

Eberhard-Karls-Universität Tübingen
Mathematisch-Naturwissenschaftliche Fakultät
Fachbereich Mathematik

DAS KALMAN-BUCY-FILTER UND SEINE KONVERGENZ
BEI DER SCHÄTZUNG VON LÖSUNGEN
GEWÖHNLICHER DIFFERENTIALGLEICHUNGEN
MIT ANWENDUNG AUF DIE ZUSTANDSSCHÄTZUNG
EINES KRAFTFAHRZEUGES

D I P L O M A R B E I T

verfasst von

Hanna Maria Walach

Zur Erlangung des Grades
einer Diplom-Mathematikerin

betreut von

PROF. DR. CHRISTIAN LUBICH
Numerical Analysis Group

in Kooperation mit der
DAIMLER AG

13. Mai 2013

Dla Mamy, Taty i Gosi

Eigenständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit mit dem Titel „Das Kalman-Bucy-Filter und seine Konvergenz bei der Schätzung von Lösungen gewöhnlicher Differentialgleichungen mit Anwendung auf die Zustandsschätzung eines Kraftfahrzeuges“ selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Hanna Maria Walach

Tübingen, den 13. Mai 2013

Vorwort

Seit der Geburtsstunde des Automobils, dem Tag, an dem Carl Benz sein selbst entwickeltes und selbst gebautes Fahrzeug als Patent angemeldet hatte, spielen Themen wie Sicherheit und Komfort in der Automobilindustrie eine wichtige Rolle. In den letzten Jahren ging die Entwicklung so weit voran, dass Menschen von Fahrerassistenzfunktionen, Fahrdynamik- und Fahrsicherheitsapplikationen in ihrem Fahrzeug unterstützt werden. Dabei ist die Kenntnis der Eigenbewegung des Fahrzeugs bei diversen Fahrmanövern von besonders großem Nutzen.

Um Kenntnis über die Eigenbewegung zu erhalten, ist das Wissen um die Zustandsgrößen, die diese beschreiben, unabdingbar. Die Eigenbewegung eines Fahrzeugs wird beschrieben durch seine Lagewinkel und Geschwindigkeiten. Beispiele für die Verwendung der Zustandsgrößen sind das Vermeiden von Kippen sowie das Nachführen von Kamera- und Radarsystemen, sodass Orientierungs- und Zielverlust bei Extremfahrmanövern vermieden werden können.

Eine direkte Messung der Zustandsgrößen ist für Serienfahrzeuge nicht möglich. Sie müssen indirekt aus den von der Fahrzeugsensorik gelieferten Messgrößen berechnet werden. Genauer handelt es sich hierbei um eine Schätzung oder Beobachtung der Zustände. Das System, welches diese Berechnung vornimmt, wird als Fahrzustandsbeobachter bezeichnet. In der vorliegenden Arbeit wird ein bestimmter Fahrzustandsbeobachter, das Kalman-Filter, thematisiert.

Es stellt sich die Frage, welche Mathematik hinter diesen praktischen Problemen steckt. Die Zustandsgrößen, die es zu bestimmen gilt, werden durch eine gewöhnliche Differentialgleichung beschrieben. In den meisten Anwendungsfällen, ist sie nichtlinear. Es gilt, diese Gleichung zu lösen, um die Zustandsgrößen zu erhalten. Da jedoch Eingangsgrößen auf das System einwirken, wird die Differentialgleichung nicht mit den gewöhnlichen numerischen Methoden gelöst, sondern mit einem Schätzalgorithmus, dem Kalman-Bucy-Filter. Es ist ein iteratives Verfahren, bestehend aus Prädiktion und Innovation. Dieser Algorithmus wird so oft wiederholt, bis er ein Ergebnis der Differentialgleichung liefert,

das heißt, bis er gegen die Zustände konvergiert.

Die Konvergenz des Kalman-Filters ist das Hauptthema dieser Arbeit.

Zu Beginn erfolgt eine Einführung in die Systemtheorie, dann eine Klärung der mathematischen Hintergründe und nachfolgend die Vorstellung des Kalman-Filters. Es schließen sich zwei Beweise für die Konvergenz des Kalman-Filters für lineare, als auch für nichtlineare Systeme und damit auch eine Stabilitätsanalyse an. Dabei wird aufgezeigt, welche Voraussetzungen das System erfüllen muss, um Konvergenz, das heißt um eine Lösung der Differentialgleichung erwarten zu können. Es ist gelungen, den Beweis der Konvergenz des Kalman-Bucy-Filters für lineare Systeme auf eine neue Weise zu führen und jenen für die Konvergenz des Kalman-Bucy-Filters für nichtlineare Systeme an einigen Stellen zu verbessern. Damit ist es möglich, genaue Aussagen über die Güte der Schätzung zu machen, was für die Einordnung der Lösung der Differentialgleichung von enormer Bedeutung ist.

Im Anschluss daran wird eine konkrete 5-dimensionale, nichtlineare Differentialgleichung vorgestellt, die die Eigenbewegung eines Fahrzeugs beschreibt. Zunächst wird gezeigt, dass das Kalman-Filter für diese Gleichung konvergiert und danach wird für ein vereinfachtes, 2-dimensionales System eine untere und eine obere Schranke für den Schätzfehler bestimmt, womit Aussagen über die Güte der Schätzung möglich sein sollen.

Danach werden die numerische Realisierung des Kalman-Filters und insbesondere die in der Automobilindustrie derzeit verwendeten Techniken und Probleme aufgeführt, bevor dann nachfolgend das 5-dimensionale System für verschiedene Fahrmanöver implementiert wird. Hier liegt der Fokus vor allem auf der Frage, wie schlecht die Startwerte höchstens sein dürfen, damit das Kalman-Filter überhaupt noch ein Ergebnis der Differentialgleichung liefert.

Die numerischen Experimente zeigen die Robustheit des Kalman-Bucy-Filters bezüglich den Startwerten auf. Damit können Aussagen über Güte und Konvergenz gemacht werden. Auch die Tatsache, dass dieses Filter in der Automobilindustrie Anwendung findet, unterstreicht seine Praktikabilität, sein hohes Maß an Stabilität und damit auch seine Geltung.

Danksagungen

An dieser Stelle möchte ich die Gelegenheit wahrnehmen, mich bei allen Personen zu bedanken, die mich bei der Erstellung meiner Diplomarbeit unterstützt haben.

Zunächst möchte ich mich bei Prof. Dr. Christian Lubich bedanken, der gerne bereit war, meine Arbeit zu betreuen.

Durch seine Zustimmung war es mir möglich, die Arbeit in Kooperation mit der Daimler AG zu verfassen.

Großer Dank gilt Prof. Dr. Jens C. Kalkkuhl, meinem Betreuer von Seiten der Daimler AG. Auf Grund seines außerordentlichen Engagements, seiner vielseitigen Unterstützung, der Bereitstellung der Literatur und auf Grund der unzähligen Stunden, in der ich professionelle Betreuung erfahren durfte, bin ich ihm besonders verbunden.

Ebenfalls bedanken möchte ich mich bei Dhia Mansour, der für mich und meine Fragen stets Zeit hatte, dessen Türe mir immer offen stand und der manchmal viel Geduld mit mir haben musste. Nicht nur für die vielen fachlichen Diskussionen bedanke ich mich bei ihm, sondern auch für die beratenden Privatgespräche.

Mein besonderer und tiefer Dank gilt meinen Eltern, die mir mein Studium ermöglicht, mir immer den Rücken gestärkt und mich in jeglicher Hinsicht unterstützt haben. Ich bedanke mich für alles, was sie bisher für mich getan haben.

Inhaltsverzeichnis

1. Systemtheoretische Grundlagen	1
1.1. Intuitives Verständnis eines Systems	1
1.2. Zustandsraumdarstellung linearer, zeitvarianter Systeme	3
1.3. Zustandsraumdarstellung nichtlinearer, zeitvarianter Systeme	4
1.3.1. Lokale Existenz und Eindeutigkeit	5
1.4. Konstruktion eines Beobachters	5
2. Mathematische Hintergründe	9
2.1. Beobachtbarkeit eines zeitvarianten Systems	9
2.1.1. Beobachtbarkeit linearer Systeme	10
2.1.2. Beobachtbarkeit nichtlinearer Systeme	10
2.1.3. Uniforme Beobachtbarkeit	12
2.2. Die Stabilität eines Systems	13
2.2.1. Bemerkung	13
2.2.2. Definition (Stabilität des Equilibriums)	14
2.2.3. Ljapunov-Stabilität	14
3. Das Kalman-Bucy-Filter	17
3.1. Der Wiener Prozess	18
3.2. Die Zustandsschätzung	18
3.3. Die Güte der Zustandsschätzung	19
3.3.1. Die Schätzfehlerkovarianzmatrix	19
3.3.2. Die Kalman-Matrix	23
3.4. Die Konvergenz des Kalman-Bucy-Filters	26
3.4.1. Behauptung	26
3.4.2. Beweis	26
4. Das Erweiterte Kalman-Bucy-Filter	29
4.1. Das Erweiterte Kalman-Bucy-Filter	30

4.2. Die Konvergenz des EKBF	30
4.2.1. Vorbemerkungen	31
4.2.2. Umformulierung des nichtlinearen Systems	31
4.2.3. Voraussetzungen	33
4.2.4. Behauptung	33
4.2.5. Beweis	34
5. Spezielles Anwendungsproblem hinsichtlich der Fahrdynamik eines Kraftfahrzeuges	49
5.1. Beweis der Konvergenz des EKBF für ein spezielles dynamisches System .	52
5.2. Ermittlung des Konvergenzradius zum Startzeitpunkt für ein konkretes 2-dimensionales System	58
5.2.1. Beweis der Konvergenz des EKBF bei Anwendung auf gegebenes 2-dimensionales System	59
5.2.2. In welchem Gebiet liegt $P(t)$ zur Konvergenz des 2-dimensionalen Systems?	61
6. Numerische Realisierung des EKF	71
6.1. Das Kalman-Filter für lineare Systeme in zeitdiskreter Form	71
6.2. Das EKF für nichtlineare Systeme in zeitdiskreter Form	77
6.3. Numerische Verfahren im EKF	78
6.3.1. Numerische Verfahren im Prädiktionsschritt	78
6.3.2. Numerische Verfahren im Innovationsschritt	79
7. Numerische Experimente	85
7.1. Geradeausfahrt mit positiver Steigung	87
7.2. Die Kreisfahrt	95
8. Zusammenfassung	105
A. Erläuterungen zur Übergangsmatrix	107
Literaturverzeichnis	111

Kapitel 1

Systemtheoretische Grundlagen

Die Beschreibung eines dynamischen Systems, das auf technischen Problemstellungen basiert, erfolgt anhand eines Zustandsraummodells in mathematischer Form. Dies ist insofern ein Vorteil, als dass es mit Hilfe mathematischer Techniken, insbesondere jener aus der numerischen Mathematik, untersucht und mit einer Implementierung realisiert werden kann.

In diesem ersten Kapitel werden wir uns mit dem Verständnis eines Systems beschäftigen, das mitunter lineare und nichtlineare, dynamische Systeme beinhaltet. Außerdem möchten wir auf den Begriff des Zustandes und auf die Zustandsraumdarstellung solcher Systeme eingehen.

Schließlich wird dann ein Beobachter konstruiert, der für den gesamten Verlauf der vorliegenden Arbeit von grundlegender Bedeutung ist.

Dieses Kapitel beruht auf Inhalten aus [GKN⁺74] sowie aus [Wen07].

Zunächst jedoch einige Leitgedanken und Ideen bezüglich eines Systems.

1.1. Intuitives Verständnis eines Systems

Ein System befindet sich in der realen Welt und steht in einer Wechselwirkung mit dieser. Die Umgebung, in der sich das System befindet, wirkt durch eine Eingangsgröße u auf dieses System ein, das heißt, die Umwelt nimmt Einfluss auf das System. Dabei können die Eingänge u einerseits bestimmte Stellgrößen sein, andererseits aber auch Störungen, die auf das System einwirken und dieses verändern. Daraufhin bearbeitet letzteres diesen Eingang und kommuniziert mit der realen Welt, indem es eine Größe y ausgibt, welche

sinngemäß als Ausgangsgröße bezeichnet wird. Der Ausgang y kann zum einen eine Größe sein, die die Umgebung des Systems beeinflusst, zum anderen jedoch auch Werte beinhalten, die sich aus Beobachtungen ergeben.

Für ein entsprechendes Verständnis veranschaulicht das folgende Schaubild diese Gegebenheiten:

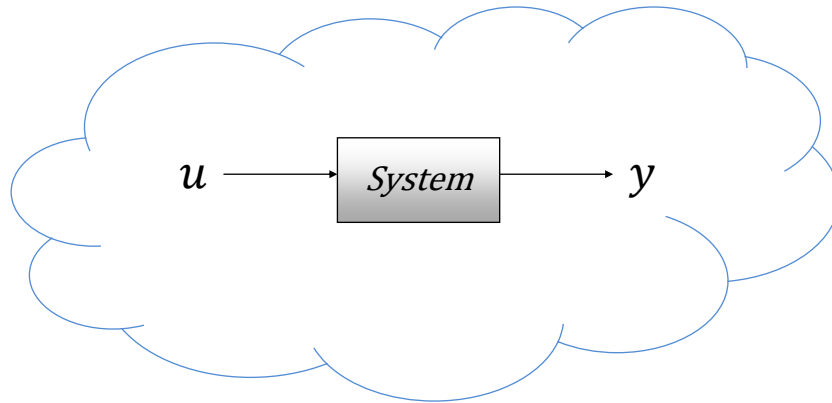


Abbildung 1.1.: Schematische Darstellung eines Systems

Ein mathematisches Modell beschreibt dabei die Korrespondenz der Eingangs- mit den Ausgangsgrößen.

Ausgangspunkt ist daher die Fragestellung: Welche Ausgänge werden bei gegebenen Eingängen aus dem System erhalten?

In der Systemtheorie wird zwischen sogenannten statischen und dynamischen Systemen unterschieden. Bei allen Systemen, die wir im folgenden betrachten werden, handelt es sich um dynamische Systeme.

Dynamische Systeme

Ein dynamisches System beschreibt die zeitliche Entwicklung eines Systems.

Es ist durch Systemeingänge u sowie Systemausgänge y charakterisiert, indem die Eingangsgrößen u auf das System Einfluss nehmen und die Ausgangsgrößen y die zeitliche Auswirkung von u beschreiben. Das dynamische System bringt diese beiden Größen in Relation.

Begriffsbestimmung (Zustand)

Der Zustand eines Systems zum Zeitpunkt t_0 wird durch das Verhalten des Systems zur Zeit $t < t_0$ beschrieben. Kennen wir das Systemverhalten, so ist das Wissen darüber ausreichend, um bei bekanntem Eingang u den Ausgang y eindeutig zu bestimmen.

Die Kenntnis über das Verhalten des Systems zu einem früheren Zeitpunkt wird anhand von Daten konkret. Diese Daten werden „Zustandsgrößen“ des Systems genannt.

Wir sehen, dass der Zustand im Rahmen der dynamischen Systeme eine entscheidende Rolle spielt. Wie dieser Zustand dargestellt werden kann, wird nun im folgenden Abschnitt thematisiert.

1.2. Zustandsraumdarstellung linearer, zeitvarianter Systeme

Ein lineares, dynamisches System kann beschrieben werden durch:

$$\begin{aligned} \dot{x}(t) &= A(t)x(t) + B(t)u(t) + G(t)w(t) && \text{Zustandsgleichung} \\ y(t) &= C(t)x(t) + v(t) && \text{Ausgangsgleichung} \\ x(t_0) &= x_0 && \text{Anfangsbedingung} \end{aligned} \quad (1.1)$$

Das Zustandsraummodell eines linearen, dynamischen Systems setzt sich zusammen aus

- einem Zustandsvektor $x(t)$
- einer Systemmatrix $A(t)$
- einem Eingangsvektor $u(t)$
- einer Eingangsmatrix $B(t)$
- einem stochastischen, normalverteilten, mittelwertfreien Systemrauschen $w(t)$
- einer Störmatrix $G(t)$
- einer Messmatrix $C(t)$
- einem stochastischen, normalverteilten, mittelwertfreien Messrauschen $v(t)$.

Hierbei sind $x(t) \in \mathbb{R}^n$, $A(t) \in \mathbb{R}^{n \times n}$, $u(t) \in \mathbb{R}^m$, $B(t) \in \mathbb{R}^{n \times m}$, $w(t) \in \mathbb{R}^p$, $G(t) \in \mathbb{R}^{n \times p}$, $y(t) \in \mathbb{R}^n$, $C(t) \in \mathbb{R}^{n \times n}$, $v(t) \in \mathbb{R}^n$.¹

Das heißt, Vektoren und Matrizen beschreiben den Zustand eines Systems anhand einer gewöhnlichen, linearen Differentialgleichung.

Das Systemrauschen $w(t)$ dient der Berücksichtigung von Modellunsicherheiten und

¹Da die Dimensionen der Vektoren und Matrizen offengelegt wurden, wird von der Vektorpfeil-Notation im weiteren Verlauf abgesehen. Vektoren werden mit Kleinbuchstaben und Matrizen mit Großbuchstaben geschrieben. Um Verwechslungen, etwa mit Skalaren, zu vermeiden, werden entsprechende Komponenten explizit als solche gekennzeichnet.

bezieht die Tatsache mit ein, dass ein reales System durch ein Modell beschrieben wird und dieses vom realen abweicht. Je komplexer das reale System, desto größer ist die Abweichung.

Das Messrauschen $v(t)$ wird in dieser Gleichung benutzt, um Störungen bei der Messung zu berücksichtigen.

In bestimmten Fällen ist die Eingangsgröße $u(t)$ mit Rauschsignalen behaftet, da sie durch Messung ermittelt wird. In diesem Fall werden die Störsignale zum Systemrauschen hinzugefügt.

Folgendes Blockdiagramm illustriert die Zustandsraummodellierung:

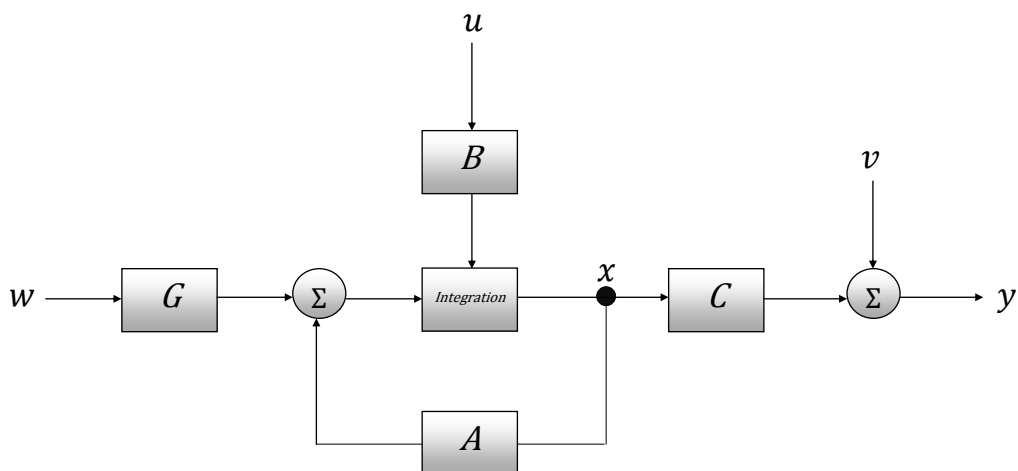


Abbildung 1.2.: Visualisierung der Zustandsgleichung eines linearen, dynamischen Systems

1.3. Zustandsraumdarstellung nichtlinearer, zeitvarianter Systeme

Die meisten Vorgänge in der Realität sind nicht von linearer Natur, sondern sie verhalten sich nichtlinear. Aus diesem Grund sind mathematische Modellierungen für Ereignisse nichtlinearer Art von besonderem Interesse.

Das folgende nichtlineare, zeitvariante Differentialgleichungssystem liefert eine mathema-

tische Darstellung eines solchen Systems:

$$\begin{aligned}\dot{x}(t) &= f(x(t), u(t)) \\ y(t) &= h(x(t), u(t)) \\ x(t_0) &= x_0,\end{aligned}\tag{1.2}$$

mit $f : D \rightarrow \mathbb{R}^n$, $h : D \rightarrow \mathbb{R}^n$, $D \subseteq \mathbb{R}^n$ und f Lipschitz-stetig.

Die Zustandsraumdarstellung setzt sich also wie im linearen Fall aus einer System- sowie einer Ausgangsgleichung zusammen; mit dem Unterschied, dass diese nichtlinear sind. Um überhaupt eine Lösung für dieses Differentialgleichungssystem zu erhalten, wird gefordert, dass f Lipschitz-stetig ist, sodass die Existenz der Lösung gewährleistet ist. Das Resultat der lokalen Existenz und Eindeutigkeit eines Systems ist wohlbekannt und wird daher nur formuliert, jedoch nicht bewiesen. Ein Beweis hierfür kann mit dem Satz von Peano und dem Satz von Picard-Lindelöf geführt werden.

1.3.1. Lokale Existenz und Eindeutigkeit

Sei $f(t, x)$ stückweise stetig in t und erfülle die Lipschitz-Bedingung

$$\forall x, y \in B = \{x \in \mathbb{R}^n : \|x - x_0\| \leq r\}, \text{ sodass } \|f(t, x) - f(t, y)\| \leq L\|x - y\|, \forall t \in [t_0, t_1].$$

Dann

$$\exists \delta > 0 : \dot{x} = f(t, x), \text{ mit } x(t_0) = x_0$$

und hat eine eindeutige Lösung in $[t_0, t_0 + \delta]$.

Was ist nun ein Beobachter und wie sieht er aus?

1.4. Konstruktion eines Beobachters

Um den Zustand eines dynamischen Systems aus den Ein- und Ausgängen und bei gegebenem mathematischem Systemmodell zu rekonstruieren, wird ein Beobachter auf dieses angesetzt.

Unter einem Beobachter ist ein mathematischer Algorithmus zu verstehen, mit dem es möglich ist, nicht messbare Größen eines realen Systems zu schätzen, um den Zustand eines dynamischen Systems zu bestimmen. Um dies zu realisieren, benötigt der Beobachter ein Zustandsraummodell, welches dem realen System möglichst genau entspricht.

Derartige Modelle zur Systembeschreibung wurden bereits in Kapitel 1.2 erläutert. Auf dieser Grundlage und unter Zuhilfenahme von Systemeingängen sowie Systemausgängen werden nicht messbare Größen des realen Systems, an dem der Anwender interessiert ist, bestimmt.

Das folgende Schaubild veranschaulicht den Grundgedanken und die Funktionsweise in groben Zügen:

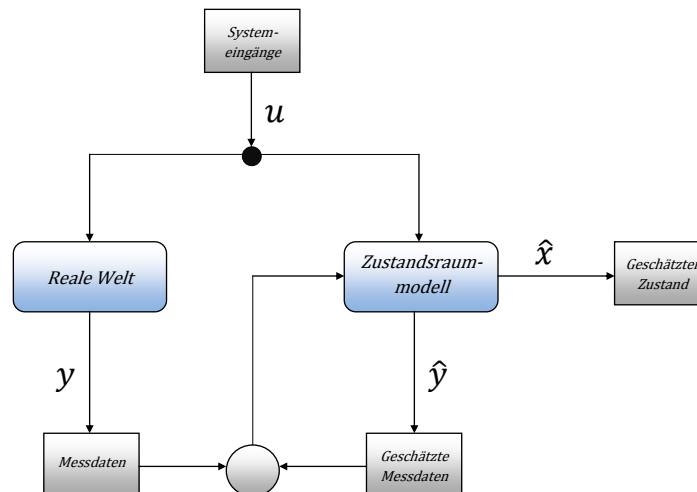


Abbildung 1.3.: Blockschaltbild eines Beobachters

Systemeingänge wirken sowohl auf das reale Modell, als auch auf das Systemmodell. Sie liefern Messdaten, die anhand des realen Systems gemessen werden, aber auch Messdaten, die anhand des Modells bestimmt werden. Da es häufig vorkommt, dass fehlerhafte Messdaten aus dem realen System vorliegen, werden diese mit Hilfe der geschätzten Messdaten verarbeitet und im Zustandsraummodell eingefügt. Daraus wird dann der Zustand des gegebenen dynamischen Systems geschätzt.

Der Gedanke hinsichtlich der Zustandsschätzung beruht auf der Idee eines Beobachters, der einem Vorhersage- und Korrekturschema gehorcht. Im Konkreten bedeutet das eine Prädiktion des Zustands und daraufhin eine Verbesserung dieser Prädiktion durch Hinzunahme der Messungen.

Mit Verzicht der Störung $w(t)$ wird zur gegebenen Zustandsgleichung, allerdings mit geschätztem Zustand $\hat{x}(t)$, die Differenz zwischen tatsächlichem und auf Grund der Zustandsschätzung vorhergesagtem Ausgang $\hat{y}(t) = C(t)\hat{x}(t)$ addiert. Das Problem der Hinzunahme der Differenz zwischen tatsächlicher und geschätzter Messung ist an dieser Stelle offensichtlich: Weichen die Messung $y(t)$ und die Schätzung der Messung $\hat{y}(t)$ stark voneinander ab, so wird ein betragsmäßig großer Wert zum Zustand addiert. Dies

impliziert eine vom realen Zustand stark abweichende Zustandsschätzung.

Um eine solche Verschlechterung zu vermeiden, wird eine Matrix $K(t)$ herangezogen. Sie vereint zwei Eigenschaften und berücksichtigt diese: Die Tatsache, dass es keine perfekte, also rauschfreie Messung gibt, sowie die Absicht, Stabilität zu gewährleisten. Daher gewichtet sie die den Einfluss der Differenz $y(t) - C(t)\hat{x}(t)$ auf die Zustandsschätzung, und zwar so, dass die Zustandsschätzung möglichst dem realen Zustand entspricht:

$$\hat{\dot{x}}(t) = A(t)\hat{x}(t) + B(t)u(t) + K(t)(y(t) - C(t)\hat{x}(t)) \quad (1.3)$$

Das folgende Blockschaltbild veranschaulicht die Struktur des Beobachters:

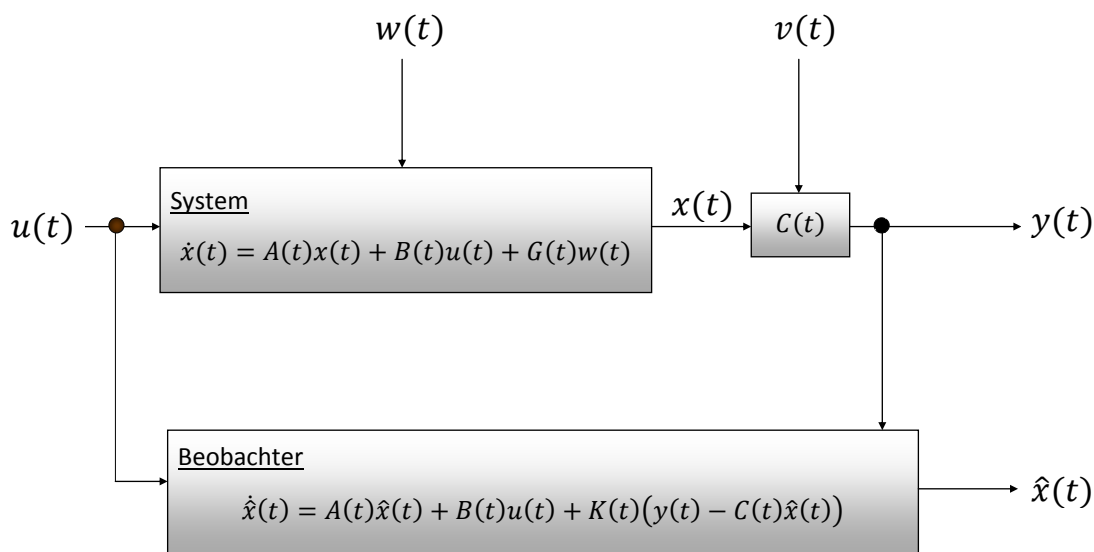


Abbildung 1.4.: Mathematisches Beobachtermodell

Die Güte dieser allgemeinen Beobachtergleichung hängt von der Wahl von $K(t)$ ab. $K(t)$ ist zunächst eine zeitabhängige Matrix, die die Differenz der Messung aus dem realen System sowie der geschätzten Messung anhand des mathematischen Systemmodells gewichtet. Durch Hinzunahme der Messwerte soll eine bessere Schätzung des Zustandes erfolgen. Daher wird eine Matrix benötigt, die, je nach Güte der Messung, diese stärker oder schwächer gewichtet, sodass sich die Messung vorteilhaft auf die Zustandsschätzung auswirkt.

Wie ist diese Matrix zu wählen, damit eine bestmögliche Schätzung erfolgt?

Eine Möglichkeit hierfür ist die sogenannte Kalman-Matrix, die wir noch bestimmen werden.

Kapitel 2

Mathematische Hintergründe

Wir haben nun die Idee und die Motivation für die allgemeine Darstellung eines Beobachters erläutert. Um jedoch einen Beobachter auf ein System anwenden zu können, gilt es zunächst einige Voraussetzungen, die an das System gestellt werden, zu überprüfen.

Weiter ist es ein Anliegen dieser Arbeit, den Beweis für die Konvergenz des Kalman-Filters zu führen. Hierfür sind einige Hilfsmittel nötig, die in diesem Kapitel formuliert und erläutert werden.

Nachdem wir in Kapitel 1 mitunter einen Beobachter konstruiert haben, beschäftigen wir uns in diesem Kapitel mit der Frage, ob und unter welchen Voraussetzungen ein System beobachtbar ist. Die Schilderungen diesbezüglich sind in Anlehnung an [Isi99] und [Wen07]. Es folgt die für den Nachweis der Konvergenz des Kalman-Filters wichtige Definition der uniformen Beobachtbarkeit, die aus [HF03] entnommen ist.

Schließlich werden wir uns mit dem Stabilitätsbegriff und insbesondere mit der Stabilitätstheorie von Ljapunov auseinandersetzen. Diese ist in [Kha92] besonders gut beschrieben.

Zu Beginn soll die Frage beantwortet werden, unter welchen Bedingungen ein System beobachtbar ist.

2.1. Beobachtbarkeit eines zeitvarianten Systems

Beobachtbarkeit ist eine Bedingung an das System, die entscheidet, ob der Zustand eines dynamischen Systems, welches auf dem mathematischen Modell basiert, aus dessen Ein- und Ausgängen eindeutig bestimmt werden kann. Folglich handelt es sich hierbei um eine Eigenschaft des Systemmodells.

Abhängig von der (Nicht-)Linearität eines gegebenen Systems ist zwischen zwei Vorgehensweisen zu unterscheiden.

2.1.1. Beobachtbarkeit linearer Systeme

Wir gehen von einem linearen System der Form (1.1) aus. Für die Formulierung des Kriteriums, durch welches dieses System beobachtet werden kann, ist zunächst folgende Definition nötig:

Definition (Beobachtbarkeitsmatrix)

Sei $A(t) \in \mathbb{R}^{n \times n}$ die Systemmatrix der Zustandsgleichung aus (1.1) und sei $C(t) \in \mathbb{R}^{p \times n}$ die Messmatrix der Ausgangsgleichung aus (1.1).

Für $L_0(t) = C(t)$, $L_1(t) = \frac{d}{dt}C(t) + C(t)A(t)$, \dots $L_k(t) = L_1(t)L_{k-1}(t)C(t)$, $k = 1, \dots, n$ heißt

$$M(t) = \begin{bmatrix} L_0(t) \\ L_1(t) \\ \vdots \\ L_{n-1}(t) \end{bmatrix}_{p \cdot n \times n}$$

Beobachtbarkeitsmatrix für das System (1.1).

Das Kalman-Kriterium

Das System (1.1) ist genau dann beobachtbar, wenn die Beobachtbarkeitsmatrix der Dimension $(p \cdot n \times n)$ vollen Rang besitzt, das heißt, wenn gilt: $rg(M(t)) = n$.

Um also zu testen, ob ein System beobachtbar ist und dementsprechend ein Beobachter darauf angesetzt werden kann, muss lediglich der Rang der zugehörigen Beobachtbarkeitsmatrix bestimmt werden.

2.1.2. Beobachtbarkeit nichtlinearer Systeme

Im Gegensatz zu linearen Systemen gibt es bei nichtlinearen Systemen eine Vielzahl an Definitionen und Sätzen, die beschreiben, wann ein gegebenes System beobachtbar ist und wie diese Beobachtbarkeit nachgewiesen werden kann.

Im Unterschied zu linearen Systemen hängt die Beobachtbarkeit nichtlinearer Systeme von den Eingängen $u(t)$ ab, die wir nun also auch beachten müssen.

Im Folgenden wollen wir die beiden Kriterien vorstellen, die in der Praxis am häufigsten

herangezogen werden.

Schließlich werden wir noch die „uniforme Beobachtbarkeit“ thematisieren, die wir im Beweis für die Konvergenz des Kalman-Filters für nichtlineare Systeme benutzen werden. Wir gehen hier vom System (1.2) aus und wollen zunächst zum Kriterium für Beobachtbarkeit dieses Systems hinführen.

Dazu stellen wir einige Vorüberlegungen an, die sich an Inhalte aus [Ada09] lehnen.

Zunächst berechnen wir die zeitlichen Ableitungen des Ausgangs $y(t)$:

$$\begin{aligned} \dot{y}(t) &= \frac{\partial h}{\partial x} f(x(t), u(t)) + \frac{\partial h}{\partial u} \dot{u}(t) =: g_1(x(t), u(t), \dot{u}(t)) \\ \ddot{y}(t) &= \frac{\partial g_1}{\partial x} f(x(t), u(t)) + \frac{\partial g_1}{\partial u} \dot{u}(t) + \frac{\partial g_1}{\partial \dot{u}} \ddot{u} =: g_2(x(t), u(t), \dot{u}(t), \ddot{u}) \\ &\vdots \\ y^{(n-1)} &= \frac{\partial g_{n-2}}{\partial x} f(x(t), u(t)) + \sum_{i=1}^{n-1} \frac{\partial g_{n-2}}{\partial u^{(i-1)}} u^{(i)} =: g_{n-1}(x(t), u(t), \dot{u}(t), \dots, u^{(n-1)}(t)) \end{aligned}$$

Damit ist es nun möglich folgende Matrix zu definieren, deren weitere Untersuchungen zur Beobachtbarkeit des Systems führen.

Sei also

$$\begin{aligned} N(t) &= \begin{bmatrix} y(t) \\ \dot{y}(t) \\ \ddot{y}(t) \\ \vdots \\ y^{(n-1)}(t) \end{bmatrix}_{n \times n} = \begin{bmatrix} h(x(t), u(t)) \\ g_1(x(t), u(t), \dot{u}(t)) \\ g_2(x(t), u(t), \dot{u}(t), \ddot{u}(t)) \\ \vdots \\ g_{(n-1)}(x(t), u(t), \dot{u}(t), \ddot{u}(t), \dots, u^{(n-1)}(t)) \end{bmatrix}_{n \times n} \\ &= s(x(t), u(t), \dot{u}(t), \ddot{u}(t), \dots, u^{(n-1)}(t)). \end{aligned}$$

Wenn es möglich ist, das Gleichungssystem $N(t) = s(x(t), u(t), \dot{u}(t), \ddot{u}(t), \dots, u^{(n-1)}(t))$ nach x zu lösen, so kann man damit globale Beobachtbarkeit des Systems (1.2) nachweisen. Hierfür ist eine Umkehrfunktion für s nötig, die jedoch in der Praxis häufig nur schwer zu bestimmen ist.

In unseren Ausführungen beschränken wir uns auf die lokale Beobachtbarkeit, die im folgenden definiert wird.

Definition (Lokale Beobachtbarkeit nichtlinearer Systeme)

Die folgende Definition ist aus [Ada09] entnommen. Sie lautet:

Sei $D \subseteq \mathbb{R}^n$ eine Umgebung von $x(t)$ und sei $E \subseteq \mathcal{C}^{n-1}$ der Unterraum des Raumes der $(n-1)$ -mal stetig differenzierbaren Funktionen mit $u(t) \in E$.

Das System (1.2) heißt lokal beobachtbar, wenn $\forall x \in D, \forall u(t) \in E$ gilt:

$$\operatorname{rg} \left(\frac{\partial s(x(t), u(t), \dot{u}(t), \ddot{u}(t), \dots, u^{(n-1)}(t))}{\partial x} \right) = \operatorname{rg} \begin{bmatrix} \frac{\partial h(x(t), u(t))}{\partial x} \\ \frac{\partial g_1(x(t), u(t), \dot{u}(t))}{\partial x} \\ \frac{\partial g_2(x(t), u(t), \dot{u}(t), \ddot{u}(t))}{\partial x} \\ \vdots \\ \frac{\partial g_{n-1}(x(t), u(t), \dot{u}(t), \dots, u^{(n-1)}(t))}{\partial x} \end{bmatrix}_{n \times n} = n$$

Für den Nachweis lokaler Beobachtbarkeit genügt es folglich den Rang der Beobachtbarkeitsmatrix

$$\left(\frac{\partial s(x(t), u(t), \dot{u}(t), \ddot{u}(t), \dots, u^{(n-1)}(t))}{\partial x} \right)$$

zu bestimmen.

2.1.3. Uniforme Beobachtbarkeit

Wir gehen weiterhin von einem nichtlinearen System (1.2) aus.

Die Idee der uniformen Beobachtbarkeit ist die Eindeutigkeit des Anfangszustandes, unabhängig von der Eingangsgröße $u(t)$. Das bedeutet, wir erhalten einen eindeutig bestimmten Anfangszustand eines Systems mit $y(t)$ und $u(t)$.

Das Konzept der uniformen Beobachtbarkeit setzt Beobachtbarkeit in obigem Sinn voraus. Das bedeutet, der Rang der Matrix, die aus den n Lie-Ableitungen besteht, muss voll sein. Die uniforme Beobachtbarkeit fordert nun weiter den Erhalt der Dimension des jeweiligen Raumes, der von der zugehörigen Lie-Ableitung aufgespannt wird. Das bedeutet, hier ist eine Unabhängigkeit der Eingänge $u(t)$ unabdingbar.

In mathematischer Sprache wird diese Idee wie folgt festgehalten:

Definition (uniforme Beobachtbarkeit)

Wir übernehmen folgende Definition von HAMMOURI und FARZA aus [HF03]:

Seien \mathcal{F}_1^u der Raum, der von den Ableitungen $\{dh_1, \dots, dh_p\}$ erzeugt wird, sowie \mathcal{F}_{k+1}^u die Räume, die für $k \geq 1$ von \mathcal{F}_k^u und der Menge der Lie-Ableitungen $\{dL_f^k(h_1), \dots, dL_f^k(h_p)\}$ erzeugt werden.

Offensichtlich gilt dann $\mathcal{F}_1^u \subset \mathcal{F}_2^u \subset \dots \subset \mathcal{F}_k^u \subset \dots$

Das System (1.2) heißt uniform beobachtbar, genau dann, wenn

- i. $\forall u, u' \in U \subset \mathbb{R}^p, \forall x \in \mathbb{R}^n : \mathcal{F}_k^u = \mathcal{F}_k^{u'}$
- ii. $\forall u \in U \subset \mathbb{R}^p, \forall x \in \mathbb{R}^n, \forall k = \{1, \dots, p\} : \dim \mathcal{F}_k^u = \beta_k^u$, d.h. für jedes k ist der zugehörige Raum \mathcal{F}_k^u von konstanter Dimension.

2.2. Die Stabilität eines Systems

Für die Untersuchung der Stabilität eines linearen, zeitinvarianten Systems gibt es verschiedene Herangehensweisen. Gewöhnlich wird hierbei die Analyse der Eigenwerte der Systemmatrix herangezogen.

In unserem Fall wird jedoch von diesem Vorgehen abgesehen.

Es liegt ein nichtlineares, zeitvariantes System (1.2) vor. Für den Nachweis der Stabilität wird an dieser Stelle die Theorie der Ljapunov-Funktionen verwendet, mit Hilfe derer es möglich ist, Stabilität eines Systems im Equilibrium zu zeigen.

Obwohl die Ljapunov-Theorie auf Grund von Stabilitätsuntersuchungen nichtlinearer Systemgleichungen entwickelt wurde, kann sie dennoch auch auf lineare Systeme angewandt werden. Diese Möglichkeit wird weiter unten in Kapitel 3.4 genutzt.

Betrachte also das nichtlineare System

$$\dot{x}(t) = f(x(t), u(t)), \quad (2.1)$$

mit $f : D \rightarrow \mathbb{R}^n$, $D \subseteq \mathbb{R}^n$ und f lokal Lipschitz-stetig.

2.2.1. Bemerkung

Angenommen $\bar{x}(t) \in D$ ist ein Equilibrium von (2.1), das heißt

$$f(\bar{x}(t), u(t)) = 0.$$

Angenommen $\bar{x}(t) \neq 0$. Definiere eine Variablentransformation

$$z(t) := x(t) - \bar{x}(t).$$

Dann ist

$$\dot{z}(t) = \dot{x}(t) = f(x(t), u(t)) = f((z + \bar{x})(t), u(t)) =: g(z(t), u(t))$$

mit $g(0, u(t)) = 0$.

Nach der Transformation hat das System (2.1) ein Equilibrium im Ursprungspunkt.

Es ist folglich möglich, jedes Equilibrium durch Variablentransformation in den Ursprung zu verschieben. Daher wird im Folgenden o.B.d.A. angenommen, dass das Equilibrium von (1.2) im Nullpunkt liegt.

Aus diesem Grund wird im weiteren Verlauf angenommen, dass $f(0, u(t)) = 0$ gilt und es wird die Stabilität der nichtlinearen Zustandsgleichung im Ursprung untersucht.

Selbiges wird für die lineare Systemgleichung angewandt.

2.2.2. Definition (Stabilität des Equilibriums)

Das Equilibrium $x(t) = 0$ von (2.1) heißt

- stabil, wenn

$$\forall \epsilon > 0 \exists \delta(\epsilon) > 0 : \|x(0)\| < \delta(\epsilon) \Rightarrow \|x(t)\| < \epsilon, \quad \forall t \geq 0$$

- asymptotisch stabil, wenn es stabil ist und wenn

$$\|x(0)\| < \delta(\epsilon) \Rightarrow \lim_{t \rightarrow \infty} x(t) = 0$$

- unstabil, falls es nicht stabil ist.

2.2.3. Ljapunov-Stabilität

Nachdem die Stabilität eines Equilibriums definiert wurde, benötigen wir nun ein Werkzeug, um diese nachweisen zu können. Folgendes Theorem liefert uns ebenjenes.

Theorem (Ljapunov, 1892)

Sei $x = 0$ ein Equilibrium für das nichtlineare System (1.2). Sei $V : D \rightarrow \mathbb{R}^n$ eine stetig differenzierbare Funktion, wobei D eine Umgebung um $x = 0$ ist, sodass

- $V(0) = 0$
- $V(x) > 0$ in $D \setminus \{0\}$
- $\dot{V}(x) \leq 0$ in D .

Dann ist das Equilibrium $x = 0$ stabil.

Beweis

Seien $\epsilon > 0$ und $r \in (0, \epsilon] : B_r = \{x \in \mathbb{R}^n : \|x\| \leq r\} \subset D$. Sei $m = \min_{\|x\|=r} V(x)$.

Wegen i. und ii. folgt $m > 0$.

Wähle ein $n \in (0, m)$, sodass $B_n = \{x \in B_r : V(x) \leq n\}$.

Zwischenbehauptung

$B_n \subseteq B_r$.

Zwischenbeweis

Angenommen B_n ist nicht vollständig in B_r enthalten. Dann $\exists p \in B_n$ mit $p \in \partial B_r$.²

Damit ist $V(p) \geq m \geq n$. Dies ist jedoch ein Widerspruch zu der Tatsache, dass $\forall x \in B_n$ gilt $V(x) \leq n$.

Damit ist $B_n \subseteq B_r$. \square

Da B_n nach Definition abgeschlossen ist, wissen wir nun mit dieser Zwischenbehauptung, dass B_n sogar kompakt ist.

Wegen iii. gilt weiter

$$\forall t \geq 0 : \dot{V}(x(t)) \leq 0 \Rightarrow V(x(t)) \leq V(x(0)) \leq n. \quad (2.2)$$

Dies wiederum impliziert, dass jede Lösung, die sich zum Zeitpunkt $t = 0$ in B_n befindet, für alle Zeiten $t \geq 0$ dort bleibt.

Wegen Theorem 1.3.1 und auf Grund der Kompaktheit von B_n folgt:

Für $x(0) \in B_n$ hat (2.1) eine eindeutige Lösung $\forall t \geq 0$.

Aus der Definition von $V(x)$ wissen wir, dass die Funktion stetig ist. Mit diesem Wissen und mit i.

$$\exists \delta > 0 : \|x\| \leq \delta \Rightarrow V(x) < n.$$

Somit sind $B_\delta \subset B_n \subset B_r$ und mit (2.2) ist

$$x(0) \in B_\delta \Rightarrow x(0) \in B_n \Rightarrow x(t) \in B_n \Rightarrow x(t) \in B_r.$$

Schließlich gilt $\forall t \geq 0$

$$\|x(0)\| < \delta \Rightarrow \|x(t)\| < r \leq \epsilon.$$

Damit ist das Equilibrium $x = 0$ stabil. \square

²Es handelt sich bei ∂B_r um den Rand von B_r . Die Schreibweise ist nicht zu verwechseln mit jener der partiellen Ableitung!

Kapitel 3

Das Kalman-Bucy-Filter

Für zeitkontinuierliche, lineare Systeme liegt das sogenannte Kalman-Bucy Filter³ vor. Es handelt sich hierbei um einen speziellen Beobachter für Systeme mit stochastischen Störungen, welche Gauß-Prozesse sind.

Das Ziel des KBF ist das Schätzen des Zustands x . Dieser Beobachter besteht aus einem Modell des Systems und einer geeigneten Rückführung der Ausgangsgrößen $y(t)$.

Das KBF hat die Form des in Kapitel 1.4 eingeführten Beobachters. Jedoch zeichnet es sich durch eine bestimmte Wahl der Matrix $K(t)$ aus. Diese Wahl von $K(t)$ führt zu einer bestmöglichen Zustandsschätzung, da sie eine Berechnung dieser mittels der mittleren quadratischen Abweichung des Schätzfehlers bewirkt.

Dieses Charakteristikum macht das KBF zu einem in der Anwendung häufig verwendeten Mittel zur Schätzung von Zuständen.

Zu Beginn dieses Kapitels werden wir uns in Kürze mit dem Wiener Prozess auseinandersetzen. Die aufgeführte Definition sowie weiterführende Inhalte hierzu sind aus [KK85] zu entnehmen.

Im Anschluss wird das KBF vorgestellt, welches den Zustand linearer, zeitvarianter Systeme schätzt. Zu dieser Thematik gibt es eine Vielzahl an Literatur. Vor allem jedoch sei hier auf das Paper [KB61] von KALMAN und BUCY hingewiesen, in welchen dieses Filter erstmals publiziert wurde. Außerdem werden weitere Informationen zum KBF aus [WB06] sowie aus [Unb02] entnommen.

Schließlich wird in diesem Kapitel der Beweis der Konvergenz des KBF geführt, der in Eigenleistung entstanden ist.

In praktischen Fällen ist die Messung des Ausgangs nicht ohne Störung zu erhalten,

³Im weiteren Verlauf mit KBF abgekürzt.

sondern sie wird von einem Rauschprozess überlagert. Dieser wurde bereits in der Zustandsgleichung aus (1.1) durch Addition von $w(t)$, welches die Störung vertritt, berücksichtigt. In den meisten Fällen ist ebendieses Rauschen der Hauptgrund für Unterschiede zwischen tatsächlichem und gemessenem Ausgang. In diesem Fall wird die Störung als sogenanntes weißes Rauschen deklariert. Um dieses besser zu verstehen und nachzuvollziehen, an welcher Stelle es im Beobachter berücksichtigt wird, werden wir uns im Folgenden zunächst mit dem Wiener Prozess beschäftigen.

3.1. Der Wiener Prozess

Sei $w(t)$, $t \geq 0$ ein stochastischer Prozess.

Es gelte für $w(t)$:

- $w(0) = 0$
- für $a, b, c, d \in \mathbb{R}$ mit $0 \leq a \leq b \leq c \leq d$ sind $w(b) - w(a)$ und $w(d) - w(c)$ unabhängige Prozesse. Das heißt, die Zuwächse sind unabhängig.
- Für jeden Zuwachs $w(b) - w(a)$ gilt $w \sim \mathcal{N}(0, \lambda |b - a|)$, wobei λ als Intensität des Prozesses bezeichnet wird. Das heißt, die Zuwächse haben eine Gauß-Verteilung mit Erwartungswert 0 und Varianz $\lambda |b - a|$.

Ein stochastischer Prozess mit diesen drei Eigenschaften ist als Wiener Prozess⁴ bekannt.

3.2. Die Zustandsschätzung

Die Gleichung zur Schätzung eines Zustandes $x(t)$ entspricht jener, der allgemeinen Beobachtergleichung:

$$\begin{aligned}\hat{x}(t) &= A(t)\hat{x}(t) + B(t)u(t) + K(t)(y(t) - C(t)\hat{x}(t)) \\ &= A(t)\hat{x}(t) + B(t)u(t) + K(t)C(t)(x(t) - \hat{x}(t))\end{aligned}\tag{3.1}$$

Ihr besonderes Merkmal ist allerdings die Wahl von $K(t)$ im Sinne der kleinsten Fehlerquadrate. Wie $K(t)$ für das KBF gewählt wird, wird im kommenden Abschnitt im Rahmen der Herleitung der Schätzfehlerkovarianz thematisiert.

⁴Der Wiener Prozess ist auch bekannt als mathematisches Modell für die Brown'sche Bewegung.

3.3. Die Güte der Zustandsschätzung

Die Natur einer Schätzung bringt mit sich, dass diese der Realität möglichst nahe ist. Ebenso verhält es sich mit der Zustandsschätzung mit Hilfe des KBF, welche dem wahren Zustand entsprechen soll. Das Ziel der Schätzung des Zustands ist daher ein möglichst kleiner Schätzfehler.

Die Güte der Zustandsschätzung wird also anhand des Gütekriteriums, und zwar der Minimierung des Schätzfehlers, charakterisiert. Es ist bereits bekannt, dass die Minimierung des quadratischen Mittels der Differenz von Zustand $x(t)$ und Schätzung $\hat{x}(t)$ die entscheidende Eigenschaft des KBF ist. Insofern halten wir folgende Optimalitätsbedingung fest:

$$\sum_{i=1}^n E \left[(x_i(t) - \hat{x}_i(t))^2 \right] \stackrel{!}{=} \min \quad (3.2)$$

Bei unbekanntem wahren Zustand und vorliegender Schätzung ist es jedoch schwer zu beurteilen, ob eine Schätzung für den jeweiligen Zweck gut oder schlecht ist. Aus diesem Grund wird hierfür ein Gütemaß in Form einer Schätzfehlerkovarianzmatrix $P(t)$ herangezogen.

Wie im weiteren Verlauf ersichtlich wird, hängt die Kovarianzmatrix $P(t)$ in entscheidender Weise von der sogenannten Kalman-Matrix $K(t)$ ab, die wir durch obige Optimalitätsbedingung bestimmen werden.

Im Folgenden erfolgt eine Herleitung dieser beiden Matrizen. Dadurch wird der Grundgedanke des Kalman-Filters, sowie die genau festgelegte Form des Algorithmus nachvollziehbar. Es wird deutlich werden, dass es sich hierbei um ein intuitives Konzept im Sinne eines Optimierungsproblems handelt.

3.3.1. Die Schätzfehlerkovarianzmatrix

Sei

$$e(t) := x(t) - \hat{x}(t)$$

der Schätzfehler, den wir betrachten werden.

Wir definieren die Kovarianzmatrix des Fehlers als eine positiv definite, symmetrische Matrix:

$$P(t) := E \left[e(t)e^T(t) \right], \quad P(t) \in \mathbb{R}^{n \times n} \quad (3.3)$$

Mit Hilfe dieser Definition können wir obige Optimierungsbedingung (3.2) auch anhand der Minimierung der Spur dieser Kovarianzmatrix ausdrücken:

$$J(t) := \text{tr}(P(t)) \stackrel{!}{=} \min \quad (3.4)$$

Mit (3.1) sowie dem Zustandsraummodell (1.1) sind wir in der Lage, die Ableitung des Fehlers nach der Zeit zu bestimmen:

$$\begin{aligned} \frac{de(t)}{dt} &= \frac{dx(t)}{dt} - \frac{d\hat{x}(t)}{dt} \\ &= A(t)x(t) + B(t)u(t) + G(t)w(t) \\ &\quad - \left(A(t)\hat{x}(t) + B(t)u(t) + K(t)(y(t) - C(t)\hat{x}(t)) \right) \\ &= (A(t) - K(t)C(t))e(t) + G(t)w(t) - K(t)v(t) \end{aligned}$$

Zur genaueren Untersuchung der beiden letzten Summanden definieren wir diese zunächst als $G(t)w(t) - K(t)v(t) =: \mu(t)$.

Bei Betrachtung der Kovarianz ergibt sich wegen den Relationen

$$\begin{aligned} E[w(t)w^T(\tau)] &= Q(t)\delta(t - \tau), \\ E[v(t)v^T(\tau)] &= R(t)\delta(t - \tau), \\ E[v(t)w^T(\tau)] &= 0, \end{aligned} \quad (3.5)$$

dass

$$\begin{aligned} E[\mu(t)\mu^T(t)] &= E\left[(G(t)w(t) - K(t)v(t))(G(\tau)w(\tau) - K(\tau)v(\tau))^T \right] \\ &= E\left[(G(t)w(t))(G(\tau)w(\tau))^T - (K(t)v(t))(G(\tau)w(\tau))^T \right. \\ &\quad \left. - (G(t)w(t))(K(\tau)v(\tau))^T + (K(t)v(t))(K(\tau)v(\tau))^T \right] \\ &= E[G(t)w(t)w^T(\tau)G^T(\tau)] - E[K(t)v(t)w^T(\tau)G^T(\tau)] \\ &\quad - E[G(t)w(t)v^T(\tau)K^T(\tau)] + E[K(t)v(t)v^T(\tau)K^T(\tau)] \\ &\stackrel{(3.5)}{=} \underbrace{(G(t)Q(t)G^T(t) + K(t)R(t)K^T(t))}_{=: Y(t)} \delta(t - \tau) \end{aligned} \quad (3.6)$$

Mit diesem Ergebnis ist ersichtlich, dass $\mu(t)$ als ein weißer, stochastischer Prozess zu verstehen ist.

Wir können nun eine Differentialgleichung für die Fehlerkovarianzmatrix $P(t) = E[e(t)e^T(t)]$ aufstellen. Hierfür ist zunächst eine Analyse der Ableitung des Schätzfehlers $e(t)$ nötig.

Betrachte also

$$\frac{de(t)}{dt} = (A(t) - K(t)C(t))e(t) + \mu(t). \quad (3.7)$$

Für die zugehörige Übergangsmatrix⁵ $\Phi(t, t_0)$ gilt

$$\begin{aligned} \frac{\partial \Phi(t, t_0)}{\partial t} &= (A(t) - K(t)C(t))\Phi(t, t_0), \\ \Phi(t_0, t_0) &= I_n. \end{aligned} \quad (3.8)$$

Damit können wir die Lösung der Gleichung bezüglich des Fehlers (3.7) angeben⁶:

$$e(t) = \Phi(t, t_0)e(t_0) + \int_{t_0}^t \Phi(t, \tau)\mu(\tau)d\tau. \quad (3.9)$$

Diese Lösung sei nun zum Zeitpunkt t_0 eine mittelwertfreie Zufallsvariable, die mit $\mu(t)$ unkorreliert ist, d.h.

$$E[e(t)\mu^T(t)] = 0. \quad (3.10)$$

Dann gilt dieses Charakteristikum für den Fehler wegen der Darstellung (3.9) auch für alle weiteren Zeiten t und die Fehlerkovarianzmatrix $P(t_1, t_2) = E[e(t_1)e^T(t_2)]$ mit $t_0 \leq t_1 \leq t_2$ hat eine quadratische Form, da $n = \dim x(t) = \dim e(t)$.

Mit der Integraldarstellung (3.9) sowie mit der Kenntnis von $P(t_1, t_2)$ erhalten wir durch Einsetzen folgendes Ergebnis für die Kovarianzmatrix des Schätzfehlers im Zeitintervall $[t_1, t_2]$:

$$\begin{aligned} P(t_1, t_2) &= E \left[\left(\Phi(t_1, t_0)e(t_0) + \int_{t_0}^{t_1} \Phi(t_1, \tau)\mu(\tau)d\tau \right) \right. \\ &\quad \cdot \left. \left(\Phi(t_2, t_0)e(t_0) + \int_{t_0}^{t_2} \Phi(t_2, \sigma)\mu(\sigma)d\sigma \right)^T \right] \\ &= E[\Phi(t_1, t_0)e(t_0)e^T(t_0)\Phi^T(t_2, t_0)] \\ &\quad + E \left[\left(\Phi(t_1, t_0)e(t_0) + \int_{t_0}^{t_1} \Phi(t_1, \tau)\mu(\tau)d\tau \right) \right. \\ &\quad \cdot \left. \int_{t_0}^{t_2} \mu^T(\sigma)\Phi(t_2, \sigma)d\sigma \right] \end{aligned}$$

⁵Die Herleitung und weitere Erläuterungen zur Übergangsmatrix sind im Anhang A nachzulesen.

⁶Diese Lösungsform wird im Anhang A thematisiert und erklärt.

$$\begin{aligned}
& + E \left[\int_{t_0}^{t_1} \Phi(t_1, \tau) \mu(\tau) d\tau \cdot e^T(t_0) \Phi^T(t_2, t_0) \right] \\
& + E \left[\int_{t_0}^{t_1} \Phi(t_1, \tau) \mu(\tau) d\tau \cdot \int_{t_0}^{t_2} \mu^T(\sigma) \Phi^T(t_2, \sigma) d\sigma \right] \\
& \stackrel{(3.10)}{=} E \left[\Phi(t_1, t_0) e(t_0) e^T(t_0) \Phi^T(t_2, t_0) \right] \\
& + E \left[\int_{t_0}^{t_1} \int_{t_0}^{t_2} \Phi(t_1, \tau) \mu(\tau) \mu^T(\sigma) \Phi^T(t_2, t_0) d\sigma d\tau \right]
\end{aligned}$$

Insofern gilt für die Fehlerkovarianz zum Zeitpunkt t in analoger Weise

$$\begin{aligned}
P(t) &= P(t, t) = E[e(t)e^T(t)] \\
&= E \left[\Phi(t, t_0) e(t_0) e^T(t_0) \Phi^T(t, t_0) \right] + E \left[\int_{t_0}^t \Phi(t, \sigma) \mu(\sigma) \mu^T(\sigma) \Phi^T(t, \sigma) d\sigma \right] \quad (3.11) \\
&\stackrel{(3.3), (3.6)}{=} \Phi(t, t_0) P(t_0) \Phi^T(t, t_0) + \int_{t_0}^t \Phi(t, \sigma) M(\sigma) \Phi^T(t, \sigma) d\sigma.
\end{aligned}$$

An dieser Stelle haben wir theoretisch eine Darstellung von $P(t)$ gefunden, mit Hilfe derer es möglich ist, Aussagen über die Güte der Schätzung zu machen. In der Praxis jedoch erweist sich die Berechnung der Schätzfehlerkovarianzmatrix in oben stehender Form als schwierig, da die Übergangsmatrix Φ immer wieder neu berechnet werden muss.

Abhilfe schafft in diesem Zusammenhang folgende Idee: Wir betrachten zunächst das Differential von $P(t)$ und berechnen daraus die Fehlerkovarianz unter Zuhilfenahme der numerischen Integration. In der Regel ist dieses Vorgehen rechentechnisch einfacher, der Rechenaufwand auf Grund der Kenntnis der Matrizen $A(t)$, $C(t)$, $K(t)$, $Y(t)$ sowie $P(t_0)$ geringer und die Bestimmung von $P(t)$ somit schneller.

Wir setzen diese Idee im Folgenden um. Hierfür wird zunächst die Ableitung von $P(t)$ nach der Zeit bestimmt:

$$\begin{aligned}
 \frac{dP(t)}{dt} &= (A(t) - K(t)C(t))\Phi(t, t_0)P(t_0)\Phi^T(t, t_0) \\
 &\quad + \Phi(t, t_0)P(t_0)\Phi^T(t, t_0)(A(t) - K(t)C(t))^T \\
 &\quad + \Phi(t, t)M(t)\Phi^T(t, t) + \int_{t_0}^t (A(t) - K(t)C(t))\Phi(t, \sigma)M(\sigma)\Phi^T(t, \sigma)d\sigma \\
 &\quad + \int_{t_0}^t \Phi(t, \sigma)M(\sigma)\Phi^T(t, \sigma)(A(t) - K(t)C(t))^T d\sigma \\
 &\stackrel{(3.8)}{=} (A(t) - K(t)C(t))\Phi(t, t_0)P(t_0)\Phi^T(t, t_0) \\
 &\quad + \int_{t_0}^t (A(t) - K(t)C(t))\Phi(t, \sigma)M(\sigma)\Phi^T(t, \sigma)d\sigma \\
 &\quad + \Phi(t, t_0)P(t_0)\Phi^T(t, t_0)(A(t) - K(t)C(t))^T \\
 &\quad + \int_{t_0}^t \Phi(t, \sigma)M(\sigma)\Phi^T(t, \sigma)(A(t) - K(t)C(t))^T d\sigma \\
 &\quad + M(t) \\
 &\stackrel{(3.11)}{=} (A(t) - K(t)C(t))P(t) + P(t)(A(t) - K(t)C(t))^T \\
 &\quad + G(t)Q(t)G^T(t) + K(t)R(t)K^T(t)
 \end{aligned} \tag{3.12}$$

Als Anfangsbedingung ergibt sich

$$P(t_0) = E[e(t_0)e^T(t_0)] = P_0.$$

Wir haben nun eine Form gefunden, mit der es anhand numerischer Integration relativ einfach ist, $P(t)$ zu bestimmen. Allerdings haben wir die oben geforderte Optimalitätsbedingung (3.4), die wir an das KBF stellen, noch nicht berücksichtigt. Insofern ist die Schätzfehlerkovarianzmatrix an dieser Stelle in unserem Sinn noch nicht optimal.

3.3.2. Die Kalman-Matrix

Wir kehren also zu unserem Optimierungsproblem zurück und werden im Folgenden die Spur der Kovarianzmatrix $P(t)$ durch entsprechende Wahl der Kalman-Matrix $K(t)$ minimieren.

Dazu betrachten wir

$$T(t) := \frac{dJ(t)}{dt} = \text{tr} \left(\frac{dP(t)}{dt} \right).$$

Um nun zunächst $T(t)$ bezüglich $K(t)$ zu minimieren, muss die notwendige Bedingung

$$\frac{dT(t)}{dK(t)} = 0$$

erfüllt sein. Bei konvexen Problemen, welche wir für unseren Zweck betrachten, ist die notwendige Bedingung gleichzeitig hinreichend.

Bemerkung

Seien $k_{ij}(t)$ mit $i, j = 1, \dots, n$ die Einträge von $K(t)$, wobei i die Zeile und j die Spalte angibt. Bezeichne

$$\begin{aligned} k_i^T(t) &= \text{i-te Zeile von } K(t) \\ l_j &= \text{j-te Spalte der Matrix } L(t), \text{ mit } L(t) \in \mathbb{R}^{n \times n}. \end{aligned}$$

Dann gelten für die Matrizen $K(t), L(t), U(t)$ mit $U(t) \in \mathbb{R}^{n \times n}$ und $U(t)$ symmetrisch, unter Zuhilfenahme der Relationen

- $tr(K(t)L(t)) = \sum_{i=1}^n k_i^T(t)l_i(t)$
- $tr(L(t)U(t)L^T(t)) = \sum_{i=1}^n k_i^T(t)U(t)l_i(t)$
- $tr(K(t)L(t)) = tr(L(t)K(t))$
- $tr(L(t)K^T(t)) = tr(K^T(t)L(t))$

folgende Formeln:

$$\text{i. } \frac{d}{dK(t)} tr(K(t)L(t)) = \frac{d}{dK(t)} \left(\sum_{i=1}^n k_i^T(t)l_i(t) \right) = \frac{d}{dK(t)} \left(\sum_{i=1}^n l_i^T(t)k_i(t) \right) = L^T(t)$$

$$\text{ii. } \frac{d}{dK(t)} tr(M(t)K^T(t)) = \frac{d}{dK(t)} \left(\sum_{i=1}^n l_i^T(t)k_i(t) \right) = \frac{d}{dK(t)} \left(\sum_{i=1}^n k_i(t)l_i(t) \right) = L(t)$$

$$\begin{aligned} \text{iii. } \frac{d}{dK(t)} tr(K(t)U(t)K^T(t)) &= \frac{d}{dK(t)} \left(\sum_{i=1}^n k_i^T(t)U(t)k_i(t) \right) \\ &= \frac{d}{dK(t)} \left(\sum_{i=1}^n k_i^T(t)k_i(t)U(t) \right) = 2K(t)U(t) \end{aligned}$$

Wir wenden die Relationen aus der Bemerkung auf unser Minimierungsproblem an und erhalten:

$$\begin{aligned}
 \frac{dT(t)}{dK(t)} &= \frac{d}{dK(t)} \left(\operatorname{tr} \left((A(t) - K(t)C(t))P(t) \right) + \operatorname{tr} \left(P(t)(A(t) - K(t)C(t))^T \right) \right. \\
 &\quad \left. + \operatorname{tr} \left(G(t)Q(t)G^T(t) \right) + \operatorname{tr} \left(K(t)R(t)K^T(t) \right) \right) \\
 &= \frac{d}{dK(t)} \operatorname{tr} \left(A(t)P(t) \right) - \frac{d}{dK(t)} \operatorname{tr} \left(K(t)C(t)P(t) \right) + \frac{d}{dK(t)} \operatorname{tr} \left(P(t)A^T(t) \right) \\
 &\quad - \frac{d}{dK(t)} \operatorname{tr} \left(P(t)C^T(t)K(t) \right) + \frac{d}{dK(t)} \operatorname{tr} \left(G(t)Q(t)G^T(t) \right) \\
 &\quad + \frac{d}{dK(t)} \operatorname{tr} \left(K(t)R(t)K^T(t) \right) \\
 &= -2P(t)C^T(t) + 2K(t)R(t)
 \end{aligned}$$

Unter Beachtung unserer Bedingung $\frac{dT(t)}{dt} = 0$ erhalten wir schließlich die Kalman-Matrix

$$\begin{aligned}
 -2P(t)C^T(t) + 2K(t)R(t) &= 0 \\
 \iff K(t) &= P(t)C^T(t)R^{-1}(t).
 \end{aligned}$$

Folglich ist die Optimalitätsbedingung für $K(t) = P(t)C^T(t)R^{-1}(t)$ erfüllt und für das Differential der Schätzfehlerkovarianzmatrix ergibt sich aus (3.12):

$$\frac{dP(t)}{dt} = A(t)P(t) + P(t)A^T(t) - P(t)C^T(t)R^{-1}(t)C(t)P(t) + G(t)Q(t)G^T(t), \quad (3.13)$$

mit der Anfangsbedingung $P(t_0) = P_0$.

Wir erhalten also eine Differentialgleichung in Matrix-Riccati Form.

Im konkreten Fall wird sie numerisch gelöst.

Schlussendlich haben wir die Gleichungen für das KBF, welches durch die Schätzfehlerkovarianzmatrix $P(t)$ im Sinne der Minimierung des Schätzfehlers optimal ist, vorliegen:

$$\begin{aligned}
 \hat{x}(t) &= A(t)\hat{x}(t) + B(t)u(t) + K(t)C(t)(x(t) - \hat{x}(t)) \\
 \dot{P}(t) &= A(t)P(t) + P(t)A^T(t) - P(t)C^T(t)R^{-1}(t)C(t)P(t) + G(t)Q(t)G^T(t) \quad (3.14) \\
 K(t) &= P(t)C^T(t)R^{-1}(t)
 \end{aligned}$$

3.4. Die Konvergenz des Kalman-Bucy-Filters

Gegeben sei das lineare System (1.1) mit dem für den zeitkontinuierlichen Fall zugehörigen Kalman-Filter (3.14).

Wir werden nun die Konvergenz des KBF in einer neuen Weise, mittels der Ljapunov-Theorie, beweisen.

Zunächst jedoch die Behauptung:

3.4.1. Behauptung

Der mit Hilfe des KBF (3.14) geschätzte Zustand $\hat{x}(t)$ des Systems (1.1) konvergiert gegen den tatsächlichen Systemzustand $x(t)$ für $t \rightarrow \infty$.

3.4.2. Beweis

Betrachte den Schätzfehler $e(t) = x(t) - \hat{x}(t)$.

Dann ist

$$\begin{aligned}\dot{e}(t) &= \dot{x}(t) - \dot{\hat{x}}(t) = A(t)(x(t) - \hat{x}(t)) - K(t)C(t)(x(t) - \hat{x}(t)) \\ &= (A(t) - K(t)C(t))e(t).\end{aligned}$$

Beweisidee

Definiere eine Funktion $V(e(t))$ des Schätzfehlers $e(t)$. Untersuche diese auf Eigenschaften einer Ljapunov-Funktion und folgere daraus Stabilität des Fehlers um den Nullpunkt. Ist der Fehler um diesen Punkt stabil, so impliziert dies die Konvergenz des geschätzten Zustandes $\hat{x}(t)$ an den tatsächlichen Zustand $x(t)$ und damit die Konvergenz des KBF.

Zum Beweis

Es ist bereits bekannt, dass $P(t)$ eine symmetrische, positiv definite Matrix ist. Insofern gilt diese Eigenschaft auch für die Inverse $P^{-1}(t)$.

Sei also

$$V(e(t)) := e^T(t)P^{-1}(t)e(t).$$

Überprüfung der Bedingungen aus dem Ljapunov-Theorem aus Kapitel 2.2.3 an eine Ljapunov-Funktion:

- Für $e(t) = 0$: $V(0) = 0 \quad \forall t > 0 \quad \checkmark$
- Aus der Voraussetzung wissen wir bereits, dass $P(t)$ positiv definit ist. Daher gilt $V(e(t)) = e^T(t)P^{-1}(t)e(t) > 0 \quad \forall t > 0 \quad \checkmark$
- Analysiere $\dot{V}(e(t)) = \frac{d}{dt}e^T(t)P^{-1}(t)e(t)$:

$$\begin{aligned}
 \dot{V}(e(t)) &= \dot{e}^T(t)P^{-1}(t)e(t) + e^T(t) \left(\frac{d}{dt}(P^{-1}(t)) \right) e(t) + e^T(t)P^{-1}(t)\dot{e}(t) \\
 &= \left((A(t) - K(t)C(t))e(t) \right)^T P^{-1}(t)e(t) \\
 &\quad + e^T(t) \left(-P^{-1}(t) \frac{d}{dt}(P(t))P^{-1}(t) \right) e(t) \\
 &\quad + e^T(t)P^{-1}(t)(A(t) - K(t)C(t))e(t) \\
 &= e^T(t)(A(t) - K(t)C(t))^T P^{-1}(t)e(t) \\
 &\quad + e^T(t) \left(-P^{-1}(t)(A(t)P(t) + P(t)A^T(t) \right. \\
 &\quad \quad \left. - P(t)C^T(t)R^{-1}(t)C(t)P(t))P^{-1}(t) \right) e(t) \\
 &\quad + e^T(t)P^{-1}(t)(A(t) - K(t)C(t))e(t) \\
 &= e^T(t) \left(A^T(t)P^{-1}(t) - C^T(t)(P(t)C^T(t)R^{-1}(t))^T P^{-1} \right. \\
 &\quad \quad \left. - P^{-1}(t)A(t) - A^T(t)P^{-1} + C^T(t)R^{-1}(t)C(t) \right. \\
 &\quad \quad \left. + P^{-1}(t)A(t) - C^T(t)R^{-1}(t)C(t) \right) e(t) \\
 &= e^T(t) \left(A^T(t)P^{-1}(t) - 2C^T(t)R^{-1}(t)C(t) + P^{-1}(t)A(t) - P^{-1}(t)A(t) \right. \\
 &\quad \quad \left. - A^T(t)P^{-1}(t) - C^T(t)R^{-1}(t)C(t) \right) e(t) \\
 &= e^T(t) \underbrace{\left(-C^T(t)R^{-1}(t)C(t) \right)}_{<0, \text{ da } R(t) \text{ nach Definition positiv definit}} e(t) \\
 &< 0 \quad \forall e \neq 0 \quad \checkmark
 \end{aligned}$$

$\implies V(e(t))$ ist eine Ljapunov-Funktion.

Da $\dot{V}(e(t)) < 0$ folgt sogar asymptotische Stabilität um $e(t) = 0$.

Das bedeutet, der Schätzfehler nähert sich für $t \rightarrow \infty$ immer weiter dem Equilibrium 0 an. Dies wiederum impliziert die Konvergenz des KBF. \square

Kapitel 4

Das Erweiterte Kalman-Bucy-Filter

Das Hauptanliegen dieses Kapitels ist das Beweisen der Konvergenz des Erweiterten Kalman-Bucy-Filters⁷ für nichtlineare Systeme. Es handelt sich hierbei um ein exponentielles Konvergenzverhalten, was eine optimale Näherung des durch das EKBF geschätzten Ergebnisses an die tatsächliche Lösung der nichtlinearen Differentialgleichung bedeutet. Unter der Voraussetzung eines nicht zu großen Fehlers $|e(t_0)| = |x(t_0) - \hat{x}(t_0)|$ der Schätzung des Anfangszustandes konvergiert der Schätzfehler des Zustandes $|e(t)| \rightarrow 0$ für $t \rightarrow \infty$.⁸

Die für dieses Kapitel verwendete Literatur beschränkt sich auf drei Angaben: Die Vorstellung des EKBF wird aus [GA01] bezogen. Für den Beweis dieses Filters wurde bereits eine Vorarbeit von GAUTHIER, HAMMOURI und OTHMAN geleistet, an der sich die Herangehensweise im vorliegenden Beweis orientiert und die in [GHO92] nachgelesen werden kann.

Der Beweis von KRENER in [Kre03] für die Konvergenz dieses Filters für nichtlineare Systeme wird hier ausgearbeitet und es werden, losgelöst von KRENER'S Beweis, Veränderungen vorgenommen, die seinen Beweis verbessern.

⁷Im weiteren Verlauf mit EKBF abgekürzt.

⁸Diese Voraussetzung wird vor allem für die in Kapitel 7 durchgeführten numerischen Experimente von besonderem Interesse sein. In der Praxis ist dies die entscheidende Bedingung, mit welcher anhand einer Implementierung Konvergenz getestet wird.

4.1. Das Erweiterte Kalman-Bucy-Filter

Das EKBF wird auf nichtlineare, zeitabhängige, kontinuierliche Differentialgleichungssysteme angewandt.

Die Zustands- und die Ausgangsgleichung sind von der bereits bekannten Form

$$\begin{aligned}\dot{x}(t) &= f(x(t), u(t)) \\ y(t) &= h(u(t)).\end{aligned}\tag{4.1}$$

Die Schätzung des Zustandes gleicht einer Innovation, im Sinne einer Verbesserung der Schätzung mit Hilfe der Ausgangsgleichung im Laufe der Zeit. Die Schätzung der Fehlerkovarianzmatrix erfordert die Lösung der Riccatischen Matrix-Differentialgleichung, die in Kapitel 3 für den linearen Fall hergeleitet wurde.

Weiterhin wird der geschätzte Zustandsstartwert $\hat{x}(0)$ für das System (4.1) aus dem vorgegebenen Startwert des Zustandes $x(0)$ geschätzt.

Damit ergeben sich schließlich folgende Gleichungen des EKBF für das nichtlineare System (4.1):

$$\begin{aligned}\dot{\hat{x}}(t) &= f(\hat{x}(t), u(t)) + P(t)C^T(t)\left(y(t) - h(\hat{x}(t), u(t))\right) \\ \dot{P}(t) &= A(t)P(t) + P(t)A^T(t) + GG^T - P(t)C^T(t)R^{-1}(t)C(t)P(t) \\ \hat{x}(0) &= \hat{x}_0 \\ P(0) &= P_0,\end{aligned}\tag{4.2}$$

wobei

$$A(t) = \left. \frac{\partial f}{\partial x} \right|_{x=\hat{x}}, \quad C(t) = \left. \frac{\partial h}{\partial x} \right|_{x=\hat{x}}.$$

4.2. Die Konvergenz des EKBF

Die Konvergenz des EKBF wurde bereits Ende der 1980er Jahre von BARAS, BENSOUSSAN und JAMES in [BBJ87] gezeigt. Dabei wurden gewisse Annahmen getätigt und es wurde vorausgesetzt, dass die Schätzung zum Startzeitpunkt nahe genug am eigentlichen Zustand ist. Allerdings erfordert dieser Beweis weitere Voraussetzungen an das System, die so stark einschränkend sind, dass dieser Beweis nahezu nur für Spezialfälle gilt.

Etwa 15 Jahre später, im Jahr 2002, gelang es dem Mathematiker KRENER einen Beweis für lokale Konvergenz des EKBF zu führen, dessen einzig wirklich restriktive Voraussetzung die uniforme Beobachtbarkeit des Systems ist. Da sein im Vergleich zu jenem von BARAS,

BENSOUSSAN und JAMES deutlich weniger Voraussetzungen an das gegebene Differentialgleichungssystem stellt, findet er großen Anklang in der Systemtheorie. Im vorliegenden Abschnitt wird dieser Beweis aufgeführt, detailliert thematisiert und an mancher Stelle verändert.

4.2.1. Vorbemerkungen

Gegeben ist das System

$$\begin{aligned}\dot{x}(t) &= f(x(t), u(t)) \\ y(t) &= h(x(t), u(t)) \\ x(0) &= x_0\end{aligned}\tag{4.3}$$

wobei $x(t) \in \mathbb{R}^n$, $u(s) \in U \subset \mathbb{R}^m$, $y \in \mathbb{R}^p$, $0 \leq s \leq t$.

Die Funktionen f, h sind in der konkreten Anwendung bekannt und es gilt $f, h \in \mathcal{C}^2$.

Gesucht ist eine Schätzung $\hat{x}(t)$, sodass $\hat{x}(t) \rightarrow x(t)$ für $t \rightarrow \infty$.

4.2.2. Umformulierung des nichtlinearen Systems

Ziel des Kapitels ist der Beweis der Konvergenz des EKBF. Dieser benötigt eine andere Form des Systems (4.3), als die vorliegende. Anhand der folgenden Definition ist die entsprechende Transformation möglich.

Betrachte nachstehendes System:

$$\dot{\zeta}(t) = f(\zeta(t), u(t))\tag{4.4}$$

$$y(t) = h(\zeta(t), u(t))\tag{4.5}$$

wobei $\zeta(t) \in \mathbb{R}^n$, $u(s) \in U \subset \mathbb{R}^m$, $y \in \mathbb{R}^p$, $0 \leq s \leq t$.

Definition

(4.4) ist uniform beobachtbar.

\implies es existieren Koordinaten x_{ij} mit $i = 1, \dots, p$ und $j = 1, \dots, l_i$, wobei $1 \leq l_1 \leq \dots \leq l_p$ und $\sum_{i=1}^p l_i = n$, sodass (4.4) diffeomorph ist zu einem System der Form

$$\begin{aligned}
 y_i &= x_{i1} + h_i(u) \\
 \dot{x}_{i1} &= x_{i2} + f_{i1}(x_1, u) \\
 \dot{x}_{i2} &= x_{i3} + f_{i2}(x_2, u) \\
 &\vdots \\
 \dot{x}_{ij} &= x_{ij+1} + f_{ij}(x_j, u) \\
 &\vdots \\
 \dot{x}_{il_i-1} &= x_{il_i} + f_{il_i-1}(x_{l_i-1}, u) \\
 \dot{x}_{il_i} &= f_{il_i}(x_{l_i}, u)
 \end{aligned} \tag{4.6}$$

mit $\bar{x}_j = (x_{11}, x_{12}, \dots, x_{1, \min\{j, l_1\}}, x_{21}, \dots, x_{pj})$.

Beachte: Die Indizes von x_j laufen von $i = 1, \dots, p$ und $k = 1, \dots, \min\{j, l_i\}$ und die Koordinaten sind so angeordnet, dass der zweite Index schneller läuft als der erste.

Für eine übersichtliche und kompakte Form von (4.6), mit der im weiteren Verlauf weitergearbeitet wird, wird dieses System anhand der im Folgenden definierten Matrizen umgeschrieben.

Seien

$$\bar{C} = \begin{bmatrix} \bar{C}_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \bar{C}_p \end{bmatrix}_{p \times n} \quad \bar{C}_i = [1 \ 0 \ 0 \ \dots \ 0]_{1 \times l_i}$$

$$\bar{A} = \begin{bmatrix} \bar{A}_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \bar{A}_p \end{bmatrix}_{n \times n} \quad \bar{A}_i = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}_{l_i \times l_i}$$

$$\bar{f}(x, u) = \begin{bmatrix} \bar{f}_1(x, u) \\ \vdots \\ \bar{f}_p(x, u) \end{bmatrix}_{n \times 1} \quad \bar{f}_i(x, u) = \begin{bmatrix} f_{i1}(x_1, u) \\ \vdots \\ f_{il_i}(x_{l_i}, u) \end{bmatrix}_{l_i \times 1} \quad \bar{h}(u) = \begin{bmatrix} h_1(u) \\ \vdots \\ h_p(u) \end{bmatrix}_{p \times 1} .$$

Mit diesen Matrizen bzw. Vektoren wird aus (4.6) das System

$$\begin{aligned} \dot{x} &= \bar{A}x + \bar{f}(x, u) \\ y &= \bar{C}x + \bar{h}(u) \end{aligned} \tag{4.7}$$

in vereinfachter Form.

4.2.3. Voraussetzungen

Setze folgende Eigenschaften voraus:

- i. Das System (4.3) sei uniform beobachtbar für beliebige Eingänge $u(t)$.
- ii. Die nichtlineare Funktion f aus (4.3) erfülle die Lipschitz-Bedingung, das heißt

$$\exists L, \text{ sodass } \forall x, \eta \in \mathbb{R}^n, u \in U : \|f_i(x_j, u) - f_i(\eta_j, u)\| \leq L \|x_j - \eta_j\|.$$

- iii. Wir wissen bereits aus Abschnitt 4.2.1, dass $f \in \mathcal{C}^2$. Da f sich unter anderem aus \bar{f} zusammensetzt, ist auch $\bar{f} \in \mathcal{C}^2$. Insofern ist die Existenz der zweiten Ableitung von \bar{f} gewährleistet und auf dieser Grundlage wird gefordert, dass diese zweite Ableitung beschränkt ist, das heißt, es gelte

$$\forall x, \eta \in \mathbb{R}^n, u \in U : \left\| \frac{\partial^2 \bar{f}}{\partial x_i \partial x_j}(x, u) \right\| \leq M, \quad 0 < M \in \mathbb{R}.$$

- iv. Die Störmatrix G sei invertierbar.
- v. Der geschätzte Zustandsvektor $\hat{x}(t)$ und die Schätzfehlerkovarianzmatrix $P(t)$ seien Lösung des EKBF (4.2).
- vi. $P(0)$ sei positiv definit.
- vii. Es gelte $|e(t_0)| \leq \epsilon$, das heißt, der Schätzfehler zum Zeitpunkt t_0 ist genügend klein, sodass die Schätzung des Startwertes nahe genug am tatsächlichen Startwert liegt.

4.2.4. Behauptung

Unter den oben gestellten Voraussetzungen an das nichtlineare Zustandsmodell sowie an das EKBF, wird Folgendes behauptet:

Die Schätzung des Zustands $\hat{x}(t)$ mittels des EKBF konvergiert bei zunehmender Zeit exponentiell gegen den tatsächlichen Zustand $x(t)$. Das heißt

$$|e(t)| = |x(t) - \hat{x}(t)| \rightarrow 0 \text{ exponentiell für } t \rightarrow \infty.$$

4.2.5. Beweis

Gemäß Abschnitt 4.2.2 nehmen wir an, dass das System

$$\begin{aligned}\dot{x} &= f(x(t), u(t)) \\ y &= h(x(t), u(t))\end{aligned}$$

beobachtbar ist und daher in die Form

$$\begin{aligned}\dot{x} &= \bar{A}x + \bar{f}(x, u) \\ y &= \bar{C}x + \bar{h}(u)\end{aligned}$$

umformuliert werden kann.

Folglich gilt

$$f(x(t), u(t)) = \bar{A}x(t) + \bar{f}(x(t), u(t))$$

und durch Anwendung der Ableitung von f nach dem Zustand $x(t)$ an der Stelle $(\hat{x}(t), u(t))$ gilt

$$\frac{\partial f}{\partial x}(\hat{x}(t), u(t)) = \bar{A} + \frac{\partial \bar{f}}{\partial x}(\hat{x}(t), u(t)).$$

Wir kürzen diese Notation durch $A(t) = \bar{A} + \tilde{A}(t)$ ab.

Ferner folgt aus obiger Umformulierung: $h(x(t), u(t)) = \bar{C}x(t) + \bar{h}(u(t))$.

In analoger Weise wird hierauf die Ableitung der Messgleichung h nach dem Zustand x an derselben Stelle (\hat{x}, u) angewandt und es folgt

$$\frac{\partial h}{\partial x}(\hat{x}, u) = \bar{C}.$$

Da jedoch $\frac{\partial h}{\partial x}(\hat{x}, u) = C(t)$ ist, erhalten wir $C(t) = \bar{C}$.

Die nun nachfolgende Beweisführung, das System zunächst in eine bestimmte Form zu transformieren und dann den Beweis anhand der transformierten Form zu führen, folgt der Idee von GAUTHIER, HAMMOURI und OTHMAN in [GHO92].

Dieser Beweis ist nicht sofort klar ersichtlich und er ist außerdem von beträchtlicher Länge. Um ihn besser zu verstehen und nachvollziehen zu können sowie auf Grund der Übersichtlichkeit, wird nun vor der Beweisführung zunächst die zu Grunde liegende Idee dargestellt.

Beweisidee

Wir zeigen, dass die Schätzfehlerkovarianzmatrix $P(t)$ nach unten und nach oben beschränkt ist. Dazu betrachten wir Minimierungsprobleme im Intervall $[0, t]$ und wenden eine suboptimale Lösung des Minimierungsproblems auf dieses an, um eine Ungleichung zu erhalten und folglich eine Abschätzung zu ermöglichen. Mit den Schranken wird ersichtlich, in welchem Bereich sich $P(t)$ befindet und es besteht die Hoffnung, dass dieser Bereich eine Umgebung um den Nullpunkt ausmacht.

Definiere danach in analoger Weise zur Beweisidee für das KBF die Funktion $V(t) = e^T(t)P^{-1}(t)e(t)$ als Ljapunov-Funktion und betrachte zugehöriges Differential $\dot{V}(t) = \frac{d}{dt}e^T(t)P^{-1}(t)e(t)$.

Weise dann mit Hilfe der zuvor getätigten Abschätzungen - vor allem mit jenen für die Fehlerkovarianzmatrix $P(t)$ - die drei Eigenschaften einer Ljapunov-Funktion nach und erhalte damit die Stabilität des Fehlers im Nullpunkt.

Dies impliziert schließlich die Konvergenz des Schätzfehlers gegen Null und folglich die Konvergenz des EKBF gegen die Lösung der gegebenen Differentialgleichung.

Zum Beweis

Für die Untersuchung des Fehlers $e(t)$ betrachten wir das Differential $\frac{d}{dt}e^T(t)P^{-1}(t)e(t)$. Insbesondere ist daher eine genauere Analyse, im Sinne der Beschränktheit von $P(t)$ nötig. Diese Matrix ist die Schätzfehlerkovarianz des Zustandes und sagt daher etwas über die Güte der Schätzung aus. Je kleiner die Kovarianz, desto besser die Schätzung.

Das EKBF berücksichtigt eben diese Minimierung der Kovarianzmatrix $P(t)$. Die Aufgabe der Minimierung von $P(t)$ ist äquivalent zur Aufgabe der Energieminimierung eines entsprechenden Energiefunktionals. Daher betrachten wir zu Beginn des Beweises ein Optimierungsproblem.

Sei hierzu

$$J(\mu) = \zeta^T(0)P(0)\zeta(0) + \int_0^t \zeta^T(s)GG^T\zeta(s) + \mu^T(s)\mu(s)ds \quad (4.8)$$

die Zielfunktion, die unter den folgenden Nebenbedingungen

$$\begin{aligned} \dot{\zeta}(s) &= -A^T(s)\zeta(s) - \bar{C}^T\mu(s) \\ \zeta(t) &= \zeta \end{aligned} \quad (4.9)$$

minimiert werden soll.

Für das Minimum von (4.8) gilt

$$J(\mu) = \min \left(\zeta^T(0)P(0)\zeta(0) + \int_0^t \zeta^T(s)GG^T\zeta(s) + \mu^T(s)\mu(s)ds \right) = \zeta^T P(t)\zeta.$$

Hierbei ist $P(t)$ die Schätzfehlerkovarianzmatrix, die Lösung des EKBF (4.2) ist.

Mit diesem Minimum wird nun weiter gearbeitet, indem es nach oben abgeschätzt wird, was letztlich - wie im Folgenden ausgeführt - eine Abschätzung für $P(t)$ impliziert. Die Wahl des entsprechenden Optimierungsproblems liegt diesem Gedanken zu Grunde. Mit anderen Worten: Es werden die benötigte Differentialgleichung sowie das Minimierungsproblem „konstruiert“, um das gewollte Ergebnis zu erhalten.

Zunächst ist daher zu zeigen: $\exists m_1 > 0$, sodass $\forall t \geq 0 : P(t) \leq m_1 I_{n \times n}$.

GAUTHIER, HAMMOURI und OTHMAN haben in [GHO92] gezeigt, dass

$$\dot{\hat{x}}(t) = \bar{A}(t)\hat{x}(t) + \bar{f}(x, u) - S(\theta)C^T(y - C\hat{x})$$

ein weiterer Beobachter für das System (4.7) ist und $S(\theta)$ die Gleichung

$$-\theta S(\theta) = \bar{A}(t)S(\theta) + S(\theta)\bar{A} - \bar{C}^T\bar{C}, \quad \theta > 0 \quad (4.10)$$

erfüllt.

Nach einigen Umformungen von (4.10) folgt die Äquivalenz dieser zur Ljapunov-Gleichung

$$\left(-\bar{A} - \frac{\theta}{2}I \right)^T S(\theta) + S(\theta) \left(-\bar{A} - \frac{\theta}{2}I \right) = -\bar{C}^T\bar{C}.$$

Da $S(\theta)$ folglich auch diese erfüllt, ist ebendiese positiv definit, für $\theta > 0$.

Weiter ist anzumerken, dass für alle Eigenwerte λ_i von $\left(-\bar{A} - \frac{\theta}{2}I \right)$ gilt: $\lambda_i = -\frac{\theta}{2}$. Dann wissen wir, dass $-\bar{A} - \frac{\theta}{2}I$ eine Stabilitätsmatrix ist.

Wegen

$$-\theta S(\theta) = \bar{A}^T S(\theta) + S(\theta)\bar{A} - \bar{C}^T\bar{C} \quad (4.11)$$

ist⁹

$$S_{i,j,\rho,\sigma}(\theta) = \frac{S_{i,j,\rho,\sigma}(1)}{\theta^{j+\sigma-1}} = \frac{(-1)^{j+\sigma}}{\theta^{j+\sigma-1}} \binom{j+\sigma-2}{j-1}.$$

Definiere nun $T(\theta) = S^{-1}(\theta)$.

Dann ist entsprechend

$$T_{i,j,\rho,\sigma}(\theta) = S_{i,j,\rho,\sigma}^{-1}(\theta) = \theta^{j+\sigma-1} S_{ij,\rho\sigma}^{-1}(1) = \theta^{j+\sigma-1} T_{i,j,\rho,\sigma}(1)$$

und $T(\theta)$ erfüllt die Matrix-Riccati-Gleichung

$$-\bar{A}^T(\theta) - T(\theta)\bar{A}^T + T(\theta)\bar{C}^T\bar{C}^T(\theta) = \theta T(\theta),$$

wie mit (4.11) leicht einzusehen ist.

Wende nun nicht das Minimum, sondern eine „etwas größere“ Lösung

$$\mu = -\bar{C}T(\theta)\xi$$

auf das Minimierungsproblem (4.8) an.

Der Gedanke hinter diesem Vorgehen ist das Erhalten einer Abschätzung für die Fehlerkovarianzmatrix $P(t)$. Sie führt zu einer oberen Schranke für $P(t)$.

⁹Um diese Implikation besser nachvollziehen zu können, hier eine Erklärung für den Fall $p=1$:

Für $p=1$ ist

$$\bar{A} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad \bar{C} = [1 \quad 0 \quad 0 \quad \cdots \quad 0]$$

Es reicht $S(1)$ zu bestimmen:

$$\bar{A}^T S(1) + S(1)\bar{A} - \bar{C}^T\bar{C} = -S(1)$$

Dann gilt für den 1. Eintrag in der S-Matrix: $-1 = -S_{11}$, also $S_{11} = 1$.

Allgemein, für den (j, σ) -Eintrag in S gilt damit

$$S_{j-1,\sigma} + S_{j,\sigma-1} = S_{j,\sigma}$$

mit der Vereinbarung, dass $S_{0,\sigma} = S_{j,0} = 1$.

Ein Vergleich dieser Rekursion mit der rekursiven Darstellung des Binomialkoeffizienten zeigt dann die hier gewünschte Relation.

Damit ist

$$\begin{aligned} & \zeta^T(0)P(0)\zeta(0) + \int_0^t \zeta^T(s)GG^T\zeta(s) + \left(-CT(\theta)\zeta(s)\right)^T \left(-CT(\theta)\zeta(s)\right) ds \\ &= \zeta^T(0)P(0)\zeta(0) + \int_0^t \zeta^T(s) \left(GG^T + T(\theta)^T C^T C T(\theta)\right) \zeta(s) ds. \end{aligned}$$

Da es sich bei diesem μ um eine suboptimale Lösung handelt, wir jedoch das Minimum, welches das Minimierungsproblem löst, schon kennen, muss das Einsetzen von μ in (4.8) größer sein als das Minimum. Dies führt zur Abschätzung

$$\zeta^T P(t) \zeta \leq \zeta^T(0)P(0)\zeta(0) + \int_0^t \zeta^T(s) \left(GG^T + T(\theta)^T C^T C T(\theta)\right) \zeta(s) ds. \quad (4.12)$$

Entsprechend verändern sich durch diese Wahl von μ auch die zugehörigen Nebenbedingungen:

$$\begin{aligned} \dot{\zeta}(s) &= -A^T(s)\zeta(s) - \bar{C}^T \mu(s) \\ &= \left(-A^T(s) + \bar{C}^T \bar{C} T(\theta)\right) \zeta(s) \\ \zeta(t) &= \zeta \end{aligned} \quad (4.13)$$

Im Folgenden wird es nun darum gehen, die rechte Seite der Ungleichung (4.12) genauer zu analysieren und abzuschätzen.

i. Betrachte $P(0)$.

Aus der Voraussetzung ist bereits bekannt, dass $P(0)$ positiv definit ist.

Daher \exists Konstante $0 < m_2 \in \mathbb{R}$, sodass $P(0) \leq m_2 I_{n \times n}$, das heißt, $P(0)$ ist nach oben beschränkt.

ii. Betrachte $GG^T + T(\theta)^T C^T C T(\theta)$.

Da es sich hierbei um quadratische Ausdrücke handelt, ist

$$GG^T + T(\theta)^T C^T C T(\theta) \geq 0,$$

also positiv semidefinit.

Außerdem \exists Konstante $0 < m_3(\theta) \in \mathbb{R}$, sodass $GG^T + T(\theta)^T C^T C T(\theta) \leq m_3(\theta) I_{n \times n}$, der Ausdruck ist folglich nach oben beschränkt.

iii. Betrachte $\zeta^T(s)\zeta(s)$, wobei $0 \leq s \leq t$.

Aus (4.13) folgt

$$\begin{aligned}
 & \frac{d}{ds} \zeta^T(s) T(\theta) \zeta(s) \\
 &= \left((-A^T(s) + \bar{C}^T \bar{C} T(\theta)) \zeta(s) \right)^T T(\theta) \zeta(s) \\
 & \quad + \zeta^T(s) T(\theta) \left((-A^T(s) + \bar{C}^T \bar{C} T(\theta)) \zeta(s) \right) \\
 &= \zeta^T(s) \underbrace{\left(\underbrace{\theta T(\theta)}_{(1)} + T(\theta) \bar{C}^T \bar{C} T(\theta) - \underbrace{\tilde{A}(s) T(\theta)}_{(2)} - T(\theta) \tilde{A}^T(s) \right)}_{(3)} \zeta(s).
 \end{aligned} \tag{4.14}$$

Wir wissen, dass $f(x(t), u(t))$ Lipschitz-stetig ist. Daher ist $\bar{A}x(t) + \bar{f}(x(t), u(t))$ ebenfalls Lipschitz-stetig, insbesondere also \bar{A} . Aus den Forderungen $\bar{f} \in \mathcal{C}^2$ sowie der Beschränktheit der zugehörigen zweiten Ableitung, sind die Voraussetzungen erfüllt, um Lipschitz-Stetigkeit der ersten Ableitung von \bar{f} zu folgern.

Daher:

$$\|A(s)\| = \|\bar{A} + \tilde{A}(s)\| \leq \|\bar{A}\| + \|\tilde{A}(s)\| = \|\bar{A}\| + \left\| \frac{\partial \bar{f}}{\partial x}(\hat{x}(s), u(s)) \right\| \leq L \tag{4.15}$$

Dies impliziert die Beschränktheit von $\tilde{A}(s)$.

Es sind $A(s) = \bar{A} + \tilde{A}(s)$ und $\|A(s)\| \leq L$. Dann:

$$\begin{aligned}
 \|A(s)\| &\leq L \\
 \iff \|A(s)\| + \|\bar{A}\| &\leq L + \|\bar{A}\|
 \end{aligned}$$

und

$$\|A(s)\| + \|\bar{A}\| \geq \|A(s) + \bar{A}\| = \|\bar{A} + \tilde{A}(s) + \bar{A}\| \geq \|\tilde{A}(s)\|.$$

Insgesamt also $\|\tilde{A}(s)\| \leq L + \|\bar{A}\|$.

Wir analysieren die rechte Seite von (4.14) in Teilschritten:

iii.i. Betrachte (1).

$$T_{ij}(\theta) = \theta^{j+i-1} T_{ij}(1) \iff \theta T_{ij}(\theta) = \theta^{j+i} T_{ij}(1)$$

iii.ii. Betrachte (2).

Wegen $T_{ij}(\theta) = \theta^{j+i-1}T_{ij}(1)$ und $\|\tilde{A}(s)\| \leq L + \|\tilde{A}\|$ gilt

$$\begin{aligned} \limsup_{s \rightarrow \infty} \frac{(\tilde{A}(s)T(\theta))_{ij}}{\theta^{j+i-1}} &= \limsup_{s \rightarrow \infty} \frac{\tilde{A}(s)\theta^{j+i-1}T_{ij}(1)}{\theta^{j+i-1}} \\ &= \limsup_{s \rightarrow \infty} \tilde{A}(s)T_{ij}(1) \leq \limsup_{s \rightarrow \infty} (L + \|\tilde{A}\|)T_{ij}(1) \leq \infty \end{aligned}$$

Insofern ist $(\tilde{A}(s)T(\theta))_{ij} \in \mathcal{O}(\theta)^{j+i-1}$

iii.iii. Betrachte (3).

Mit Hilfe von i.i. und i.ii. kann der Klammerausdruck (3) nach oben abgeschätzt werden:

$$\underbrace{\theta T(\theta)}_{\text{iii.i. } \theta^{j+1}T(1) > 0} + \underbrace{T(\theta)\tilde{C}^T\tilde{C}T(\theta)}_{> 0} - \underbrace{\tilde{A}(s)T(\theta)}_{\text{iii.ii. } \in \mathcal{O}(\theta)^{j+i-1}} - \underbrace{T(\theta)\tilde{A}^T(s)}_{\text{iii.i. } \theta^{j+1}T(1) > 0}$$

Folglich gilt $\forall \alpha > 0 \exists \theta$, wobei θ genügend groß ist, sodass

$$\theta T(\theta) + T(\theta)\tilde{C}^T\tilde{C}T(\theta) - \tilde{A}(s)T(\theta) - T(\theta)\tilde{A}^T(s) \geq \alpha I_{n \times n} > 0 \quad (4.16)$$

beziehungsweise

$$-\theta T(\theta) - T(\theta)\tilde{C}^T\tilde{C}T(\theta) + \tilde{A}(s)T(\theta) + T(\theta)\tilde{A}^T(s) \leq -\alpha I_{n \times n}. \quad (4.17)$$

Aus (4.16) folgt schließlich

$$\frac{d}{ds} \tilde{\zeta}^T(s)T(\theta)\tilde{\zeta}(s) \geq \tilde{\zeta}^T(s)\alpha I_{n \times n}\tilde{\zeta}(s) \geq 0.$$

Folglich ist $\tilde{\zeta}^T(s)T(\theta)\tilde{\zeta}(s)$ monoton steigend, also

$$\tilde{\zeta}^T(s)T(\theta)\tilde{\zeta}(s) \leq \tilde{\zeta}^T(t)T(\theta)\tilde{\zeta}(t)$$

Rückwärtsintegration der Gleichung (4.14) in der Zeit und Berücksichtigung der Ungleichung (4.17) ergibt für $0 \leq s \leq t$:

$$\tilde{\zeta}^T(s)T(\theta)\tilde{\zeta}(s) \leq \tilde{\zeta}^T(t)e^{\alpha(s-t)}\tilde{\zeta}(t) = \tilde{\zeta}^T\zeta e^{\alpha(s-t)} \quad (4.18)$$

Für $\zeta^T(s)\zeta(s) \exists 0 < m_4(\theta) \in \mathbb{R}$, sodass

$$\zeta^T(s)\zeta(s) \leq m_4(\theta)\zeta^T(s)T(\theta)\zeta(s) \quad (4.19)$$

und mit (4.18) folgt dann

$$\zeta^T(s)\zeta(s) \leq m_4(\theta)e^{\alpha(s-t)}\zeta^T\zeta.$$

Hiermit haben wir die gesuchte Abschätzung für $\zeta^T(s)\zeta(s)$ gefunden.

Einfügen der erhaltenen Abschätzungen für einzelne Terme aus (4.12) in ebendiese Ungleichung ergibt somit die nachfolgende obere Schranke $m_1 I_{n \times n}$ für $P(t)$:

$$\begin{aligned} \zeta^T P(t) \zeta &\leq m_2 m_4(\theta) e^{-\alpha t} \zeta^T \zeta + \int_0^t m_3(\theta) I_{n \times n} m_4(\theta) e^{\alpha(s-t)} \zeta^T \zeta ds \\ &\leq m_2 m_4(\theta) \zeta^T \zeta + m_3(\theta) m_4(\theta) \int_{-\infty}^t e^{\alpha(s-t)} \zeta^T \zeta ds \\ &\leq \underbrace{m_4(\theta) \left(m_2 + \frac{m_3(\theta)}{\alpha} \right)}_{=: m_1} \zeta^T \zeta \end{aligned} \quad (4.20)$$

Es wurde also eine obere Schranke für die Schätzfehlerkovarianzmatrix $P(t)$ gefunden. Um ihr Verhalten noch besser zu verstehen und um zu zeigen, dass sie bei zunehmender Zeit nicht divergiert, sondern wenn möglich klein bleibt, gilt es im weiteren Verlauf des Beweises eine untere Schranke für $P(t)$ zu finden beziehungsweise zu zeigen, dass es eine solche gibt.

Das Vorgehen gleicht hierbei dem des Aufzeigens der oberen Schranke für $P(t)$.

Definiere zunächst $Q(t) := P^{-1}(t)$, $Q(t) \in \mathbb{R}^{n \times n}$.

Zu zeigen: \exists eine Konstante $0 < m_5 \in \mathbb{R}$ sodass $\forall t \geq 0 : Q(t) \leq m_5 I_{n \times n}$.

Damit wäre eine untere Schranke für $P(t)$ gefunden, denn dann wäre

$$P^{-1}(t) \leq m_5 I_{n \times n} \iff P(t) \geq \frac{1}{m_5} I_{n \times n}.$$

Zunächst setzen wir Q in das EKBF (4.2) ein und erhalten

$$\begin{aligned}\dot{Q}(t) &= -A^T(t)Q(t) - Q(t)A(t) - Q(t)GG^TQ(t) + \bar{C}^T\bar{C} \\ Q(0) &= P^{-1}(0).\end{aligned}\tag{4.21}$$

Die Struktur des EKBF (4.2) bleibt weiterhin erhalten und $Q(t)$ löst diese Differentialgleichung. Daher wird der Beweis auf dieser Grundlage fortgeführt.

Wie schon für die Abschätzung von $P(t)$ nach oben, betrachten wir auch hier ein Minimierungsproblem der folgenden Gestalt:

$$\zeta^T(0)Q(0)\zeta(0) + \int_0^t \zeta^T(s)\bar{C}^T\bar{C}\zeta(s) + \mu^T(s)\mu(s)ds\tag{4.22}$$

mit den Nebenbedingungen

$$\begin{aligned}\dot{\zeta}(s) &= A(s)\zeta(s) + G\mu(s) \\ \zeta(t) &= \zeta\end{aligned}\tag{4.23}$$

Es ist

$$\min J'(\mu) = \min \left(\zeta^T(0)Q(0)\zeta(0) + \int_0^t \zeta^T(s)\bar{C}^T\bar{C}\zeta(s) + \mu^T(s)\mu(s)ds \right) = \zeta^T Q(t)\zeta$$

und $Q(t)$ ist Lösung des EKBF (4.21).

Dieses Minimum soll nun nach oben abgeschätzt werden, um die obere Schranke für $Q(t)$ beziehungsweise die untere Schranke für $P(t)$ zu erhalten.

Da es sich hier um das Minimum handelt, wird nun die suboptimale Lösung

$$\mu(s) = G^{-1}(\alpha I_{n \times n} - A(s))\zeta(s)$$

des Minimierungsproblems (4.22) auf ebenjenes angewandt.

Dies impliziert:

$$\begin{aligned}\zeta^T Q(t)\zeta &\leq \zeta^T(0)Q(0)\zeta(0) \\ &+ \int_0^t \zeta^T(s) \left(\bar{C}^T\bar{C} + (\alpha I_{n \times n} - A^T(s))(GG^T)^{-1}(\alpha I_{n \times n} - A(s)) \right) \zeta(s)ds\end{aligned}\tag{4.24}$$

Anpassung der nach Wahl von μ veränderten Nebenbedingungen liefert:

$$\begin{aligned}\dot{\xi}(s) &= A(s)\xi(s) + (\alpha I_{n \times n} - A(s))\xi(s) \\ \xi(t) &= \zeta\end{aligned}\tag{4.25}$$

Für das Erhalten einer oberen Schranke von $Q(t)$ wird nun die rechte Seite von (4.24) genauer untersucht.

a. Betrachte $Q(0)$.

Aus der Voraussetzung ist bekannt, dass $P(0)$ positiv definit ist. Folglich gilt dies auch für $Q(0)$.

Weiterhin \exists Konstante $0 < m_6 \in \mathbb{R}$, sodass $Q(0)$ beschränkt ist:

$$Q(0) \leq m_6 I_{n \times n}$$

b. Betrachte den Faktor $\bar{C}^T \bar{C} + (\alpha I_{n \times n} - A^T(s))(GG^T)^{-1}(\alpha I_{n \times n} - A(s))$.

Wir wissen aus (4.15), dass $A(s) \leq L$ und ebenso $A^T(s) \leq L$. Damit:

$$\bar{C}^T \bar{C} + (\alpha - L)(GG^T)^{-1}(\alpha - L) = \bar{C}^T \bar{C} + (\alpha - L)^2 (GG^T)^{-1}$$

Es handelt sich hierbei um quadratische Ausdrücke, weshalb der Term $\bar{C}^T \bar{C} + (\alpha - L)^2 (GG^T)^{-1} \geq 0$ ist.

Ferner \exists Konstante $0 < m_7 \in \mathbb{R}$ sodass $\bar{C}^T \bar{C} + (\alpha - L)^2 (GG^T)^{-1}$ beschränkt ist:

$$\bar{C}^T \bar{C} + (\alpha - L)^2 (GG^T)^{-1} \leq m_7 I_{n \times n}$$

c. Betrachte $\xi(s)$.¹⁰

Setze gewähltes $\mu(s)$ in (4.25) ein und erhalte:

$$\begin{aligned}\dot{\xi}(s) &= A(s)\xi(s) + G \left(G^{-1}(\alpha I_{n \times n} - A(s))\xi(s) \right) \\ &= \alpha \xi(s) \\ \xi(t) &= \zeta\end{aligned}$$

Die Differentialgleichung $\dot{\xi}(s) = \alpha \xi(s)$ hat die analytische Lösung $\xi(s) = e^{\alpha(s-t)} \zeta$.

Schließlich werden alle in a. - c. erhaltenen Abschätzungen für die Terme aus (4.24) sowie die Kenntnis von $\xi(s)$ und $\xi(t)$ in entsprechende Ungleichung eingefügt und es ergibt

¹⁰An dieser Stelle ist keine Analyse von $\frac{d}{ds} \xi^T(s) \xi(s)$ nötig, da $\xi(s)$ nur von s abhängt und durch die Wahl von $\mu(s)$ zu einer linearen Differentialgleichung wird. Aus dieser ist es möglich $\xi(s)$ analytisch zu bestimmen.

sich die obere Schranke $m_5 I_{n \times n}$ für $Q(t)$:

$$\begin{aligned}
 \zeta^T Q(t) \zeta &\leq \zeta^T(0) Q(0) \zeta(0) \\
 &\quad + \int_0^t \zeta^T(s) \left(\bar{C}^T \bar{C} + (\alpha I_{n \times n} - A^T(s))(GG^T)^{-1}(\alpha I_{n \times n} - A(s)) \right) \zeta(s) ds \\
 &\leq e^{-2\alpha t} \zeta^T Q(0) \zeta + \int_0^t e^{2\alpha(s-t)} \zeta^T (\bar{C}^T \bar{C} + (\alpha + L)^2 (GG^T)^{-1}) \zeta ds \\
 &\leq \zeta^T Q(0) \zeta + (\bar{C}^T \bar{C} + (\alpha + L)^2 (GG^T)^{-1}) \int_0^\infty e^{2\alpha(s-t)} \zeta^T \zeta ds \\
 &\leq \underbrace{\left(m_6 + \frac{m_7}{2\alpha} \right)}_{m_5} \zeta^T \zeta
 \end{aligned} \tag{4.26}$$

Damit ist die obere Schranke für $Q(t)$ und insbesondere die gesuchte untere Schranke für $P(t)$ gefunden. Die Schätzfehlerkovarianzmatrix ist damit beschränkt.

Die gewonnenen Ergebnisse werden nun benutzt, um schließlich die Konvergenz des EKBF nachzuweisen.

Betrachte dazu das nichtlineare Differentialgleichungssystem (4.7)

$$\begin{aligned}
 \dot{x}(t) &= \bar{A}x + \bar{f}(x, u) \\
 y(t) &= \bar{C}x + \bar{h}(x, u) \\
 x(0) &= x_0.
 \end{aligned}$$

Seien $x(t), u(t), y(t)$ Lösungen hierfür und sei $\hat{x}(t)$ Lösung des EKBF (4.2).

Definiere $e(t) := x(t) - \hat{x}(t)$, $e_0 := x_0 - \hat{x}_0$.

Wir möchten nun Stabilität des Schätzfehlers $e(t)$ nachweisen. Hierfür wird die Theorie der Ljapunov-Funktionen herangezogen, mit Hilfe derer Stabilität nachgewiesen werden kann.

Definiere also

$$V(e(t)) := e^T(t) Q(t) e(t)$$

als Kandidat für eine Ljapunov-Funktion des Fehlers $e(t)$.

Wie oben im Ljapunov-Theorem aus Kapitel 2.2.3 gesehen, wird $V(e(t))$ nun auf entsprechende Eigenschaften überprüft:

- $e(t) = 0 : V(0) = 0 \quad \forall t > 0 \quad \checkmark$
- Da $P(t)$ positiv definit ist, gilt:

$$V(e(t)) := e^T(t)Q(t)e(t) > 0 \quad \forall t > 0 \quad \checkmark$$

- Es ist

$$\begin{aligned} \frac{d}{dt}e^T(t)Q(t)e(t) &= -\underbrace{e^T(t)(\bar{C}^T\bar{C} + Q(t)GG^TQ(t))e(t)}_{(*)} \\ &\quad + 2e^T(t)Q(t)\underbrace{\left(\bar{f}(x(t), u(t)) - \bar{f}(\hat{x}(t), u(t)) - \tilde{A}(t)e(t)\right)}_{(**)} \end{aligned}$$

Schätze (*) mit Hilfe dem Wissen, dass $\bar{C}^T\bar{C} \geq I$, $GG^T \geq m_8I$, wobei $0 < m_8 \in \mathbb{R}$, und $P(t) \leq m_1I_{n \times n}$ wie folgt ab:

$$\begin{aligned} e^T(t)(\bar{C}^T\bar{C} + Q(t)GG^TQ(t))e(t) &\geq e^T(t)(Q(t)GG^TQ(t))e(t) \\ &\geq m_8e^T(t)(Q(t)Q(t))e(t) \\ &\geq \frac{m_8}{m_1^2}|e(t)|^2 \\ -e^T(t)(\bar{C}^T\bar{C} + Q(t)GG^TQ(t))e(t) &\leq -\frac{m_8}{m_1^2}|e(t)|^2 \end{aligned}$$

Betrachte nun (**).

Anwenden des Mittelwertsatzes für vektorwertige Funktionen ergibt folgendes Ergebnis:

$$\begin{aligned} &\left(\bar{f}(x(t), u(t)) - \bar{f}(\hat{x}(t), u(t)) - \tilde{A}(t)e(t)\right) \\ &= \int_0^1 \int_0^1 r \frac{\partial^2 \bar{f}}{\partial x_i \partial x_j}(\hat{x}(t) + rse(t), u(t))e_i(t)e_j(t)dsdr \end{aligned}$$

Mit BARAS, BENSOUSSAN, JAMES in [BBJ87] gilt damit für $0 < M \in \mathbb{R}$:

$$\int_0^1 \int_0^1 r \frac{\partial^2 \bar{f}}{\partial x_i \partial x_j}(\hat{x}(t) + rse(t), u(t))e_i(t)e_j(t)dsdr \leq M|e(t)|^2$$

Alle Abschätzungen zusammengefasst und an entsprechender Stelle eingefügt liefert mit $0 < M' \in \mathbb{R}$:

$$\begin{aligned}
 \frac{d}{dt}e^T(t)Q(t)e(t) &= -e^T(t)(\bar{C}^T\bar{C} + Q(t)GG^TQ(t))e(t) \\
 &\quad + 2e^T(t)Q(t)\left(\bar{f}(x(t), u(t)) - \bar{f}(\hat{x}(t), u(t)) - \bar{A}(t)e(t)\right) \\
 &\stackrel{(*), (**)}{\leq} -\frac{m_8}{m_1^2}|e(t)|^2 + 2e^T(t)Q(t)M|e(t)|^2 \\
 &\stackrel{Q(t) \leq m_5 I}{\leq} -\frac{m_8}{m_1^2}|e(t)|^2 + m_5M'|e(t)|^3
 \end{aligned}$$

Da nun $\frac{1}{m_5}Q(t) \leq I_{n \times n}$ und $m_1P(t)^{-1} \stackrel{P(t) \leq m_1 I_{n \times n}}{\geq} P(t)P(t)^{-1} = I_{n \times n}$, gilt weiter

$$\begin{aligned}
 -\frac{m_8}{m_1^2}|e(t)|^2 + m_5M'|e(t)|^3 &\leq -\frac{m_8}{m_1^2m_5}e^T(t)Q(t)e(t) + m_5M'|e(t)|^3 \left(m_1Q(t)\right)^{\frac{3}{2}} \\
 &= -\frac{m_8}{m_1^2m_5}e^T(t)Q(t)e(t) + m_5M' \left(m_1e^T(t)Q(t)e(t)\right)^{\frac{3}{2}} \\
 &= e^T(t)Q(t)e(t) \left(-\frac{m_8}{m_1^2m_5} \right. \\
 &\quad \left. + m_5M'm_1^{\frac{3}{2}} \underbrace{\left(e^T(t)Q(t)e(t)\right)^{\frac{1}{2}}}_{(***)} \right).
 \end{aligned}$$

In dieser Form ist es nicht möglich zu entscheiden, ob $\frac{d}{dt}e^T(t)Q(t)e(t) \leq 0$ ist. Daher müssen wir eine Forderung an (***) stellen.

Zunächst halten wir jedoch fest, dass für

$$\left(e^T(t)Q(t)e(t)\right)^{\frac{1}{2}} < \frac{m_8}{2m_1^{\frac{7}{2}}m_5^2M'}$$

schließlich gilt

$$\frac{d}{dt}e^T(t)Q(t)e(t) \leq -\frac{m_8}{2m_1^2m_5}e^T(t)Q(t)e(t) < 0. \quad \checkmark$$

Damit haben wir eine Ljapunov-Funktion des Schätzfehlers gefunden.

Wegen der Wahl von m_1 für die obere Schranke von $P(t)$ liegt sogar exponentielle Konvergenz vor. Wir erhalten

$$e^T(t)Q(t)e(t) \rightarrow 0 \text{ für } t \rightarrow \infty.$$

Dies liefert Stabilität des Schätzfehlers $e(t)$ im Equilibrium $x = 0$.
Zusammengefasst gilt schlussendlich unter der Bedingung

$$(e^T(0)Q(0)e(0))^{\frac{1}{2}} < \frac{m_8}{2m_1^{\frac{7}{2}}m_5^2M'}$$

lokale Konvergenz des EKBF.

□

Kapitel 5

Spezielles Anwendungsproblem hinsichtlich der Fahrdynamik eines Kraftfahrzeuges

Die in diesem Kapitel verwendete Literatur entstammt vor allem einem internen technischen Bericht [Kal09] der DAIMLER AG. Hieraus entnommen wurden insbesondere die Bewegungsgleichungen, die Festlegung der Beobachterzustände sowie die zu Grunde liegende Differentialgleichung, die den Zustand eines Fahrzeugs mit Hilfe von Lagewinkeln und Geschwindigkeiten beschreibt. Für weitere Informationen zur Eigenbewegung eines Fahrzeugs und seinen Kenngrößen sei an dieser Stelle auf den technischen Bericht [RKHO07] der ROBERT BOSCH GMBH hingewiesen.

In Eigenarbeit wird im zweiten und dritten Teil dieses Kapitels die Konvergenz dieses 5-dimensionalen und auch eines sich daraus ergebenden 2-dimensionalen Systems in Anlehnung an den Beweis in Kapitel 4.2 gezeigt. Folglich liegen hier die Arbeiten [GHO92], [HF03] sowie [Kre03] zu Grunde.

Schließlich wird eine konkrete untere und obere Schranke für die Schätzfehlerkovarianzmatrix des 2-dimensionalen Systems erstmalig in dieser Arbeit angegeben, wobei eine Orientierung an dem hierfür konstruktiven Beweis aus Kapitel 4.2 erfolgt.

In vielen Bereichen, vor allem jedoch innerhalb der Fahrdynamik hinsichtlich von Fahrerassistenz- und Fahrsicherheitssystemen, ist das Wissen um die Eigenbewegung eines Kraftfahrzeuges von besonderem Interesse. Dies rührt daher, dass sich Rückschlüsse aus der Eigenbewegung auf den Fahrzustand schließen lassen.

Für die Bestimmung des Fahrzustands eines Autos, ist die Eigenbewegung erforderlich,

die sich wiederum durch Zustandsgrößen beschreiben lässt.

In unserem Fall sind dies

- die Geschwindigkeiten v_x in Längs-, v_y in Quer- und v_z in Vertikalrichtung, sowie
- die Lagewinkel in der Darstellung als Eulerwinkel: der Gierwinkel Ψ , der Nickwinkel Θ und der Wankwinkel Φ .

Es ist nicht möglich, diese Zustandsgrößen durch Messung zu erhalten. Daher werden messbare Größen herangezogen, um aus diesen schließlich die Zustandsgrößen und daraus dann die Fahrzeugeigenbewegung zu bestimmen. Als Fahrzustandsbeobachter wird das System bezeichnet, welches diese Berechnung vornimmt.

Die entsprechenden Messwerte werden mittels Sensoren, Kamera oder Radar ermittelt.¹¹ Mittels Sensoren werden die Wankrate ω_x , die Nickrate ω_y sowie die Gierrate ω_z , die die Rotationsgeschwindigkeit um die jeweilige körperfeste Achse angeben, erfasst.

Augenscheinlich hängen die unbekanntes Lagewinkel mit den bekannten Drehraten sowie die gesuchten Geschwindigkeiten mit den gemessenen Beschleunigungen zusammen. Aber wie?

Die folgenden Bewegungsgleichungen beschreiben die jeweiligen Beziehungen:

$$\dot{\Theta} = \omega_y \cdot \cos \Phi - \omega_z \cdot \sin \Phi$$

$$\dot{\Phi} = \omega_x + \omega_y \cdot \sin \Phi \cdot \tan \Theta + \omega_z \cdot \cos \Phi \cdot \tan \Theta$$

$$\dot{\Psi} = \omega_y \cdot \frac{\sin \Phi}{\cos \Theta} + \omega_z \cdot \frac{\cos \Phi}{\cos \Theta}$$

und

$$\dot{v}_x = \omega_z \cdot v_y - \omega_y \cdot v_z + a_x + g \cdot \sin \Theta$$

$$\dot{v}_y = -\omega_z \cdot v_x + \omega_x \cdot v_z + a_y - g \cdot \sin \Phi \cdot \cos \Theta$$

$$\dot{v}_z = \omega_y \cdot v_x - \omega_x \cdot v_y + a_z - g \cdot \cos \Phi \cdot \cos \Theta$$

Wir legen nun die Zustandsgleichungen fest und definieren:

$$x_1 = \sin \Theta$$

$$x_2 = \sin \Phi \cdot \cos \Theta$$

$$x_3 = v_x$$

$$x_4 = v_y$$

$$x_5 = v_z$$

¹¹In der weiter unten aufgeführten Anwendung werden lediglich Fahrzeugsensoren benötigt.

Seien $0 < \alpha < 1$, $0 < \beta < 1$.

Die folgenden Betrachtungen dieses Systems schränken wir auf jenes Gebiet ein, welches durch $\{x \in \mathbb{R}^5 : \alpha \leq x_1^2 + x_2^2 \leq \beta\}$, beschrieben wird.

Auf unserem Breitengrad liegt die Erdbeschleunigung bei $9.81m/s^2$.

Ableiten der Zustände unter Beachtung der Bewegungsgleichungen, der Erdbeschleunigung $g = 9.81 \frac{m}{s^2}$ sowie der Relation $\cos \Theta \cdot \cos \Phi = \sqrt{1 - x_1^2 - x_2^2}$ ergibt das nichtlineare Differentialgleichungssystem

$$\begin{aligned}
 \dot{x}_1 &= \omega_y \cdot \sqrt{1 - x_1^2 - x_2^2} - \omega_z \cdot x_2 \\
 \dot{x}_2 &= \omega_x \cdot \sqrt{1 - x_1^2 - x_2^2} + \omega_z \cdot x_1 \\
 \dot{x}_3 &= \omega_z \cdot x_4 - \omega_y \cdot x_5 + g \cdot x_1 + a_x \\
 \dot{x}_4 &= -\omega_z \cdot x_3 + \omega_x \cdot x_5 - g \cdot x_2 + a_y \\
 \dot{x}_5 &= \omega_y \cdot x_3 - \omega_x \cdot x_4 - g \cdot \sqrt{1 - x_1^2 - x_2^2} + a_z,
 \end{aligned} \tag{5.1}$$

welches Ausgangspunkt unserer Betrachtungen im weiteren Verlauf sein wird. In verallgemeinerter Schreibweise liegt also eine Zustandsgleichung der Form

$$f : \mathbb{R}^5 \rightarrow \mathbb{R}^5, \quad \dot{x} = f(x, u)$$

vor, wobei $x^T = [\sin \Theta, \sin \Phi \cos \Theta, v_x, v_y, v_z]$, $u^T = [\omega_x, \omega_y, \omega_z, a_x, a_y, a_z]$ und f ist nichtlinear, vor.

Weiterhin benötigen wir für das EKBF eine Messgleichung, die im Allgemeinen die Form

$$h : \mathbb{R}^5 \rightarrow \mathbb{R}^3, \quad y = h(x, u),$$

im Besonderen jedoch die Form

$$y = C \cdot x, \text{ mit } C \in \mathbb{R}^{p \times n},$$

$$C = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

hat. Sie ergibt sich dadurch, dass die Geschwindigkeiten gemessen werden können.

Für eine kürzere Schreibweise bezeichnen wir die Nichtlinearitäten mit $\rho := \sqrt{(1 - x_1^2 - x_2^2)}$.

5.1. Beweis der Konvergenz des EKBF für ein spezielles dynamisches System

Es liegt nun also die Situation vor, dass obiges System (5.1) gegeben ist und wir darauf das EKBF anwenden.

Wir haben bereits in 4.2 gesehen, welche Voraussetzungen ein nichtlineares System, wie auch das vorliegende, erfüllen muss, sodass das darauf angewandte EKBF konvergiert. Eine entscheidende Eigenschaft, die vom Differentialgleichungssystem gewährleistet werden muss, ist die uniforme Beobachtbarkeit.

Um also Konvergenz des EKBF für dieses spezielle System nachzuweisen, werden wir im Folgenden das gegebene System (5.1) unter anderem auf uniforme Beobachtbarkeit überprüfen. Hierbei ziehen wir die entsprechende Definition aus Kapitel 2.1.3 heran und weisen für unsere Zwecke nur die Rückrichtung der Äquivalenz nach.

Bemerkung

Es ist bereits bekannt, dass ein System uniform beobachtbar ist, wenn es unabhängig von den Systemeingängen u gewisse Eigenschaften erfüllt. Insbesondere soll das EKBF für beliebige Eingänge konvergieren.

Angenommen, es liegen keine Systemeingänge u vor, das heißt $u = 0$. Für diesen trivialen Fall ist bei Überprüfung der Unabhängigkeit der Eingänge mittels der Definition aus Kapitel 2.1.3 nichts zu zeigen, da diese von vornherein gegeben ist.¹²

Daher gehen wir im Folgenden davon aus, dass $u \neq 0$.

Wir überprüfen nun unser System auf jene Eigenschaften, die für den Beweis der Konvergenz des EKBF vorausgesetzt werden:

i. Überprüfung auf uniforme Beobachtbarkeit.

Der gegebene Ausgangsvektor $h(x)$ ist dreidimensional $\implies k = \{1, 2, 3\}$.

Damit ist

- $k = 1$:

Bestimme \mathcal{F}_1^u :

$$\begin{aligned}\mathcal{F}_1^u &= \text{span} \left\{ \frac{d}{dx}h_1, \frac{d}{dx}h_2, \frac{d}{dx}h_3 \right\} \\ &= \text{span} \left\{ (0 \ 0 \ 1 \ 0 \ 0), (0 \ 0 \ 0 \ 1 \ 0), (0 \ 0 \ 0 \ 0 \ 1) \right\}\end{aligned}$$

¹²Eine explizite Überprüfung auf Erfüllung der Eigenschaften aus der Definition bestätigt dies.

$\implies \mathcal{F}_1^u$ ist unabhängig von $u \implies \mathcal{F}_1^u = \mathcal{F}_1^{u'}$.

Bestimme $\dim \mathcal{F}_1^u$:

$$\dim \mathcal{F}_1^u = \dim \operatorname{span} \{(0 \ 0 \ 1 \ 0 \ 0), (0 \ 0 \ 0 \ 1 \ 0), (0 \ 0 \ 0 \ 0 \ 1)\} = 3.$$

Aufgrund der Unabhängigkeit des Aufspans von u sowie von x bleibt die Dimension von $\mathcal{F}_1^u \forall x, \forall u$ konstant.

- $k = 2$:

Bestimme \mathcal{F}_2^u :

$$\begin{aligned} \mathcal{F}_2^u &= \mathcal{F}_1^u + \operatorname{span} \left\{ \frac{d}{dx} L_f^1(h_1), \frac{d}{dx} L_f^1(h_2), \frac{d}{dx} L_f^1(h_3) \right\} \\ &= \mathcal{F}_1^u + \operatorname{span} \left\{ \frac{d}{dx} ((0 \ 0 \ 1 \ 0 \ 0)f(x, u)), \frac{d}{dx} ((0 \ 0 \ 0 \ 1 \ 0)f(x, u)), \right. \\ &\quad \left. \frac{d}{dx} ((0 \ 0 \ 0 \ 0 \ 1)f(x, u)) \right\} \\ &= \mathcal{F}_1^u + \operatorname{span} \left\{ (g \ 0 \ 0 \ \omega_z \ -\omega_y), (0 \ -g \ -\omega_z \ 0 \ \omega_x), \right. \\ &\quad \left. \begin{pmatrix} gx_1 & gx_2 & \omega_y & -\omega_x & 0 \\ \rho & \rho & & & \end{pmatrix} \right\} \end{aligned}$$

Wir sehen, dass die Vektoren, die \mathcal{F}_2^u aufspannen, voneinander linear unabhängig sind. Daher wird derselbe Raum \mathcal{F}_2 durch beliebige u, u' und $\forall x \in \mathbb{R}^n$ erzeugt, das heißt $\mathcal{F}_2^u = \mathcal{F}_2^{u'}$.

Bestimme $\dim \mathcal{F}_2^u$:

Mit dem vorliegenden Aufspann von \mathcal{F}_2^u ist die Dimension von \mathcal{F}_2^u wegen der Unabhängigkeit aller vorliegenden Vektoren leicht zu sehen. Da der Aufspann also unabhängig von u sowie von x ist, bleibt seine Dimension $\forall x, \forall u$ konstant.

Es ist $\dim \mathcal{F}_2^u = 3 + 3 = 6$.

- $k = 3$. Bestimme \mathcal{F}_3^u :

$$\begin{aligned}
 \mathcal{F}_3^u &= \mathcal{F}_2^u + \text{span} \left\{ \frac{d}{dx} L_f^2(h_1), \frac{d}{dx} L_f^2(h_2), \frac{d}{dx} L_f^2(h_3) \right\} \\
 &= \mathcal{F}_2^u + \text{span} \left\{ \frac{d}{dx} ((g \ 0 \ 0 \ \omega_z \ -\omega_y) f(x, u)), \right. \\
 &\quad \frac{d}{dx} ((0 \ -g \ \omega_z \ 0 \ \omega_x) f(x, u)), \\
 &\quad \left. \frac{d}{dx} \left(\left(\frac{gx_1}{\rho} \ \frac{gx_2}{\rho} \ \omega_y \ -\omega_x \ 0 \right) f(x, u) \right) \right\} \\
 &= \mathcal{F}_2^u + \text{span} \left\{ \left(\frac{-2g\omega_y x_1}{\rho}, \frac{-2g\omega_y x_2}{\rho} - 2g\omega_z, -\omega_z^2 - \omega_y^2, \right. \right. \\
 &\quad \left. \left. \omega_x \omega_y, \omega_x \omega_z \right), \right. \\
 &\quad \left(\frac{2g\omega_x x_1}{\rho}, \frac{2g\omega_x x_2}{\rho}, \omega_x \omega_y, \omega_z^2 - \omega_x^2, \omega_y \omega_z \right), \\
 &\quad \left. \left(2g\omega_y, 2g\omega_x, \omega_x \omega_z, \omega_y \omega_z, -\omega_y^2 - \omega_x^2 \right) \right\}
 \end{aligned}$$

Auch \mathcal{F}_3^u wird also von linear unabhängigen Vektoren aufgespannt. Dabei bleibt \mathcal{F}_3^u der gleiche Raum, unabhängig von u, x . Das bedeutet:

$$\mathcal{F}_3^u = \mathcal{F}_3^{u'} \quad \forall x \in \mathbb{R}^n, u, u' \in U.$$

Bestimme $\dim \mathcal{F}_3^u$:

Auf Grund der Unabhängigkeit der Vektoren, die \mathcal{F}_3^u erzeugen, ergibt sich für die Dimension:

$$\dim \mathcal{F}_3^u = 6 + 3 = 9. \text{ Sie bleibt konstant } \forall x \in \mathbb{R}^n, \forall u \in U.$$

Schließlich sehen wir, dass das gegebene Differentialgleichungssystem (5.1) die geforderten Eigenschaften erfüllt und folglich ist es uniform beobachtbar.

- ii. Überprüfung auf Lipschitz-Stetigkeit.

Aus den Voraussetzungen aus 4.2 ist abzulesen, wie diese zu überprüfen ist.

Sei

$$\|f(x, u) - f(\eta, u)\|_1^2 = \sum_{i=1}^5 |\dot{x}_i - \dot{\eta}_i|^2.$$

Es ist leicht zu sehen, dass $\forall i = 1, \dots, 5$ gilt:

$$|\dot{x}_i - \dot{\eta}_i| \leq L \|x - \eta\|_1^2$$

So etwa für $i = 1$:

$$\begin{aligned} & \left| \left(\omega_y \sqrt{1 - x_1^2 - x_2^2} - \omega_z x_2 \right) - \left(\omega_y \sqrt{1 - \eta_1^2 - \eta_2^2} - \omega_z \eta_2 \right) \right| \\ & \leq |\omega_y| \underbrace{\left| \sqrt{1 - x_1^2 - x_2^2} - \sqrt{1 - \eta_1^2 - \eta_2^2} \right|}_{(*)} + |\omega_z| |\eta_2 - x_2| \end{aligned}$$

Weitere Abschätzung des nichtlinearen Terms (*) liefert:

$$\begin{aligned} & \left| \sqrt{1 - x_1^2 - x_2^2} - \sqrt{1 - \eta_1^2 - \eta_2^2} \right| \cdot \left| \sqrt{1 - x_1^2 - x_2^2} + \sqrt{1 - \eta_1^2 - \eta_2^2} \right| \\ & = |(x_1^2 - x_2^2) - (\eta_1^2 + \eta_2^2)| \\ & \leq |(x_1 + x_2) - (\eta_1 + \eta_2)| \cdot |(x_1 + x_2) + (\eta_1 + \eta_2)|, \end{aligned}$$

folglich

$$\begin{aligned} & \left| \sqrt{1 - x_1^2 - x_2^2} - \sqrt{1 - \eta_1^2 - \eta_2^2} \right| \\ & \leq |(x_1 + x_2) - (\eta_1 + \eta_2)| \cdot \underbrace{\frac{|(x_1 + x_2) + (\eta_1 + \eta_2)|}{\left| \sqrt{1 - x_1^2 - x_2^2} + \sqrt{1 - \eta_1^2 - \eta_2^2} \right|}}_{(**)} \end{aligned}$$

Betrachte nun (**). Mit $\alpha \leq |x_1| \leq \beta$, $\alpha \leq |x_2| \leq \beta$ ergibt sich folgende Abschätzung für (**):

$$\frac{|(x_1 + x_2) + (\eta_1 + \eta_2)|}{\left| \sqrt{1 - x_1^2 - x_2^2} + \sqrt{1 - \eta_1^2 - \eta_2^2} \right|} \leq \frac{4\beta}{2\alpha} = \frac{2\beta}{\alpha}$$

Wir haben also eine Lipschitz-Konstante $\frac{2\beta}{\alpha}$ gefunden.

Definiere nun $|\omega_y| =: c_1$, $|\omega_z| =: c_2$, $\frac{2\beta}{\alpha} =: c_3$.

Dann gilt weiter

$$\begin{aligned} & |\omega_y| \left| \sqrt{1 - x_1^2 - x_2^2} - \sqrt{1 - \eta_1^2 - \eta_2^2} \right| + |\omega_z| |\eta_2 - x_2| \\ & \leq \underbrace{c_1 \cdot c_3}_{=: L_1} |(\eta_1 + \eta_2) - (x_1 + x_2)| + c_2 |\eta_2 - x_2| \\ & \leq L_1 |(\eta_1 + \eta_2) - (x_1 + x_2)| + c_2 (|\eta_2 - x_2| + |\eta_1 - x_1|) \\ & \leq L_1 (|x_1 - \eta_1| + |x_2 - \eta_2|) + c_2 (|\eta_2 - x_2| + |\eta_1 - x_1|) \\ & = L_2 \sum_{i=1}^2 |x_i - \eta_i| \leq L_2 \sum_{i=1}^5 |x_i - \eta_i| \\ & = L_2 \|x - \eta\|_1, \end{aligned}$$

folglich

$$\begin{aligned} & \left| \left(\omega_y \sqrt{1 - x_1^2 - x_2^2} - \omega_z x_2 \right) - \left(\omega_y \sqrt{1 - \eta_1^2 - \eta_2^2} - \omega_z \eta_2 \right) \right| = |\dot{x}_1 - \dot{\eta}_1| \\ & \leq L_2 \|x - \eta\|_1 \\ & \implies |\dot{x}_i - \dot{\eta}_i|^2 \leq L_2^2 \|x - \eta\|_1^2. \end{aligned}$$

Durch ähnliches Vorgehen ist diese Abschätzung für $i = 2, 3, 4, 5$ leicht nachzurechnen. Damit erhalten wir

$$\|f(x, u) - f(\eta, u)\|_1^2 \leq L^2 \|x - \eta\|_1^2$$

und schließlich

$$\|f(x, u) - f(\eta, u)\|_1 \leq L \|x - \eta\|_1$$

Es ist also die Lipschitz-Stetigkeit des Systems (5.1) $\forall x \in \mathbb{R}^5 : \alpha \leq x_1^2 + x_2^2 \leq \beta$ nachgewiesen.

iii. Überprüfung auf Beschränktheit der 2. Ableitung.

Da es sich beim gegebenen Differentialgleichungssystem (5.1), um ein solches handelt, das in der Praxis Anwendung findet, wollen wir also den Fokus auf die Praxisorientierung legen. Daher werden unsere Betrachtungen auf gewisse Bedingungen eingeschränkt.

Es handelt sich um ein mechanisches System, sodass die vorliegenden Größen aus physikalischen Gründen beschränkt sind.¹³ Insbesondere nehmen wir an, dass für die Drehraten gilt: $\omega_i \in \mathbb{R}$, $i = x, y, z$, also, dass sie insbesondere beschränkt sind.¹⁴

Weiterhin gehen wir auf Grund der Wurzelterme im Nenner einzelner Einträge auch hier davon aus, dass

$$x_1^2 + x_2^2 \leq \delta, \quad \delta \in (0, 1),$$

um Singularitäten zu vermeiden.

Unter diesen Bedingungen bestimmen wir nun zunächst die erste Ableitung von (5.1):

¹³Beispielsweise ist in der Praxis die Längsgeschwindigkeit eines Fahrzeuges beschränkt.

¹⁴Typischerweise gilt für die Fahrdynamik $\omega_i < 300^\circ/\text{s}$, $i = x, y, z$.

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{-\omega_y x_1}{\rho} & \frac{-\omega_y x_2}{\rho} - \omega_z & 0 & 0 & 0 \\ \frac{-\omega_x x_1}{\rho} + \omega_z & \frac{-\omega_x x_2}{\rho} & 0 & 0 & 0 \\ g & 0 & 0 & \omega_z & -\omega_y \\ 0 & -g & -\omega_z & 0 & \omega_x \\ \frac{g x_1}{\rho} & \frac{g x_2}{\rho} & \omega_y & -\omega_x & 0 \end{bmatrix}$$

Wir sehen also, dass $\frac{\partial f}{\partial x}$ existiert und obige Gestalt hat.

Die Form von $\frac{\partial^2 f}{\partial x_i \partial x_j}$, $i, j \in 1, \dots, 5$ ist leicht nachzurechnen. Es ergibt sich:

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \begin{bmatrix} \frac{-\omega_y \rho - \omega_y x_1^2 \rho^{-1}}{\rho^2} & \frac{-\omega_y x_1 x_2}{\rho^3} & 0 & 0 & 0 \\ \frac{-\omega_x x_1 x_2}{\rho^3} & \frac{-\omega_x \rho - \omega_x x_2^2 \rho^{-1}}{\rho^2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Um nun zu sehen, dass $A' := \left\| \frac{\partial^2 f}{\partial x_i \partial x_j} \right\|$ beschränkt ist, bedienen wir uns der Norm

$$\|A'\| = \left(\sum_{i=1}^5 \sum_{j=1}^5 |a'_{ij}|^2 \right)^{1/2} :$$

$$\begin{aligned} \|A'\|^2 &= \sum_{i=1}^5 \sum_{j=1}^5 |a'_{ij}|^2 \\ &= \left| \frac{\omega_y \rho + \omega_y x_1^2 \rho^{-1}}{\rho^2} \right| + \left| \frac{\omega_y x_1 x_2}{\rho^3} \right| + \left| \frac{\omega_x x_1 x_2}{\rho^3} \right| + \left| \frac{\omega_x \rho + \omega_x x_2^2 \rho^{-1}}{\rho^2} \right| \\ &\leq \left| \frac{1}{\rho} \right| \cdot |\omega_y| + \left| \frac{1}{\rho^3} \right| \cdot |\omega_y x_1^2| + \left| \frac{1}{\rho^3} \right| \cdot |\omega_x x_1 x_2| + \left| \frac{1}{\rho} \right| \cdot |\omega_x| + \left| \frac{1}{\rho^3} \right| \cdot |\omega_x x_2^2| \end{aligned}$$

Wegen $x_1^2 + x_2^2 \leq \delta$ und da die Drehraten $\omega_x, \omega_y, \omega_z \in \mathbb{R}$ sind, gilt also:

$$\left| \frac{1}{\rho} \right| \leq c_1, \quad |\omega_y| = c_2, \quad |\omega_y x_1^2| \leq c_3, \quad |\omega_x x_1 x_2| \leq c_4, \quad |\omega_x| = c_5, \quad |\omega_x x_2^2| \leq c_6$$

Insbesondere liegt bei diesen Termen Beschränktheit vor und damit ist

$$\|A'\|^2 \leq m' \implies \left\| \frac{\partial^2 f}{\partial x_i \partial x_j} \right\| \leq m, \quad \text{für } 0 < m, m' \in \mathbb{R}.$$

Wir halten also fest, dass die zweite Ableitung von $f(x, u)$ beschränkt ist.

iv. Überprüfung der Störmatrix G im EKBF auf Invertierbarkeit.

Im vorliegenden Anwendungsbeispiel gehen wir davon aus, dass $G = I_{5 \times 5}$. Insofern ist

die Invertierbarkeit von G gegeben.

v. Überprüfung von $P(0)$ auf positive Definitheit.

Diese Größe ist frei bestimmbar. Daher wählen wir $P(0)$ so, dass die Schätzfehlerkovarianzmatrix zum Startzeitpunkt positiv definit ist. Damit ist auch diese Eigenschaft gewährleistet.

Ebenso verhält es sich mit dem Schätzfehler zu diesem Zeitpunkt. Wir bestimmen $P(0)$ nicht nur so, dass sie positiv definit ist, sondern auch so, dass der Schätzfehler $|e(t_0)| \leq \epsilon$ erfüllt, sodass die Konvergenz des EKBF für gegebenens System gewährleistet wird.

Es ist also gezeigt, dass das dynamische System (5.1) die notwendigen Bedingungen erfüllt, sodass mittels des in Kapitel 4.2 aufgeführten Beweises nachgewiesen ist, dass das EKBF für unser gegebenes System konvergiert.

5.2. Ermittlung des Konvergenzradius zum Startzeitpunkt für ein konkretes 2-dimensionales System

In diesem Abschnitt beschäftigen wir uns mit einem System, welches aus den ersten beiden Zustandsgleichungen aus (5.1) zusammengesetzt ist:

$$\begin{aligned} \dot{x}_1 &= \omega_y \cdot \sqrt{1 - x_1^2 - x_2^2} - \omega_z \cdot x_2 \\ \dot{x}_2 &= \omega_x \cdot \sqrt{1 - x_1^2 - x_2^2} + \omega_z \cdot x_1 \end{aligned} \quad (5.2)$$

Wir haben also

$$g : \mathbb{R}^2 \rightarrow \mathbb{R}^2, \quad \dot{x} = g(x, u),$$

mit $x^T = [\sin \Theta, \sin \Phi \cos \Theta]$, $u^T = [\omega_x, \omega_y, \omega_z]$ und g ist nichtlinear, vorliegen.

Die Messgleichung

$$k : \mathbb{R}^2 \rightarrow \mathbb{R}^2, \quad y = k(x, u),$$

hat die Form

$$y = D \cdot x, \quad \text{mit } D = I_{2 \times 2}.$$

Auch hier bezeichnen wir $\rho := \sqrt{1 - x_1^2 - x_2^2}$.

Seien $0 < \alpha, \beta < 1$.

Die folgenden Betrachtungen beschränken sich auf den „Ring“, der durch $\{x \in \mathbb{R}^2 : \alpha \leq \|x\| \leq \beta\}$ beschrieben wird.

5.2.1. Beweis der Konvergenz des EKBF bei Anwendung auf gegebenes 2-dimensionales System

Um den Konvergenzradius zum Startzeitpunkt bestimmen zu können, das heißt, um die Schätzfehlerkovarianzmatrix $P(0)$ auf ein bestimmtes Gebiet einschränken zu können, muss zunächst gewährleistet sein, dass das EKF für das System (5.2) konvergiert. Hierfür gehen wir wie schon in Kapitel 5.1 vor, indem wir die erforderlichen Eigenschaften von $g(x, u)$ überprüfen:

i. Überprüfung auf uniforme Beobachtbarkeit.

$k(x)$ ist zweidimensional $\implies k = \{1, 2\}$.

Damit ist

- $k = 1$:

Bestimme \mathcal{G}_1^u :

$$\mathcal{G}_1^u = \text{span} \left\{ \frac{d}{dx} k_1, \frac{d}{dx} k_2 \right\} = \text{span} \{(1 \ 0), (0 \ 1)\}$$

$\implies \mathcal{G}_1^u$ ist unabhängig von $u \implies \mathcal{G}_1^u = \mathcal{G}_1^{u'}$.

Bestimme $\dim \mathcal{G}_1^u$:

$$\dim \mathcal{G}_1^u = \dim \text{span} \{(1 \ 0), (0 \ 1)\} = 2.$$

Aufgrund der Unabhängigkeit des Aufspans von u sowie von x bleibt die Dimension von $\mathcal{G}_1^u \forall x, \forall u$ konstant.

- $k = 2$:

Bestimme \mathcal{G}_2^u :

$$\begin{aligned} \mathcal{F}_2^u &= \mathcal{F}_1^u + \text{span} \left\{ \frac{d}{dx} L_g^1(k_1), \frac{d}{dx} L_g^1(k_2) \right\} \\ &= \mathcal{F}_1^u + \text{span} \left\{ \frac{d}{dx} ((1 \ 0)g(x, u)), \frac{d}{dx} ((0 \ 1)g(x, u)) \right\} \\ &= \mathcal{F}_1^u + \text{span} \left\{ \left(-\frac{\omega_y x_1}{\rho} \quad -\frac{\omega_y x_2}{\rho} - \omega_z \right), \left(-\frac{\omega_x x_1}{\rho} + \omega_z \quad -\frac{\omega_x x_2}{\rho} \right) \right\} \end{aligned}$$

Wir sehen, dass die Vektoren, die \mathcal{G}_2^u aufspannen, voneinander linear unabhängig sind. Daher wird derselbe Raum \mathcal{G}_2 durch beliebige u, u' und $\forall x \in \mathbb{R}^n$ erzeugt

$$\implies \mathcal{G}_2^u = \mathcal{G}_2^{u'}.$$

Bestimme $\dim \mathcal{G}_2^u$:

Mit dem vorliegenden Aufspann von \mathcal{G}_2^u ist die Dimension von \mathcal{G}_2^u wegen der Unabhängigkeit aller vorliegenden Vektoren leicht zu sehen. Die Dimension des Aufspans bleibt $\forall x, \forall u$ konstant.

$$\text{Es ist } \dim \mathcal{G}_2^u = 2 + 2 = 4.$$

Schließlich folgt die uniforme Beobachtbarkeit von (5.2).

ii. Überprüfung auf Lipschitz-Stetigkeit.

Ein Vergleich des 5-dimensionalen Systems (5.1) mit dem 2-dimensionalen System (5.2) zeigt, dass die ersten beiden Zeilen identisch sind.

Auch hier definieren wir

$$\|g(x, u) - g(\eta, u)\|_1^2 = \sum_{i=1}^2 |\dot{x}_i - \dot{\eta}_i|^2.$$

Die Lipschitz-Stetigkeit für $i = 1$ wurde bereits in 5.1 nachgewiesen. Durch leichte Rechnung lässt sie sich auch für $i = 2$ nachweisen.

Dann ist sofort ersichtlich, dass $\forall i = 1, 2$ gilt:

$$|\dot{x}_i - \dot{\eta}_i| \leq L \|x - \eta\|_1^2$$

Schließlich erhalten wir also die Lipschitz-Stetigkeit für das System (5.2) mit

$$\|g(x, u) - g(\eta, u)\|_1^2 \leq L \|x - \eta\|_1^2.$$

iii. Überprüfung auf Beschränktheit der zweiten Ableitung. Wie schon in 5.1, gehen wir aus denselben Gründen von folgenden Einschränkungen auf unser System aus:

$$|\omega_i| < \gamma, \quad i = x, y, z, \quad \gamma \in \mathbb{R}, \quad x_1^2 + x_2^2 \leq \delta, \quad \delta \in (0, 1).$$

Damit ist

$$\frac{\partial g}{\partial x} = \begin{bmatrix} -\frac{\omega_y x_1}{\rho} & -\frac{\omega_y x_2}{\rho} - \omega_z \\ -\frac{\omega_x x_1}{\rho} + \omega_z & -\frac{\omega_x x_2}{\rho} \end{bmatrix}$$

und

$$\frac{\partial^2 g}{\partial x_i \partial x_j} = \begin{bmatrix} \frac{-\omega_y \rho - \omega_y x_1^2 \rho^{-1}}{\rho^2} & \frac{-\omega_y x_1 x_2}{\rho^3} \\ \frac{-\omega_x x_1 x_2}{\rho^3} & \frac{-\omega_x \rho - \omega_x x_2^2 \rho^{-1}}{\rho^2} \end{bmatrix}.$$

Wir sehen also, dass $\frac{\partial^2 g}{\partial x_i \partial x_j}$, $i, j = 1, 2$ existiert.

Bei genauem Vergleich von $\frac{\partial f^2}{\partial x_i \partial x_j}$ mit $\frac{\partial g^2}{\partial x_i \partial x_j}$ sehen wir, dass die Einträge für $i, j = 1, 2$ die gleichen sind.

Sei $B' := \left\| \frac{\partial g^2}{\partial x_i \partial x_j} \right\|$.

Insbesondere gilt dann für die Normen:

$$\left\| \frac{\partial g^2}{\partial x_i \partial x_j} \right\| = \left(\sum_{i=1}^2 \sum_{j=1}^2 |b'_{ij}|^2 \right)^{1/2} = \left(\sum_{i=1}^5 \sum_{j=1}^5 |a'_{ij}|^2 \right)^{1/2} \leq M, \quad 0 < M \in \mathbb{R}$$

Damit ist die zweite Ableitung von $g(x, u)$ beschränkt.

iv. Überprüfung der Störmatrix G im EKBF auf Invertierbarkeit.

Wenn wir das EKBF auf das System (5.2) anwenden, so legen wir $G = I_{2 \times 2}$ fest. Insofern ist die Invertierbarkeit von G gegeben.

v. Überprüfung von $P(0)$ auf positive Definitheit.

Diese Größe ist wie schon im 5-dimensionalen Fall frei bestimmbar. Daher wählen wir $P(0)$ so, dass diese Matrix positiv definit ist. Damit ist auch diese Eigenschaft gewährleistet.

Weiterhin bestimmen wir $P(0)$ auch so, dass der Schätzfehler $|e(t_0)| \leq \epsilon$ erfüllt, sodass die Konvergenz des EKBF für gegebenens System gewährleistet wird.

Das System (5.2) erfüllt die erforderlichen Voraussetzungen. Dies impliziert die Konvergenz des EKF bei Anwendung auf dieses dynamische System.

Da wir zeigen konnten, dass das EKF konvergiert, ist es nun möglich einen Einzugsbereich zu bestimmen, der besagt, wie weit eine Schätzung des Anfangszustands $\hat{x}(0)$ vom tatsächlichen Anfangszustand $x(0)$ weg sein darf, um überhaupt noch Konvergenz, also eine Lösung des System (5.2) zu erhalten.

Wir möchten uns im Folgenden damit befassen, dieses Gebiet näher zu bestimmen.

5.2.2. In welchem Gebiet liegt $P(t)$ zur Konvergenz des 2-dimensionalen Systems?

In Anlehnung an den Beweis der Konvergenz des EKBF in Kapitel 4.2 suchen wir eine untere und eine obere Schranke für die Schätzfehlerkovarianzmatrix $P(t)$. Dabei bedienen wir uns derselben Notation.

Nachdem nachgewiesen wurde, dass das EKBF für das System (5.2) konvergiert, ist bekannt, dass eine Transformation dieses Systems in die Form (4.6) existiert.

Wir widmen uns nun der Aufgabe, eine solche zu finden.

Zunächst benötigen wir entsprechende Variablen x_{ij} mit $i = 1, \dots, p$ und $j = 1, \dots, l_i$, wobei

$$1 \leq l_1 \leq \dots \leq l_p \quad \text{und} \quad \sum_{i=1}^p l_i = n.$$

In unserem Fall sind $p = 2$ wegen $y \in \mathbb{R}^2$ und $n = 2$ wegen $x \in \mathbb{R}^2$.

Außerdem ist $\bar{x}_1 = (x_{11}, x_{21})$.

Folglich gilt wegen $\sum_{i=1}^2 l_i = 2$, dass $l_1 = 1, l_2 = 1$ und die Matrizen zur Transformation haben folgende Gestalt:

$$\bar{C} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \bar{A} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \bar{g}(x, u) = \begin{bmatrix} \bar{g}_1(x, u) \\ \bar{g}_2(x, u) \end{bmatrix} = \begin{bmatrix} g_{11}(\bar{x}_1, u) \\ g_{21}(\bar{x}_1, u) \end{bmatrix}, \quad k(u) = \begin{bmatrix} k_1(u) \\ \vdots \\ k_2(u) \end{bmatrix}$$

Da \bar{A} die Nullmatrix ist und wegen $\bar{C} = C$, ist mit der Form

$$\begin{aligned} \dot{x} &= \bar{A}x + \bar{g}(x, u) \\ y &= k(x) = \bar{C}x \end{aligned}$$

des transformierten Systems sofort einzusehen, dass dieses dem gegebenen System (5.2) gleicht. Insofern ist $\bar{g}(x, u) = g(x, u)$.

Bevor wir mit der Bestimmung der unteren und der oberen Schranke von $P(t)$ beginnen, halten wir noch die bereits berechnete partielle Ableitung des Vektors $g(x, u)$ nach den Zuständen fest:

$$A(t) := \frac{\partial g}{\partial x} \begin{bmatrix} -\frac{\omega_y x_1}{\rho} & -\frac{\omega_y x_2}{\rho} - \omega_z \\ -\frac{\omega_x x_1}{\rho} + \omega_z & -\frac{\omega_x x_2}{\rho} \end{bmatrix}$$

Auf Grund der Gleichheit $g(x, u) = \bar{g}(x, u)$ gilt für die Ableitungen ebenfalls Gleichheit:

$$\bar{A}(t) = \frac{\partial \bar{g}}{\partial x} = \frac{\partial g}{\partial x} = A(t).$$

Gesucht ist nun zunächst eine obere Schranke m_1 für $P(t)$.

Sei $GG^T = I_{2 \times 2}$.

Wir betrachten das Minimierungsproblem

$$J(\mu) = \zeta^T(0)P(0)\zeta(0) + \int_0^t \zeta^T(s)\zeta(s) + \mu^T(s)\mu(s)ds,$$

welches unter den Nebenbedingungen

$$\begin{aligned} \dot{\zeta}(s) &= -A^T(s)\zeta(s) - \mu(s)I_{2 \times 2} \\ \zeta(t) &= \zeta \end{aligned}$$

minimiert werden soll. Für das Minimum gilt

$$J(\mu) = \zeta^T P(t) \zeta.$$

Sei nun $\theta > 0$ und sei $S(\theta)$ die Lösung der Matrixgleichung

$$\bar{A}S(\theta) + S(\theta)\bar{A} - \bar{C}^T\bar{C} = -\theta S(\theta). \quad (5.3)$$

Ausnutzen der Kenntnis über \bar{A} , sowie \bar{C} und Umformung liefert

$$I_{2 \times 2} = \theta S(\theta) \implies S_{11}(\theta) = S_{22}(\theta) = \theta^{-1}, \quad S_{12}(\theta) = S_{21}(\theta) = 0.$$

Wir halten fest, dass $S(\theta)$ positiv definit ist.

Weiter ist wegen (5.3) folgende Relation leicht nachzurechnen, da im vorliegenden Fall $i, \rho = 1, 2$ und $j, \sigma = 1$ sind:

$$S_{i,j,\rho,\sigma}(\theta) = \frac{S_{i,j,\rho,\sigma}(1)}{\theta^{j+\sigma-1}}$$

Im weiteren Verlauf betrachten wir nun die Inverse von $S(\theta)$ und definieren sie als $T(\theta) := S^{-1}(\theta)$.

Dann ist $T_{i,j,\rho,\sigma}(\theta) = \theta T_{i,j,\rho,\sigma}(1)$.

Aus (5.3) erfüllt $T(\theta)$ folgende Gleichung:

$$-\bar{A}T(\theta) - T(\theta)\bar{A}^T + T(\theta)\bar{C}^T\bar{C}T(\theta) = \theta T(\theta)$$

Auch an dieser Stelle setzen wir die uns bekannten Matrizen \bar{A} und \bar{C} ein und erhalten folgendes Ergebnis:

$$T(\theta) = \theta I_{2 \times 2} \quad (5.4)$$

Ein besonderes Augenmerk wird auf den Parameter θ gelegt, da er für die obere Schranke für $P(t)$ benötigt und daher später noch genau bestimmt werden wird.

Vorher wenden wir die suboptimale Lösung $\mu = -T(\theta)\zeta$ auf unser Minimierungsproblem an, und erhalten

$$\zeta^T P(t) \zeta \leq \tilde{\zeta}^T(0) P(0) \tilde{\zeta}(0) + \int_0^t \tilde{\zeta}^T(s) (I_{2 \times 2} + T(\theta)^2) \tilde{\zeta}(s) ds$$

mit den Nebenbedingungen

$$\begin{aligned} \dot{\tilde{\zeta}}(s) &= \left(-A^T(s) + T(\theta) \right) \tilde{\zeta}(s) \\ \tilde{\zeta}(t) &= \zeta. \end{aligned}$$

Weiter ist

$$\frac{d}{ds} \tilde{\zeta}^T(s) T(\theta) \tilde{\zeta}(s) = \tilde{\zeta}^T(s) \underbrace{\left(2\theta^2 I_{2 \times 2} - \theta I_{2 \times 2} \tilde{A}(s) + \theta I_{2 \times 2} \tilde{A}^T(s) \right)}_{(*)} \tilde{\zeta}(s).$$

Wir wollen nun den Ausdruck (*) betrachten.

Dazu stellen wir fest, dass für Lipschitz-stetiges $g(x, u)$, $|\tilde{A}(s)| \leq L$ gilt. Folglich wissen wir, dass jeder Eintrag in der Matrix $\theta I_{2 \times 2} \tilde{A}(s)$ von der Größenordnung $\mathcal{O}(\theta)$.

Für (*) bedeutet dies: $\exists \alpha$ beliebig, sodass für ein $\theta(\alpha)$ groß genug die Ungleichung

$$\begin{aligned} 2\theta^2 I_{2 \times 2} - \theta I_{2 \times 2} \tilde{A}(s) + \theta I_{2 \times 2} \tilde{A}^T(s) &\geq \alpha I_{2 \times 2} \\ \iff (2\theta^2 - \alpha) I_{2 \times 2} - \theta \left(\tilde{A}(s) + \tilde{A}^T(s) \right) &\geq 0 \quad (\text{positiv definit}) \end{aligned} \quad (5.5)$$

erfüllt ist. Das Ziel ist das Auffinden einer oberen Schranke $m_1 I_{2 \times 2}$ für $P(t)$, mit $0 < m_1 \in \mathbb{R}$. Aus (4.20) wissen wir bereits, dass sie die allgemeine Form

$$m_1 I_{2 \times 2} = m_4(\theta) \left(m_2 + \frac{m_3(\theta)}{\alpha} \right) I_{2 \times 2}$$

hat, wobei m_2 und $m_3(\theta)$ dadurch charakterisiert sind, dass für diese beiden

$$\begin{aligned} P(0) &\leq m_2 I_{2 \times 2}, \\ (1 + \theta^2) I_{2 \times 2} &\leq m_3(\theta) I_{2 \times 2} \end{aligned}$$

gilt und $0 < m_2 \in \mathbb{R}$, $0 < m_3(\theta) \in \mathbb{R}$, $0 < m_4(\theta) \in \mathbb{R}$.

Um nun m_1 für unseren Fall bestimmen zu können, benötigen wir ein θ sowie ein α . Da α jedoch beliebig ist, verfolgen wir folgende Idee:

Zunächst muss gewährleistet sein, dass die Matrix $(2\theta^2 - \alpha) I_{2 \times 2} - \theta (\tilde{A}(s) + \tilde{A}^T(s))$ positiv definit ist, das heißt, dass die Eigenwerte $\lambda_1, \lambda_2 \geq 0$ sind. Wir fordern also, dass diese Bedingung erfüllt wird. Da θ in dieser Matrix die einzig noch unbekannte Größe ist, fragen wir also, welches θ die positive Definitheit gewährleistet.

Hierbei wählen wir den kleineren der beiden Eigenwerte, denn wenn dieser für ein noch zu bestimmendes θ größer oder gleich Null ist, so ist dies auch der größere Eigenwert.

Mit Erhalt von θ , kennen wir auch $m_3(\theta) = 1 + \theta^2$. Da m_2 frei wählbar ist, definieren wir die obere Schranke von $P(t)$ dann als eine Funktion, die von α abhängt:

$$f(\alpha) := m_4(\theta) \left(m_2 + \frac{m_3(\theta)}{\alpha} \right)$$

Wir werden dann das Minimum dieser Funktion suchen. Dieser Gedanke rührt daher, dass das Minimum die kleinstmögliche obere Schranke m_1 ist, die für jedes beliebige andere

α gültig ist. Das heißt m_1 kann größer werden, aber da wir $P(t)$ durch die kleinste obere Schranke dominieren, ist die Ungleichung $P(t) \leq m_1 I_{2 \times 2}$ in jedem Fall erfüllt.

Beginnen wir also mit der Bestimmung der Eigenwerte. Hierfür ziehen wir die Matrix

$$\begin{bmatrix} 2\theta^2 - \alpha + \frac{\theta}{\rho}(2\omega_y x_1) & \frac{\theta}{\rho}(\omega_y x_2 - \omega_x x_1) \\ \frac{\theta}{\rho}(\omega_y x_2 - \omega_x x_1) & 2\theta^2 - \alpha + \frac{\theta}{\rho}(2\omega_x x_2) \end{bmatrix}$$

heran. Ihre Eigenwerte ergeben sich als

$$\begin{aligned} \lambda_1 &= 2\theta^2 - \alpha + \frac{\theta}{\rho} \left(\omega_y x_1 + \omega_x x_2 - \sqrt{(\omega_x x_2 - \omega_y x_1)^2 + (\omega_y x_2 - \omega_x x_1)^2} \right) \\ \lambda_2 &= 2\theta^2 - \alpha + \frac{\theta}{\rho} \left(\omega_y x_1 + \omega_x x_2 + \sqrt{(\omega_x x_2 - \omega_y x_1)^2 + (\omega_y x_2 - \omega_x x_1)^2} \right). \end{aligned}$$

Offensichtlich ist $\lambda_1 \leq \lambda_2$. Daher ziehen wir λ_1 für die Untersuchung von θ heran. Das heißt, wir werden nun die Frage beantworten, für welches θ , $\lambda_1 \geq 0$ ist. Dafür setzen wir $\lambda_1 = 0$ und lösen nach θ auf. Das Ergebnis ist

$$\begin{aligned} \theta_1 &= \frac{1}{4\rho} \left(-\sqrt{8\alpha\rho^2 + \left(\sqrt{(\omega_x x_1 - \omega_y x_2)^2 + (\omega_x x_2 - \omega_y x_1)^2} - (\omega_x x_2 + \omega_y x_1) \right)^2} \right. \\ &\quad \left. + \sqrt{(\omega_x x_1 - \omega_y x_2)^2 + (\omega_x x_2 - \omega_y x_1)^2} - (\omega_x x_2 + \omega_y x_1) \right) \\ \theta_2 &= \frac{1}{4\rho} \left(\sqrt{8\alpha\rho^2 + \left(\sqrt{(\omega_x x_1 - \omega_y x_2)^2 + (\omega_x x_2 - \omega_y x_1)^2} - (\omega_x x_2 + \omega_y x_1) \right)^2} \right. \\ &\quad \left. + \sqrt{(\omega_x x_1 - \omega_y x_2)^2 + (\omega_x x_2 - \omega_y x_1)^2} - (\omega_x x_2 + \omega_y x_1) \right). \end{aligned}$$

Wir schlussfolgern also aus der Ungleichung (5.5), dass diese erfüllt ist, wenn θ die Bedingung

$$\theta_1 \leq \theta \leq \theta_2$$

erfüllt.

Wir wählen daher o.B.d.A. $\theta := \theta_2$. Damit können wir $m_3(\theta)$ genau bestimmen. Die Konstante m_2 hingegen ist frei wählbar, sodass dieser Parameter nicht zu untersuchen ist. Für die vollständige Bestimmung von $f(\alpha)$ gilt es nun noch, $m_4(\theta)$ zu bestimmen.

Betrachten wir hierfür die Ungleichung (4.19) aus dem Beweis des EKBF in Kapitel 4.2.5:

$$\tilde{\zeta}^T(s)\tilde{\zeta}(s) \leq m_2(\theta)\tilde{\zeta}^T(s)T(\theta)\tilde{\zeta}(s)$$

Aus (5.4) wissen wir bereits, dass $T(\theta) = \theta I_{2 \times 2}$. Folglich muss $m_4(\theta)$ die Ungleichung

$$1 \leq m_4(\theta)\theta$$

erfüllen.

Dem Beweis der Konvergenz des EKBF für eine allgemeine und nicht konkret gegebene Funktion genügt diese Eigenschaft von $m_4(\theta)$. Da wir jedoch eine explizite obere Schranke für $P(t)$ angeben möchten, stehen wir nun vor der Wahl eines $m_4(\theta)$, welches diese Bedingung erfüllt.

Sei also

$$m_4(\theta) \geq \frac{1}{\theta} + c, \quad 0 < c \in \mathbb{R}.$$

Um eine aussagekräftige Abschätzung für $P(t)$ zu bekommen, wählen wir $m_4(\theta) = \frac{1}{\theta} + c$. Damit haben wir das kleinstmögliche $m_4(\theta)$ gewählt und dieses führt zur gewünschten Abschätzung. Würden wir ein größeres $m_4(\theta)$ wählen, so wäre die Ungleichung $P(t) \leq m_1 I_{2 \times 2}$ zwar immernoch erfüllt; sie würde jedoch zu einer unnötig „großen“ Schranke führen.

Indem wir $m_4(\theta) = \frac{1}{\theta} + c$ wählen, erhalten wir folgende Gestalt für $f(\alpha)$:

$$f(\alpha) = \left(\frac{1}{\theta} + c\right) \left(m_2 + \frac{1 + \theta^2}{\alpha}\right)$$

In einem weiteren Schritt wollen wir nun also die Frage beantworten, ob es ein $\alpha > 0$ gibt, für welches $f(\alpha)$ minimal wird.

Für unsere Zwecke genügt es, die in θ von α unabhängigen Terme als konstant zu betrachten: Sei also $d := \left(\sqrt{(\omega_x x_1 - \omega_y x_2)^2 + (\omega_x x_2 - \omega_y x_1)^2} - (\omega_x x_2 + \omega_y x_1)\right)^2$.

Dann ist

$$\theta = \frac{1}{4\rho} \left(\sqrt{8\alpha\rho^2 + d} + \sqrt{d}\right)$$

und schließlich

$$f(\alpha) = \left(\frac{1}{\frac{1}{4\rho} \left(\sqrt{8\alpha\rho^2 + d} + \sqrt{d}\right)} + c\right) \left(m_2 + \frac{1 + \left(\frac{1}{4\rho} \left(\sqrt{8\alpha\rho^2 + d} + \sqrt{d}\right)\right)^2}{\alpha}\right).$$

Betrachten wir den Kurvenverlauf von $f(\alpha)$, so sehen wir, dass der Graph einen „ $\frac{1}{\alpha}$ “-ähnlichen Verlauf hat. Insofern ist kein Minimum in $(0, \infty)$ zu erwarten.

Daher untersuchen wir den Kurvenverlauf mittels der Grenzwerte und erhalten:

$$\begin{aligned}\lim_{\alpha \rightarrow 0} f(\alpha) &\rightarrow \infty \\ \lim_{\alpha \rightarrow \infty} f(\alpha) &= cm_2\end{aligned}$$

Offensichtlich hat $f(\alpha)$ kein Minimum. Dies ist auf die Form von θ und schließlich auf jene der Differentialgleichung (5.2) zurückzuführen.

An dieser Stelle ist es also nicht möglich ein Minimum aus $f(\alpha)$ direkt anzugeben.

Um dieses Problem zu lösen, verfolgen wir nun also die Idee, den Grenzwert $\lim_{\alpha \rightarrow \infty} f(\alpha)$ als Minimum zu wählen. Der Gedanke liegt darin begründet, dass wir dann zwar kein „echtes“ Minimum haben, aber uns insofern Abhilfe schaffen, dass wir einen Wert nehmen, der den kleinsten Funktionswerten von $f(\alpha)$ sehr nahe kommt.

Sei also $m_1 = cm_2$, dann ist

$$P(t) \leq m_1 I_{2 \times 2} \iff P(t) \leq cm_2 I_{2 \times 2}.$$

Mit dieser Wahl gilt für beliebiges $\alpha > 0$:

$$P(t) \leq cm_2 I_{2 \times 2}$$

Das heißt, $cm_2 I_{2 \times 2}$ ist die kleinstmögliche obere Schranke von $P(t)$.

Obwohl also kein Minimum existiert, haben wir dennoch eine obere Schranke für die Schätzfehlerkovarianzmatrix $P(t)$ gefunden.

Wir widmen uns nun der Aufgabe, eine untere Schranke für $P(t)$, das heißt, eine obere Schranke für $Q(t) := P^{-1}(t)$ zu finden und nennen sie $m_5 I_{2 \times 2}$.

Da $P(t)$ aus dem EKBF die Gleichung (4.2) erfüllt, folgern wir, dass $Q(t)$ entsprechend die Gleichung

$$\dot{Q}(t) = -A^T(t)Q(t) - Q(t)A(t) - Q^2(t) + I_{2 \times 2}$$

erfüllt.

Außerdem gilt $Q(0) = P^{-1}(0) > 0$.

Gesucht ist also ein $0 < m_5 \in \mathbb{R}$, sodass $\forall t > 0 : Q(t) \leq m_5 I_{2 \times 2}$. Dazu betrachten wir folgendes Minimierungsproblem:

$$J'(\mu) = \zeta^T(0)Q(0)\zeta(0) + \int_0^t \zeta^T(s)\zeta(s) + \mu^T(s)\mu(s)ds$$

mit den Nebenbedingungen

$$\begin{aligned}\dot{\zeta}(s) &= A(s)\zeta(s) + \mu(s)I_{2 \times 2} \\ \zeta(t) &= \zeta.\end{aligned}$$

Für das Minimum gilt

$$J'(\mu) = \zeta^T Q(t) \zeta.$$

Wir kennen nun die optimale Lösung dieses Minimierungsproblems, wenden jedoch eine suboptimale Lösung $\mu(s) = (\alpha I_{2 \times 2} - A(s)) \zeta(s)$ an, um eine Ungleichung zu erhalten und damit schließlich eine Abschätzung zu ermöglichen. Damit bekommen wir mit den gleichen Argumenten wie für die Abschätzung in (4.26):

$$\begin{aligned}\zeta^T Q(t) \zeta &\leq \zeta^T(0)Q(0)\zeta(0) \\ &\quad + \int_0^t \zeta^T(s)\zeta(s) + ((\alpha I_{2 \times 2} - A(s)) \zeta(s))^T (\alpha I_{2 \times 2} - A(s)) \zeta(s) ds \\ &= \zeta^T(0)Q(0)\zeta(0) \\ &\quad + \int_0^t \zeta^T(s) \left(I_{2 \times 2} + (\alpha I_{2 \times 2} - A(s))^T (\alpha I_{2 \times 2} - A(s)) \right) \zeta(s) ds \\ &\leq e^{-2\alpha t} \zeta^T Q(0) \zeta + \int_0^t e^{2\alpha(s-t)} \zeta \left(I_{2 \times 2} + (\alpha + L)^2 I_{2 \times 2} \right) \zeta ds\end{aligned}$$

Die Schätzfehlerkovarianzmatrix $P(0)$ ist zum Startzeitpunkt frei wählbar; insofern gilt dies auch für ihre Inverse $Q(t)$. Das heißt, $\exists 0 < m_6 \in \mathbb{R}$, sodass $Q(0) \leq m_6 I_{2 \times 2}$. Ferner ist $I_{2 \times 2} + (\alpha + L)^2 I_{2 \times 2}$ wegen $0 < \alpha \in \mathbb{R}$ und wegen $0 < L \in \mathbb{R}$ ebenfalls beschränkt; das heißt $\exists 0 < m_7 \in \mathbb{R}$, sodass $I_{2 \times 2} + (\alpha + L)^2 I_{2 \times 2} \leq m_7 I_{2 \times 2}$.

Daraus resultiert

$$\zeta^T Q(t) \zeta \leq \left(m_6 + \frac{m_7}{2\alpha} \right) \zeta^T \zeta =: m_5 \zeta^T \zeta.$$

Wie oben, fragen wir uns auch an dieser Stelle, ob es ein $\alpha > 0$ gibt, für welches die obere Schranke $m_5 I_{2 \times 2}$ minimal wird?

Hierfür definieren wir

$$g(\alpha) := m_6 + \frac{m_7}{2\alpha} = m_6 + \frac{1 + (\alpha + L)^2}{2\alpha}$$

und suchen das Minimum dieser Funktion.

Einige leichte Rechnungen führen zu dem Ergebnis, dass tatsächlich ein Minimum existiert und $g(\alpha)$ dieses in $m_6 + \frac{1 + (\sqrt{L^2+1}+L)^2}{2\sqrt{L^2+1}}$ annimmt.

Wir wählen $m_6 + \frac{1 + (\sqrt{L^2+1}+L)^2}{2\sqrt{L^2+1}} I_{2 \times 2}$ als obere Schranke für $Q(t)$. Folglich wissen wir, dass

$$Q(t) \leq m_6 + \frac{1 + (\sqrt{L^2+1}+L)^2}{2\sqrt{L^2+1}} I_{2 \times 2} \quad \forall \alpha > 0.$$

Schließlich fassen wir beide Schranken für $P(t)$ zusammen und erhalten:

$$m_6 + \frac{1 + (\sqrt{L^2+1}+L)^2}{2\sqrt{L^2+1}} I_{2 \times 2} \leq P(t) \leq cm_2 I_{2 \times 2}$$

Das bedeutet, so lange die Schätzfehlerkovarianzmatrix $P(t)$ in diesem Bereich liegt, erhalten wir Konvergenz des EKBF für das gegebene System (5.2). Damit ist auch festgelegt, wie groß die Schätzfehler höchstens sein dürfen, damit das EKBF überhaupt konvergiert. Für die praktische Anwendung hat dieses Wissen besondere Relevanz, um ein gewünschtes Ergebnis des Algorithmus, also um eine Lösung der Differentialgleichung (5.2) zu erhalten.

Kapitel 6

Numerische Realisierung des EKF

Bisher haben wir das EKBF in zeitkontinuierlicher Form kennengelernt. Jedoch wird - vor allem auf Grund der Implementierung - in der Praxis eine diskrete Systemdarstellung benötigt, um vorliegende Probleme mit dem Computer lösen zu können.

Die für diese Thematisierung verwendeten Inhalte werden aus [BH97] und aus [GA01] entnommen.

Für eine numerische Betrachtung einer solchen Aufgabenstellung ist es jedoch nicht nur notwendig, das gegebene System in diskreter Form vorliegen zu haben, sondern auch das Verfahren, das zu seiner Lösung führen soll, in ebendieser Form zu benutzen.

Die hier vorgestellten Verfahren sind auch diejenigen, die im folgenden Kapitel zur Implementierung verwendet werden. Die Inhalte entstammen dem internen technischen Bericht [Kal09] der DAIMLER AG, sodass aufgezeigt wird, welche Verfahren dort benutzt werden. Einzig der Bierman-Algorithmus ist in diesem technischen Bericht nicht in detaillierter Form aufzufinden, weshalb hierfür auf die Originalschrift von Bierman [Bie77], in der er seinen Algorithmus vorstellt, zurückgegriffen wird.

6.1. Das Kalman-Filter für lineare Systeme in zeitdiskreter Form

Im Gegensatz zum zeitkontinuierlichen Erweiterten Kalman-Bucy Filter, unterscheidet sich das Erweiterte Kalman-Filter¹⁵, dadurch, dass es aus zwei Schritten besteht. Zunächst schätzt es den Systemzustand in der Zeit unter Zuhilfenahme des geschätzten Zustands zu einem früheren Zeitpunkt und verbessert diese Schätzung anschließend im zweiten Schritt, indem es vorliegende Messwerte berücksichtigt und einbezieht.

¹⁵Kalman hat das Kalman-Filter für Systeme in zeitdiskreter Form entwickelt. Daher werden wir das Erweiterte Kalman-Filter es im Folgenden mit EKF abkürzen.

Die Motivation und die Grundidee des zeitdiskreten Kalman-Filters gleichen jener des zeitkontinuierlichen. Daher folgt die Herleitung dieser beiden derselben Idee: Gesucht ist ein Schätzalgorithmus, der die Schätzfehler im quadratischen Mittel minimiert.

Wir haben bereits eine ausführliche Herleitung der Gleichungen für den zeitkontinuierlichen Fall in Kapitel 3 thematisiert. Daher wird der Algorithmus des zeitdiskreten Kalman-Filters abgekürzt vorgestellt.

Anwendung der Euler-Diskretisierung auf das zeitkontinuierliche System (1.1) ergibt mit der Annahme, dass die Eingangsgrößen u innerhalb der Zeitintervalle konstant sind und mit den Matrizen A_k, B_k, C_k, G_k das zeitdiskrete System:

$$\begin{aligned} x_{k+1} &= A_k x_k + B_k u_k + G_k w_k && \text{Zustandsgleichung} \\ y_k &= C_k x_k + v_k && \text{Ausgangsgleichung} \end{aligned} \quad (6.1)$$

x_k bezeichnet den Systemzustand und v_k sowie w_k sind jeweils als unkorrelierte, mittelwertfreie Rauschprozesse zu verstehen, das heißt

$$\begin{aligned} E[w_k] &= 0, \quad E[v_k] = 0 \\ E[w_k^T w_m] &= Q_k \delta(k - m) \\ E[v_k^T v_m] &= R_k \delta(k - m) \\ E[w_k^T v_m] &= 0. \end{aligned} \quad (6.2)$$

Außerdem ist der Zustand zum Startzeitpunkt k_0 normalverteilt und wir wählen die Schätzfehlerkovarianzmatrix P zum Zeitpunkt k_0 folgendermaßen:

$$P_0 = E[(x_{k_0} - E[x_{k_0}])(x_{k_0} - E[x_{k_0}])^T]$$

Schließlich nehmen wir noch an, dass das System- und das Messrauschen mit dem Anfangszustand jeweils unkorreliert sind:

$$E[x_{k_0}^T w_k] = 0, \quad E[x_{k_0}^T v_k] = 0, \quad \text{für } k > k_0 \quad (6.3)$$

Im Vergleich zum zeitkontinuierlichen Fall, in welchem der Schätzalgorithmus aus einer Modellgleichung des Zustandes und einer Nutzung der Ausgangsgleichung besteht, setzt sich das Kalman-Filter für den zeitdiskreten Fall aus zwei Schritten zusammen: Die Prädiktion und die Innovation.

Die Prädiktion des Systemzustands orientiert sich an der gegebenen Zustandsgleichung und schätzt den Zustand anhand dieser. Dabei bedient sie sich dem Wissen über den Zustand zum Zeitpunkt k ohne die Ausgangsgleichung einzubeziehen und schätzt damit den Zustand des Systems zum Zeitpunkt $k + 1$. Daher sprechen wir an dieser Stelle von

einer Extrapolation des Systemzustands.

Wir kennzeichnen die Prädiktion des Zustands mit \hat{x}_{k+1}^- . Das Minuszeichen im Exponenten drückt aus, dass die Ausgangsgleichung nicht berücksichtigt wird.

Der Innovationsschritt korrigiert die Prädiktion des Zustandes, indem er zum prädizierten Systemzustand \hat{x}_k^- die Messwerte aus der Ausgangsgleichung y_k berücksichtigt. Daraus erklärt sich die Bezeichnung x_k^+ des Zustandes im Korrekturschritt.

Was passiert jedoch, wenn die Messung y_k stark verrauscht und damit von Messfehlern behaftet ist? In diesem Fall wäre eine Innovation im negativen Sinn zu erwarten, denn anstatt die prädizierte Schätzung zu verbessern, würde sie verschlechtert werden und das Kalman-Filter würde schließlich divergieren. Abhilfe schafft hier die Kalman-Matrix K_k , die den Einfluss der Messwerte auf den Innovationsschritt entscheidend gewichtet. Sind die Eingänge in das System nicht ausschlaggebend stark von Messrauschen behaftet, so werden diese als vertrauenswürdig angesehen und daher stärker gewichtet, um positiven Einfluss auf die prädizierte Zustandsschätzung zu nehmen, das bedeutet, sie zu verbessern. Sind die Rauschprozesse jedoch so stark, dass sie die Messung überlagern, so hat die Kalman-Matrix eine solche Form, die die Auswirkung auf die Schätzung dämpft, also schwach gewichtet.

Wir werden nun auf diese beiden Schritte und auf die Darstellung der Kalmanmatrix eingehen.

Dazu definieren wir im Vorfeld den Prädiktions- sowie den Innovationsfehler¹⁶ bezüglich der Schätzung des Zustands als

$$\begin{aligned} e_k^- &:= x_k - \hat{x}_k^- \\ e_k^+ &:= x_k - \hat{x}_k^+ \end{aligned} \quad (6.4)$$

Der Prädiktionsschritt

Mit Hilfe obiger Anmerkungen ist die Prädiktion des Systemzustandes aus (6.1) leicht nachvollziehbar:

$$\hat{x}_{k+1}^- = A_k \hat{x}_k^+ + B_k u_k + G_k w_k \quad (6.5)$$

Das diskrete Kalman-Filter bestimmt die Güte dieser prädizierten Größe ebenfalls im Prädiktionsschritt. Dabei werden zunächst der Fehler dieser Prädiktion und anschließend dessen Kovarianz betrachtet.

Sei also e_k^- der Fehler, der durch die Prädiktion auftritt. Dann gilt mit (6.1) und (6.5) für

¹⁶Das jeweilige Minus- sowie das Pluszeichen im Exponenten des Fehlers sind in gleicher Weise zu verstehen wie schon bei der Kennzeichnung der Schätzungen. Daher werden diese Fehler in der Literatur häufig auch als a priori beziehungsweise a posteriori Schätzfehler bezeichnet.

den a priori Fehler:

$$\begin{aligned} e_{k+1}^- &= x_{k+1} - \hat{x}_{k+1}^- = x_{k+1} = A_k x_k + B_k u_k + G_k w_k - (A_k \hat{x}_k^+ + B_k u_k + G_k w_k) \\ &= A_k e_k^+ + G_k w_k \end{aligned}$$

Wie schon erwähnt, ziehen wir nun die Kovarianz dieses Fehlers als Maß für die Genauigkeit der Schätzung heran.

Sei $P_k^- := E[e_k^-(e_k^-)^T]$ die Schätzfehlerkovarianzmatrix. Dann gilt

$$\begin{aligned} P_{k+1}^- &= E[e_{k+1}^-(e_{k+1}^-)^T] = E[(A_k e_k^+ + G_k w_k)(A_k e_k^+ + G_k w_k)^T] \\ &= E[A_k e_k^+ (e_k^+)^T A_k^T] + E[A_k e_k^+ (w_k)^T G_k^T] \\ &\quad + E[G_k w_k (e_k^+)^T A_k] + E[G_k w_k (w_k)^T G_k^T]. \end{aligned} \quad (6.6)$$

Der Erwartungswert tangiert nur das Systemrauschen w_k und das Messrauschen v_k . Beim Betrachten des Fehlers e_k^+ ist jedoch ersichtlich, dass hiervon ebenfalls der Erwartungswert gebildet werden kann, da er unter anderem von diesen Rauschprozessen abhängt.

Wie wir im nächsten Kapitel „Der Innovationsschritt“ sehen werden, hängt e_k^+ nur vom Systemrauschen zu vorangegangenen Zeitpunkten ab, also von w_m für $m = k_0, \dots, k-1$ sowie vom Messrauschen v_p für $p = k_0, \dots, k$. Ein etwas genauerer Blick auf den Fehler macht deutlich, dass er außerdem nur noch vom Schätzfehler zum Startzeitpunkt k_0 mit $e_{k_0}^+ = x_{k_0} - \hat{x}_{k_0}^+$ abhängt.

Folglich können wir den Erwartungswert im zweiten Summanden aus (6.6) als Linearkombination aus $E[(x_{k_0} - \hat{x}_{k_0}^+)^T w_k]$, $E[w_m^T w_k]$ für $k_0 \leq m \leq k-1$ und $E[v_m^T v_k]$ für $k_0 + 1 \leq m \leq k$ schreiben. Selbiges gilt für den dritten Summanden, wobei hier die Terme der Linearkombination zu transponieren sind.

Mit (6.2) und (6.3) fallen daher der zweite sowie der dritte Summand weg und für die Schätzfehlerkovarianzmatrix im Prädiktionsschritt ergibt sich:

$$P_{k+1}^- = A_k P_k^+ A_k^T + G_k Q_k G_k^T \quad (6.7)$$

Mit dieser Kovarianzmatrix sind die Gleichungen für den Prädiktionsschritt des Kalman-Filters nun vollständig:

$$\begin{aligned} \hat{x}_{k+1}^- &= A_k \hat{x}_k^+ + B_k u_k + G_k w_k \\ P_{k+1}^- &= A_k P_k^+ A_k^T + G_k Q_k G_k^T \end{aligned}$$

Der Innovationsschritt

Wir wissen bereits, dass es sich beim Innovationsschritt um eine Verbesserung der Zustandsprädiktion handelt. Dies geschieht anhand von Addition einer Gewichtung des Einflusses der Messgleichung zur Zustandsprädiktion. Es findet also ein Übergang der Schätzung x_{k+1}^- nach x_{k+1}^+ statt. In mathematischer Schreibweise ausgedrückt bedeutet dies:

$$\hat{x}_{k+1}^+ = \hat{x}_{k+1}^- + K_{k+1}(y_{k+1} - C_{k+1}\hat{x}_{k+1}^-) \quad (6.8)$$

Die Kalman-Matrix K_{k+1} ist an dieser Stelle noch unbestimmt. Die optimale Wahl dieser Gewichtsmatrix, wird noch zu bestimmen sein.

Zunächst jedoch betrachten wir die Schätzfehlerkovarianzmatrix $P_{k+1}^+ := [(e_{k+1}^+)^T e_{k+1}^+]$. Ihre Herleitung folgt derselben Idee wie jener von P_{k+1}^- . Da die Vorgehensweise und die Argumentation an dieser Stelle analog sind, beschränken wir uns darauf, die Kovarianzmatrix ohne Weiteres anzugeben. Sie hat die Gestalt

$$P_{k+1}^+ = (I - K_{k+1}C_{k+1})P_{k+1}^-(I - K_{k+1}C_{k+1})^T + K_{k+1}R_{k+1}K_{k+1}^T, \quad (6.9)$$

wobei K_{k+1} noch immer unbekannt ist.

Wir wollen nun eine Bedingung an K_{k+1} stellen, anhand derer es möglich sein wird, eine im Sinne der Minimierung der Schätzfehlerkovarianzmatrix optimale Darstellung von K_{k+1} zu erhalten.

Wie im zeitkontinuierlichen Fall ist dies analog zur Bedingung

$$\text{tr}(P_{k+1}^+) \stackrel{!}{=} \min.$$

Wenn wir nun die Matrixrelationen aus der Bemerkung in 3.3.2 anwenden und die für Minima notwendigen Ableitungen von $\text{tr}(P_{k+1}^+)$ berechnen, so erhalten wir¹⁷

$$\begin{aligned} -2(I - K_{k+1}C_{k+1})P_{k+1}^-C_{k+1}^T + 2K_{k+1}R_{k+1} &= 0 \\ P_{k+1}^-C_{k+1}^T(C_{k+1}P_{k+1}^-C_{k+1}^T + R_{k+1})^{-1} &= K_{k+1} \end{aligned} \quad (6.10)$$

Wir haben dann eine Darstellung der Kalman-Matrix K_{k+1} gefunden, die es uns erlaubt, das Kalman-Filter als optimal zu bezeichnen, da mit dieser Wahl von K_{k+1} der Schätzfehler minimiert wird.

Der aufmerksame Leser sieht, wie die Gewichtung durch diese Matrix zu verstehen ist:

¹⁷Die Vorgehensweise ist völlig analog zu jener in Kapitel 3.3 und kann dort nachgelesen werden.

Wie oben schon beschrieben, werden Einflüsse der Messdaten je nach Güte stärker oder schwächer gewichtet, um eine Verbesserung zu erlangen sowie eine Verschlechterung der Zustandsschätzung zu verhindern. Sind also die Messdaten stark verrauscht und eine Schätzung des Zustands nur schwer möglich, so sind die Einträge in der Kovarianzmatrix R_{k+1} der Messstörungen entsprechend groß. Für uns bedeutet dies, dass wir den Einfluss der Messdaten zu ebendiesem Zeitpunkt nicht zu sehr gewichten möchten. Die Kalman-Matrix erfüllt genau dies durch ihre Darstellung und insbesondere durch die Inverse der Summe $(C_{k+1}P_{k+1}^-C_{k+1}^T + R_{k+1})$, da die Einträge in K_{k+1} klein werden und die Addition dieser verrauschten Messdaten in der Innovation x_{k+1}^+ des Zustands nicht zu sehr ins Gewicht fällt. In diesem Fall vertrauen wir eher der Prädiktion x_{k+1}^- , was bei Betrachtung der Gleichung (6.8) auch rechentechnisch nachvollzogen werden kann.

Umgekehrt, bei wenig verrauschten Messdaten, wird die Kovarianzmatrix R_{k+1} betragsmäßig kleine Einträge haben, sodass die Kalman-Matrix die Messdaten entsprechend stärker gewichtet und die Schätzung mit Hilfe der Ausgänge verbessert wird.

Setzen wir dieses Ergebnis nun in die vorläufige Form (6.9) der Schätzfehlerkovarianzmatrix P_{k+1}^+ ein, so erhalten wir schließlich:

$$\begin{aligned}
P_{k+1}^+ &= (-P_{k+1}^- - K_{k+1}C_{k+1}P_{k+1}^-) \cdot (-C_{k+1}^TK_{k+1}^T + I) + K_{k+1}R_{k+1}K_{k+1}^T \\
&= P_{k+1}^- - P_{k+1}^-C_{k+1}^TK_{k+1}^T - K_{k+1}C_{k+1}P_{k+1}^- + K_{k+1}(C_{k+1}P_{k+1}^-C_{k+1}^T + R_{k+1})K_{k+1}^T \\
&\stackrel{(6.10)}{=} P_{k+1}^- - P_{k+1}^-C_{k+1}^TK_{k+1}^T - K_{k+1}C_{k+1}P_{k+1}^- \\
&\quad + \left(P_{k+1}^-C_{k+1}^T(C_{k+1}P_{k+1}^-C_{k+1}^T + R_{k+1})^{-1}\right) \cdot (C_{k+1}P_{k+1}^-C_{k+1}^T + R_{k+1})K_{k+1}^T \\
&= (I - K_{k+1}C_{k+1})P_{k+1}^-
\end{aligned} \tag{6.11}$$

Wir sehen also, dass die Schätzfehlerkovarianzmatrix P_{k+1}^+ im Innovationsschritt von jener durch die Prädiktion erhaltenen Matrix P_{k+1}^- abhängt und diese gleichzeitig durch die Kalman-Matrix K_{k+1} sowie durch die Eingangsmatrix C_{k+1} gewichtet wird.

Letztlich halten wir folgende, hergeleitete Gleichungen aus dem Prädiktions- sowie dem Innovationsschritt fest:

Prädiktion

$$\begin{aligned}
\hat{x}_{k+1}^- &= A_k\hat{x}_k^+ + B_k u_k + G_k w_k \\
P_{k+1}^- &= A_k P_k^+ A_k^T + G_k Q_k G_k^T
\end{aligned}$$

Innovation

$$\begin{aligned}
\hat{x}_{k+1}^+ &= \hat{x}_{k+1}^- + K_{k+1}(y_{k+1} - C_{k+1}\hat{x}_{k+1}^-) \\
P_{k+1}^+ &= (I - K_{k+1}C_{k+1})P_{k+1}^- \\
K_k &= P_{k+1}^-C_{k+1}^T(C_{k+1}P_{k+1}^-C_{k+1}^T + R_{k+1})^{-1}
\end{aligned}$$

Diese beschreiben das „lineare“ Kalman-Filter für den zeitdiskreten Fall.

Der Algorithmus beginnt mit der Prädiktion des Systemzustands sowie der zugehörigen Schätzfehlerkovarianzmatrix und geht dann über in den Innovationsschritt, wo die

Kalman-Matrix bestimmt wird und mit ihm eine Verbesserung der Schätzung im vorigen Schritt stattfindet. Danach wird der Zeitschritt erhöht, der Algorithmus beginnt wieder mit der Prädiktion und läuft auf diese Weise fort.

6.2. Das EKF für nichtlineare Systeme in zeitdiskreter Form

Wir wissen bereits aus Kapitel 4.1, dass das EKF auf nichtlinearen System- und Messmodellen beruht.

Gemäß (4.1) haben diese im zeitdiskreten Fall die Form

$$\begin{aligned}x_{k+1} &= f(x_k, u_k, w_k) \\ y_k &= h(u_k, v_k).\end{aligned}$$

Die Idee des EKF folgt im Grunde genommen jener des Kalman-Filters für lineare Systemmodelle. Um diesen Gedanken nachzuvollziehen, muss der Anwendung des Algorithmus eine Linearisierung der Systemgleichungen vorangehen, um ein lineares System vorliegen zu haben und dann den Algorithmus anwenden zu können.

Die Anwendung des EKF auf solche Systeme mit der Prädiktion der Zustände und der Berechnung der prädizierten Messgrößen ist mittels des gegebenen nichtlinearen Systemmodells direkt möglich:

$$\begin{aligned}\hat{x}_{k+1}^- &= f(\hat{x}_k^+, u_k, 0) \\ \hat{y}_{k+1} &= h(\hat{x}_{k+1}^-, 0)\end{aligned}$$

Um hingegen die Schätzfehlerkovarianzmatrix zu bestimmen, ist im Vorfeld in jedem Zeitschritt eine Linearisierung der Zustands- und Ausgangsgleichung um den zugehörigen Arbeitspunkt x^- beziehungsweise y^- nötig. Daher werden hierfür folgende Jacobimatrizen der nichtlinearen Funktionen f und h benötigt:

$$A_k = \frac{\partial f}{\partial x}(\hat{x}_k, u_k, 0), \quad W_k = \frac{\partial f}{\partial w}(\hat{x}_k, u_k, 0), \quad H_k = \frac{\partial h}{\partial x}(\hat{x}_k, 0), \quad V_k = \frac{\partial h}{\partial v}(\hat{x}_k, 0)$$

Die Grundidee, die hinter den Darstellungen der Gleichungen, die einer vorangehenden Linearisierung bedürfen, gleicht derjenigen des EKF für den zeitkontinuierlichen Fall. Deshalb werden im Folgenden nur die Gleichungen des Algorithmus des zeitdiskreten EKF angegeben:

Prädiktion

$$\begin{aligned}\hat{x}_{k+1}^- &= f(\hat{x}_k^+, u_k, 0) \\ P_{k+1}^- &= A_{k+1} P_k^+ A_{k+1}^T + W_{k+1} Q_k W_{k+1}^T\end{aligned}$$

Innovation

$$\begin{aligned}\hat{x}_{k+1}^+ &= \hat{x}_{k+1}^- + K_{k+1} (y_{k+1} - h(\hat{x}_{k+1}^-, 0)) \\ P_{k+1}^+ &= (I - K_{k+1} H_{k+1}) P_{k+1}^- \\ K_{k+1} &= P_{k+1}^- H_{k+1}^T (H_{k+1} P_{k+1}^- H_{k+1}^T + V_{k+1} R_k V_{k+1}^T)^{-1}\end{aligned}$$

Dabei beschreiben die Matrizen A_{k+1} sowie H_{k+1} die Änderung des Systems beziehungsweise den Zusammenhang zwischen Systemzustand und Messung. V_{k+1} und W_{k+1} wiederum sind die zum System- sowie zum Messrauschen gehörenden Störmatrizen, die den Einfluss der Störungen auf das System sowie die Messung repräsentieren.

6.3. Numerische Verfahren im EKF

Wie im zeitkontinuierlichen Fall, hängen die Jacobimatrizen zur Linearisierung der Modellgleichungen auch im zeitdiskreten von der Zeit ab. Daher enthalten sie in jedem Zeitschritt verschiedene Einträge. Insofern müssen die Jacobimatrizen in jedem Zeitschritt neu berechnet werden. Unter Umständen kann dies entsprechend der nichtlinearen Zustands- sowie Ausgangsgleichung zu erheblichem numerischen Aufwand führen.

Wir werden nun im Einzelnen auf die numerische Realisierung des EKF eingehen.

6.3.1. Numerische Verfahren im Prädiktionsschritt

Die Prädiktion besteht aus zwei Gleichungen. Wir widmen uns zunächst der Zustandsschätzung.

Numerische Realisierung der Zustandsschätzung

Es ist bereits bekannt, dass diese auf dem nichtlinearen Systemmodell $x_{k+1} = f(x_k, u_k, w_k)$ basiert, wobei x_{k+1} als Ableitung zum Zeitpunkt $t = k + 1$ zu verstehen ist. Daher gilt es, diese nichtlineare Funktion zu integrieren, was numerisch realisiert wird.

Ist die Schrittweite τ genügend klein, so liefert das explizite Euler-Verfahren ein ausreichend genaues Integrationsergebnis. Allerdings ist zu beachten, dass die Fehlerordnung mit $\mathcal{O}(\tau)$ bei diesem Verfahren niedrig ist.

Um den Rechenaufwand zu verringern, ist in vielen Fällen eine größere Schrittweite wünschenswert. Aus diesem Grund wird in den meisten Fällen das wohlbekanntere Adams-Bashforth-Verfahren 4. Ordnung zur numerischen Integration herangezogen. Es hat mit

$\mathcal{O}(h^4)$ eine deutlich höhere Fehlerordnung und gewährleistet dadurch ein genaueres numerisches Integrationsergebnis.¹⁸

Numerische Realisierung der Bestimmung der Schätzfehlerkovarianz

Wir möchten numerische Stabilität beim Kalman-Filter haben. Erreichen können wir dies durch entsprechende Faktorisierungen unserer Matrizen. Im Folgenden gehen wir daher von einer Schätzfehlerkovarianzmatrix P in faktorisierter Form aus. Dies ist legitim, da wir bereits wissen, dass P symmetrisch und damit quadratisch ist. Sei also

$$P = UDU^T$$

die faktorisierte Form von P , wobei U eine obere Dreiecksmatrix und D eine Diagonalmatrix ist. Für die Faktorisierung gibt es mehrere Möglichkeiten; beispielsweise die Cholesky- oder die UD-Zerlegung. Die UD-Zerlegung, die auch als modifizierte Cholesky-Zerlegung bekannt ist, ist rechentechnisch weniger aufwändig. Aus diesem Grund realisieren wir sie mittels dem Bierman-Algorithmus, der in [Bie77] nachzulesen ist. Er zeichnet sich durch eine numerisch effektive Methode zur Berechnung einer solchen Faktorisierung auszeichnet. BIERMAN hatte die Idee, die Basis, welche die Schätzfehlerkovarianzmatrix aufspannt, in orthogonaler Form beizubehalten.

Ferner sei Q eine Diagonalmatrix.

Damit ergibt sich für die Prädiktion der Schätzfehlerkovarianzmatrix folgende Faktorisierungsformel nach BIERMAN:

$$P_{k+1}^- = \begin{bmatrix} A_k U_k & G_k \end{bmatrix} \cdot \begin{bmatrix} D_k & 0 \\ 0 & Q_k \end{bmatrix} \cdot \begin{bmatrix} A_k U_k & G_k \end{bmatrix}^T = WDW^T = U_{k+1}^- D_{k+1}^- (U_{k+1}^-)^T$$

Die Lösung erfolgt durch modifizierte, gewichtete Gram-Schmitt-Orthogonalisierung, die wir auf W anwenden:

$$W = U_{k+1}^- M, \quad MDM^T = D_{k+1}^-$$

6.3.2. Numerische Verfahren im Innovationsschritt

In diesem Schritt werden die Zustandsschätzung aus dem vorangegangenen Prädiktionsschritt und die UD-faktorisierte Schätzfehlerkovarianzmatrix, ebenfalls aus der Prädiktion, gleichzeitig verbessert. Dieses Vorgehen erfolgt sequentiell, sozusagen „Zeile für Zeile“

¹⁸Derzeit wird in der Automobilindustrie vorwiegend eine Schrittweite von $\tau = 0.01\text{sec}$ für dieses Mehrschrittverfahren benutzt. Begründet ist dies darin, dass eine kleinere Schrittweite aus technischen Gründen, insbesondere auf Grund der Messtechnik, nicht möglich ist.

für die skalaren Messgrößen y_i , $i = 1, \dots, l$.

Hier wird der Algorithmus von BIERMAN angewandt. Bevor wir die numerischen Verfahren im Innovationsschritt diskutieren, erfolgt zunächst eine Erläuterung der Technik von Bierman.

Der Bierman-Algorithmus

Der Bierman-Algorithmus dient der UDU^T -Faktorisierung der Schätzfehlerkovarianzmatrix P im Innovationsschritt.

Er ist deshalb günstig, da er keine Wurzelberechnungen beinhaltet, die eventuell zu numerischen Unsicherheiten führen können. Damit ist er ein numerisch stabiles und effizientes Verfahren zur Faktorisierung, das leicht zu implementieren ist.

Wir gehen also von den Gleichungen

$$\begin{aligned} P_{k+1}^+ &= (I - K_{k+1}H_{k+1})P_{k+1}^- \\ K_{k+1} &= P_{k+1}^- H_{k+1}^T (H_{k+1}P_{k+1}^- H_{k+1}^T + V_{k+1}R_k V_{k+1}^T)^{-1} \end{aligned}$$

aus Kapitel 6.2 aus und erhalten:

$$P_{k+1}^+ = P_{k+1}^- - P_{k+1}^- H_{k+1}^T (H_{k+1}P_{k+1}^- H_{k+1}^T + V_{k+1}R_k V_{k+1}^T)^{-1} H_{k+1}P_{k+1}^-$$

BIERMAN erläutert im zweiten Kapitel in [Bie77], wie es möglich ist, eine a priori Schätzung x_{k+1}^- sowie eine a priori Schätzfehlerkovarianz P_{k+1}^- mit einer skalaren Messgleichung $y_k = a_k^T x_k + v_k$ zu kombinieren, um eine Verbesserung der Zustands- sowie der Fehlerkovarianzschätzung im Innovationsschritt zu erhalten. Wir werden dies zwar benutzen, jedoch nicht weiter erläutern, sodass der interessierte Leser auf die angegebene Literatur verwiesen wird.

Da wir also skalare Messgrößen vorliegen haben, bezeichnen wir die Einträge der Messmatrix H^T mit a , sowie

$$H_{k+1}P_{k+1}^- H_{k+1}^T + V_{k+1}R_k V_{k+1}^T =: v.$$

Wir definieren weiter:

$$\begin{aligned} s &:= (U_{k+1}^-)^T a, & s^T &= (s_1, s_2, \dots, s_n) \\ z &:= D_{k+1}^- s, & z_i &= d_{k+1}^- s_i \quad \text{für } i = 1, \dots, n \end{aligned}$$

Anwendung der modifizierten Cholesky-Zerlegung auf P_{k+1}^+ und auf P_{k+1}^- liefert folgende Form dieser Matrizen:

$$\begin{aligned} P_{k+1}^+ &= U_{k+1}^+ D_{k+1}^+ (U_{k+1}^+)^T \\ P_{k+1}^- &= U_{k+1}^- D_{k+1}^- (U_{k+1}^-)^T \end{aligned} \quad (6.12)$$

Damit ist

$$\begin{aligned} P_{k+1}^+ &= U_{k+1}^- \left(D_{k+1}^- - \underbrace{D_{k+1}^- (U_{k+1}^-)^T H_{k+1}^-}_{=z} \left(\underbrace{H_{k+1} P_{k+1}^- H_{k+1}^T + V_{k+1} R_k V_{k+1}^T}_{=v} \right)^{-1} \right. \\ &\quad \left. \cdot \left(D_{k+1}^- (U_{k+1}^-)^T H_{k+1}^- \right)^T \right) (U_{k+1}^-)^T \\ &= U_{k+1}^- \left(D_{k+1}^- - z \alpha^{-1} z^T \right) (U_{k+1}^-)^T. \end{aligned}$$

Die Idee für das weitere Vorgehen ist nun die, eine Faktorisierung für den Ausdruck in der Klammer, also für $D_{k+1}^- - z v^{-1} z^T$ zu bestimmen. Diese wird mit dem Bierman-Algorithmus realisiert. Er wird in jedem Zeitschritt angewandt und hat im $(k+1)$ -ten Zeitschritt die Form:

$$v_0 = 0; \quad j = 1, \dots, n.$$

$$\begin{aligned} v_j &= \sum_{m=1}^j D_{mm} s_r^2 + V_{k+1} R_k V_{k+1}^T \\ \bar{D}_{jj} &= D_{jj} v_{j-1} v_j^{-1} \\ \bar{U}_{ij} &= \begin{cases} -D_{ii} v_{j-1}^{-1} s_i s_j & \text{für } i = 1, 2, \dots, j-1 \\ 1 & \text{für } i = j \\ 0 & \text{für } i = j+1, j+2, \dots, n \end{cases} \end{aligned} \quad (6.13)$$

Mit dem Bierman-Algorithmus erhalten wir also

$$D_{k+1}^- - z \alpha^{-1} z^T = \bar{U}_{k+1} \bar{D}_{k+1} \bar{U}_{k+1}^T.$$

Wir fassen dieses Ergebnis mit dem aus (6.12) zusammen:

$$U_{k+1}^+ D_{k+1}^+ (U_{k+1}^+)^T = P_{k+1}^+ = (U_{k+1}^- \bar{U}_{k+1}) \bar{D}_{k+1} \left(\bar{U}_{k+1}^T (U_{k+1}^-)^T \right)^T$$

Da das Produkt von oberen Dreiecksmatrizen wiederum eine solche Form hat, erhalten wir damit also P_{k+1}^+ in UD-faktorisierter Form, mit

$$\begin{aligned} U_{k+1}^+ &= U_{k+1}^- \bar{U}_{k+1}, \\ D_{k+1}^+ &= \bar{D}_{k+1}. \end{aligned}$$

Das eigentliche Faktorisierungsproblem wird also auf jenes der Faktorisierung von $D_{k+1}^- - z\alpha^{-1}z^T$ reduziert.

Wir kehren nun zur Klärung der numerischen Verfahren, die im Innovationsschritt des Kalman-Filters vorkommen, zurück.

Seien

$$U^- = U_0, \quad D^- = D_0$$

die Komponenten der faktorisierten Kovarianzmatrix P vor der Innovation und

$$U^+ = U_l, \quad D^+ = D_l$$

die Komponenten nach der Innovation über l Messgrößen.

Dann kann die Formel für die Innovation der Schätzfehlerkovarianzmatrix für die i -te skalare Größe wiederum mit Verweis auf BIERMAN angegeben werden:

$$\begin{aligned} P_i &= U_i D_i U_i^T = U_{i-1} D_{i-1} U_{i-1}^T - \frac{U_{i-1} D_{i-1} U_{i-1}^T a_i^T a_i U_{i-1} D_{i-1} U_{i-1}^T}{a_i U_{i-1} D_{i-1} U_{i-1}^T a_i^T + r_i} \\ &= U_{i-1} \left(D_{i-1} - \frac{D_{i-1} s_i s_i^T D_{i-1}}{s_i^T D_{i-1} s_i + r_i} \right) U_{i-1}^T, \end{aligned}$$

wobei $s_i = U_{i-1}^T a_i^T$ und a_i sind die skalaren Größen der Messmatrix H .

Es ist aus numerischen Stabilitätsgründen weiterhin unser Anliegen, die Matrix P_i in faktorisierter Form vorliegen zu haben. Ein Blick auf ihre obige Form zeigt jedoch, dass der Faktor $\left(D_{i-1} - \frac{D_{i-1} s_i s_i^T D_{i-1}}{s_i^T D_{i-1} s_i + r_i} \right)$ dieser gewünschten Darstellung nicht entspricht.

Holen wir dies mit dem oben beschriebenen Bierman-Algorithmus nach, so ergibt sich

$$D_{i-1} - \frac{D_{i-1} s_i s_i^T D_{i-1}}{r_i + s_i^T D_{i-1} s_i} = \bar{U}_i \bar{D}_i \bar{U}_i^T.$$

Danach wird die neue obere Dreiecksmatrix mit Hilfe von

$$U_i = \bar{U}_{i-1} \bar{D}_i$$

berechnet.

Schließlich haben wir eine wünschenswerte Darstellung der Schätzfehlerkovarianzmatrix mittels numerischer Verfahren gefunden.

Die Kalman-Matrix sowie der Zustandsvektor ergeben sich dann unter Einsetzen der Fehlerkovarianzmatrix durch einfache Rechnung.

Das folgende Schaubild visualisiert die Schritte und verschafft einen Überblick über die bisherigen Ausführungen zu den Rechenschritten im EKF.

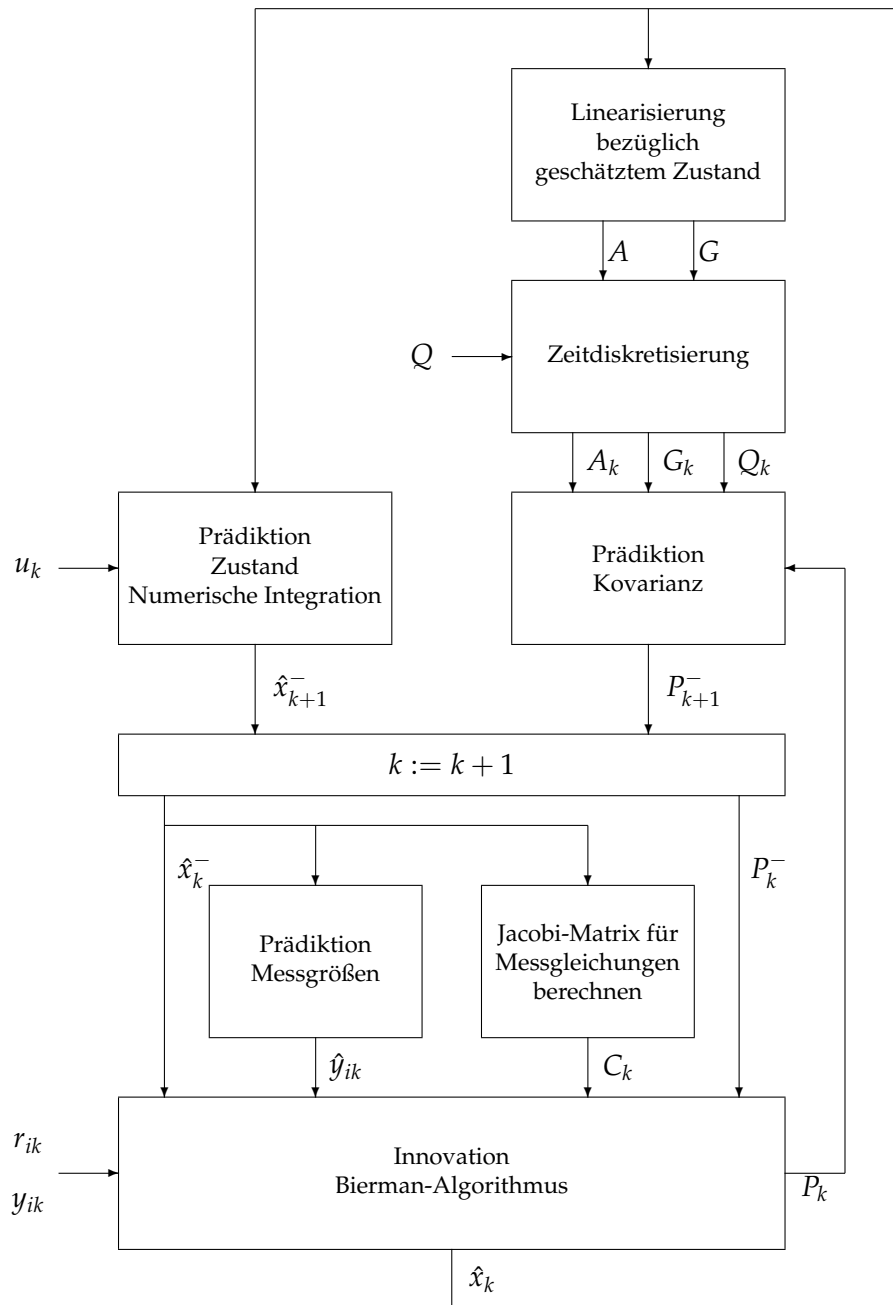


Abbildung 6.1.: Rechenschritte für das EKF

Kapitel 7

Numerische Experimente

In diesem Kapitel werden wir mit Hilfe einer Implementierung die Konvergenz des EKF aus Kapitel 6 untersuchen.

Hierfür betrachten wir einige ausgewählte Fahrmanöver und ziehen dazu das Differentialgleichungssystem (5.1) heran. Mit dem EKF werden wir dann bestimmte Größen des Fahrzeugs schätzen, die dessen Zustand bei diesen Fahrten beschreiben.

Außerdem werden wir den Einzugsbereich für die Fahrmanöver näherungsweise bestimmen. Das heißt, wir wollen die Frage beantworten, wie groß die Differenz zwischen Anfangsdaten und Referenzgrößen höchstens sein darf, damit das EKF überhaupt konvergiert. Insbesondere handelt es sich hierbei also um die Frage, wie groß der Schätzfehler zum Startzeitpunkt maximal sein darf.

Die für dieses Kapitel verwendete Literatur beschränkt sich auf den bereits mit [Kal09] zitierten internen technischen Bericht der DAIMLER AG, aus dem die Form der 5-dimensionalen Differentialgleichung entnommen wurde.

Alle Fahrmanöver sind selbst konstruiert und die numerischen Experimente in Eigenarbeit durchgeführt worden, sodass die Ergebnisse diesbezüglich erstmals in dieser Arbeit auftreten.

Konkret sind die zu schätzenden Größen, wie bereits in Kapitel 5 aufgeführt, durch

Nickwinkel Θ

Wankwinkel Φ

Längsgeschwindigkeit v_x

Quergeschwindigkeit v_y

Vertikalgeschwindigkeit v_z

beschrieben.

Folgende Grafik¹⁹ zeigt die Drehraten sowie die Beschleunigungen bezüglich der fahrzeugfesten Koordinaten und sorgt für ein besseres Verständnis der Fahrzeugbewegung, die mittels der genannten fünf Größen beschrieben wird:

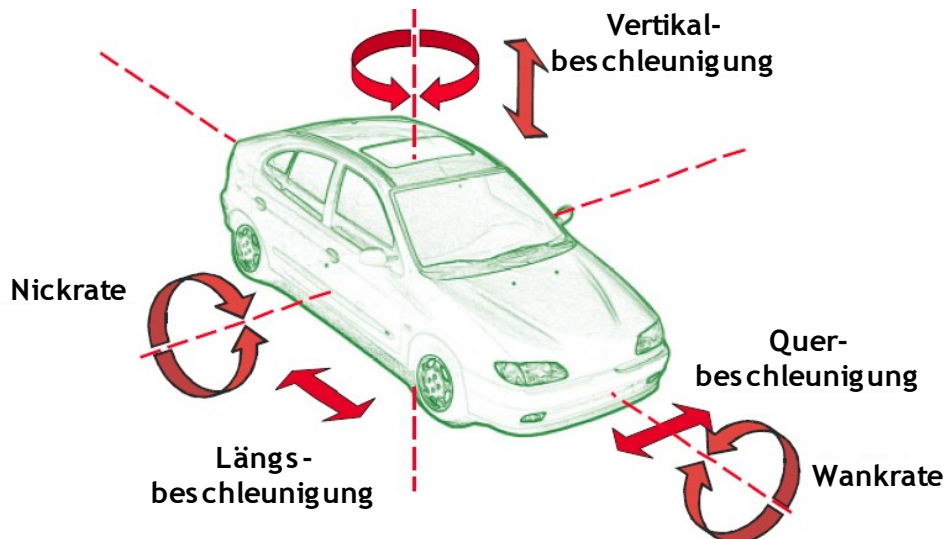


Abbildung 7.1.: Visualisierung der Fahrzeugbewegung durch Gierraten und Beschleunigungen

Die angegebenen Lagewinkel beziehen sich auf das fahrzeugfeste Koordinatensystem und sind durch Eulerwinkel beschrieben.

Mit dem EKF wollen wir also die fünf oben genannten Größen bestimmen. Dazu legen wir folgende Beobachterzustände fest, die wir dann mit Hilfe der Differentialgleichung (5.1) lösen:

$$x_1 = \sin(\Theta)$$

$$x_2 = \sin(\Phi)\cos(\Theta)$$

$$x_3 = v_x$$

$$x_4 = v_y$$

$$x_5 = v_z$$

Es ist anzumerken, dass die Referenzgeschwindigkeit des Fahrzeugs während der Fahrmanöver konstant bleibt.

¹⁹Abbildungsquelle: [Kal09].

Außerdem beschränken wir uns in unseren Experimenten auf das Zeitintervall $T = [0; 100]$ in Sekunden.

Die Zeitschrittweite beträgt $\tau = 0.01$ sec.

Die numerischen Verfahren, die zur Realisierung des EKF implementiert werden, wurden bereits in Kapitel 6.3 aufgeführt, sodass sie an dieser Stelle nicht mehr beschrieben werden, sondern nur auf ebendieses Kapitel verwiesen wird.

7.1. Geradeausfahrt mit positiver Steigung

Das erste Fahrmanöver, das wir untersuchen werden, ist eine Geradeausfahrt mit $m = 10\%$ Steigung. Das Fahrzeug fährt mit der Geschwindigkeit $v_x = 10m/s$ beziehungsweise $v_x = 36km/h$.

Welche Größen nehmen bei dieser Fahrt überhaupt einen Wert an und welche Gleichungen ergeben sich für diese Fahrt?

Für die Drehraten gilt:

$$\omega_x = \omega_y = \omega_z = 0$$

Auf Grund der Geradeausfahrt, wankt das Fahrzeug nicht nach links oder rechts, sodass für den Wankwinkel ebenfalls gilt: $\Phi = 0$. Dann:

$$\Phi = 0 \implies x_2 = 0$$

Für den Nickwinkel Θ hingegen gilt aus x_1 :

$$x_1 = \sin(\text{atan}(0.1))$$

Aus der Differentialgleichung (5.1), die uns zu Grunde liegt, gilt für die Beschleunigungen nach Einsetzen der bekannten Parameter:

$$a_x = -9.81 \cdot x_1 \quad (\text{aus } \dot{x}_3)$$

$$a_y = 0 \quad (\text{aus } \dot{x}_4)$$

$$a_z = 9.81 \sqrt{1 - x_1^2} \quad (\text{aus } \dot{x}_5)$$

Um zu sehen, dass das EKF gegen den Zustand des Fahrzeugs bei dieser Fahrsituation konvergiert, werden wir die Anfangsdaten bei der Implementierung stören. Insbesondere orientieren wir uns an den Referenzdaten $v_x = 10m/s$ und $x_1 = \sin(\text{atan}(0.1))$ und geben

von diesen verschiedene Anfangsdaten ein. Es ist zu erwarten, dass je größer der Fehler zum Startzeitpunkt ist, das EKF mehr Zeit benötigt, um ein Ergebnis in einem vom Anwender festgelegten Toleranzbereich zu liefern.

Die Steigung bezeichnen wir mit m und geben diese in Prozent an.

Den trivialen Fall, das heißt die Eingabe der korrekten Anfangsdaten, lassen wir aussen vor.

Nachfolgend sehen wir, wie sich das EKF bei Veränderung der Längsgeschwindigkeit sowie der Steigung verhält.

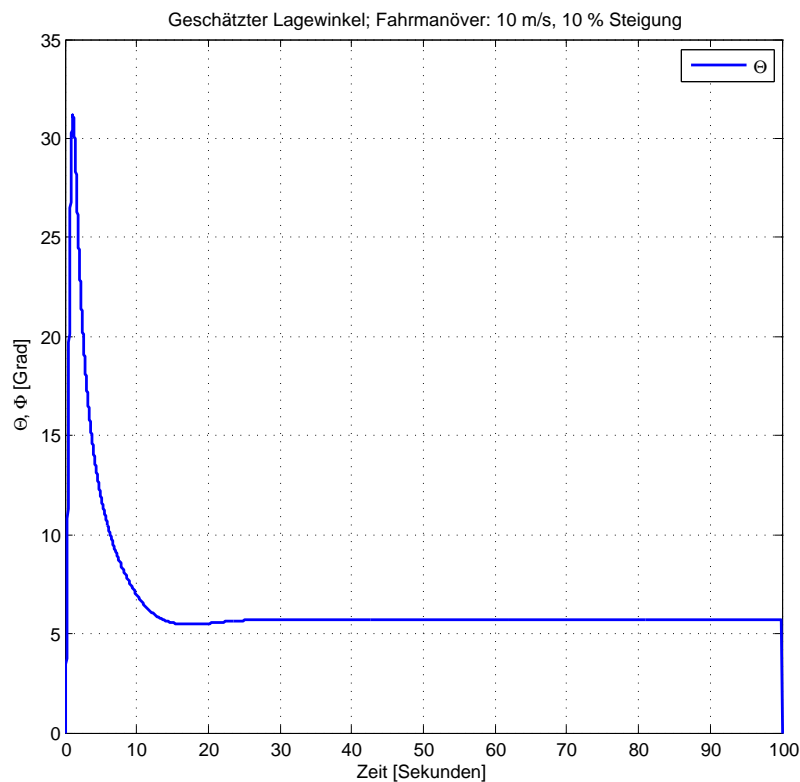


Abbildung 7.2.: Schätzung der Lagewinkel durch das EKF bei Startdaten

$$v_x = 0 \text{ m/s}, m = 0\%$$

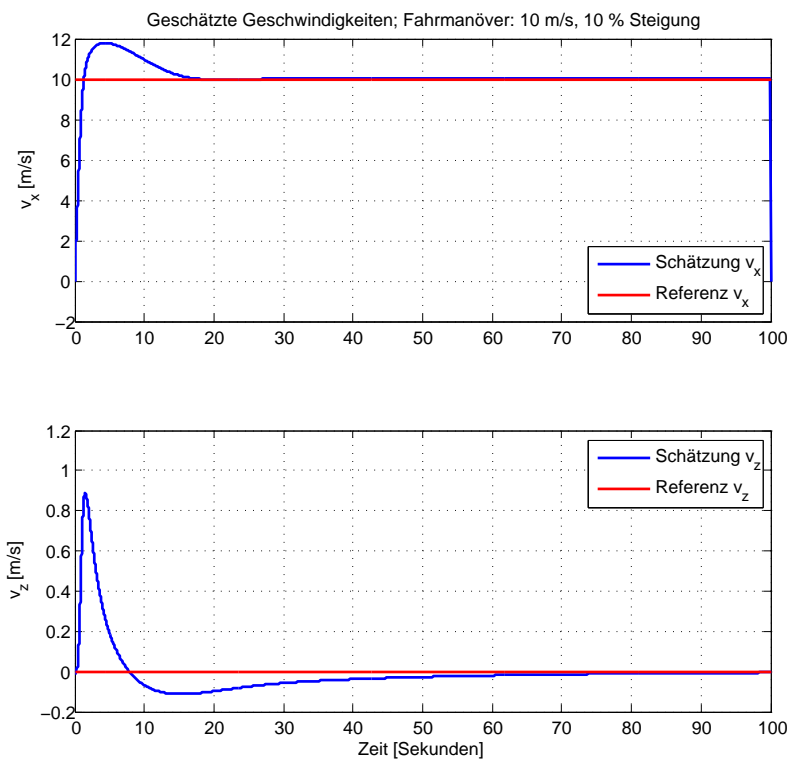


Abbildung 7.3.: Schätzung der Geschwindigkeiten durch das EKF bei Startdaten $v_x = 0\text{ m/s}$, $m = 0\%$

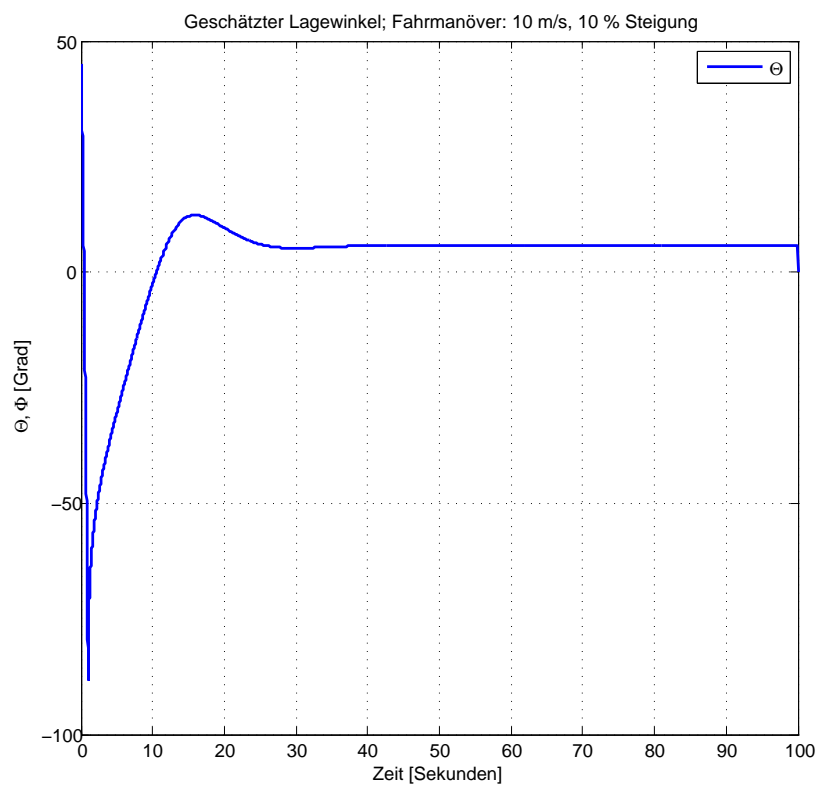


Abbildung 7.4.: Schätzung der Lagewinkel durch das EKF bei Startdaten
 $v_x = 42m/s, m = 100\%$

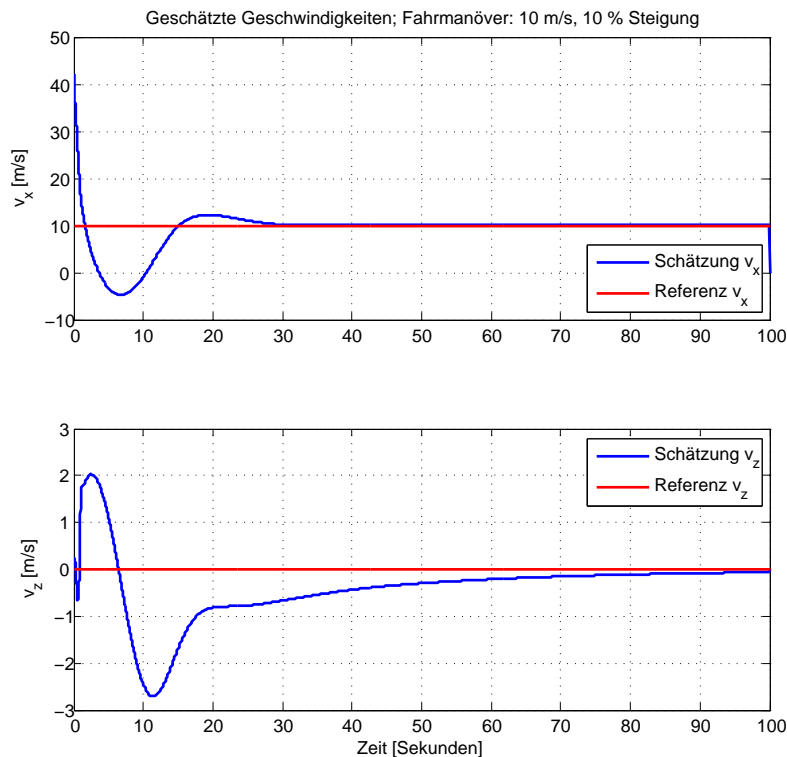


Abbildung 7.5.: Schätzung der Geschwindigkeiten durch das EKF bei Startdaten $v_x = 42\text{m/s}$, $m = 100\%$

Wir sehen, dass der Nickwinkel Θ bei diesem Fahrmanöver auf ca. $\Theta = 6^\circ$ geschätzt wird. Die Referenzgrößen von Längsgeschwindigkeit sowie Vertikalgeschwindigkeit werden vom EKF innerhalb eines Toleranzbereichs geschätzt, sodass wir hier von Konvergenz sprechen.

Ein Vergleich der jeweiligen Kurvenverläufe in den Grafiken zeigt, dass sich das EKF den Referenzgrößen langsamer nähert, je schlechter die Anfangsdaten als Schätzwerte gewählt werden. Die Startdaten, die von den Referenzdaten betragsmäßig weiter weg liegen als im ersten Fall, beeinflussen den Schätzalgorithmus noch stärker in negativer Weise.

Auffällig ist hier der Einfluss der Anfangsdaten auf die Vertikalgeschwindigkeit v_z . Zu Beginn des Schätzalgorithmus befindet sie sich bei 0m/s und konvergiert schließlich auch gegen diesen Wert. Jedoch schlägt sie kurz nach Beginn des Algorithmus sowohl in die positive, als auch in die negative Richtung aus. Dies ist darauf zurückzuführen, dass die Vertikalgeschwindigkeit bei diesem Fahrmanöver zwar Null ist; sie hängt aber, wie in (5.1) einzusehen ist, von der Steigung ab, die wiederum den Referenzwert zu Beginn des Algorithmus nicht annimmt. Erst mit der Konvergenz des EKF gegen den Lagewinkel, konvergiert auch die Schätzung der Vertikalgeschwindigkeit gegen ihren Referenzwert. Weiter sehen wir, dass die Vertikalgeschwindigkeit betragsmäßig desto größere Werte

annimmt, je größer die Differenz zwischen Anfangsdaten und Referenzgrößen ist. Für alle drei Zustandsgrößen ist klar erkennbar, dass die Konvergenzgeschwindigkeit des EKF deutlich von der Wahl der Anfangsdaten abhängt: Je schlechter diese gewählt werden, desto langsamer konvergiert es, das bedeutet, desto mehr Zeit benötigt es, um sich einer Lösung zu nähern.

Bei Betrachtung der beiden Peaks in den Grafiken, die den Lagewinkel veranschaulichen, sehen wir, dass auch dieser desto weiter ausschlägt, je schlechter die Anfangsdaten gewählt werden. In der Schätzung des Lagewinkels in Abbildung 7.4 sehen wir, dass der Peak nahezu bei -90° liegt. Auf Grund der Nichteindeutigkeit der Eulerwinkel würde eine Überschreitung dieses Wertes zu Divergenz führen. Der Algorithmus wäre dann von der Referenzgröße „zu weit weg“ und könnte sich nicht mehr „einpendeln“.

Wir sehen an der folgenden Abbildung ein Beispiel für Divergenz des EKF:

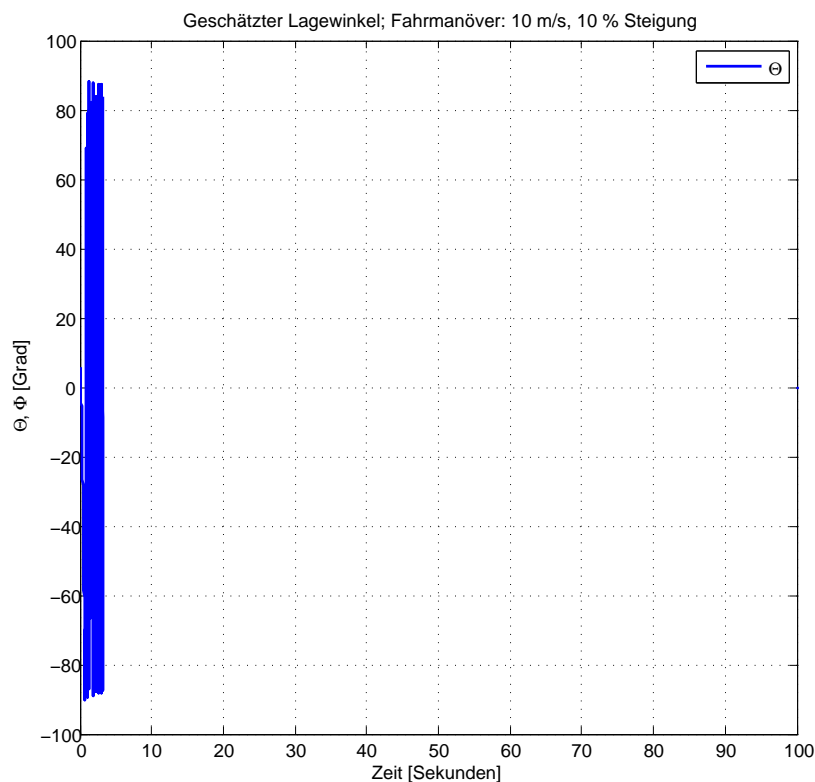


Abbildung 7.6.: Schätzung der Lagewinkel durch das EKF bei Startdaten
 $v_x = 40\text{m/s}$, $m = 10\%$

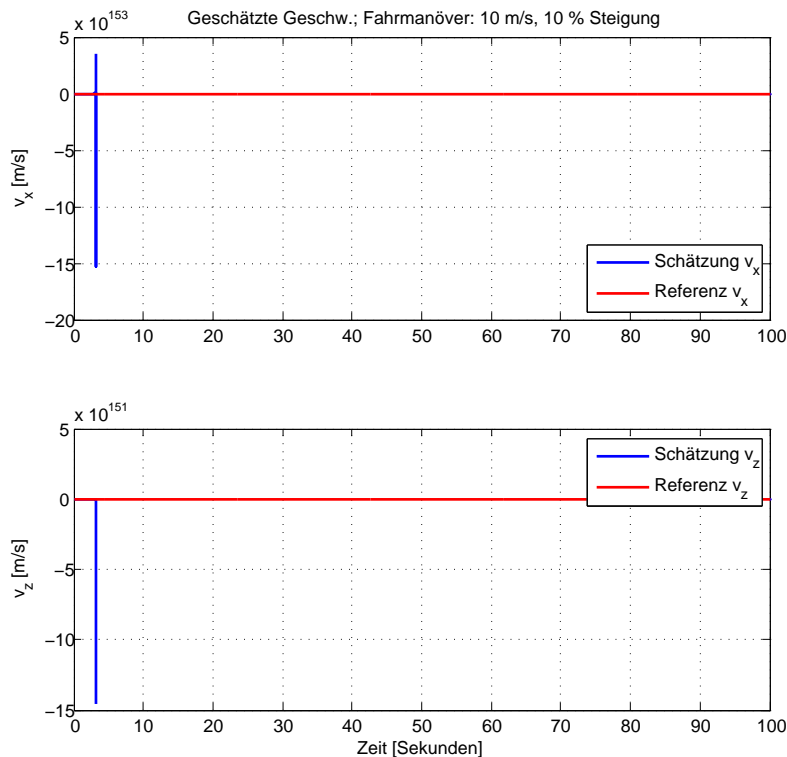


Abbildung 7.7.: Schätzung der Geschwindigkeiten durch das EKF bei Startdaten
 $v_x = 40\text{m/s}$, $m = 10\%$

Obwohl die Steigung gleich zu Beginn den Referenzwert von 10% hat, divergiert das EKF bereits bei einer Geschwindigkeit von ca. 144km/h . Wir sehen ganz deutlich, dass die Schätzung für den Lagewinkel auf Grund der hohen Fahrgeschwindigkeit stark oszilliert und sich nicht auf den Referenzwert einstellen kann.

Die Schlussfolgerung, dass das EKF selbst dann nicht in jedem Fall konvergiert, wenn eine der beiden Größen bereits zu Beginn der Referenzgröße entspricht, ist an dieser Stelle voreilig und falsch. Denn drehen wir die Übereinstimmung mit den echten Daten um und wählen nun $v_x = 10\text{m/s}$ als Startgeschwindigkeit und eine beliebige Steigung $m \in [0, \infty)$ in Prozent als Anfangswert, so konvergiert das EKF für beliebiges m . Würde m eine für die Praxis unrealistische und für den Anwender eine auf Grund ihrer Größe unvorstellbare Zahl annehmen, so würde das EKF dennoch konvergieren. Der einzige Fall der Divergenz bei richtiger Wahl der Anfangsgeschwindigkeit, der jedoch nur von theoretischer Natur ist, ist der einer Steilwand, das heißt, wenn die Steigung $m = \infty$ wäre.

Es ist also möglich, an mehreren Stellen Größen zu verändern und dennoch Konvergenz des EKF erwarten zu können.

An dieser Stelle tritt eine interessante Frage auf: Wie stark und in welche Richtung können

wir die Anfangsdaten verändern beziehungsweise wie weit können wir uns mit den Startdaten von den Referenzgrößen weg bewegen, um überhaupt noch Konvergenz des EKF erwarten zu können?

Wir wollen also die Frage nach dem Einzugsgebiet der frei zu wählenden Anfangsdaten an der folgenden Abbildung anschaulich machen. Die obere sowie die untere Schranke der Startdaten sind blau markiert und alle Werte dazwischen, die als Startwerte gewählt werden können, sind in roter Schraffur dargestellt:

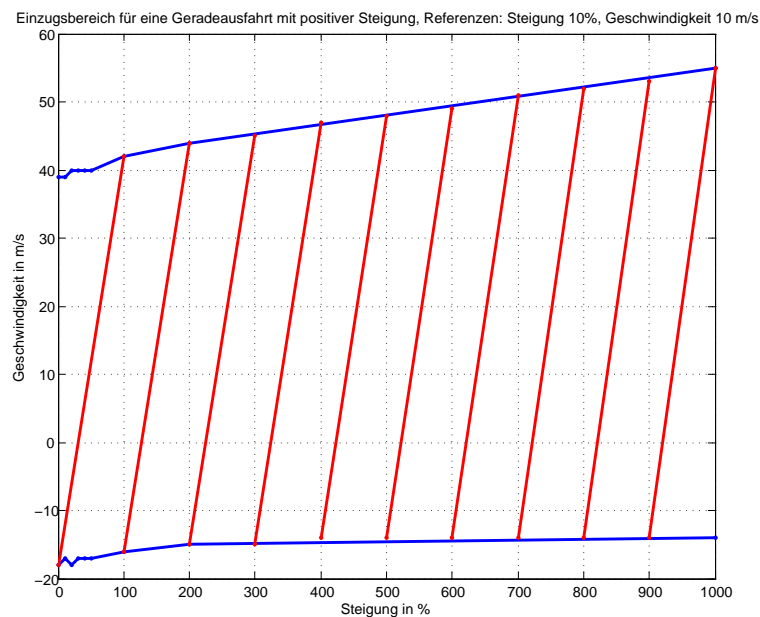


Abbildung 7.8.: Einzugsbereich der Anfangsdaten

Die Länge des Intervalls, das die untere und die obere Grenze bezüglich der Geschwindigkeit bei wachsender Steigung ausmacht, beträgt zu Beginn 57m/s und ist bei der Steigung $m = 1000\%$ auf 70m/s gestiegen. Jedoch wird dieses als positiv zu bewertende Ergebnis dadurch eingebüßt, dass der Algorithmus bei wachsender Steigung zum Startzeitpunkt deutlich länger rechnet, also mehr Zeit benötigt, um ein Ergebnis zu liefern.

Für wachsende Steigung ist demnach ein linearer Verlauf der unteren sowie der oberen Schranke der Geschwindigkeiten zu erwarten.

Für die Praxis sind Steigungen von bis zu 100% realistisch und die Analyse daher bis zu dieser Grenze ausreichend. Hier sehen wir, dass das Geschwindigkeitsintervall beim vorliegenden Fahrmanöver ca. 30m/s nach oben und -27m/s nach unten beträgt. Wir haben also einen Einzugsbereich von ca. 100km/h von der Referenzgröße weg, und zwar in beide Richtungen, was durchaus beeindruckend ist und auf Robustheit des EKF für dieses Fahrmanöver schließt.

Im nächsten Kapitel werden wir weitere Schätzgrößen hinzunehmen, indem wir unser Fahrzeug im Kreis fahren lassen.

7.2. Die Kreisfahrt

Es liegt nun folgendes Fahrmanöver vor: Ein Fahrzeug fährt mit $v_t = 28m/s$, das heißt mit ca. $100km/h$ in einem Kreis mit $r = 80m$ Radius gegen den Uhrzeigersinn. Das Zeitintervall, welches wir betrachten ist $T = [0; 100]$ in Sekunden.

Die Vorgehensweise ist dieselbe wie im vorigen Abschnitt. Der Fokus liegt auch hier auf der Konvergenz des implementierten EKF für diese Fahrsituation und der Frage, für welche Bedingungen dieses divergiert. Auch wollen wir einen Einzugsbereich der Anfangsdaten bestimmen.

Vorab jedoch werden wir festhalten, welche Größen eine Rolle spielen und welche bei diesem Manöver den Wert Null annehmen.

Für die Tangentialgeschwindigkeit gilt

$$v_t = \sqrt{v_x^2 + v_y^2}.$$

Wichtig ist der Gedanke, dass es sich bei v_t nicht um die Geschwindigkeit des Fahrzeugs in x -Richtung, sondern um die Tangentialgeschwindigkeit entlang des Kreises handelt. Es ist zu erwarten, dass sich die Längsgeschwindigkeit des Fahrzeugs der Tangentialgeschwindigkeit annähert, je größer der Radius des Kreises ist und sich betragsmäßig von ihr entfernt, je kleiner dieser wird.

Für die Drehraten gilt

$$w_x = w_y = 0.$$

Das Fahrzeug bewegt sich mit seiner „Nase“ nicht nach oben oder nach unten, das heißt es hat keinen Nickwinkel, damit ist

$$\Theta = 0 \implies x_1 = 0.$$

Ebenso fährt es nicht auf einer geneigten Fahrbahn, weshalb für den Wankwinkel ebenfalls gilt:

$$\Phi = 0 \implies x_2 = 0$$

Setzen wir die Kenntnis um die Drehraten und um die Winkel in die Differentialgleichung (5.1) ein, so erhalten wir für die Beschleunigungen:

$$a_x = -\omega_z v_y$$

$$a_y = \omega_z v_x$$

$$a_z = 9.81$$

Das Fahrzeug erfährt nur die Erdbeschleunigung in z-Richtung und es hat keine Vertikalgeschwindigkeit.

Wie sind jedoch die anderen beiden Geschwindigkeiten in diesem Fahrmanöver zu modellieren?

Aus Längs- und Quergeschwindigkeit ergibt sich der Schwimmwinkel β in Grad, der die Richtung der Fahrzeugbewegung bezüglich seiner Längsachse beschreibt. Führt das Fahrzeug im Kreis, so stellt sich ein Schwimmwinkel ein, der durch folgende Gleichung gegeben ist:

$$\beta = \frac{l_H}{r} - SG \frac{v_t^2}{r}$$

l_H bezeichnet den halbierten Abstand von Vorder- zu Hinterachse, die bei unserem Fahrzeug $l_H = 3.165m$ beträgt. Daher liegt der Nullpunkt des körperfesten Koordinatensystems an der Stelle

$$l_h = \frac{3.165}{2}m = 1.5825m.$$

Den Schwimmwinkelgradienten, der eine Kenngröße des Fahrzeugs ist, bezeichnen wir hier mit SG . Er lässt sich durch

$$SG = \frac{0.2\pi}{180}$$

berechnen.

Für die Längsgeschwindigkeit und für die Quergeschwindigkeit des Fahrzeugs ergeben sich mit dem Schwimmwinkel folgende Gleichungen:

$$x_3 = v_x = v_t \cos(\beta)$$

$$x_4 = v_y = v_t \sin(\beta)$$

Setzen wir die uns nun bekannten Größen in die entsprechenden Gleichungen ein und implementieren das EKF für dieses Fahrmanöver, so erhalten wir unter beliebiger Wahl der Anfangsdaten folgendes Verhalten des EKF:

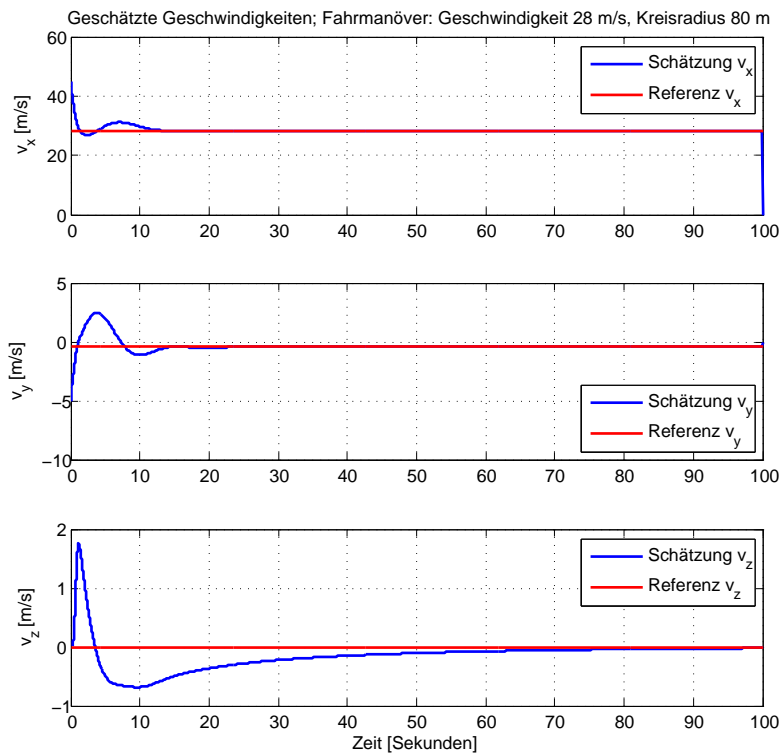


Abbildung 7.9.: Konvergenz bei Anfangsdaten $v_t = 45\text{m/s}$, $r = 50\text{m}$

Die Grafik zeigt, dass das EKF bei den Anfangsdaten $v_t = 45\text{m/s}$, $r = 50\text{m}$ konvergiert. Die Längsgeschwindigkeit kennen wir bereits, sie wird auf etwa $v_x = 28\text{m/s}$ geschätzt. Für die Vertikalgeschwindigkeit ergibt sich $v_y = -0.4039\text{m/s}$, das heißt das Fahrzeug hat eine Geschwindigkeit aus dem Kreis heraus; in unserem Fall, da unser Fahrzeug gegen den Uhrzeigersinn fährt, nach rechts.

Auch hier ist auffällig, dass das Fahrzeug keine Vertikalgeschwindigkeit hat. Das EKF schätzt jedoch eine solche im Verlauf des Zeitintervalls, bis es schließlich gegen den richtigen Wert, also gegen Null konvergiert. Es ist deutlich erkennbar, dass sich die geschätzte Lösung der wahren nur sehr langsam nähert und nahezu das gesamte Zeitintervall benötigt, um zu konvergieren.

Aus (5.1) können wir die Begründung entnehmen. Die Beschleunigung a_z beeinflusst die Vertikalgeschwindigkeit. Weiter wirken die anderen beiden Geschwindigkeiten und, wie wir im Anschluss sehen werden, auch die Lagewinkel auf v_z ein, da sie zunächst mit einer betragsmäßig großen Differenz zum Referenzwert geschätzt werden. Erst wenn diese sich dem Grenzwert nähern, nähert sich auch die Vertikalgeschwindigkeit der Null.

Ein weiteres interessantes Phänomen sind die Winkel.

Wir haben bereits festgestellt, dass das Fahrzeug bei diesem Fahrmanöver weder einen

Nick-, noch einen Wankwinkel hat.

Lassen wir jedoch unser Fahrzeug in unserem Programm fahren, so erhalten wir als Schätzung des EKF:

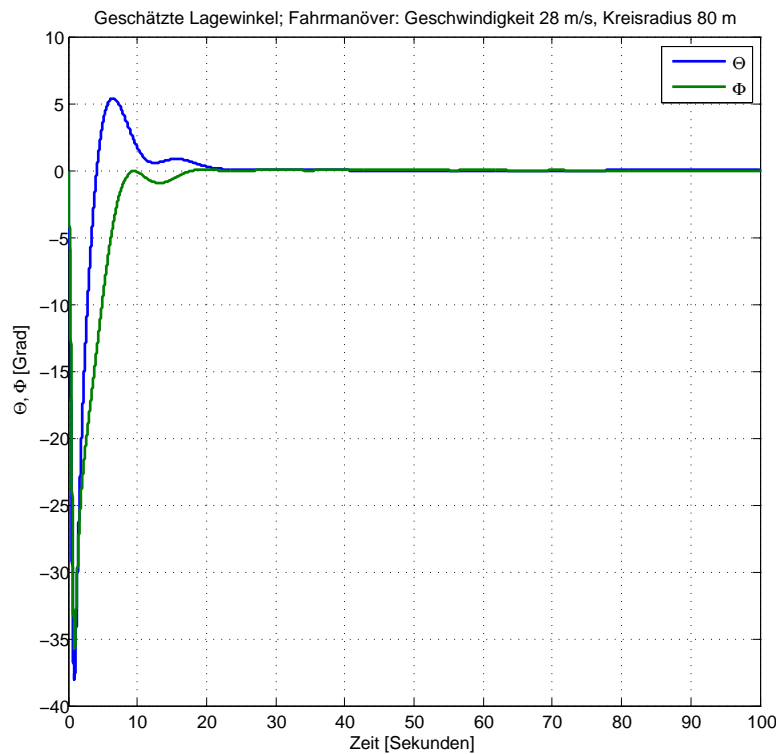


Abbildung 7.10.: Konvergenz bei Anfangsdaten $v_t = 45\text{m/s}$, $r = 50\text{m}$

Wie auch schon die Vertikalgeschwindigkeit, stellen sich beim EKF Lagewinkel ein, die jedoch im Lauf der Zeit gegen Null konvergieren und das EKF diesen Wert als Ergebnis liefert. Der Grund dafür ist der sich in jedem Zeitschritt ändernde Zustand des Fahrzeugs. Je nachdem wie die Geschwindigkeiten geschätzt werden, richten sich danach auch die Lagewinkel aus. Wie oben schon erwähnt wurde, wirken diese, wenn sie von Null verschiedene Werte annehmen, auf die anderen Größen. Insbesondere ist dies mit ein Grund dafür, dass die Vertikalgeschwindigkeit nicht konstant auf Null bleibt.

Gegen Ende des Zeitintervalls, wenn das EKF die Zustandsgrößen schätzt, die sich den Referenzgrößen immer weiter annähern, schätzt es dann auch die Lagewinkel auf ihre Referenzgrößen.

In einem nächsten Experiment erhöhen wir nun die Anfangsdaten, die wir dem Algorithmus für den Startzeitpunkt liefern, und legen einen Radius von $r = 2000\text{m}$ fest; für die

Tangentialgeschwindigkeit geben wir $v_t = 55\text{ m/s}$ vor.
Dies ergibt

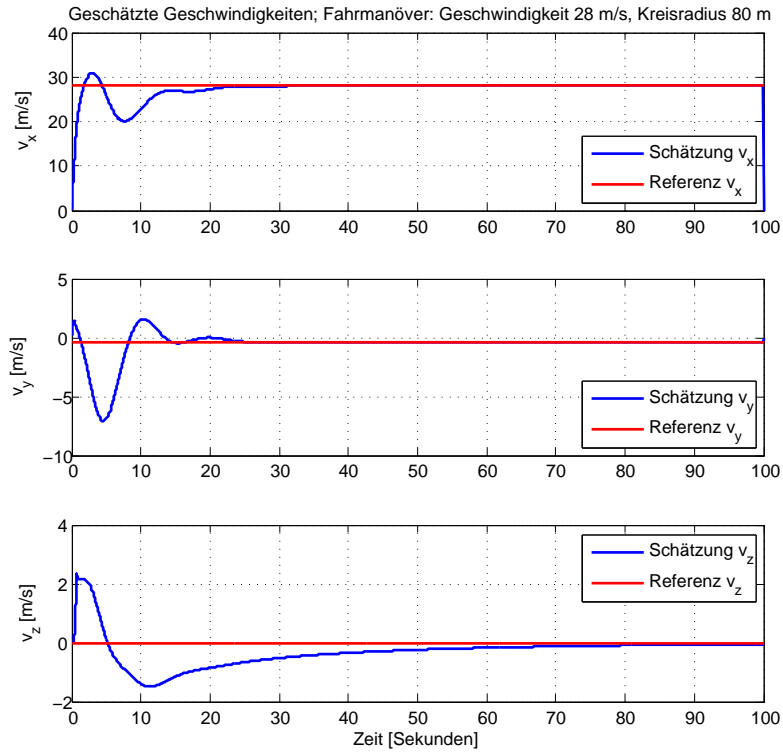


Abbildung 7.11.: Konvergenz bei Anfangsdaten $v_t = 0\text{ m/s}$, $r = 2000\text{ m}$

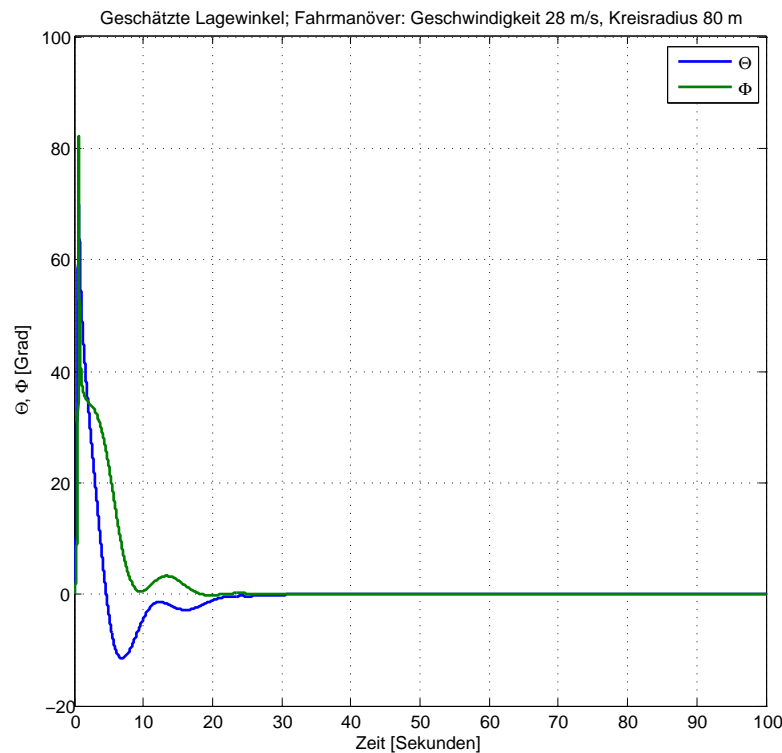


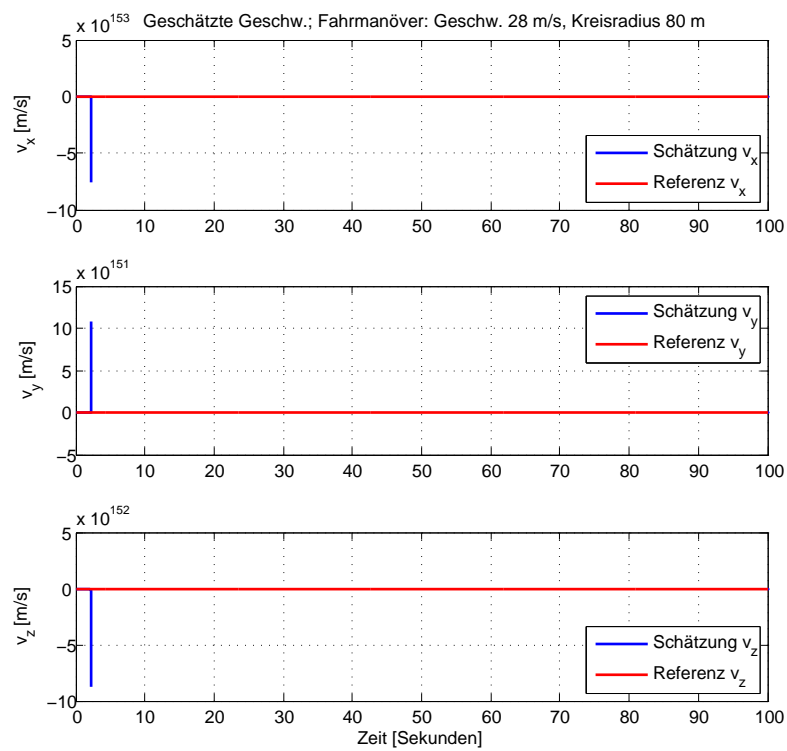
Abbildung 7.12.: Konvergenz bei Anfangsdaten $v_t = 0\text{m/s}$, $r = 2000\text{m}$

Auch die Lagewinkel verändern im Laufe des Schätzalgorithmus ihren Wert, bis sie sich dann wieder auf die Referenzgrößen „einpendeln“.

Außerdem ist wie auch aus den Schätzergebnissen bei der Geradeausfahrt mit Steigung aus Kapitel 7.1 ersichtlich, dass sich das EKF langsamer an den Referenzwert nähert, je weiter entfernt sich die Anfangsdaten betragsmäßig von der Lösung der Differentialgleichung, die dem Referenzwert entspricht, befinden. Bei einer solchen Wahl sehen wir auch, dass die Schätzungen der Geschwindigkeiten und auch der Lagewinkel immer stärker oszillieren, bis sie schließlich konvergieren.

Es gibt jedoch Fälle, in denen die Schätzungen so sehr oszillieren, dass keine Konvergenz zu erwarten ist. Der Peak in Abbildung 7.12 deutet darauf hin, dass sich dieses Fahrmanöver mit den entsprechenden Anfangsdaten an der Grenze des möglichen Einzugsbereichs für Konvergenz befindet.

Wir sehen im Folgenden ein Beispiel für Divergenz, bei der Nick- wie auch Wankwinkel betragsmäßig die 90° -Schranke überschreiten:

Abbildung 7.13.: Divergenz bei Anfangsdaten $v_t = 57\text{m/s}$, $r = 100\text{m}$

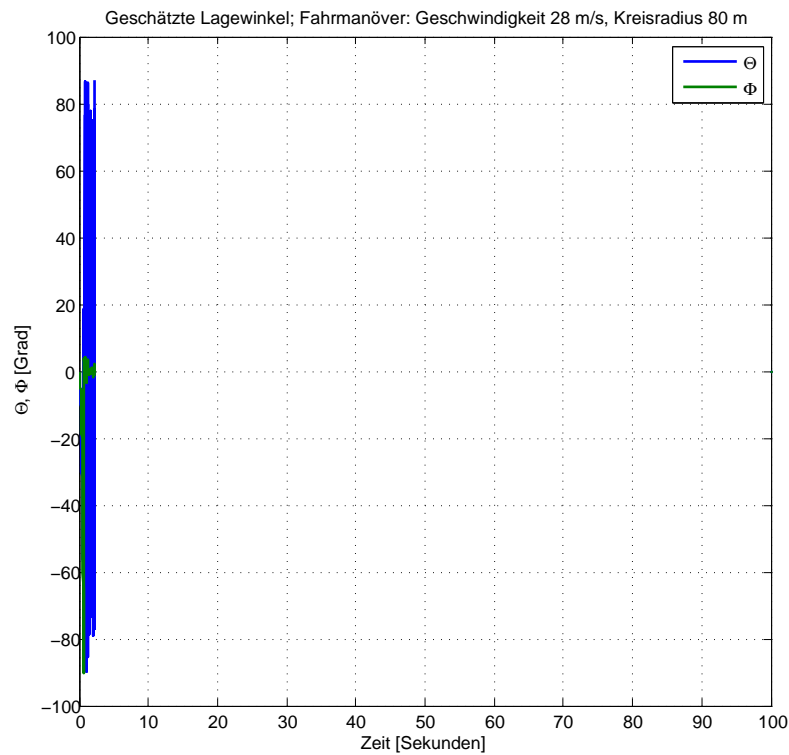


Abbildung 7.14.: Divergenz bei Anfangsdaten $v_t = 57\text{m/s}$, $r = 100\text{m}$

Der Nickwinkel Θ nimmt den Wert von -90° an. Dies ist der Grund dafür, dass das EKF kein Schätzergebnis unter gegebenen Daten zum Zeitpunkt Null liefert.

Um dieses Problem zu umgehen, fragen wir uns nun, in welchen Fällen das EKF konvergiert, das heißt Schätzergebnisse liefert?

Dies ist die Frage nach dem Einzugsgebiet der Anfangsdaten. Mehrere numerische Experimente führen auf den folgenden Bereich:

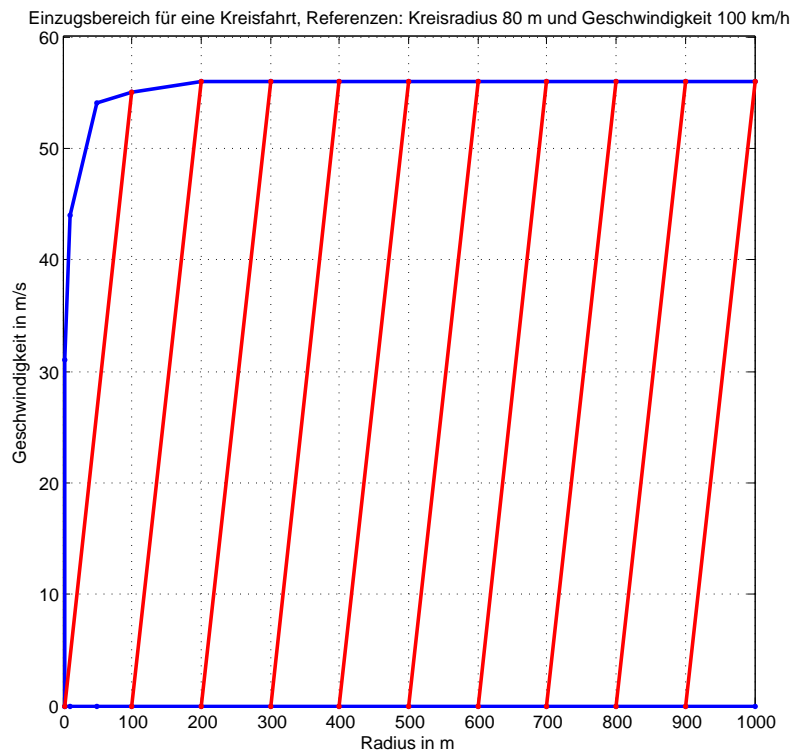


Abbildung 7.15.: Einzugsbereich für die Kreisfahrt mit Kreisradius $80m$ und Tangentialgeschwindigkeit $28m/s$

Dabei sind die obere sowie die untere Schranke der Startdaten blau und alle Werte dazwischen, die als Startwerte gewählt werden können, rot schraffiert visualisiert.

Der Einzugsbereich ist zu Beginn sehr gering. Das EKF konvergiert überhaupt erst ab den Anfangsdaten:

$$v_t = 31m/s, \quad r = 2m$$

Das Geschwindigkeitsintervall, in dem wir Konvergenz erwarten können, bleibt bis zum Kreisradius von etwa $10m$ gering, bis es dann jedoch ab einem Kreisradius von $50m$ schon bei etwa $54m/s$ liegt und dann ab dem Radius von $200m$ für alle größeren Radien konstant bei $56m/s$ bleibt. Das bedeutet, für einen Kreisradius von mehr als $200m$ ist unser Einzugsgebiet bezüglich der Geschwindigkeit bei $[0;56]$ in m/s . Folglich können wir bei diesem Fahrmanöver die Anfangsdaten als etwa $28m/s$ unterhalb der Referenzgröße und oberhalb ebenfalls von etwa $28m/s$ wählen. Das bedeutet einen „Spielraum“ von etwa 100 km/h in beide Richtungen.

Dieses Ergebnis ist erstaunlich und lässt Schlüsse auf die Robustheit des EKF bei diesem Fahrmanöver gegenüber schlechten Anfangsdaten ziehen.

Kapitel 8

Zusammenfassung

Das Kalman-Bucy-Filter ist ein iterativer Algorithmus, der die Lösung eines Differentialgleichungssystems schätzt. Damit stellt er eine Alternative zu den in der mathematischen Anwendung häufig verwendeten Verfahren dar.

Die vorliegende Arbeit hat die Konvergenz des Kalman-Bucy-Filters als zentrales Thema. Dabei geht es hier nicht nur um den Beweis seiner Konvergenz, sondern auch um die Frage, unter welchen Voraussetzungen es eine Lösung für gegebene Differentialgleichung liefert. Insbesondere charakterisieren die zu erfüllenden Bedingungen die Form derjenigen Differentialgleichungen, auf die das Kalman-Bucy-Filter angewandt werden kann. In dieser Arbeit haben wir uns auf die Untersuchung der Kalman-Bucy-Filters für gewöhnliche Differentialgleichungen fokussiert.

Dabei haben wir zunächst geklärt, was unter einem (dynamischen) System zu verstehen ist, wie es mit seiner Umwelt kommuniziert und wie dieses Verhältnis anhand eines Zustandsraummodells dargestellt werden kann. Weiter haben wir uns mit einem Beobachtermodell beschäftigt, mit dem es möglich ist, das Verhalten des Systems zu beschreiben. Danach haben wir die mathematischen Hintergründe und Grundlagen als Vorarbeit für die nachfolgenden Kapitel erläutert. Wir haben gezeigt, wann es möglich ist, einen Beobachter an ein System anzusetzen. Auch haben wir gesehen, dass es diverse Formen und Definitionen der Beobachtbarkeit eines Systems gibt. Dabei war die uniforme Beobachtbarkeit für diese Arbeit entscheidend.

Nachdem wir dann den Stabilitätsbegriff im Sinne von Ljapunov thematisiert hatten, haben wir das Kalman-Bucy-Filter für lineare Systeme hergeleitet und seine Werkzeuge, die Zustandsschätzung, die Kovarianzmatrix für Aussagen über die Güte der Schätzung und die für dieses Filter charakteristische Kalman-Matrix erklärt. In diesem Kapitel ist es gelungen, den Beweis der Konvergenz des Filters für lineare Systeme auf einem neuen

Weg zu führen.

Von größerem Interesse ist jedoch der Beweis für das Erweiterte Kalman-Bucy-Filter, da es Lösungen für nichtlineare Systeme, die in der Anwendung häufiger vorkommen, schätzt. Im vierten Kapitel haben wir den von KRENER bereits bestehenden Beweis ausgearbeitet und an mancher Stelle modifiziert.

Danach ist es in zwei Anwendungsbeispielen gelungen zu zeigen, dass das EKF für ein gegebenes 5-dimensionales und auch 2-dimensionales System konvergiert. Für das 2-dimensionale System konnten wir sogar ein Gebiet angeben, in dem sich die Schätzfehlerkovarianz befinden muss, um eine Lösung für diese Differentialgleichung zu erhalten. Vorbereitend für das letzte Kapitel dieser Arbeit, wurde die zeitdiskrete Form des EKF erklärt. Es unterscheidet sich vom zeitkontinuierlichen dadurch, dass es aus einem Prädiktions- und einem Innovationsschritt besteht. Auch die numerischen Verfahren, mit denen die Implementierung realisiert wird, wurden in diesem Kapitel vorgestellt. Schließlich haben wir einige numerische Experimente durchgeführt und damit die Praktikabilität des EKF aufgezeigt.

Von besonderer Bedeutung ist die Frage nach den Startwerten für eine Differentialgleichung. Wir haben gesehen, wie sich das EKF bei gestörten Anfangsdaten verhält und können folgern, dass es selbst mit vom realen Anfangszustand stark abweichenden Startwerten dennoch konvergiert, das heißt eine Lösung für die betrachteten Fahrmanöver liefert. Es ist uns letztlich auch gelungen, Einzugsbereiche festzulegen, die genau bestimmen, wie die Startdaten zu wählen sind, um Konvergenz des EKF zu erhalten.

Die Größe dieser Einzugsbereiche deutet darauf hin, dass das EKF eine gute Alternative für das Erhalten von Lösungen gewöhnlicher Differentialgleichungen ist.

Abschließend noch zwei Gedanken über diese Arbeit hinaus:

- Weiterführend wäre eine Untersuchung hinsichtlich alternativer numerischer Verfahren, die das implementierte EKF verwendet, interessant
- In den letzten Jahren wurden verschiedene Varianten dieses Kalman-Filters entwickelt. Eine Analyse dieser Varianten hinsichtlich ihrer Verwendung für von gewöhnlichen Differentialgleichungen verschiedene Systeme, etwa parabolische oder hyperbolische, könnte das Spektrum der hierfür verwendeten Lösungsverfahren erweitern

Anhang A

Erläuterungen zur Übergangsmatrix

Dieser Anhang beruht auf Inhalten aus [Unb02] und aus [Kar07].

Wir werden uns mit der Darstellung der Übergangsmatrix und ihren Eigenschaften auseinandersetzen.

Dazu betrachten wir zunächst die homogene Gleichung

$$\frac{dx(t)}{dt} = A(t)x(t) \quad (\text{A.1})$$

mit stetigen Einträgen in $A(t)$.

Seien e_i mit $i = 1, \dots, n$ die Einheitsvektoren und sei

$$\phi_i(t, t_0) = \begin{bmatrix} \phi_{1i}(t, t_0) \\ \phi_{2i}(t, t_0) \\ \vdots \\ \phi_{ni}(t, t_0) \end{bmatrix}$$

zusammen mit der Anfangsbedingung $\phi_i(t_0, t_0) = e_i$ die Lösung der homogenen Gleichung (A.1).

Behauptung

Die Vektoren aus $\phi_i(t, t_0)$, die Lösungen für (A.1) sind, sind zu jedem Zeitschritt t voneinander linear unabhängig.

Beweis

Angenommen, die Lösungsvektoren sind voneinander abhängig, das heißt, \exists Konstanten κ_i mit $i = 1, \dots, n$, wobei für mindestens eine Konstante $\kappa_i \neq 0$ gilt und \exists ein Zeitpunkt τ sodass

$$\sum_{i=1}^n \kappa_i \phi_i(\tau, t_0) = 0.$$

Da $\phi_i(t, t_0)$ Lösung von (A.1) ist, sind dies auch alle Linearkombinationen $\sum_{i=1}^n \kappa_i \phi_i(t, t_0)$ dieser Lösung der homogenen Gleichung.

Dann gilt aber auf Grund der Eindeutigekeit der Lösung $x(t) \equiv 0$ und damit ist die Anfangsbedingung $x(t_0) = 0$.

Durch Einsetzen der Lösung $\phi_i(t_0, t_0) = e_i$ zum Startzeitpunkt t_0 gilt also

$$\sum_{i=1}^n \kappa_i \phi_i(t_0, t_0) = \sum_{i=1}^n \kappa_i e_i = 0.$$

Es ist jedoch bekannt, dass die Einheitsvektoren e_i voneinander linear unabhängig sind. Da weiterhin zumindest ein $\kappa_i \neq 0$, muss gelten:

$$\sum_{i=1}^n \kappa_i \phi_i(t_0, t_0) = \sum_{i=1}^n \kappa_i e_i \neq 0$$

Insofern führt unsere Annahme zum Widerspruch und wir erhalten das Ergebnis der linearen Unabhängigkeit der Lösungsvektoren. \square

Folglich wissen wir, dass jede beliebige Lösung der homogenen Gleichung (A.1) als Linearkombination der Lösungen $\phi_i(t, t_0)$ ausgedrückt werden kann:

Sei $x(t)$ mit $x(t_0) = [x_1(t_0), \dots, x_n(t_0)]^T$ eine Lösung von (A.1), dann gilt für die Lösung zum Zeitpunkt t_0 :

$$x(t_0) = \sum_{i=1}^n x_i(t_0) e_i$$

und für die Lösungsfunktion

$$x(t) = \sum_{i=1}^n x_i(t_0) \phi_i(t, t_0).$$

Auf Grund der Eindeutigkeit der Lösung gilt also

$$\sum_{i=1}^n x_i(t_0)\phi_i(t, t_0) = x(t).$$

In Matrixschreibweise sieht die Lösung $x(t)$ folgendermaßen aus:

$$x(t) = \underbrace{\begin{bmatrix} \phi_{11}(t, t_0) & \dots & \phi_{1n}(t, t_0) \\ \vdots & & \vdots \\ \phi_{n1}(t, t_0) & \dots & \phi_{nn}(t, t_0) \end{bmatrix}}_{=: \Phi(t, t_0)} \cdot x(t_0)$$

$\Phi(t, t_0)$ bezeichnen wir an dieser Stelle als *Übergangsmatrix* des Lösungssystems.

Es ist klar, dass die Spalten der Übergangsmatrix die Lösungsvektoren der homogenen Gleichung (A.1) sind. Daher erfüllt $\Phi(t, t_0)$ die Matrix-Differentialgleichung

$$\frac{\partial}{\partial t} \Phi(t, t_0) = A(t)\Phi(t, t_0)$$

mit der Anfangsbedingung $\Phi(t_0, t_0) = I_n$.

Damit bilden die Lösungsvektoren eine Basis beziehungsweise ein Fundament, mit dem es möglich ist, weitere Lösungen eines gegebenen Systems anzugeben. Daher ist unsere Übergangsmatrix in der Literatur häufig unter dem Namen *Fundamentalmatrix* besser bekannt.

Im Regelfall muss diese numerisch ermittelt werden.

Sei nun die inhomogene Gleichung

$$\frac{dx(t)}{dt} = A(t)x(t) + B(t)u(t) \tag{A.2}$$

mit bekannter Anfangsbedingung $x(t_0)$ gegeben.

Wir wollen im Folgenden das Prinzip der Variation der Konstanten heranziehen, um dann unter Berücksichtigung der Lösung $x(t) = \Phi(t, t_0)x(t_0)$ der homogenen Gleichung (A.1) eine Lösung für diese inhomogene Gleichung zu finden: Betrachte zunächst

$$x(t) = \Phi(t, t_0)f(t)$$

und die entsprechende Ableitung nach der Zeit

$$\frac{dx(t)}{dt} = \frac{\partial \Phi(t, t_0)}{\partial t} f(t) + \Phi(t, t_0) \frac{df(t)}{dt}.$$

Dann gilt für die gegebene inhomogene Gleichung mit Berücksichtigung aus dem homo-

genen Fall, dass $\Phi(t, t_0)$ die Differentialgleichung $\frac{\partial \Phi(t, t_0)}{\partial t} = A(t)\Phi(t, t_0)$ löst:

$$\begin{aligned}\frac{\partial \Phi(t, t_0)}{\partial t} f(t) + \Phi(t, t_0) \frac{df(t)}{dt} &= A(t)x(t) + B(t)u(t) \\ \left(\frac{\partial \Phi(t, t_0)}{\partial t} - A(t)\Phi(t, t_0) \right) f(t) + \Phi(t, t_0) \frac{df(t)}{dt} &= B(t)u(t) \\ \frac{df(t)}{dt} &= \Phi^{-1}(t, t_0)B(t)u(t)\end{aligned}$$

Integration dieses Differentials und Einsetzen in die inhomogene Gleichung ergibt ihre Lösung

$$x(t) = \Phi(t, t_0) \cdot \int_{t_0}^t \Phi^{-1}(\sigma, t_0)B(\sigma)u(\sigma)d\sigma$$

mit $x(t_0) = 0$.

Wir führen nun die Lösungen der homogenen Gleichung mit der letztgefundenen zusammen und erhalten schließlich die Lösung für die inhomogene Gleichung (A.2):

$$x(t) = \Phi(t, t_0)x(t_0) + \Phi(t, t_0) \cdot \int_{t_0}^t \Phi^{-1}(\sigma, t_0)B(\sigma)u(\sigma)d\sigma$$

Es ist wohlbekannt, dass folgende Eigenschaft für die Inverse der Fundamentalmatrix, die wegen $\det(\Phi(t, t_0)) \neq 0$ existiert, gilt: $\Phi^{-1}(t_1, t_0) = \Phi(t_0, t_1)$, für zwei Zeitpunkte $t_0 \leq t_1$. Unter dieser Bedingung können wir die Lösung der inhomogenen Gleichung schließlich vereinfachen:

$$x(t) = \Phi(t, t_0)x(t_0) + \int_{t_0}^t \Phi(t, \sigma)B(\sigma)u(\sigma)d\sigma$$

Literaturverzeichnis

- [Ada09] ADAMY, Jürgen: *Nichtlineare Regelungen*. Springer, 2009
- [BBJ87] BARAS, John S. ; BENSOUSSAN, Alain ; JAMES, Matthew R.: Dynamic Observer as Asymptotic Limits of Recursive Filters: Special Cases. In: *SIAM J on Applied Mathematics* 48(5) (1987), S. 1147–1158
- [BH97] BROWN, Robert G. ; HWANG, Patrick Y. C.: *Introduction to Random Signals and Applied Kalman Filtering*. Bd. 3. John Wiley and Sons, Inc., 1997
- [Bie77] BIERMAN, Gerald J.: *Factorization Methods for Discrete Sequential Estimation*. Academic Press, 1977
- [GA01] GREWAL, Mohinder S. ; ANDREWS, Angus P.: *Kalman Filtering: Theory and Practice Using MATLAB*. 2. John Wiley and Sons, Inc., 2001
- [GHO92] GAUTHIER, Jean-Paul A. ; HAMMOURI, Hassan ; OTHMAN, Sami: A Simple Observer for Nonlinear Systems. Applications to Bioreactors. In: *IEEE Transactions on Automatic Control* 37(6) (1992), S. 875–880
- [GKN⁺74] GELB, Arthur ; KASPER, Joseph F. J. ; NASH, Raymond A. J. ; PRICE, Charles F. ; SUTHERLAND, Arthur A. J. ; GELB, Arthur (Hrsg.): *Applied Optimal Estimation*. The M.I.T. Press, 1974
- [HF03] HAMMOURI, Hassan ; FARZA, Mondher: Nonlinear Observers for Locally Uniformly Observable Systems. In: *ESAIM: Control, Optimisation and Calculus of Variations* 9 (2003), Nr. 1, S. 353–370
- [Isi99] ISIDORI, Alberto: *Nonlinear Control Systems*. 2. Springer, 1999
- [Kal09] KALKKUHL, Jens C.: Technischer Bericht / Daimler AG. 2009. – Forschungsbericht
- [Kar07] KARRIS, Steven T.: *Numerical Analysis: Using MATLAB and Excel*. Orchard Publications, 2007

- [KB61] KALMAN, Rudolph E. ; BUCY, Richard S.: New Results in Linear Filtering and Prediction Theory. In: *Journal of Basic Engineering* 83 (1961), Nr. 3, S. 95–108
- [Kha92] KHALIL, Hassan K. ; GRIFFIN, John (Hrsg.): *Nonlinear Systems*. Macmillan, 1992
- [KK85] KNOBLOCH, Hans W. ; KWAKERNAAK, Huibert: *Lineare Kontrolltheorie*. Springer, 1985
- [Kre03] KRENER, Arthur J.: The Convergence of the Extended Kalman Filter. In: RANTZER, Anders (Hrsg.) ; BYRNES, Christopher I. (Hrsg.): *Directions in Mathematical Systems Theory and Optimization*. Springer, 2003, S. 173–182
- [RKHO07] REIM, Andreas ; KLIER, Willy ; HILLENBRAND, Stefan ; OTTERBEIN, Stefan: Design and Implementation of a Vehicle Observer for Sideslip Angle Computation with Vehicle and Tire Models / Robert Bosch GmbH. 2007. – Technischer Bericht
- [Unb02] UNBEHAUEN, Rolf: *Systemtheorie 1: Allgemeine Grundlagen, Signale und lineare Systeme im Zeit- und Frequenzbereich. Mit 148 Aufgaben*. Bd. 1. Oldenbourg Verlag, 2002
- [WB06] WELCH, Greg ; BISHOP, Gary: An Introduction to the Kalman Filter. In: *Department of Computer Science, University of North Carolina at Chapel Hill, NC* (2006)
- [Wen07] WENDEL, Jan: *Integrierte Navigationssysteme: Sensordaten, GPS und Inertiale Navigation*. Oldenbourg Verlag, 2007