, 2. Aufl.
- [ZURMÜHL und FALK, 1984] ZURMÜHL, R. und S. FALK (1984). *Matrizen und ihre Anwendungen für Mathematiker, Physiker und Ingenieure Teil 1: Grundlagen*. Springer, Berlin, etc., 5. Aufl.