



UNIVERSITÄT PADERBORN

Die Universität der Informationsgesellschaft

Fakultät für Elektrotechnik, Informatik und Mathematik

Institut für Mathematik

Nichtlineare Erweiterung des Kalman-Filters auf der Grundlage des Bayes-Filters

Bachelorarbeit

zur Erlangung des akademischen Grades

Bachelor of Science

von
DOMENIK SLOTTA

Betreut durch:

Professor Dr. Michael Dellnitz

Jun.-Prof. Dr. Sebastian Peitz

Paderborn, 9. Mai 2022

Inhaltsverzeichnis

1	Einleitung	1
2	Problembeschreibung	4
2.1	Modellbeschreibung	4
2.2	Beschreibung durch Wahrscheinlichkeitsdichten	5
2.3	Zustandsschätzung und das Filterproblem	6
3	Herleitung der Filter	7
3.1	Der Bayes-Filter	7
3.1.1	Bayes-Filter Algorithmus	8
3.1.2	Diskussion	9
3.2	Das Kalman-Filter	9
3.2.1	Kalman-Filter Algorithmus	12
3.2.2	Diskussion	12
3.3	Allgemeiner Gauß Filter	13
3.3.1	Diskussion	15
3.4	Das Erweiterte Kalman-Filter	15
3.4.1	Nichtlineare Transformation EKF	15
3.4.2	EKF Algorithmus	17
3.4.3	Diskussion	17
3.5	Unscented Kalman-Filter	18
3.5.1	Unscented Transformation	18
3.5.2	UKF Algorithmus	20
3.6	Gauß-Hermite Kalman-Filter	20
3.6.1	Gauß-Hermite-Integration	20
4	Vergleich der Filter	24
4.1	Vergleich von EKF und UKF	24
4.1.1	Diskussion	30
4.2	Vergleich mit weiteren Filtern	31
4.3	Anwendung und empirischer Vergleich	32
5	Zusammenfassung	37
	Bibliography	39

Einleitung

Die Zustandsschätzung ist ein zentrales Problem der Robotik, der Positionsbestimmung, in industriellen Systemen und vielen weiteren Anwendungsgebieten. In dieser Arbeit werden bekannte Verfahren der Zustandsschätzung, sogenannte Filter, für nichtlineare Modelle hergeleitet und verglichen.

Messungen durch Sensoren sind aufgrund verschiedener Einflüsse immer fehlerbehaftet. Kombiniert man diese verrauschten Beobachtungen mit einem Modell, welches beispielsweise die Zustände zum Zeitpunkt der nächsten Messung vorhersagt, so kann man die Genauigkeit der Schätzung erheblich verbessern. Als Beispiel für ein solches Modell ist die Positionsbestimmung von Flugzeugen durch ein Radar.

Nehmen wir an, ein Flugzeug fliegt mit einer konstanten Höhe von 10 Kilometern im zweidimensionalen Raum, wie in Abbildung 1.1 dargestellt.

An der Position $(x, y) = (50, 0)$ befindet sich eine Radaranlage, welche die Distanz zum Flugzeug und den Winkel α misst.

Aus diesen beiden Messdaten könnten wir die Position bestimmen. Allerdings werden die Sensordaten durch zufälliges Rauschen gestört und die zusätzliche Information, dass das Flugzeug auf einer bestimmten Höhe fliegt, werden dabei nicht beachtet.

Wir beschreiben zunächst die Bewegung des Flugzeugs durch ein einfaches Bewegungsmodell. Als Zustand des Flugzeuges geben wir die horizontale Position, Geschwindigkeit und Beschleunigung an. Nach der klassischen Mechanik ist erhalten wir das folgende zeitdiskrete Modell mit konstanter Beschleunigung [MD17]:

$$\begin{pmatrix} pos_{k+1} \\ vel_{k+1} \\ acc_{k+1} \end{pmatrix} = \begin{pmatrix} 1 & T & T^2/2 \\ 0 & 1 & T \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} pos_k \\ vel_k \\ acc_k \end{pmatrix}$$

Kennen wir den aktuellen Zustand, können wir diesen für den nächsten Zeitpunkt vorhersagen. Mit einem weiteren Modell wollen wir nun bestimmen, welche Messungen für die aktuelle Position des Flugzeuges beobachtet werden müssten. Aus der Position berechnen wir die Distanz und den Winkel durch

$$\begin{pmatrix} d_k \\ \alpha_k \end{pmatrix} = \begin{pmatrix} \sqrt{10^2 + (pos_k - 50)^2} \\ \arctan(\frac{pos_k - 50}{10}) \end{pmatrix}.$$

Hier ist auch die Information über die Flughöhe enthalten.

Beide Modelle sind zufälliger Störungen ausgesetzt. Im Systemmodell haben wir keine

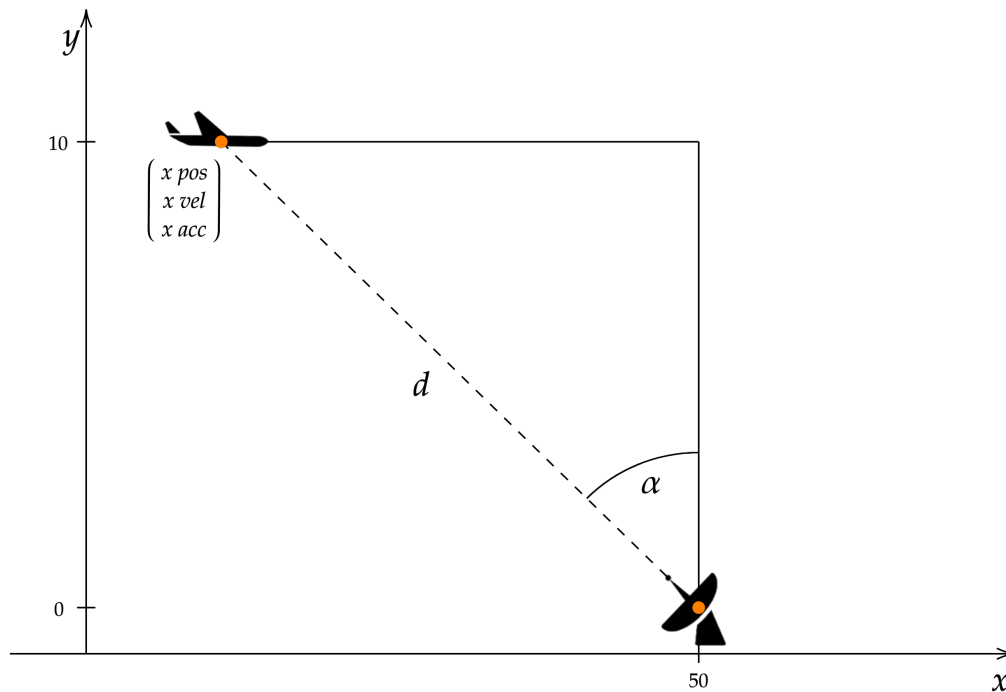


Abbildung 1.1: Zustandsbestimmung eines Flugzeuges durch Radarmessungen

Wetterbedingungen und keinen Luftwiderstand bedacht und das Radar kann die Position nur bis auf eine bekannte Messungenauigkeit bestimmen. Diese Störungen modellieren wir durch Zufallsvariablen, sodass unsere Vorhersagen keine einzelnen Punkte, sondern Verteilungen sind.

Konzentriert sich die Verteilung des Zustandes nun auf eine kleine Fläche, sind wir uns sicher, dass wir Position, Geschwindigkeit und Beschleunigung mit hoher Genauigkeit kennen. Das Systemmodell sagt nun den Zustand zum nächsten Zeitpunkt voraus und mit dem Beobachtungsmodell erhalten wir die erwartete Messung. Seien die Radarmessungen verhältnismäßig starken Störungen ausgesetzt und die neuen Messdaten liegen weit von der vorhergesagten Messung entfernt. Dann gehen wir von einer fehlerhaften Beobachtung aus und schätzen den tatsächlichen Zustand nah am vorhergesagten Wert. Nach den Modellfehlern und der abweichenden Messung sind wir uns jedoch nicht mehr ganz so sicher in unserer Schätzung, also die Verteilung konzentriert sich nicht mehr auf einen kleinen Raum.

Nachdem wir das Filterproblem genauer beschrieben haben, werden wir auf Grundlage der Bayes-Statistik ein Verfahren zur optimalen Kombination von Vorhersagen, Messdaten und bekannten Störprozessen ausarbeiten. Die theoretische Lösung, welche im Allgemeinen nicht exakt berechnet werden kann, liefert für lineare Modelle mit normalverteilten Zufallsvariablen das Kalman-Filter. Indem wir die Wahrscheinlichkeitsverteilungen auch in allgemeineren Fällen als normalverteilt annehmen, leiten wir weitere Filter mit einer ähnlichen Struktur ab. Diese Erweiterungen des Kalman-Filters zählen zu den wichtigsten Filter-Algorithmen.

Im Verlauf der Arbeit werden die Funktionsweisen der Filter veranschaulicht und jewei-

KAPITEL 1. EINLEITUNG

lige Vor- und Nachteile werden basierend auf theoretischen und empirischen Ergebnissen diskutiert.

Problembeschreibung

Systeme wie in der Einleitung können durch ein Hidden Markov Modell beschrieben werden. Der nächste Abschnitt definiert so ein Modell und führt wichtige Begriffe ein. Danach wird dieses Modell durch Wahrscheinlichkeitsdichten verallgemeinert. Mit der Wahrscheinlichkeitstheoretischen Beschreibung spezifizieren wir dann das Filterproblem.

2.1 Modellbeschreibung

Betrachte diskrete Zeiträume I , mit $I = \{1, \dots, T\}$ für ein $T \in \mathbb{N}$ oder mit $I = \mathbb{N}$. Schreibe $I_0 = I \cup \{0\}$. Sei eine zeitdiskrete Beschreibung des Zustands, das Systemmodell, wie folgt gegeben:

$$x_{k+1} = f_{k+1}(x_k, v_{k+1}) \quad (2.1)$$

$$x_0 \sim \pi_0(x) \quad (2.2)$$

Dabei sei $f_k: \mathbb{R}^d \times \mathbb{R}^{n_v} \rightarrow \mathbb{R}^{n_x}$ eine Funktion, welche x_k , den Systemzustand zum Zeitpunkt $k \in I_0$, und das Systemrauschen v_{k+1} auf den nachfolgenden Zustand x_{k+1} abbildet. Das Systemrauschen wird durch die unabhängige Folge von \mathbb{R}^{n_v} -wertigen Zufallsvariablen $\{v_k, k \in I\}$ modelliert. Zusammen mit der bekannten Zufallsvariablen x_0 , gegeben durch die zugehörige Dichtefunktion π_0 , erhalten wir eine Folge von Zufallsvariablen $\{x_k, k \in I_0\}$. Realisierungen dieser Zufallsvariablen sind der exakte, unbeobachtete Zustand des Systems. Die Aufgabe der Zustandsschätzung ist, diesen Zustand durch indirekte oder teilweise Beobachtungen zu schätzen. Beobachtungen bzw. Messungen sind durch das folgende Beobachtungsmodell gegebene Zufallsvariablen $\{y_k, k \in I\}$:

$$y_k = h_k(x_k, w_k) \quad (2.3)$$

$h_k: \mathbb{R}^{n_x} \times \mathbb{R}^{n_w} \rightarrow \mathbb{R}^{n_y}$ ist eine Funktion, welche den Zustand x_k und das Messrauschen w_k auf die Beobachtung bzw. Messung zum Zeitpunkt k , y_k , abbildet. Hier wird das Messrauschen durch die unabhängige Folge von \mathbb{R}^{n_w} -wertigen Zufallsvariablen $\{w_k, k \in I\}$ modelliert. Messwerte von Sensoren entsprechen Realisierungen dieser Zufallsvariablen, basierend auf dem echten und unbeobachteten Systemzustand. Außerdem nehmen wir immer an, dass die beiden Rauschprozesse unabhängig voneinander und unabhängig von x_0 sind.

2.2 Beschreibung durch Wahrscheinlichkeitsdichten

Die in Abschnitt 2.1 betrachtete Situation lässt sich auch probabilistisch formulieren. Sei das zeitdiskrete Systemmodell durch

$$x_{k+1} \sim p(x_{k+1}|x_k) \quad (2.4)$$

$$x_0 \sim \pi_0(x) \quad (2.5)$$

und das Beobachtungsmodell durch

$$y_k \sim p(y_k|x_k) \quad (2.6)$$

gegeben.

Das probabilistische Modell besteht also aus Folgen von Wahrscheinlichkeitsverteilungen, welche die Zustände und Beobachtungen charakterisieren. Die Dichtefunktion π_0 des Zustands zum initialen Zeitpunkt $k = 0$ wird als bekannt vorausgesetzt. Wir nehmen an, dass x_k ein Markov-Prozess ist, also dass

$$p(x_{k+1}|x_{0:k}, y_{1:k}) = p(x_{k+1}|x_k) \quad (2.7)$$

gilt. Vorhersagen des Zustandes x_{k+1} durch das Systemmodell sollen also unabhängig von vergangenen Systemzuständen und Beobachtungen sein, falls der aktuelle Systemzustand x_k bekannt ist. In der Modellbeschreibung als dynamisches System (2.1) ist diese Eigenschaft erfüllt, da f_k nicht von Messungen abhängt und das Systemrauschen eine unabhängige Folge von Zufallsvariablen ist.

Außerdem soll die Beobachtung y_k nur vom aktuellen Zustand x_k abhängen und nicht von vorherigen Systemzuständen oder Beobachtungen. Formal fordern wir also, dass

$$p(y_k|x_{0:k}, y_{1:k-1}) = p(y_k|x_k) \quad (2.8)$$

gilt. Auch diese Eigenschaft ist in dem Beobachtungsmodell (2.3) erfüllt, denn y_k ist eine Funktion von x_k und der unabhängigen Zufallsvariablen w_k . Erfüllen Systeme die Eigenschaft (2.7), werden sie auch *Hidden Markov Modelle* (erster Ordnung) genannt. Bild 2.1 veranschaulicht dieses **HMM** unter Berücksichtigung der Eigenschaft (2.8).

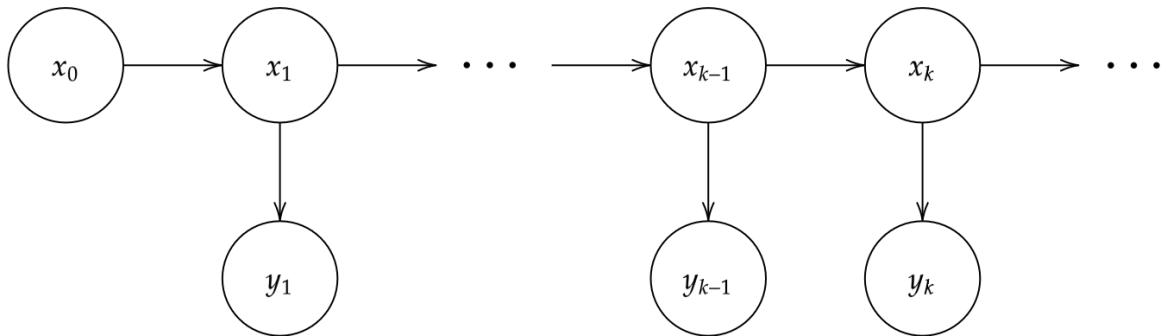


Abbildung 2.1: Visualisierung des in dieser Arbeit verwendeten Hidden Markov Modells

Wir setzen in dieser Arbeit immer voraus, dass alle betrachteten Dichtefunktionen existieren und wir die gesuchten Größen mit dem gewählten Modell berechnen können. Mit

dieser Voraussetzung können wir das Bayes-Filter auf elementare Weise herleiten. Typischerweise lassen sich die Ergebnisse durch Wahrscheinlichkeitsverteilungen verallgemeinern. Wenn wir in dieser Arbeit von Verteilungen sprechen, sollen diese auch immer eine Dichtefunktion besitzen. Dafür nötige Voraussetzungen sind für einige Modelle in [EAM08] zu finden.

2.3 Zustandsschätzung und das Filterproblem

Wir wollen die Realisierungen der Zufallsvariablen x_k , das heißt die tatsächliche Trajektorie der Systemzustände, auf eine optimale Weise schätzen. Allerdings sind nur die beobachteten Größen y_k bis zu einem Zeitpunkt $k = m$ mit $m \in \mathbb{N}$ verfügbar. Diese Messungen und das Modell aus Abschnitt 2.2 sind alle bekannten Informationen, die wir über das echte System haben. Damit können wir die bedingten Dichtefunktionen $p(x_k|y_{1:m})$ errechnen. Mit dieser Dichte kann eine Punktschätzung für den Zustand berechnet werden. Falls $m > k$ gilt, sprechen wir von einer *smoothing*-Schätzung. Hier werden auch spätere Messungen berücksichtigt, mit welchen wir die Schätzung zum Zeitpunkt k verbessern können. Diese Schätzprobleme treten beispielsweise auf, wenn eine komplette Zeitreihe der Messdaten vorliegt und wir die Trajektorie schätzen wollen.

Für $m < k$ heißt die Berechnung der obigen Dichte *Prädiktion*. Ist spezieller $m = k - 1$, nennen wir $p(x_k|y_{1:k-1})$ *a priori* Dichte. Es wird hier auf Grundlage von vergangenen Messungen eine Vorhersage für den nachfolgenden Zustand gemacht. Im Fall $m = k$ heißt die Berechnung der *a posteriori* Dichtefunktion $p(x_k|y_{1:k})$ *Filterung*. Wie im Einführungsbeispiel sind wir häufig nur am aktuellen Zustand interessiert.

Das Filterproblem ist also die Schätzung des Zustandes zum Zeitpunkt k basierend auf allen bisherigen Messungen bis einschließlich zum Zeitpunkt k .

Verfahren zur Berechnung der a posteriori Dichte werden *Filter* genannt.

Herleitung der Filter

Im folgenden wird zunächst der allgemeine *Bayes-Filter* (**BF**), die theoretische rekursive Lösung des Filter-Problems, hergeleitet. Es wird gezeigt, dass das *Kalman-Filter* (**KF**) ein spezieller Bayes-Filter für lineare Modelle mit additivem normalverteilten Rauschen ist. Durch die ausschließlich normalverteilten Dichten im **KF** erhalten wir eine effiziente Berechnung der Zustandsschätzung. Indem wir im nichtlinearen Fall die Fehler als normalverteilt annehmen, erhalten wir eine Reihe wichtiger approximativer Verfahren, welche effizient umgesetzt werden können. Genauer werden das *Erweiterte Kalman-Filter* (**EKF**), das *Unscented Kalman-Filter* (**UKF**) und das *Gauss-Hermite Kalman-Filter* (**GHKF**) als Erweiterungen des **KF** für nichtlineare Modelle hergeleitet.

3.1 Der Bayes-Filter

Sei das in Kapitel 2.2 vorgestellte Modell gegeben. Unser Ziel ist die rekursive Berechnung der a priori und a posteriori Dichten. Hier seien der Einfachheit halber die Dichtefunktionen im Nenner immer größer null. Dazu betrachten wir zunächst die gemeinsame bedingte Dichte $p(x_{k-1}, x_k | y_{1:k-1})$.

Nach der Definition bedingter Dichten und unter Verwendung der Markov-Eigenschaft (2.7) im letzten Schritt gilt:

$$\begin{aligned} p(x_k, x_{k-1} | y_{1:k-1}) &= \frac{p(x_k, x_{k-1}, y_{1:k-1})}{p(y_{1:k-1})} \cdot \frac{p(x_{k-1}, y_{1:k-1})}{p(x_{k-1}, y_{1:k-1})} \\ &= p(x_k | x_{k-1}, y_{1:k-1}) \cdot p(x_{k-1} | y_{1:k-1}) \\ &= p(x_k | x_{k-1}) \cdot p(x_{k-1} | y_{1:k-1}) \end{aligned} \quad (3.1)$$

Betrachtet man nun die bedingte Randdichte dieser gemeinsamen Dichte und benutzt die gezeigte Gleichheit (3.1), erhält man die *Chapman-Kolmogorov Gleichung*:

$$\begin{aligned} p(x_k | y_{1:k-1}) &= \int_{\mathbb{R}^{n_x}} p(x_k, x_{k-1} | y_{1:k-1}) dx_{k-1} \\ &= \int_{\mathbb{R}^{n_x}} p(x_k | x_{k-1}) \cdot p(x_{k-1} | y_{1:k-1}) dx_{k-1} \end{aligned}$$

Diese Berechnung der a priori Dichte $p(x_k | y_{1:k-1})$ bildet den Prädiktionsschritt des Bayes-Filters. Der linke Term im Integral entspricht dem Systemmodell und der rechte Term ist

die a posteriori Dichte zum Zeitpunkt $k - 1$. Die a priori Dichte repräsentiert unser Wissen über den Zustand zur Zeit k auf Grundlage aller Beobachtungen bis zum Zeitpunkt $k - 1$. Durch die zusätzliche Information der nächsten Beobachtung $\mathbf{y}_k = y_k$, können wir unser Wissen über den Zustand x_k im Korrekturschritt verbessern. Zur Berechnung wenden wir die Definition bedingter Dichten an und erweitern wie folgt:

$$p(x_k|y_{1:k}) = \frac{p(x_k, y_{1:k})}{p(y_{1:k})} = \frac{p(x_k)}{p(x_k)} \cdot \frac{p(x_k, y_{1:k-1})}{p(x_k, y_{1:k-1})} \cdot \frac{p(y_{1:k-1})}{p(y_{1:k-1})} \cdot \frac{p(x_k, y_k)}{p(x_k, y_k)}$$

Nach Umsortieren und mit der Notation $p(y_{1:k}) = p(y_k, y_{1:k-1})$ fassen wir die Terme wieder zu bedingten Dichten zusammen:

$$\begin{aligned} p(x_k|y_{1:k}) &= \frac{p(y_k|x_k, y_{1:k-1}) \cdot p(y_k|x_k) \cdot p(x_k|y_{1:k-1})}{p(y_k|y_{1:k-1}) \cdot p(y_k|x_k)} \\ &= \frac{p(y_k|x_k, y_{1:k-1}) \cdot p(x_k|y_{1:k-1})}{p(y_k|y_{1:k-1})} \end{aligned}$$

Diese Umformung entspricht der direkten Anwendung des namensgebenden *Satz von Bayes* in einer hinreichend allgemeinen Form. Mit Gleichung (2.8) und der Unabhängigkeit der Rauschprozesse erhalten wir $p(y_k|x_k, y_{1:k-1}) = p(y_k|x_k)$. Folglich kürzen wir zu:

$$p(x_k|y_{1:k}) = \frac{p(y_k|x_k) \cdot p(x_k|y_{1:k-1})}{p(y_k|y_{1:k-1})}$$

Die Normalisierungskonstante im Nenner ist dabei durch

$$\eta^{-1} := p(y_k|y_{1:k-1}) = \int_{\mathbb{R}^{n_x}} p(y_k|x_k) \cdot p(x_k|y_{1:k-1}) dx_k$$

gegeben. Der erste Faktor ist durch das Beobachtungsmodell (2.6) gegeben und der zweite Faktor ist die im Prädiktionsschritt berechnete a priori Dichte. Sei die Verteilung des Anfangszustands x_0 durch die Dichte π_0 gegeben.

3.1.1 Bayes-Filter Algorithmus

Der Bayes-Filter besteht nun aus der wiederholten Vorhersage der a priori Dichte und dem Korrekturschritt bei Eingang neuer Messungen:

Algorithmus 1 Bayes-Filter

▷ **Initialisierung:**

$$p(x_0|y_{1:0}) = \pi_0(x_0)$$

for $k \in I$ **do**

▷ **Prädiktion:**

$$p(x_k|y_{1:k-1}) = \int_{\mathbb{R}^{n_x}} p(x_k|x_{k-1}) \cdot p(x_{k-1}|y_{1:k-1}) dx_{k-1}$$

▷ **Korrektur:**

$$p(x_k|y_{1:k}) = \eta \cdot p(y_k|x_k) \cdot p(x_k|y_{1:k-1})$$

3.1.2 Diskussion

Die exakte Lösung des Filterproblems hat einen entscheidenden Nachteil. Im Allgemeinen können Dichtefunktionen nicht durch endlich viele Parameter beschrieben werden. Oft benutzt man daher die Dichte der multivariaten Normalverteilung als Approximation. Diese wird durch den Mittelwert und die Kovarianzmatrix vollständig charakterisiert. Ein weiterer Ansatz ist die Approximation durch eine große Anzahl von Zufallsstichproben. Zudem ist die Berechnung des Integrals im Prädiktionsschritt für beliebige Dichten schwierig und eine geschlossene analytische Lösung ist nicht vorhanden. Durch die Annahme von normalverteilten a priori und a posteriori Verteilungen lassen sich schwierige Berechnungen umgehen, da man nur den Mittelwert und die Kovarianzmatrix berechnen muss. In den meisten Anwendungen interessieren wir uns außerdem nicht für die exakte Verteilung, sondern für nur eine Punktschätzung des Zustands. Falls wir die a posteriori Dichte kennen, möchten wir für den Zustand durch einen „bestmöglichen“ Wert angeben, um damit zum Beispiel eine optimale Steuerung des Systems zu berechnen. Eine häufig verwendete und intuitive Wahl ist das Maximum von $p(x_k|y_{1:k})$. Die Punktschätzung $\hat{x}_{k,map} := \arg \max_{x_k} p(x_k|y_{1:k})$ ist die *maximum a posteriori* (map) Schätzung. Der Erwartungswert der a posteriori Dichte liefert die sogenannte *minimum mean square error* (mmse) Schätzung. Für normalverteilte Dichten sind die beiden genannten Punktschätzungen identisch.

Falls das System- und Mess- Modell linear sind und die initiale Verteilung von x_0 sowie das System- und Mess- Rauschen normalverteilt sind, kann die exakte Lösung des Bayes-Filters berechnet werden.

3.2 Das Kalman-Filter

Seien das folgende zeit-diskrete und lineare Systemmodell

$$x_{k+1} = A_{k+1}x_k + B_{k+1}u_{k+1} + L_{k+1}v_{k+1} \quad (3.2)$$

$$x_0 \sim \mathcal{N}(m_{x_0}, P_{x_0}) \quad (3.3)$$

und das lineare Beobachtungsmodell

$$y_k = C_k x_k + M_{k+1} w_k \quad (3.4)$$

gegeben.

Das Systemrauschen v_k und das Messrauschen w_k seien Folgen von unabhängigen, normalverteilten und mittelwertfreien Zufallsvariablen mit bekannten Kovarianzmatrizen Q_k und R_k in $\mathbb{R}^{d \times d}$. d bezeichnet dabei wieder die Länge des Zustandsvektors x_k . Die Systemfunktion A_k und die Matrizen M_k und L_k , die den Einfluss der Rauschprozesse auf den Zustand und die Messung modellieren, sind dabei invertierbare Matrizen in $\mathbb{R}^{d \times d}$. Die Matrix $C_k \in \mathbb{R}^{n_y \times d}$ ist die Messfunktion und soll vollen Rang haben. Die Eingangsmatrix $B_k \in \mathbb{R}^{d \times n_u}$ gibt an, wie der Eingangsvektor $u_k \in \mathbb{R}^{n_u}$ den Systemzustand beeinflusst. Oft werden L_k und M_k als Identitätsmatrix gewählt. Mit den gewählten Annahmen, können wir das lineare Modell mittels Wahrscheinlichkeitsdichten beschreiben. Da das **KF** auf verschiedene Weisen herleiten werden kann, unterscheiden sich die Voraussetzungen an das Modell oft.

Bevor wir den Filter-Algorithmus herleiten, halten wir verwendete Eigenschaften normalverteilter Zufallsvariablen fest.

KAPITEL 3. HERLEITUNG DER FILTER

Für eine affine Transformation $y = Ax + b$ eines \mathbb{R}^d -wertigen Zufallsvektors $x \sim \mathcal{N}(m, P)$ mit der regulären Matrix $A \in \mathbb{R}^{d \times d}$ und $b \in \mathbb{R}^d$ gilt

$$y \sim \mathcal{N}(Am + b, APA^T).$$

Aus [Bis06, Kapitel 2.3] entnehmen wir mit veränderter Notation die folgenden Resultate: Für Zufallsvariablen x und y mit

$$\begin{aligned} x &\sim \mathcal{N}(m, P) \\ y|x &\sim \mathcal{N}(Cx + b, R) \end{aligned}$$

ist die gemeinsame Verteilung durch

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} m \\ Cm + b \end{pmatrix}, \begin{pmatrix} P & PC^T \\ CP & CPC^T + R \end{pmatrix}\right) \quad (3.5)$$

gegeben. Als Marginalverteilung erhalten wir

$$y \sim \mathcal{N}(Cm + b, CPC^T + R). \quad (3.6)$$

Für eine gemeinsame Verteilung mit

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} A & C \\ C^T & B \end{pmatrix}\right)$$

können wir die bedingte Verteilung von x gegeben y wie folgt berechnen:

$$x|y \sim \mathcal{N}(a + CB^{-1}(y - b), A - CB^{-1}C^T). \quad (3.7)$$

Unser Ziel ist es, die a priori und a posteriori Dichten zu berechnen. Dazu nutzen wir das lineare Modell aus. Das Systemmodell (3.2) liefert für ein festes x_{k-1}

$$p(x_k|x_{k-1}) = \mathcal{N}(A_k x_{k-1} + B_k u_k, L_k Q_k L_k^T). \quad (3.8)$$

Analog gilt auch mit dem Beobachtungsmodell (3.4)

$$p(y_k|x_k) = \mathcal{N}(C_k x_k, M_k R_k M_k^T). \quad (3.9)$$

Wir nehmen an, dass die a posteriori Verteilung für den vorherigen Zeitschritt

$$p(x_{k-1}|y_{1:k-1}) = \mathcal{N}(m_{k-1}, P_{k-1})$$

bekannt ist. Die bedingte Verteilung von x_k , gegeben x_{k-1} und allen bisherigen Beobachtungen, ist wegen der Markov-Eigenschaft (2.7) und Gleichung (3.8)

$$\begin{aligned} p(x_k|x_{k-1}, y_{1:k-1}) &= p(x_k|x_{k-1}) \\ &= \mathcal{N}(A_k x_{k-1} + B_k u_k, L_k Q_k L_k^T) \end{aligned}$$

Mit den letzten beiden Gleichungen und (3.6) erhalten wir die a priori Verteilung

$$p(x_k|y_{1:k-1}) = \mathcal{N}(A_k m_{k-1} + B_k u_k, A_k P_{k-1} A_k^T + L_k Q_k L_k^T). \quad (3.10)$$

KAPITEL 3. HERLEITUNG DER FILTER

Der Prädiktionsschritt des Kalman-Filters entspricht demnach der Berechnung des Mittelwertes m_k^- und der Kovarianzmatrix P_k^- der a priori Verteilung, gegeben durch

$$\begin{aligned} m_k^- &= A_k m_{k-1} + B_k u_k \\ P_k^- &= A_k P_{k-1} A_k^T + L_k Q_k L_k^T. \end{aligned}$$

Für den Korrekturschritt wollen wir uns die Verteilung von $p(x_k, y_k | y_{1:k-1})$ anschauen. Damit wir die Gleichung (3.5) verwenden können, müssen wir $p(x_k | y_{1:k-1})$ und $p(y_k | x_k, y_{1:k-1})$ kennen. Der erste Term ist die im Prädiktionsschritt berechnete a priori Verteilung (3.10) welche kürzer als $\mathcal{N}(m_k^-, P_k^-)$ geschrieben wird. Mit den Eigenschaften des Hidden Markov Modells (2.7), (2.8) und der Gleichung (3.9) folgt

$$\begin{aligned} p(y_k | x_k, y_{1:k-1}) &= p(y_k | x_k) \\ &= \mathcal{N}(C_k x_k, M_k R_k M_k^T). \end{aligned}$$

Nun erhalten durch (3.5) die gesuchte Verteilung

$$p(x_k, y_k | y_{1:k-1}) = \mathcal{N} \left(\begin{pmatrix} m_k^- \\ C_k m_k^- \end{pmatrix}, \begin{pmatrix} P_k^- & P_k^- C_k^T \\ C_k P_k^- & C_k P_k^- C_k^T + M_k R_k M_k^T \end{pmatrix} \right).$$

Nun berechnen wir nach Gleichung (3.7) die a posteriori Verteilung $p(x_k | y_k, y_{1:k-1}) = p(x_k | y_{1:k})$ und schreiben diese als $\mathcal{N}(m_k, P_k)$ mit

$$\begin{aligned} m_k &= m_k^- + P_k^- C_k^T (C_k P_k^- C_k + M_k R_k M_k^T)^{-1} (y_k - C_k m_k^-) \\ P_k &= P_k^- - P_k^- C_k^T (C_k P_k^- C_k + M_k R_k M_k^T)^{-1} C_k P_k^-. \end{aligned}$$

Wir teilen diese langen Terme auf, sodass wir eine klassische Form des Kalman-Filters erhalten. Außerdem können wir die Rechenschritte nach der Umformung besser interpretieren. Zunächst wird die Kovarianzmatrix von x_k und y_k mittels

$$P_{x_k y_k} = P_k^- C_k^T$$

berechnet. Außerdem taucht die durch

$$P_{y_k} = C_k P_k^- C_k^T + M_k R_k M_k^T$$

berechnete Kovarianzmatrix von y_k im Korrekturschritt auf.

Wir schreiben $\hat{y}_k = C_k m_k^-$ für den Erwartungswert der Beobachtung und die sogenannte *Kalman-Verstärkung* ist wie folgt definiert:

$$K_k = P_{x_k y_k} P_{y_k}^{-1}$$

Die Differenz $(y_k - \hat{y}_k)$ wird *Innovation* genannt. Dass die neu definierten Größen tatsächlich den entsprechenden Erwartungswerten und Kovarianzmatrizen entsprechen, kann leicht durch die vorangegangenen Rechenschritte verifiziert werden. Mit diesen Bezeichnungen lauten Mittelwert und Kovarianzmatrix der a posteriori Verteilung also

$$\begin{aligned} m_k &= m_k^- + K_k (y_k - \hat{y}_k) \\ P_k &= P_k^- - K_k P_{y_k} K_k^T. \end{aligned}$$

Wir bemerken, dass die a posteriori Verteilung des vorherigen Zeitschrittes $k - 1$ als normalverteilt angenommen wurde. Wir konnten durch die konkrete Berechnung folgern, dass dann erst die a priori Verteilung zum aktuellen Zeitpunkt k und damit dann auch die a posteriori Verteilung $p(x_k | y_{1:k})$ normalverteilt sind. Da wir mit dem normalverteilten Zustand x_0 starten, ist unsere Annahme nach vollständiger Induktion immer erfüllt.

3.2.1 Kalman-Filter Algorithmus

Durch die rekursive Berechnung des Prädiktions- und des Korrekturschrittes erhalten wir den folgenden Algorithmus.

Algorithmus 2 Kalman-Filter

▷ *Initialisierung:* ◁

$$m_0 = m_{x_0}$$

$$P_0 = P_{x_0}$$

for $k \in I$ **do**

▷ *Prädiktion:* ◁

$$m_k^- = A_k m_{k-1} + B_k u_k$$

$$P_k^- = A_k P_{k-1} A_k^T + L_k Q_k L_k^T$$

▷ *Korrektur:* ◁

$$\hat{y}_k = C_k m_k^-$$

$$P_{x_k y_k} = P_k^- C_k^T$$

$$P_{y_k} = C_k P_k^- C_k^T + M_k R_k M_k^T$$

$$K_k = P_{x_k y_k} P_{y_k}^{-1}$$

$$m_k = m_k^- + K_k (y_k - \hat{y}_k)$$

$$P_k = P_k^- - K_k P_{y_k} K_k^T$$

3.2.2 Diskussion

Für das betrachtete lineare Modell (3.2)-(3.4) entspricht das Kalman-Filter der exakten Lösung des Bayes-Filters. Da für normalverteilte Dichtefunktionen sowohl die **map**-Schätzung als auch die **mmse**-Schätzung dem Mittelwert m_k im Korrekturschritt entsprechen, werden selten andere Punktschätzungen berechnet. Allerdings werden für die Prädiktion und die Korrektur immer die vollständigen Dichten geschätzt.

Aus anderen Herleitungen geht hervor, dass das **KF** ein optimaler *lineare Filter* für **mmse**-Schätzungen mit beliebig verteilte Rauschprozessen ist [Sim06, Kal60]. Eine weitere gute Eigenschaft ist, dass die Kovarianzmatrizen nicht vom den Beobachtungen abhängen. Die Kovarianzen können also schon vor Eingang eines neuen Messwertes berechnet werden. Für einige Systeme konvergieren die Matrizen und statt der Berechnung werden die konstanten Matrizen in sogenannten *Steady-State* Filtern eingesetzt [Sim06, Kapitel 4]. Systeme in der echten Welt verhalten sich jedoch in der Regel nichtlinear. Falls lineare Modelle das System nicht gut beschreiben können und man genauere Modelle braucht, muss man Filter für nichtlineare Modelle entwerfen. Aufgrund der einfachen Struktur und der hohen Effizienz des Kalman-Filters wollen wir diesen für nichtlineare Modelle erweitern.

3.3 Allgemeiner Gauß Filter

Seien nun ein zeit-diskretes nichtlineares Systemmodell der Form

$$x_{k+1} = f_{k+1}(x_k, v_{k+1}) \quad (3.11)$$

$$x_0 \sim \mathcal{N}(m_{x_0}, P_{x_0}) \quad (3.12)$$

und ein nichtlineares Beobachtungsmodell

$$y_k = h_k(x_k, w_k) \quad (3.13)$$

gegeben.

Spezieller als im ursprünglichen Modell (2.1)-(2.3) ist der initiale Zustand durch eine Normalverteilung gegeben. Außerdem seien das Systemrauschen v_k und das Messrauschen w_k Folgen von unabhängigen, normalverteilten und mittelwertfreien Zufallsvariablen mit bekannten Kovarianzmatrizen Q_k und R_k .

Nachdem wir für das lineare Modell eine exakte und effiziente Lösung für das Filterproblem gefunden haben, möchten wir Algorithmen für die schwierigeren nichtlinearen Modelle entwickeln. Unsere Strategie ist es, die Rechenschritte des **KF**s zu verallgemeinern. Zuvor wurde das lineare Modell zur konkreten Berechnung der Erwartungswerte und Kovarianzmatrizen genutzt. Stattdessen können wir die gesuchten Größen auch durch bedingte Erwartungswerte ersetzen. Folgen wir damit der Herleitung aus Abschnitt 3.2, erhalten wir wieder ein Filter-Algorithmus. Da wir das **KF** schon in eine geeignete Form gebracht haben, ersetzen wir an dieser Stelle nur die Terme für die Erwartungswerte und Kovarianzmatrizen.

Bisher haben wir als Notation für Normalverteilungen $\mathcal{N}(m, P)$ verwendet. Für eine normalverteilte Zufallsvariable x bezeichnen wir die zugehörige Wahrscheinlichkeitsdichte mit $\mathcal{N}(x|m, P)$. Es gilt also

$$\mathcal{N}(x|m, P) = \frac{1}{\sqrt{(2\pi)^d |P|}} \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}(x-m)^T P^{-1}(x-m)\right) dx.$$

Dabei ist $|P|$ die Determinante von P . Die Berechnungen im Prädiktionsschritt basieren auf der a posteriori Dichte des vorangegangenen Korrekturschrittes. Daher ersetzen wir zum Beispiel

$$\begin{aligned} m_k^- &= A_k m_{k-1} + B_k u_k \\ P_k^- &= A_k P_{k-1} A_k^T + L_k Q_k L_k^T \end{aligned}$$

durch

$$\begin{aligned} m_k^- &= \int_{\mathbb{R}^d \times \mathbb{R}^{n_v}} f_k(x_{k-1}, v_k) \\ &\quad \cdot \mathcal{N}(x_{k-1}|m_{k-1}, P_{k-1}) \cdot \mathcal{N}(v_k|0, Q_k) dx_{k-1} dv_k \\ P_k^- &= \int_{\mathbb{R}^d \times \mathbb{R}^{n_v}} (f_k(x_{k-1}, v_k) - m_{k-1})(f_k(x_{k-1}, v_k) - m_{k-1})^T \\ &\quad \cdot \mathcal{N}(x_{k-1}|m_{k-1}, P_{k-1}) \cdot \mathcal{N}(v_k|0, Q_k) dx_{k-1} dv_k \end{aligned}$$

Im Korrekturschritt gehen wir analog vor. Hier werden die als Integrale angegebenen Erwartungswerte allerdings durch die a priori Verteilung aus dem Prädiktionsschritt bedingt.

KAPITEL 3. HERLEITUNG DER FILTER

Die restlichen Gleichungen können dem Algorithmus 3 entnommen werden.

Wir nehmen nun an, dass alle a priori und a posteriori Dichten normalverteilt sind. Im Gauß Filter werden die gesuchten Mittelwerte und Kovarianzmatrizen allgemein durch Integrale angegeben. Durch diese Vereinfachung erhalten wir ein approximatives Verfahren. Bisher haben wir ausgenutzt, dass affine Transformation normalverteilter Zufallsvariablen wieder normalverteilt sind. Für nichtlineare Systeme gilt das jedoch nicht. Wir verlassen uns also darauf, dass wir die Filter-Verteilungen hinreichend gut mittels einer Normalverteilung beschrieben werden kann.

In der Praxis funktioniert diese Annahme jedoch aus mehreren Gründen recht gut. Einerseits werden die Rauschprozesse oft als normalverteilt angenommen. Ein Grund dafür liefert der Zentrale Grenzwertsatz, denn auf reale Systeme wirken eine Reihe unbekannte Störungen, die in Summe ungefähr normalverteilt sind. Beispiele für solche Störungen sind elektromagnetisches Rauschen in Sensoren und Wettereinflüsse auf ein Flugzeug. Zusätzlich wissen wir nach dem Satz von Taylor, dass sich Funktionen für kleine Abweichungen beziehungsweise kleinen Schrittweiten annähernd linear Verhalten. Dementsprechend sollte die transformierte Zufallsvariable gut durch den Erwartungswert und der Kovarianz beschrieben werden.

Algorithmus 3 Gauß Filter

▷ *Initialisierung:* ◁

$$m_0 = m_{x_0}$$

$$P_0 = P_{x_0}$$

for $k \in I$ **do**

▷ *Prädiktion:* ◁

$$\begin{aligned} m_k^- &= \int_{\mathbb{R}^d \times \mathbb{R}^{n_v}} f_k(x_{k-1}, v_k) \\ &\quad \cdot \mathcal{N}(x_{k-1} | m_{k-1}, P_{k-1}) \cdot \mathcal{N}(v_k | 0, Q_k) dx_{k-1} dv_k \\ P_k^- &= \int_{\mathbb{R}^d \times \mathbb{R}^{n_v}} (f_k(x_{k-1}, v_k) - m_{k-1})(f_k(x_{k-1}, v_k) - m_{k-1})^T \\ &\quad \cdot \mathcal{N}(x_{k-1} | m_{k-1}, P_{k-1}) \cdot \mathcal{N}(v_k | 0, Q_k) dx_{k-1} dv_k \end{aligned}$$

▷ *Korrektur:* ◁

$$\begin{aligned} \hat{y}_k &= \int_{\mathbb{R}^d \times \mathbb{R}^{n_w}} h_k(x_k, w_k) \\ &\quad \cdot \mathcal{N}(x_k | m_k^-, P_k^-) \cdot \mathcal{N}(w_k | 0, R_k) dx_k dw_k \\ P_{x_k y_k} &= \int_{\mathbb{R}^d \times \mathbb{R}^{n_w}} (x_k - m_k^-)(h_k(x_k, w_k) - \hat{y}_k)^T \\ &\quad \cdot \mathcal{N}(x_k | m_k^-, P_k^-) \cdot \mathcal{N}(w_k | 0, R_k) dx_k dw_k \\ P_{y_k} &= \int_{\mathbb{R}^d \times \mathbb{R}^{n_w}} (h_k(x_k, w_k) - \hat{y}_k)(h_k(x_k, w_k) - \hat{y}_k)^T \\ &\quad \cdot \mathcal{N}(x_k | m_k^-, P_k^-) \cdot \mathcal{N}(w_k | 0, R_k) dx_k dw_k \\ K_k &= P_{x_k y_k} P_{y_k}^{-1} \\ m_k &= m_k^- + K_k(y_k - \hat{y}_k) \\ P_k &= P_k^- - K_k P_{y_k} K_k^T \end{aligned}$$

3.3.1 Diskussion

Das allgemeine Gauß Filter, auch manchmal *Gaussian Assumed Density* Filter genannt [May82, Kapitel 12], reduziert die Entwicklung weiterer Filter auf die Approximation der oben beschriebenen Erwartungswerte. In der Regel können wir die Integrale nicht direkt ausrechnen. Eine Möglichkeit ist, diese mithilfe der *Monte-Carlo Integration* zu berechnen.

In Abschnitt 3.6.1 stellen wir die Gauss-Hermite Integration vor. Darin wird die spezielle Struktur der Integrale ausgenutzt. Des Weiteren erhalten wir das **EKF** und das **UKF**, welche die gesuchten Größen effizient schätzen.

Das Verfahren kann außerdem genutzt werden um die Klasse der Gauß Filter, also alle darauf basierenden Verfahren, mit anderen Filtern zu vergleichen. Theoretische Erkenntnisse und Modifikationen können gegebenenfalls auf konkrete Filter-Algorithmen übertragen werden.

3.4 Das Erweiterte Kalman-Filter

Gegeben sei das durch Gleichungen (3.11)-(3.13) nichtlineare Modell. Aus der Taylorreihenentwicklung geht hervor, dass sich stetig differenzierbare Funktionen für kleine Abweichungen annähernd linear verhalten. Mit dieser Idee leiten wir das meistgenutzte Filter für nichtlineare Modelle, das Erweiterte Kalman-Filter her.

3.4.1 Nichtlineare Transformation EKF

Wir schauen uns die nichtlineare Transformation einer \mathbb{R}^{n_x} -wertigen Zufallsvariablen $x \sim \mathcal{N}(m, P)$ der Form $y = g(x, q)$ an. $g: \mathbb{R}^{n_x} \times \mathbb{R}^{n_q} \rightarrow \mathbb{R}^{n_y}$ ist dabei eine nichtlineare Funktion und $q \in \mathbb{R}^{n_q}$ ist eine durch $q \sim \mathcal{N}(0, Q)$ gegebene, von x unabhängige Zufallsvariable. Im folgenden wird gezeigt, wie wir mithilfe der Taylorreihenentwicklung eine numerisch einfach zu berechnende Approximation des Erwartungswertes und der Kovarianzmatrix von y erhalten.

Dazu sei $x = m + \tilde{x}$ mit $\tilde{x} \sim \mathcal{N}(0, P)$. Vorausgesetzt die Funktion g ist stetig differenzierbar folgt mit dem Satz von Taylor für kleine Werte \tilde{x} und q

$$\begin{aligned} g(x, q) &= g(m + \tilde{x}, q) \\ &\approx g(m, 0) + G_x(m, 0)\tilde{x} + G_q(m, 0)q. \end{aligned}$$

Wir schreiben hier $G_x(m, 0)$ und $G_q(m, 0)$ für die durch

$$\begin{aligned} [G_x(m, 0)]_{i,j} &= \left. \frac{\partial g^{(i)}(x, q)}{\partial x_j} \right|_{(x,q)=(m,0)} \\ [G_q(m, 0)]_{i,k} &= \left. \frac{\partial g^{(i)}(x, q)}{\partial q_k} \right|_{(x,q)=(m,0)} \end{aligned}$$

gegebenen Ableitungen von g , mit $i = 1, \dots, n_y$, $j = 1, \dots, n_x$ und $k = 1, \dots, n_q$.

Die Berechnung des Erwartungswertes von y mit dieser linearen Approximation liefert eine Normalverteilung, die für kleine (\tilde{x}, q) näherungsweise der Verteilung von y entspricht.

KAPITEL 3. HERLEITUNG DER FILTER

Es gilt

$$\begin{aligned}
 E[y] &= E[g(x, q)] \approx E[g(m, 0) + G_x(m, 0)\tilde{x} + G_q(m, 0)q] \\
 &= g(m, 0) + G_x(m, 0) \underbrace{E[\tilde{x}]}_{=0} + G_q(m, 0) \underbrace{E[q]}_{=0} \\
 &= g(m, 0)
 \end{aligned} \tag{3.14}$$

Die Näherung durch die Taylorreihe wird im Folgenden ohne die Erwähnung kleiner \tilde{x} und q benutzt. Für die Kovarianzmatrix von y gilt

$$\begin{aligned}
 \text{Cov}[y] &= E[(y - E[y])(\cdots)^T] \\
 &\approx E[(g(m, 0) + G_x(m, 0)\tilde{x} + G_q(m, 0)q - g(m, 0))(\cdots)^T] \\
 &= E[(G_x(m, 0)\tilde{x} + G_q(m, 0)q)(\cdots)^T] \\
 &= E[G_x(m, 0)\tilde{x}(G_x(m, 0)\tilde{x})^T + G_x(m, 0)\tilde{x}(G_q(m, 0)q)^T \\
 &\quad + G_q(m, 0)q(G_x(m, 0)\tilde{x})^T + G_q(m, 0)q(G_q(m, 0)q)^T] \\
 &= G_x(m, 0) E[\tilde{x}\tilde{x}^T] G_x(m, 0)^T + G_x(m, 0) \underbrace{E[\tilde{x}q^T]}_{=0} G_q(m, 0)^T \\
 &\quad + G_q(m, 0) \underbrace{E[q\tilde{x}^T]}_{=0} G_x(m, 0)^T + G_q(m, 0) E[qq^T] G_q(m, 0)^T \\
 &= G_x(m, 0) P G_x(m, 0)^T + G_q(m, 0) Q G_q(m, 0)^T
 \end{aligned} \tag{3.15}$$

Dabei haben wir ausgenutzt, dass x und damit auch \tilde{x} unabhängig von q ist. Ähnlich gilt für die Kovarianzmatrix von x und y

$$\begin{aligned}
 \text{Cov}[x, y] &\approx E[(x - m)(g(m, 0) + G_x(m, 0)\tilde{x} + G_q(m, 0)q)^T] \\
 &= \underbrace{E[x]}_{=m} g(m, 0)^T + \underbrace{E[x\tilde{x}^T]}_{=P} G_x(m, 0)^T + \underbrace{E[xq^T]}_{=0} G_q(m, 0)^T \\
 &\quad - m g(m, 0)^T - m \underbrace{E[\tilde{x}^T]}_{=0} G_x(m, 0)^T - m \underbrace{E[q^T]}_{=0} G_q(m, 0)^T \\
 &= P G_x(m, 0)^T.
 \end{aligned} \tag{3.16}$$

Um nun das Erweiterte Kalman-Filter zu erhalten, ersetzen wir in der Prädiktion die Funktion g durch das Systemmodell f_k . Die Ableitungen sind mit verkürzter Notation durch

$$\begin{aligned}
 F_k &= \left. \frac{\partial f_k(x, v)}{\partial x} \right|_{(x,v)=(m_{k-1},0)} \\
 L_k &= \left. \frac{\partial f_k(x, v)}{\partial v} \right|_{(x,v)=(m_{k-1},0)}
 \end{aligned}$$

gegeben. Für die Ableitungen der Funktion h_k schreiben wir

$$\begin{aligned}
 H_k &= \left. \frac{\partial h_k(x, w)}{\partial x} \right|_{(x,w)=(m_k^-,0)} \\
 M_k &= \left. \frac{\partial h_k(x, w)}{\partial w} \right|_{(x,w)=(m_k^-,0)}
 \end{aligned}$$

3.4.2 EKF Algorithmus

Durch Einsetzen der passenden Funktionen, Mittelwerte und Kovarianzen in die Gleichungen (3.14)-(3.16) für das System- und das Beobachtungsmodell erhalten wir das folgende Verfahren.

Algorithmus 4 Extended Kalman-Filter

▷ *Initialisierung:* ◁

$$m_0 = m_{x_0}$$

$$P_0 = P_{x_0}$$

for $k \in I$ **do**

▷ *Prädiktion:* ◁

$$m_k^- = f_k(m_{k-1}, 0)$$

$$P_k^- = F_k P_{k-1} F_k^T + L_k Q_k L_k^T$$

▷ *Korrektur:* ◁

$$\hat{y}_k = h_k(m_k^-, 0)$$

$$P_{x_k y_k} = P_k^- H_k^T$$

$$P_{y_k} = H_k P_k^- H_k^T + M_k R_k M_k^T$$

$$K_k = P_{x_k y_k} P_{y_k}^{-1}$$

$$m_k = m_k^- + K_k (y_k - \hat{y}_k)$$

$$P_k = P_k^- - K_k P_{y_k} K_k^T$$

3.4.3 Diskussion

Als erstes beachten wir, dass wir nach der Approximation ein lineares System erhalten. Folgen wir damit der Herleitung des **KF**s, erhalten wir den selben Algorithmus. Wenn wir zusätzlich eine Steuerung u_k des Systems explizit als $x_k = f_k(x_{k-1}, u_k, v_k)$ angeben, enthält der somit entstandene Algorithmus auch eine Steuermatrix.

Gegenüber dem **KF** haben wir hier den Nachteil, dass man zur Berechnung der Kovarianzmatrizen die Ableitungen an den prädizierten und den korrigierten Mittelwerten m_k^- bzw. m_k benötigt. Man muss also immer auf die entsprechenden Werte warten und kann die Kovarianzmatrizen nicht vorher berechnen. In manchen Anwendungen ist es schwierig oder unmöglich die Ableitungen zu berechnen, sodass das **EKF** nicht eingesetzt werden kann.

Die lokalen Ableitungen an potentiell schlechten Schätzungen führen zu weiteren Fehlern, sodass das **EKF** in manchen Fällen divergiert. Ein weiteres Problem ist, dass die Unsicherheit des Zustands typischerweise unterschätzt wird. Dadurch bleibt die Kalman-Verstärkung klein und die Messungen werden nicht stark genug berücksichtigt.

Nichtsdestotrotz ist das **EKF** das Standardwerkzeug für die Zustandsschätzung für nicht-lineare Systeme. Gründe dafür sind die einfache und bekannte **KF**-Struktur und viele gute Ergebnisse in der Praxis. Durch die bekannte Struktur und der Tatsache, dass für das **KF** viele exakte Resultate und Anpassungen für verbesserte Robustheit und höhere Genauigkeit vorliegen, können entsprechende Anpassungen oft direkt oder zumindest ähnlich übernommen werden. Außerdem wird das **EKF** schon seit der Veröffentlichung des **KF** 1960 verwendet, weshalb es viele Forschungsergebnisse und praktische Erfahrung gibt. Gerade die langjährige Erfahrung ist für den Filterentwurf wichtig, denn Parameter der neuen Algorithmen müssen für den Einsatz in echten Systemen angepasst werden. Auch die modellierten Unsicherheiten haben unterschiedliche Auswirkungen auf Genauigkeit und

Stabilität.

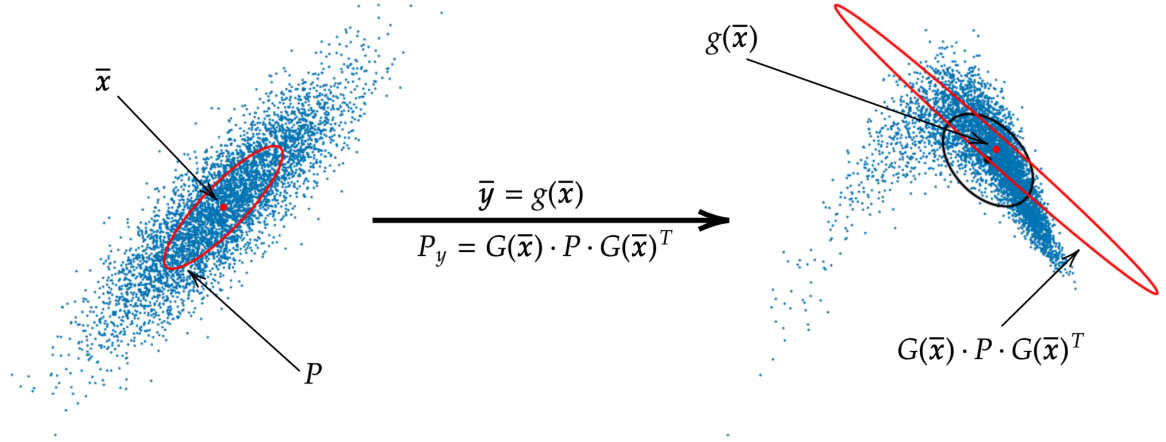


Abbildung 3.1: Veranschaulichung der linearisierten Transformation einer normalverteilten Zufallsvariablen aus dem EKF (rot). Die tatsächlichen Größen sind in Schwarz eingezeichnet.

Im Bild 3.1 ist links eine normalverteilte Zufallsvariable als Punktwolke und durch einen Erwartungswert und der Kovarianz-Ellipse dargestellt. Die durch ein Pendel gegebene Transformation verhält sich aufgrund der großen Schrittweite stark nichtlinear. Man kann hier erkennen, dass die lokale Approximation zu einer schlechten Schätzung der transformierten Zufallsvariable führt.

3.5 Unscented Kalman-Filter

Eine weitere Möglichkeit die nichtlineare Transformation einer Normalverteilung zu betrachten liefert die *Unscented Transformation* (**UT**). Im **EKF** haben wir die Funktion approximiert, welche die Zufallsvariablen transformiert.

Die in [JUDW95] vorgestellte Idee ist, dass es einfacher sein sollte eine Wahrscheinlichkeitsverteilung, statt eine allgemeine nichtlineare Funktion, z.B. mit der Taylorreihenentwicklung, zu approximieren. Dazu wird eine Menge von Punkten, sogenannte *Sigma-Punkte*, gewählt, deren Stichprobenmittel und Stichprobenvarianz den geschätzten Statistiken des Zustands entspricht. Diese Sigma-Punkte werden dann einzeln durch die nichtlineare Funktion g abgebildet und anschließend werden die benötigten Statistiken der neuen Punktwolke berechnet.

Wir stellen eine Version der Unscented Transformation vor und leiten daraus das **UKF** ab. Die Wahl der Sigma-Punkte und der Gewichte wird aus der Analyse in Abschnitt 4.1 ersichtlich.

3.5.1 Unscented Transformation

Genau wie bei **EKF** betrachten wir die Transformation der \mathbb{R}^d -wertigen Zufallsvariablen $x \sim \mathcal{N}(m, P)$ der Form $y = g(x)$ mit der nichtlinearen Funktion $g: \mathbb{R}^d \rightarrow \mathbb{R}^n$. Die Menge

KAPITEL 3. HERLEITUNG DER FILTER

\mathcal{X} bestehend aus $2d + 1$ Sigma-Punkten definieren wir für $i = 1, \dots, d$ wie folgt:

$$\begin{aligned}\mathcal{X}_0 &:= m \\ \mathcal{X}_i &:= m + \sqrt{d + \kappa} \left[\sqrt{P} \right]_i \\ \mathcal{X}_{i+n} &:= m - \sqrt{d + \kappa} \left[\sqrt{P} \right]_i\end{aligned}$$

Dabei ist $\left[\sqrt{P} \right]_i$ die i -te Spalte der durch $\sqrt{P} \cdot \sqrt{P}^T = P$ definierten Matrix \sqrt{P} . Für positiv definite, symmetrische Matrizen P berechnen wir \sqrt{P} mithilfe der Cholesky-Zerlegung. Außerdem wollen wir die Sigma-Punkte folgendermaßen gewichten:

$$\begin{aligned}W_0 &:= \frac{\kappa}{d + \kappa} \\ W_i &:= \frac{1}{2(d + \kappa)}\end{aligned}$$

mit $i = 1, \dots, 2d$. Offensichtlich ist das Stichprobenmittel dieser Punkte gleich m und die Stichprobenvarianz ist P , denn

$$\begin{aligned}P_x &= \sum_{i=0}^{2d} W_i \cdot (\mathcal{X}_i - m)(\mathcal{X}_i - m)^T \\ &= \frac{1}{2(d + \kappa)} \sum_{i=1}^{2d} (\sqrt{d + \kappa} \left[\sqrt{P} \right]_i)(\sqrt{d + \kappa} \left[\sqrt{P} \right]_i)^T \\ &= \frac{2}{2} \sum_{i=1}^d \left[\sqrt{P} \right]_i \left[\sqrt{P} \right]_i^T \\ &= P.\end{aligned}$$

Die Sigma-Punkte werden nun durch die Funktion g abgebildet und wir erhalten die Menge der transformierten Sigma-Punkte \mathcal{Y} via

$$\mathcal{Y}_i := g(\mathcal{X}_i).$$

Aus diesen Punkten können wir die gesuchten Statistiken für unseren Filter berechnen. Die Approximationen an die tatsächlichen Werte für Mittelwert und Kovarianzmatrizen lauten

$$\begin{aligned}\mathbb{E}[y] &\simeq \bar{y}_u := \sum_{i=0}^{2d} W_i \cdot \mathcal{Y}_i \\ &= \frac{1}{d + \kappa} \left(\kappa \cdot \mathcal{Y}_0 + \frac{1}{2} \cdot \sum_{i=1}^{2d} \mathcal{Y}_i \right) \\ \text{Cov}[x, y] &\simeq \sum_{i=0}^{2d} W_i \cdot (\mathcal{X}_i - m)(\mathcal{Y}_i - \bar{y}_u)^T \\ &= \frac{1}{d + \kappa} \left(\kappa \cdot (\mathcal{X}_0 - m)(\mathcal{Y}_0 - \bar{y}_u)^T + \frac{1}{2} \cdot \sum_{i=1}^{2d} (\mathcal{X}_i - m)(\mathcal{Y}_i - \bar{y}_u)^T \right).\end{aligned}$$

Für $\text{Cov}[y]$ tauscht man an entsprechenden Stellen \mathcal{X}_i durch \mathcal{Y}_i und m durch \bar{y}_u aus.

3.5.2 UKF Algorithmus

Wir schreiben den Algorithmus 5 für ein Modell der Form

$$\begin{aligned} x_0 &\sim \mathcal{N}(m_{x_0}, P_{x_0}) \\ x_{k+1} &= f_{k+1}(x_k) + v_{k+1} \\ y_k &= h_k(x_k) + w_k \end{aligned}$$

mit $x_k, v_k \in \mathbb{R}^d$ und $y_k, w_k \in \mathbb{R}^n$.

Auf Vor- und Nachteile sowie Besonderheiten des **UKF** gehen wir danach im Vergleich mit dem **EKF** in Abschnitt 4.1 ein.

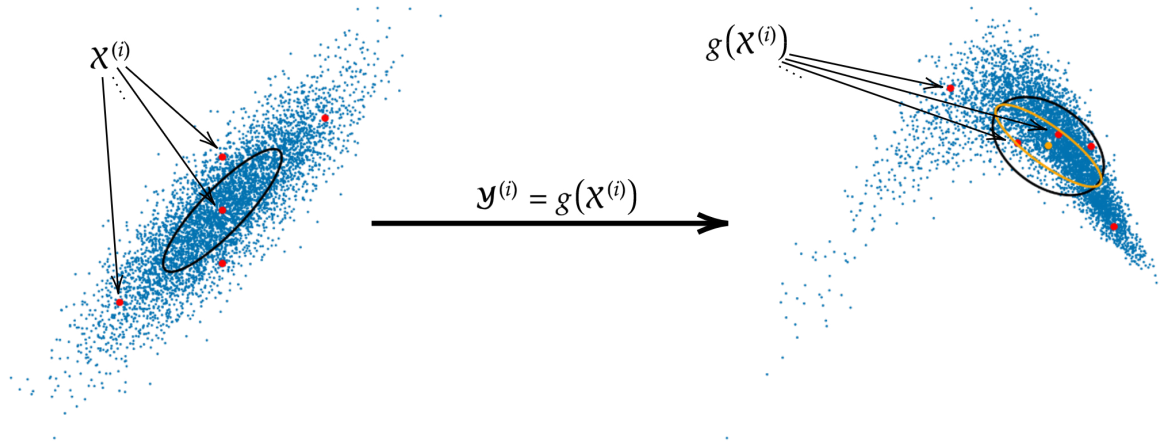


Abbildung 3.2: Veranschaulichung der *Unscented Transformation* einer normalverteilten Zufallsvariablen aus dem UKF. In Rot sind die Sigma-Punkte und in Schwarz die tatsächlichen Größen eingezeichnet. Der in der UT berechnete Mittelwert und die Kovarianzen sind orange gefärbt.

Abbildung 3.2 zeigt ähnlich wie in der Visualisierung zum Kalman-Filter eine nicht-lineare Transformation. Statt der Ellipse sind hier die mit der Kovarianzmatrix berechneten Sigma-Punkte eingezeichnet. Da die Funktion g nun nicht nur lokal ausgewertet wird, wird der Erwartungswert genau bestimmt und die Schätzung der Kovarianzmatrix auf der rechten Seite ist besser als durch die Linearisierung.

3.6 Gauß-Hermite Kalman-Filter

Durch die Gauss-Hermite Integration können die Integrale aus dem Algorithmus 3 approximiert werden. Der Algorithmus wird dann als Gauß-Hermite Kalman-Filter bezeichnet [Sär13].

3.6.1 Gauß-Hermite-Integration

Die Gauß-Hermite Integration ist ein Verfahren zur Approximation von Integralen der Form $\int_{\mathbb{R}} f(x) e^{-x^2} dx$. Dazu wird die Funktion f an N Stellen ξ_i ausgewertet und eine

Algorithmus 5 Unscented Kalman-Filter

 ▷ *Initialisierung:*

$$m_0 = m_{x_0}$$

$$P_0 = P_{x_0}$$

for $k \in I$ **do**

 ▷ *Prädiktion:*

$$\mathcal{X}_{k-1}^{(0)} = m_{k-1}$$

for $i = 1, \dots, d$ **do**

$$\begin{bmatrix} \mathcal{X}_{k-1}^{(i)} = m_{k-1} + \sqrt{d + \kappa} \begin{bmatrix} \sqrt{P_{k-1}} \end{bmatrix}_i \\ \mathcal{X}_{k-1}^{(i+d)} = m_{k-1} - \sqrt{d + \kappa} \begin{bmatrix} \sqrt{P_{k-1}} \end{bmatrix}_i \end{bmatrix}$$

for $i = 1, \dots, 2d$ **do**

$$\begin{bmatrix} \hat{\mathcal{X}}_k^{(i)} = f_k(\mathcal{X}_{k-1}^{(i)}) \end{bmatrix}$$

$$m_k^- = \sum_{i=0}^{2d} W_i \cdot \hat{\mathcal{X}}_k^{(i)}$$

$$P_k^- = \sum_{i=0}^{2d} W_i \cdot \left(\hat{\mathcal{X}}_k^{(i)} - m_k^- \right) \cdot \left(\hat{\mathcal{X}}_k^{(i)} - m_k^- \right)^T + Q_k$$

 ▷ *Korrektur:*

$$\mathcal{X}_k^{(0)-} = m_k^-$$

for $i = 1, \dots, d$ **do**

$$\begin{bmatrix} \mathcal{X}_k^{(i)-} = m_k^- + \sqrt{d + \kappa} \begin{bmatrix} \sqrt{P_k^-} \end{bmatrix}_i \\ \mathcal{X}_k^{(i+d)-} = m_k^- - \sqrt{d + \kappa} \begin{bmatrix} \sqrt{P_k^-} \end{bmatrix}_i \end{bmatrix}$$

for $i = 1, \dots, 2d$ **do**

$$\begin{bmatrix} \hat{\mathcal{Y}}_k^{(i)} = h_k(\mathcal{X}_{k-1}^{(i)-}) \end{bmatrix}$$

$$\bar{y} = \sum_{i=0}^{2d} W_i \cdot \hat{\mathcal{Y}}_k^{(i)}$$

$$P_{x_k y_k} = \sum_{i=0}^{2d} W_i \cdot \left(\mathcal{X}_k^{(i)-} - m_k^- \right) \cdot \left(\hat{\mathcal{Y}}_k^{(i)} - \bar{y}_k \right)^T$$

$$P_{y_k} = \sum_{i=0}^{2d} W_i \cdot \left(\hat{\mathcal{Y}}_k^{(i)} - \bar{y}_k \right) \cdot \left(\hat{\mathcal{Y}}_k^{(i)} - \bar{y}_k \right)^T + R_k$$

$$K_k = P_{x_k y_k} \cdot P_{y_k}^{-1}$$

$$m_k = m_k^- + K_k \cdot (y_k - \bar{y})$$

$$P_k = P_k^- - K_k \cdot P_{y_k} \cdot K_k^T$$

KAPITEL 3. HERLEITUNG DER FILTER

gewichtete Summe approximiert das Integrale

$$\int_{\mathbb{R}} f(x) e^{-x^2} dx \approx \sum_{i=1}^N w_i \cdot f(\xi_i).$$

Die Stellen ξ_i an denen f ausgewertet wird, sind die Nullstellen des Hermite-Polynoms $H_N(x)$ und die Gewichte werden durch $w_i = 2^{N-1} N! \sqrt{\pi} \cdot N^{-2} \cdot H_{N-1}(\xi_i)^{-2}$ berechnet. Die Zahl $N \in \mathbb{N}$ ist ein wählbarer Parameter. Nehmen wir die Zustände als normalverteilt an, tauchen in dem Filterproblem Integrale der Form

$$\int_{\mathbb{R}^d} g(x) \mathcal{N}(x|m, P) dx = \frac{1}{\sqrt{(2\pi)^d |P|}} \int_{\mathbb{R}^d} g(x) \exp\left(-\frac{1}{2}(x-m)^T P^{-1}(x-m)\right) dx$$

auf. Um die Gauss-Hermite Integration zu verwenden, benutzen wir die Substitution $x = m + \sqrt{2P}z$ und erhalten so

$$\begin{aligned} & \frac{1}{\sqrt{(2\pi)^d |P|}} \int_{\mathbb{R}^d} g(x) \exp\left(-\frac{1}{2}(x-m)^T P^{-1}(x-m)\right) dx \\ &= \frac{1}{\sqrt{\pi^d}} \int_{\mathbb{R}^d} g\left(m + \sqrt{2P}z\right) \exp\left(-z^T z\right) dz \\ &= \frac{1}{\sqrt{\pi^d}} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} g\left(m + \sqrt{2P}z\right) e^{-z_1^2} \cdots e^{-z_d^2} dz_1 \cdots dz_d \\ &\approx \sum_{i_1, \dots, i_d=1}^N w_{i_1} \cdots w_{i_d} \cdot g\left(m + \sqrt{2P}(\xi_{i_1} \cdots \xi_{i_d})^T\right) \end{aligned}$$

Das Verfahren ist exakt für Polynomfunktionen bis zum Grad $2N - 1$ und benötigt zur Berechnung N^d Funktionsauswertungen.

Die ersten beiden Hermite-Polynome lauten $H_0(x) = 1$ und $H_1(x) = x$. Alle weiteren Polynome berechnet man mit der Rekursion

$$H_{p+1}(x) = xH_p(x) - pH_{p-1}(x).$$

Werden die Störungen nicht als additives Rauschen, sondern wie im Abschnitt 3.3 modelliert, formen wir die zu berechnenden Terme als Doppelintegrale.

In Abbildung 3.3 sehen wir, dass die Stützstellen aus der Integral-Approximation so gewählt sind, dass die gesamte Normalverteilung repräsentieren. Dadurch wird auch die transformierte Zufallsvariable gut überdeckt, sodass Erwartungswert und Kovarianzmatrix exakt berechnet werden.

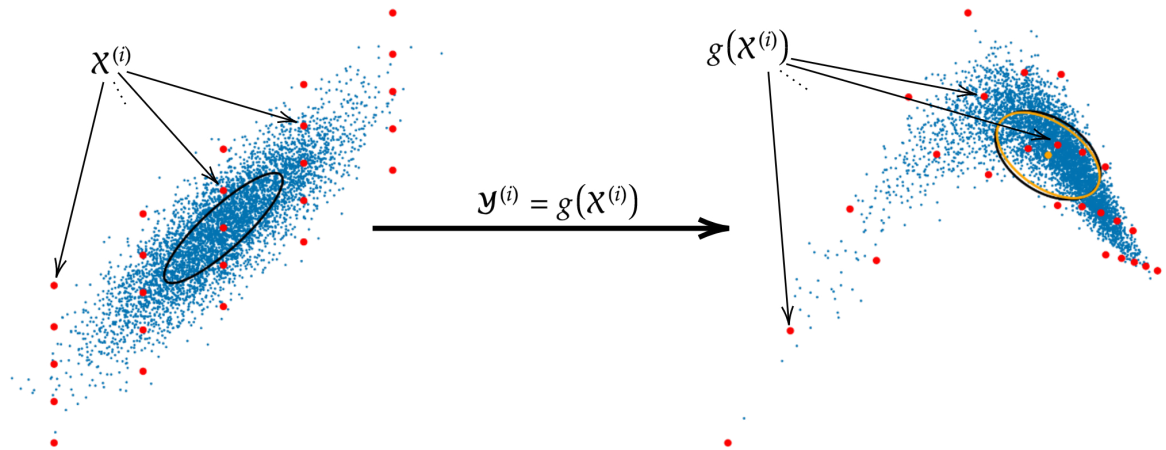


Abbildung 3.3: Veranschaulichung der nichtlinearen Transformation einer normalverteilten Zufallsvariablen im GHKF. Rote Punkte sind die zur Berechnung des Erwartungswertes und der Kovarianzmatrix verwendeten Stützstellen. Tatsächliche Größen sind Schwarz und die berechneten Werte in Orange eingezeichnet.

Vergleich der Filter

In diesem Kapitel werden verschiedene Filter verglichen. Wir fangen mit einem Vergleich des **EKF** und des **UKF** an indem wir die Approximationen der Erwartungswerte und Kovarianzmatrizen analysieren. Es folgt eine Reihe an Beispielen und Argumenten, welche zeigen wie sich der theoretische Unterschied auswirken kann. Außerdem wird erläutert, warum die theoretische Analyse nur ein Anhaltspunkt für die Filtergenauigkeit ist.

4.1 Vergleich von EKF und UKF

Der Argumentation von Julier und Uhlmann [JU96] und [Sim06, Kapitel 14] folgend, vergleichen wir nun das **EKF** mit dem **UKF**. Dazu analysieren wir die Genauigkeit der Approximation von Erwartungswert und Kovarianzmatrix der transformierten Zufallsvariablen $y = g(x)$.

Wie zuvor sei $x \sim \mathcal{N}(\bar{x}, P)$. Wir beachten, dass wir durch diese einfache Form der Transformation auch Schlussfolgerungen über Modelle mit additivem und nicht-additivem Rauschen ziehen können. Denn es gilt für die von x unabhängige ZV $q \in \mathbb{R}^n$ mit $q \sim \mathcal{N}(0, Q)$

$$\begin{aligned} \mathbb{E}[g(x) + q] &= \mathbb{E}[g(x)] + \underbrace{\mathbb{E}[q]}_{=0} \\ \text{Cov}[g(x) + q] &= \text{Cov}[g(x)] + \text{Cov}[g(x), q] + \text{Cov}[q, g(x)] + \text{Cov}[q] \\ &= \text{Cov}[g(x)] + Q \end{aligned}$$

Für eine Transformation der Form $y = g(x, q)$, mit $\mathbb{R}^{n_q} \ni q \sim \mathcal{N}(0, Q)$ und x wie zuvor, betrachten wir den erweiterten Zustand

$$x_a := \begin{pmatrix} x \\ q \end{pmatrix} \quad \text{mit} \quad x_a \sim \mathcal{N}\left(\begin{pmatrix} \bar{x} \\ 0 \end{pmatrix}, \begin{pmatrix} P & 0 \\ 0 & Q \end{pmatrix}\right)$$

Mit $y = g(x_a^{(x)}, x_a^{(q)})$ befinden wir uns dann in der gewünschten Situation. $x_a^{(x)}$ bezeichnet die ersten d und $x_a^{(q)}$ die letzten n_q Komponenten des erweiterten Zustandes x_a .

Zur Analyse schauen wir uns die Taylorreihenentwicklung von g um den Mittelwert \bar{x} an. Um die Notation zu vereinfachen definieren wir $\tilde{x} := x - \bar{x}$, sodass $\tilde{x} \sim \mathcal{N}(0, P)$. Die

KAPITEL 4. VERGLEICH DER FILTER

durch

$$D_{\tilde{x}}^j g(\bar{x}) = \left(\sum_{k=1}^d \tilde{x}_k \frac{\partial}{\partial x_k} \right)^j g(x) \Big|_{x=\bar{x}}$$

gegebene Ableitung kürzen wir mit $D^j g$ bzw. mit Dg für die erste Ableitung ab. Für hinreichend stetig differenzierbare Funktionen gilt nach dem Satz von Taylor

$$y = g(x) = g(\bar{x}) + Dg + \frac{1}{2!} D^2 g + \frac{1}{3!} D^3 g + \frac{1}{4!} D^4 g + HOT$$

In diesem Kapitel bezeichnet HOT (*higher order terms*) alle Terme fünfter und höherer Ordnung (bezüglich \tilde{x}). Für den Erwartungswert von y gilt damit

$$\bar{y} := E[y] = \underbrace{E[g(\bar{x})]}_{=g(\bar{x})} + E[Dg] + \frac{1}{2!} E[D^2 g] + \frac{1}{3!} E[D^3 g] + \frac{1}{4!} E[D^4 g] + E[HOT] \quad (4.1)$$

Bei ungeraden Ableitungen $D^{2j+1}g$ sieht man nach ausmultiplizieren, dass in jedem Summanden ein Produkt von $2j+1$ Komponenten von \tilde{x} auftaucht. Da der Zufallsvektor \tilde{x} mittelwertfrei und normalverteilt ist, gilt $p_{\tilde{x}}(x) = p_{\tilde{x}}(-x)$. Der Erwartungswert von einer Funktion $f(x)$ mit $f(-x) = -f(x)$ ist

$$\begin{aligned} E[f(x)] &= \int_{\mathbb{R}^d} f(x) p(x) dx \\ &= \int_{\mathbb{R}_{\geq 0}^d} f(x) p(x) dx + \int_{\mathbb{R}_{\leq 0}^d} f(x) p(x) dx \\ &= \int_{\mathbb{R}_{\geq 0}^d} f(x) p(x) dx + \int_{\mathbb{R}_{\geq 0}^d} \underbrace{f(-x) p(-x)}_{=-f(x) p(x)} dx \\ &= 0 \end{aligned} \quad (4.2)$$

Da $\prod_{i=1}^{2j+1} \tilde{x}_{k_i} = - \prod_{i=1}^{2j+1} -\tilde{x}_{k_i}$ gilt, fallen alle Terme ungerader Ordnung in Gleichung (4.1) weg und es gilt

$$\bar{y} = g(\bar{x}) + \frac{1}{2!} E[D^2 g] + \frac{1}{4!} E[D^4 g] + E[HOT] \quad (4.3)$$

Für die Kovarianzmatrix von y schreiben wir zunächst

$$\begin{aligned} (y - \bar{y})(y - \bar{y})^T &= \left(g(\bar{x}) + Dg + \frac{1}{2!} D^2 g + \frac{1}{3!} D^3 g + \frac{1}{4!} D^4 g + HOT \right. \\ &\quad \left. - (g(\bar{x}) + \frac{1}{2!} E[D^2 g] + \frac{1}{4!} E[D^4 g] + E[HOT]) \right) (\cdots)^T \\ &= Dg \cdot (Dg + \frac{1}{2!} D^2 g + \frac{1}{3!} D^3 g - \frac{1}{2!} E[D^2 g])^T \\ &\quad + \frac{1}{2!} D^2 g \cdot (Dg + \frac{1}{2!} D^2 g - \frac{1}{2!} E[D^2 g])^T + \frac{1}{3!} D^3 g \cdot Dg^T \\ &\quad - \frac{1}{2!} E[D^2 g] \cdot (Dg + \frac{1}{2!} D^2 g - \frac{1}{2!} E[D^2 g])^T + HOT \end{aligned}$$

KAPITEL 4. VERGLEICH DER FILTER

Hier wurden bereits alle Terme fünfter Ordnung und höher in *HOT* zusammengefasst. Bilden des Erwartungswertes und ausmultiplizieren liefert

$$\text{Cov}[y] = \mathbb{E}[(y - \bar{y})(y - \bar{y})^T] \quad (4.4)$$

$$\begin{aligned} &= \mathbb{E}[Dg \cdot Dg^T] + \frac{1}{2!} \underbrace{\mathbb{E}[Dg \cdot D^2 g^T]}_{=0} + \frac{1}{3!} \mathbb{E}[Dg \cdot D^3 g^T] \\ &\quad - \frac{1}{2!} \underbrace{\mathbb{E}[Dg]}_{=0} \cdot \mathbb{E}[D^2 g]^T + \frac{1}{2!} \underbrace{\mathbb{E}[D^2 g \cdot Dg^T]}_{=0} + \frac{1}{2!2!} \mathbb{E}[D^2 g \cdot D^2 g^T] \\ &\quad - \frac{1}{2!2!} \mathbb{E}[D^2 g] \cdot \mathbb{E}[D^2 g]^T + \frac{1}{3!} \mathbb{E}[D^3 g \cdot Dg^T] - \frac{1}{2!} \mathbb{E}[D^2 g] \cdot \underbrace{\mathbb{E}[Dg]^T}_{=0} \\ &\quad - \underbrace{\frac{1}{2!2!} \mathbb{E}[D^2 g] \cdot \mathbb{E}[D^2 g]^T + \frac{1}{2!2!} \mathbb{E}[D^2 g] \cdot \mathbb{E}[D^2 g]^T}_{=0} + HOT \\ &= \mathbb{E}[Dg \cdot Dg^T] + \frac{1}{3!} \mathbb{E}[Dg \cdot D^3 g^T] + \frac{1}{2!2!} \mathbb{E}[D^2 g \cdot D^2 g^T] \\ &\quad - \frac{1}{2!2!} \mathbb{E}[D^2 g] \cdot \mathbb{E}[D^2 g]^T + \frac{1}{3!} \mathbb{E}[D^3 g \cdot Dg^T] + HOT \end{aligned} \quad (4.5)$$

Dabei sind wieder einige Terme nach Gleichung (4.2) gleich null. Für den ersten Summanden in (4.5) gilt außerdem

$$\begin{aligned} \mathbb{E}[Dg \cdot Dg^T] &= \mathbb{E} \left[\left(\sum_{i=1}^d \tilde{x}_i \frac{\partial g(x)}{\partial x_i} \bigg|_{x=\bar{x}} \right) (\cdots)^T \right] \\ &= \mathbb{E} \left[\sum_{i=1}^d \sum_{j=1}^d \tilde{x}_i \frac{\partial g(x)}{\partial x_i} \frac{\partial g(x)}{\partial x_j}^T \tilde{x}_j^T \right] \\ &= \sum_{i,j=1}^d G_i \mathbb{E}[\tilde{x}_i \tilde{x}_j^T] G_j^T \\ &= GPG^T \end{aligned} \quad (4.6)$$

,wobei G_i als die i -te Spalte der Jacobimatrix von g an der Stelle \bar{x} definiert ist. Damit erhalten wir die alternative Schreibweise für die Kovarianzmatrix

$$\begin{aligned} \text{Cov}[y] &= GPG^T + \frac{1}{3!} \left(\mathbb{E}[Dg \cdot D^3 g^T] + \mathbb{E}[D^3 g \cdot Dg^T] \right) \\ &\quad + \frac{1}{2!2!} \left(\mathbb{E}[D^2 g \cdot D^2 g^T] - \mathbb{E}[D^2 g] \cdot \mathbb{E}[D^2 g]^T \right) + HOT \end{aligned} \quad (4.7)$$

Im **EKF** werden $\bar{y}_e = g(\bar{x})$ und $\text{Cov}_e[y] = GPG^T$ verwendet. Der Erwartungswert enthält also Fehler zweiter Ordnung und die Kovarianzmatrix enthält Fehler vierter Ordnung. Nun schauen wir uns die Taylorreihenentwicklung in der Unscented Transformation an. Zunächst seien für $i = 1, \dots, d$ $\sigma^{(i)} := \sqrt{d + \kappa} \left[\sqrt{P} \right]_i$ und die Sigma-Punkte sind definiert als $\mathcal{X}^{(0)} = \bar{x}$, $\mathcal{X}^{(i)} = \mathcal{X}^{(0)} + \sigma^{(i)}$ und $\mathcal{X}^{(i+n)} = \mathcal{X}^{(0)} - \sigma^{(i)}$. Hier wird der Mittelwert von y durch $\bar{y}_u = \sum_{i=0}^{2d} W_i \mathcal{Y}_i$ berechnet. Es gilt für hinreichen stetig differenzierbare Funktionen

KAPITEL 4. VERGLEICH DER FILTER

g nach dem Satz von Taylor

$$\begin{aligned}\sum_{i=0}^{2d} W_i \mathcal{Y}_i &= \sum_{i=0}^{2d} W_i g(\mathcal{X}_i) \\ &= \sum_{i=0}^{2d} W_i \left(g(\bar{x}) + D_{\sigma^{(i)}} g(\bar{x}) + \frac{1}{2!} D_{\sigma^{(i)}}^2 g(\bar{x}) \right. \\ &\quad \left. + \frac{1}{3!} D_{\sigma^{(i)}}^3 g(\bar{x}) + \frac{1}{4!} D_{\sigma^{(i)}}^4 g(\bar{x}) + HOT \right)\end{aligned}$$

Die Summe der Gewichte W_i ist 1, weshalb wir den Term $g(\bar{x})$ vor die Summe schreiben können. \mathcal{X}_0 ist gleich \bar{x} , also ist $\sigma^{(0)} = 0$ und wir erhalten

$$\sum_{i=0}^{2d} W_i \mathcal{Y}_i = g(\bar{x}) + \sum_{i=1}^{2d} W_i \left(D_{\sigma^{(i)}} g(\bar{x}) + \frac{D_{\sigma^{(i)}}^2 g(\bar{x})}{2!} + \frac{D_{\sigma^{(i)}}^3 g(\bar{x})}{3!} + \frac{D_{\sigma^{(i)}}^4 g(\bar{x})}{4!} + HOT \right) \quad (4.8)$$

Man beachte, dass

$$\sigma^{(i+d)} := \mathcal{X}^{(i+d)} - \bar{x} = -\sigma^{(i)} \quad \text{für } i = 1, \dots, d \quad (4.9)$$

gilt. Aus dieser Symmetrie folgt

$$\begin{aligned}D_{\sigma^{(i+d)}}^{2j+1} g(\bar{x}) &= \left(\sum_{i=1}^d \sigma^{(i+d)} \frac{\partial}{\partial x_k} \right)^{2j+1} g(x) \Big|_{x=\bar{x}} \\ &= \left(\sum_{i=1}^d -\sigma^{(i)} \frac{\partial}{\partial x_k} \right)^{2j+1} g(x) \Big|_{x=\bar{x}} \\ &= - \left(\sum_{i=1}^d \sigma^{(i)} \frac{\partial}{\partial x_k} \right)^{2j+1} g(x) \Big|_{x=\bar{x}} \\ &= -D_{\sigma^{(i)}}^{2j+1} g(\bar{x})\end{aligned} \quad (4.10)$$

für $j \in \mathbb{N}_0$, falls die entsprechende Ableitung existiert. Wir erhalten nun mit $W_i = W_{i+d}$

$$\begin{aligned}\sum_{i=1}^{2d} W_i D_{\sigma^{(i)}}^{2j+1} g(\bar{x}) &= \sum_{i=1}^d W_i D_{\sigma^{(i)}}^{2j+1} g(\bar{x}) + \sum_{i=1}^d W_i D_{\sigma^{(i+d)}}^{2j+1} g(\bar{x}) \\ &= 0\end{aligned}$$

, d.h. alle ungeraden Ableitungen in (4.8) fallen weg. Des Weiteren stellen wir fest, dass

$$\begin{aligned}\sum_{i=1}^{2d} W_i \frac{D_{\sigma^{(i)}}^2 g(\bar{x})}{2!} &= \frac{1}{2!} \sum_{i=1}^{2d} W_i \left(\sum_{k=1}^d \sigma_k^{(i)} \frac{\partial}{\partial x_k} \right)^2 g(x) \Big|_{x=\bar{x}} \\ &= \frac{1}{2!} \sum_{i=1}^{2d} W_i \sum_{k=1}^d \sum_{j=1}^d \sigma_k^{(i)} \sigma_j^{(i)} \frac{\partial^2}{\partial x_k \partial x_j} g(x) \Big|_{x=\bar{x}} \\ &= \frac{1}{2!} \sum_{k,j=1}^d \sum_{i=1}^{2d} W_i \sigma_k^{(i)} \sigma_j^{(i)} \frac{\partial^2}{\partial x_k \partial x_j} g(x) \Big|_{x=\bar{x}}\end{aligned}$$

KAPITEL 4. VERGLEICH DER FILTER

Wir sehen wieder mit der Symmetrie aus (4.9) und $W_i = W_{i+d}$, dass

$$\begin{aligned} \sum_{i=1}^{2d} W_i \sigma_k^{(i)} \sigma_j^{(i)} &= \sum_{i=1}^d W_i \sigma_k^{(i)} \sigma_j^{(i)} + \sum_{i=1}^d W_{i+d} \sigma_k^{(i+d)} \sigma_j^{(i+d)} \\ &= \sum_{i=1}^d W_i \sigma_k^{(i)} \sigma_j^{(i)} + \sum_{i=1}^d W_i (-\sigma_k^{(i)}) (-\sigma_j^{(i)}) \\ &= 2 \cdot \sum_{i=1}^d W_i \sigma_k^{(i)} \sigma_j^{(i)} \end{aligned}$$

gilt. Setzen wir $\sigma_k^{(i)} = \sqrt{(d+\kappa)}[\sqrt{P}]_{k,i}$ und $W_i = \frac{1}{2(d+\kappa)}$ (für $i = 1, \dots, d$) ein, erhalten wir

$$\begin{aligned} 2 \cdot \sum_{i=1}^d W_i \sigma_k^{(i)} \sigma_j^{(i)} &= 2 \cdot \sum_{i=1}^d \underbrace{W_i \cdot (d+\kappa)}_{=1/2} [\sqrt{P}]_{k,i} [\sqrt{P}]_{j,i} \\ &= P_{k,j} \end{aligned} \tag{4.11}$$

Zusammen haben wir nun

$$\begin{aligned} \sum_{i=1}^{2d} W_i \frac{D_{\sigma^{(i)}}^2 g(\bar{x})}{2!} &= \frac{1}{2!} \sum_{k,j=1}^d P_{k,j} \frac{\partial^2}{\partial x_k \partial x_j} g(x) \Big|_{x=\bar{x}} \\ &= \frac{1}{2!} \sum_{k,j=1}^d \mathbb{E}[\tilde{x}_k \tilde{x}_j] \frac{\partial^2}{\partial x_k \partial x_j} g(x) \Big|_{x=\bar{x}} \\ &= \frac{1}{2!} \mathbb{E}[D^2 g] \end{aligned} \tag{4.12}$$

gezeigt. Damit erhalten wir den in der Unscented Transformation berechneten Erwartungswert

$$\bar{y}_u = g(\bar{x}) + \frac{1}{2!} \mathbb{E}[D_{\bar{x}}^2 g(\bar{x})] + \sum_{i=1}^{2d} W_i \left(\frac{D_{\sigma^{(i)}}^4 g(\bar{x})}{4!} + HOT \right) \tag{4.13}$$

Der Erwartungswert stimmt also bis zur dritten Ordnung mit dem exakten Erwartungswert aus (4.3) überein. Im Gegensatz zur Approximation im **EKF** wurde hier auch in der Berechnung des Erwartungswertes die Kovarianzmatrix P beachtet.

Nun schauen wir uns die Kovarianzmatrix aus der Unscented Transformation an. Dazu stellen wir zunächst fest, dass

$$\begin{aligned} \mathcal{Y}_i - \bar{y}_u &= g(\mathcal{X}_i) - \bar{y}_u \\ &= \left(g(\bar{x}) + D_{\sigma^{(i)}} g(\bar{x}) + \frac{D_{\sigma^{(i)}}^2 g(\bar{x})}{2!} + \frac{D_{\sigma^{(i)}}^3 g(\bar{x})}{3!} + \frac{D_{\sigma^{(i)}}^4 g(\bar{x})}{4!} + HOT \right) \\ &\quad - \left(g(\bar{x}) + \sum_{j=1}^{2d} W_j \left(\frac{D_{\sigma^{(j)}}^2 g(\bar{x})}{2!} + \frac{D_{\sigma^{(j)}}^4 g(\bar{x})}{4!} + HOT \right) \right) \end{aligned}$$

KAPITEL 4. VERGLEICH DER FILTER

gilt. Dann ist die Kovarianzmatrix

$$\begin{aligned}
\text{Cov}_u[y] &= \mathbb{E}_u[(\mathcal{Y}_i - \bar{y}_u)(\mathcal{Y}_i - \bar{y}_u)^T] := \sum_{i=0}^{2d} W_i (\mathcal{Y}_i - \bar{y}_u)(\mathcal{Y}_i - \bar{y}_u)^T \\
&= \sum_{i=1}^{2d} W_i \left(D_{\sigma(i)} g(\bar{x}) + \frac{1}{2!} D_{\sigma(i)}^2 g(\bar{x}) + \frac{1}{3!} D_{\sigma(i)}^3 g(\bar{x}) + \frac{1}{4!} D_{\sigma(i)}^4 g(\bar{x}) \right. \\
&\quad \left. - \sum_{j=1}^{2d} W_j \left(\frac{1}{2!} D_{\sigma(j)}^2 g(\bar{x}) + \frac{1}{4!} D_{\sigma(j)}^4 g(\bar{x}) \right) + HOT \right) (\cdots)^T \\
&= \sum_{i=1}^{2d} W_i \left[D_{\sigma(i)} g(\bar{x}) (D_{\sigma(i)} g(\bar{x}))^T + \frac{1}{2!} D_{\sigma(i)} g(\bar{x}) (D_{\sigma(i)}^2 g(\bar{x}))^T \right. \\
&\quad + \frac{1}{3!} D_{\sigma(i)} g(\bar{x}) (D_{\sigma(i)}^3 g(\bar{x}))^T + \frac{1}{2!} D_{\sigma(i)} g(\bar{x}) \left(\sum_{j=1}^{2d} W_j D_{\sigma(j)}^2 g(\bar{x}) \right)^T \\
&\quad + \frac{1}{2!} D_{\sigma(i)}^2 g(\bar{x}) (D_{\sigma(i)} g(\bar{x}))^T + \frac{1}{2!2!} D_{\sigma(i)}^2 g(\bar{x}) (D_{\sigma(i)}^2 g(\bar{x}))^T \\
&\quad - \frac{1}{2!2!} D_{\sigma(i)}^2 g(\bar{x}) \left(\sum_{j=1}^{2d} W_j D_{\sigma(j)}^2 g(\bar{x}) \right)^T + \frac{1}{3!} D_{\sigma(i)}^3 g(\bar{x}) (D_{\sigma(i)} g(\bar{x}))^T \\
&\quad - \frac{1}{2!} \left(\sum_{j=1}^{2d} W_j D_{\sigma(j)}^2 g(\bar{x}) \right) (D_{\sigma(i)} g(\bar{x}))^T - \frac{1}{2!2!} \left(\sum_{j=1}^{2d} W_j D_{\sigma(j)}^2 g(\bar{x}) \right) (D_{\sigma(i)}^2 g(\bar{x}))^T \\
&\quad \left. + \frac{1}{2!2!} \left(\sum_{j=1}^{2d} W_j D_{\sigma(j)}^2 g(\bar{x}) \right) \left(\sum_{j=1}^{2d} W_j D_{\sigma(j)}^2 g(\bar{x}) \right)^T \right] + HOT.
\end{aligned}$$

Dabei sind die ungeraden Terme nach der Summation wieder wegen der Symmetrie (4.9) gleich null. Weiter folgt

$$\begin{aligned}
\text{Cov}_u[y] &= \sum_{i=1}^{2d} W_i \left[D_{\sigma(i)} g(\bar{x}) (D_{\sigma(i)} g(\bar{x}))^T \right. \\
&\quad + \sum_{i=1}^{2d} W_i \left[\frac{1}{3!} D_{\sigma(i)} g(\bar{x}) (D_{\sigma(i)}^3 g(\bar{x}))^T + \frac{1}{2!2!} D_{\sigma(i)}^2 g(\bar{x}) (D_{\sigma(i)}^2 g(\bar{x}))^T \right. \\
&\quad \left. + \frac{1}{3!} D_{\sigma(i)}^3 g(\bar{x}) (D_{\sigma(i)} g(\bar{x}))^T - \frac{1}{2!2!} D_{\sigma(i)}^2 g(\bar{x}) \left(\sum_{j=1}^{2d} W_j D_{\sigma(j)}^2 g(\bar{x}) \right)^T \right] \\
&\quad - \frac{1}{2!2!} \sum_{i=1}^{2d} W_i \left[\left(\sum_{j=1}^{2d} W_j D_{\sigma(j)}^2 g(\bar{x}) \right) (D_{\sigma(i)}^2 g(\bar{x}))^T \right] \\
&\quad \left. + \frac{1}{2!2!} \sum_{i=1}^{2d} W_i \left[\left(\sum_{j=1}^{2d} W_j D_{\sigma(j)}^2 g(\bar{x}) \right) \left(\sum_{j=1}^{2d} W_j D_{\sigma(j)}^2 g(\bar{x}) \right)^T \right] + HOT.
\end{aligned}$$

Wegen

$$\sum_{i=1}^{2d} W_i \left[\left(\sum_{j=1}^{2d} W_j D_{\sigma(j)}^2 g(\bar{x}) \right) (D_{\sigma(i)}^2 g(\bar{x}))^T \right] = \left(\sum_{j=1}^{2d} W_j D_{\sigma(j)}^2 g(\bar{x}) \right) \left(\sum_{i=1}^{2d} W_i D_{\sigma(i)}^2 g(\bar{x}) \right)^T$$

und da die Summe der Gewichte W_i eins ist, heben sich die letzten beiden Summanden auf und wir erhalten schließlich

$$\begin{aligned} \text{Cov}_u[y] &= \sum_{i=1}^{2d} W_i \left[D_{\sigma^{(i)}} g(\bar{x}) (D_{\sigma^{(i)}} g(\bar{x}))^T \right] \\ &\quad + \sum_{i=1}^{2d} W_i \left[\frac{1}{3!} D_{\sigma^{(i)}} g(\bar{x}) (D_{\sigma^{(i)}}^3 g(\bar{x}))^T + \frac{1}{2!2!} D_{\sigma^{(i)}}^2 g(\bar{x}) (D_{\sigma^{(i)}}^2 g(\bar{x}))^T \right. \\ &\quad \left. + \frac{1}{3!} D_{\sigma^{(i)}}^3 g(\bar{x}) (D_{\sigma^{(i)}} g(\bar{x}))^T - \frac{1}{2!2!} D_{\sigma^{(i)}}^2 g(\bar{x}) \left(\sum_{j=1}^{2d} W_j D_{\sigma^{(j)}}^2 g(\bar{x}) \right)^T \right] + HOT \end{aligned} \quad (4.14)$$

Es bleiben also der erste Summand und einige Terme vierter Ordnung. Für diesen ersten Summanden gilt ähnlich wie in (4.6)

$$\begin{aligned} \sum_{i=1}^{2d} W_i \left[D_{\sigma^{(i)}} g(\bar{x}) (D_{\sigma^{(i)}} g(\bar{x}))^T \right] &= \sum_{i=1}^{2d} W_i \left[\sum_{j,k=1}^d \frac{\partial g(x)}{\partial x_j} \tilde{x}_j \tilde{x}_k^T \frac{\partial g(x)}{\partial x_k} \right] \\ &= \sum_{i=1}^{2d} W_i \left[\sum_{j,k=1}^d G_j P_{j,k} G_k^T \right] \\ &= G P G^T \end{aligned} \quad (4.15)$$

Nach Einsetzen von (4.15) in (4.14) und weiterem Umordnen erhalten wir

$$\begin{aligned} \text{Cov}_u[y] &= G P G^T + \frac{1}{3!} \sum_{i=1}^{2d} W_i \left[D_{\sigma^{(i)}} g(\bar{x}) (D_{\sigma^{(i)}}^3 g(\bar{x}))^T + D_{\sigma^{(i)}}^3 g(\bar{x}) (D_{\sigma^{(i)}} g(\bar{x}))^T \right. \\ &\quad \left. + \frac{1}{2!2!} D_{\sigma^{(i)}}^2 g(\bar{x}) (D_{\sigma^{(i)}}^2 g(\bar{x}))^T - D_{\sigma^{(i)}}^2 g(\bar{x}) \left(\sum_{j=1}^{2d} W_j D_{\sigma^{(j)}}^2 g(\bar{x}) \right)^T \right] + HOT \end{aligned} \quad (4.16)$$

Ein Vergleich mit der Taylorreihenentwicklung der echten Kovarianz (4.7) zeigt, dass die Terme bis zur dritten Ordnung übereinstimmen. Die Erwartungswerte in (4.7) entsprechen bis auf eine Ausnahme augenscheinlich den gewichteten Stichprobenmitteln in (4.16). Wir haben allerdings keine Aussagen über die Genauigkeit dieser Approximationen getroffen. In [JUDW00] werden Fehler höherer Ordnung weiter untersucht. Es wird gezeigt, dass die Wahl $\kappa = 3 - d$ genannte Fehler vierter Ordnung für normalverteilte Zustände minimiert. Man kann insgesamt erwarten, dass das **UKF** gegenüber dem **EKF** eine genauere Kovarianzmatrix liefert, obwohl beide Berechnungen nur für lineare Transformationen exakt sind [Sim06]. Wir beachten an dieser Stellen, dass die Kovarianz bis zur vierten Ordnung exakt der Taylorreihenentwicklung entsprechen muss, um für quadratische Transformationen exakt zu sein.

4.1.1 Diskussion

In diesem Kapitel haben wir zunächst die Taylorreihenentwicklung einer nichtlinearen Funktion g und damit Erwartungswert und Kovarianz der transformierten Zufallsvariablen z mit $z = g(x)$ angeschaut. Ein Vergleich mit der Berechnung dieser beiden Größen

im **EKF** zeigt, dass im Erwartungswert Fehler zweiter Ordnung und bei der Kovarianz Fehler vierter Ordnung auftreten. Dieses Resultat ist nicht überraschend, da das **EKF** gerade durch Abschneiden höherer Terme der Taylorreihenentwicklung von g an der Stelle \bar{x} hergeleitet wurde. Die Umformungen der Berechnung des Erwartungswertes mit der **UT** (4.13) hat einen Vergleich mit der Taylorreihenentwicklung des exakten Erwartungswertes (4.3) ermöglicht. Dieser hat gezeigt, dass der Erwartungswert für Polynomfunktionen dritten Grades exakt berechnet wird. Diese Verbesserung gegenüber dem **EKF** ist maßgeblich, da der prädizierte Mittelwert in den meisten weiteren Berechnungen verwendet wird. In (4.10) bis (4.12) konnten wir sehen, wie die spezielle Struktur der Gewichte W_i und Sigma-Punkte \mathcal{X}_i ausgenutzt wurde, um diese Verbesserung zu erzielen. Ein wichtiger Unterschied zum **EKF** ist die Abhängigkeit der Sigma-Punkte von der Kovarianzmatrix. In der theoretischen Betrachtung wird die exakte Matrix P verwendet. Im Filter wird jedoch die zuvor geschätzte Kovarianzmatrix verwendet. Durch diese Abweichung gilt Gleichung (4.11) nicht mehr und es treten wie im **EKF** Fehler zweiter Ordnung auf. Im nächsten Kapitel gehen wir weiter auf Schwierigkeiten des theoretischen Vergleiches ein. Die Umformungen zur Kovarianz-Berechnung in der **UT** zeigt zunächst, dass die Kovarianzschätzung, genau wie die Schätzung im **EKF**, Fehler vierter Ordnung enthält. Die Tatsache, dass in der Berechnung für das **EKF** in Gleichungen (3.14) und (3.15) keine höheren Terme berücksichtigt werden, jedoch in der **UT** viele weitere Terme höherer Ordnung mithilfe der Sigma-Punkte erfasst werden, lässt eine höhere Genauigkeit erwarten [JUDW00, Sim06]. Über die Kovarianzen von Zustand und Beobachtung können ähnliche Aussagen getroffen werden, die hier nicht gezeigt wurden. Des Weiteren begründen die Ausführungen zur **UT** die Wahl der $2d + 1$ Sigma-Punkte und der zugehörigen Gewichte in Abschnitt 3.5.1. Die konkrete Wahl der Gewichte wurde dabei nur in (4.11) ausgenutzt. Man erahnt, dass andere Sigma-Punkte und Gewichte ähnliche Approximationen liefern. Eine Klasse von Filtern, die Erwartungswert und Kovarianz mithilfe deterministischer Sigma-Punkte schätzen, nennt man auch *Sigma-Punkt Kalman-Filter* [VDM04].

4.2 Vergleich mit weiteren Filtern

In Kapitel 4.1 haben wir zum Vergleichen der Filter die Transformation einer Zufallsvariablen betrachtet. Das **EKF** und das **UKF** sind beides approximative Gauß Filter. Als solche können wir die Genauigkeit direkt mit einem Gauß-Hermite Kalman-Filter vergleichen. Die Gauß-Hermite Integration berechnet alle Erwartungswerte aus dem **GF** exakt bis zur Ordnung $2N - 1$. Da wir für die Integration N^d Sigma-Punkte benötigen, wobei d die Länge des Zustandsvektors ist, kann das **GHKF** nur in kleinen Modellen angewendet werden. Außerdem können wir die theoretische Lösung des Gauß Filters offline annähern und so nützliche Informationen zum verbessern eines effizienteren Filters erhalten.

Die Berechnung der Filterverstärkung K_k und der korrigierten Kovarianzmatrix sind bei den Filtern, welche normalverteilte Zufallsvariablen annehmen, identisch. Da diese Annahme für nichtlineare Systeme nie exakt ist, ist die Filterverstärkung nicht optimal und die Kovarianzmatrix enthält weitere Fehler. Für kleine Abstraten und Modelle die sich fast linear verhalten, sind diese Fehler oft klein und die Filter funktionieren gut. Verhält sich das Modell stark nichtlinear, wird die Annahme verletzt und die entstehende Abweichung kann die zuvor betrachteten Fehler dominieren. Damit lässt sich erklären, warum das **UKF** nicht immer bessere Ergebnisse als das **EKF** erzielt.

Für reale Systeme ist neben der Genauigkeit auch die Laufzeit der Filter ein entscheidendes Kriterium. In jedem Prädiktions- und Korrektur-Zyklus müssen die Funktionen f und

h ausgewertet werden. Das **EKF** wertet diese Funktionen jeweils nur einmal aus. In den anderen vorgestellten Filter werden die Funktionen entsprechend der Anzahl der verwendeten Sigma-Punkte oder Partikel ausgewertet. Im **UKF** hängt Anzahl von der Länge des Zustandsvektors ab, während die Anzahl von Partikeln stark variieren kann und für jedes Modell passend gewählt werden muss. Ein Nachteil des **EKF** ist die Berechnung der Kovarianzen, denn hier müssen Ableitungen an den geschätzten Zuständen ausgewertet werden. Wie stark sich dies auf die Laufzeit auswirkt, hängt stark vom Modell ab. Im Extremfall ist die Verwendung von Ableitungen ein Ausschlusskriterium für das **EKF**. Im **UKF** ist neben der Funktionsauswertungen das Invertieren der Kovarianzmatrizen für die Laufzeit signifikant.

Zusammengefasst ist das **EKF** am effizientesten, dicht gefolgt vom **UKF**. Ob die höhere Genauigkeit relevant oder sogar notwendig ist, ist immer modellabhängig. Falls die Verteilungen nicht gut durch Normalverteilungen, oder einigen anderen speziellen Verteilungen, approximiert wird, verwendet man oft ein Partikel Filter.

Approximative Gauß Filter, die aufgrund ihrer Struktur auch manchmal erweiterte Kalman-Filter genannt werden zählen zu den wichtigsten Filter Algorithmen. Neben der Tatsache, dass Zustände in der Natur oft annähernd normalverteilt sind, können wir auch andere Verteilungen bzw. Dichtefunktionen mittels einer Linearkombination von normalverteilten Dichten approximieren. In den sogenannten *Mixed Gaussian* Filtern können die einzelnen Normalverteilungen durch die vorgestellten Gauß Filter prädictiert und korrigiert werden [IX00]. Die Schwierigkeit liegt dann darin, die Anzahl der Summanden zu wählen und die Gewichte zu bestimmen. Durch diese Annäherung allgemeiner Dichten wird der Anwendungsbereich der **GF** nochmals erheblich vergrößert.

4.3 Anwendung und empirischer Vergleich

Die in Kapitel 3 vorgestellten wurden in Abschnitten 4.1 und 4.2 auf theoretischer Grundlage und der gegebenen Rechenvorschriften verglichen. Im Folgenden vergleichen wir die Schätzgüte und die verwendete Rechenzeit anhand eines einfachen Beispiels.

Dazu schauen wir uns das Van-der-Pol System an. Dieses ist durch das dynamische System

$$\begin{aligned}\dot{x}_1 &= x_1 \\ \dot{x}_2 &= \mu \cdot (1 - x_1^2)x_2 - x_1\end{aligned}\tag{4.17}$$

gegeben. Durch das explizite Euler-Verfahren erhalten wir ein einfaches diskretes Modell. Der Systemzustand wird in jedem Zeitschritt durch additives, normalverteiltes Rauschen v_k mit $v_k \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.0025 & 0 \\ 0 & 0.01 \end{pmatrix}\right)$ und mit einer Abtastrate $T = 0.1$ gestört.

Zusammen erhalten wir das Systemmodell:

$$\begin{pmatrix} x_{1,k+1} \\ x_{2,k+1} \end{pmatrix} = \underbrace{\begin{pmatrix} x_{1,k} + T \cdot x_{1,k} \\ x_{2,k} + T \cdot (\mu(1 - x_{1,k}^2)x_{2,k} - x_{1,k}) \end{pmatrix}}_{f_k(x_k)} + v_k$$

Für dieses Beispiel wird die erste Komponente x_1 des Zustandsvektors direkt gemessen. Die Beobachtungen werden ebenfalls durch additives und normalverteiltes Rauschen $w_k \sim$

KAPITEL 4. VERGLEICH DER FILTER

$\mathcal{N}(0, 1)$ gestört. Das Beobachtungsmodell lautet also

$$y_k = \underbrace{x_{1,k}}_{h_k(x_k)} + w_k$$

Als initiale Schätzung wählen wir $x_0 \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.5 & 0 \\ 0 & 1 \end{pmatrix}\right)$. Hier wurde als Mittelwert ein beliebiger Punkt in der Nähe der üblichen Trajektorie des Van-der-Pol Systems gewählt. Außerdem wurde eine Kovarianzmatrix mit hinreichend großen Varianzen gewählt. Dadurch werden die Filterschätzungen sich zunächst stärker den Messdaten anpassen, bis die geschätzte Unsicherheit kleiner wird. Diese erste Phase der Zustandsschätzung wird auch *Initialisierungsphase* genannt.

Die simulierte Trajektorie wurde mit einem Runge-Kutta Verfahren fünfter Ordnung simuliert. Dabei wurde der Zustand in jedem Zeitschritt mit dem angegebenen Systemrauschen gestört. Dadurch ist das für die Filter verwendete Modell nicht exakt und stellt neben dem Rauschen eine weitere Fehlerquelle dar. Da wir für reale Systeme auch oft keine exakten Modelle verwenden können, stellen wir ein realistisches Szenario dar.

Für die Filter musste aufgrund der genannten Fehler die Kovarianzmatrix des Systemrauschen mit dem Faktor 1.5 und die Kovarianz des Messrauschen mit dem Faktor 1.1 vergrößert werden.

Als Filterparameter wurden für das **UKF** $\kappa = 1$, für das **GHKF** $N = 10$ und für das Partikel-Filter wurden 500 Partikel und $N_{eff} = 125$ gewählt.

Das Bild 4.1 zeigt exemplarisch die ersten 80 Zeitschritte des Systemverlaufs zusammen mit Schätzungen verschiedener Filter.

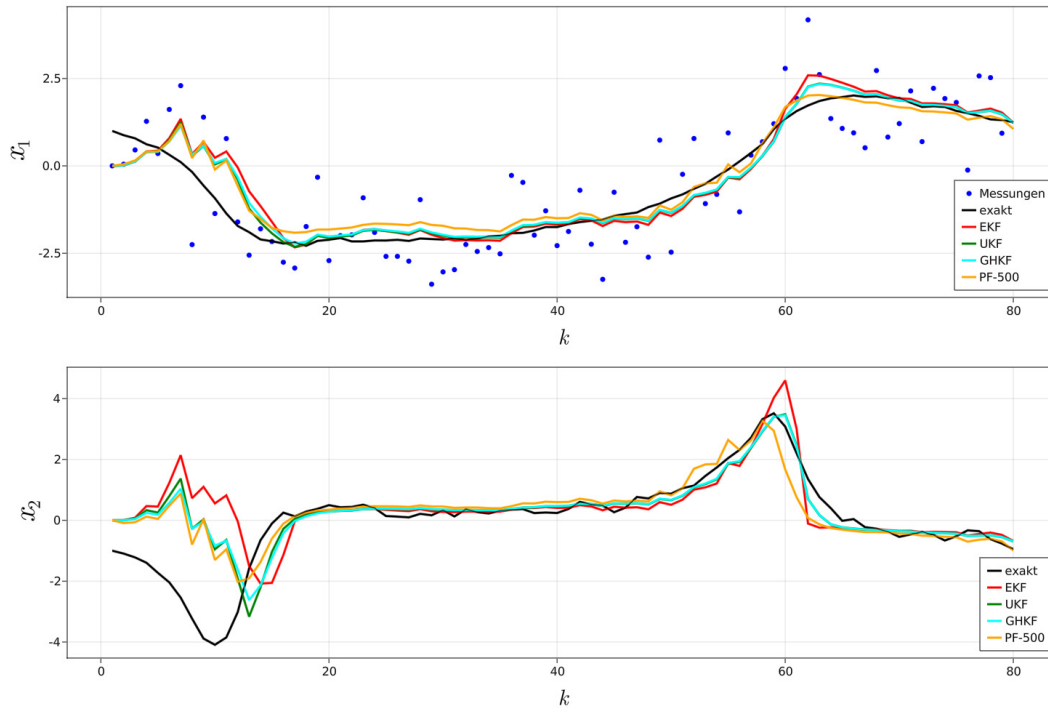


Abbildung 4.1: Initialisierungsphase der Filter für das Van-der-Pol System

Die echte Trajektorie startet bei $x = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Wir sehen, dass alle Filter, trotz der stark

KAPITEL 4. VERGLEICH DER FILTER

verrauschten Messungen, die echte Trajektorie nach circa 20 Zeitschritten gut schätzen. In weiteren Simulationen ist aufgefallen, dass das **EKF** für manche Anfangswerte deutlich größere Abweichungen in der Initialisierungsphase aufweist.

Bild 4.2 zeigt die durchschnittliche Abweichung der Filterschätzungen zum tatsächlichen Zustand.

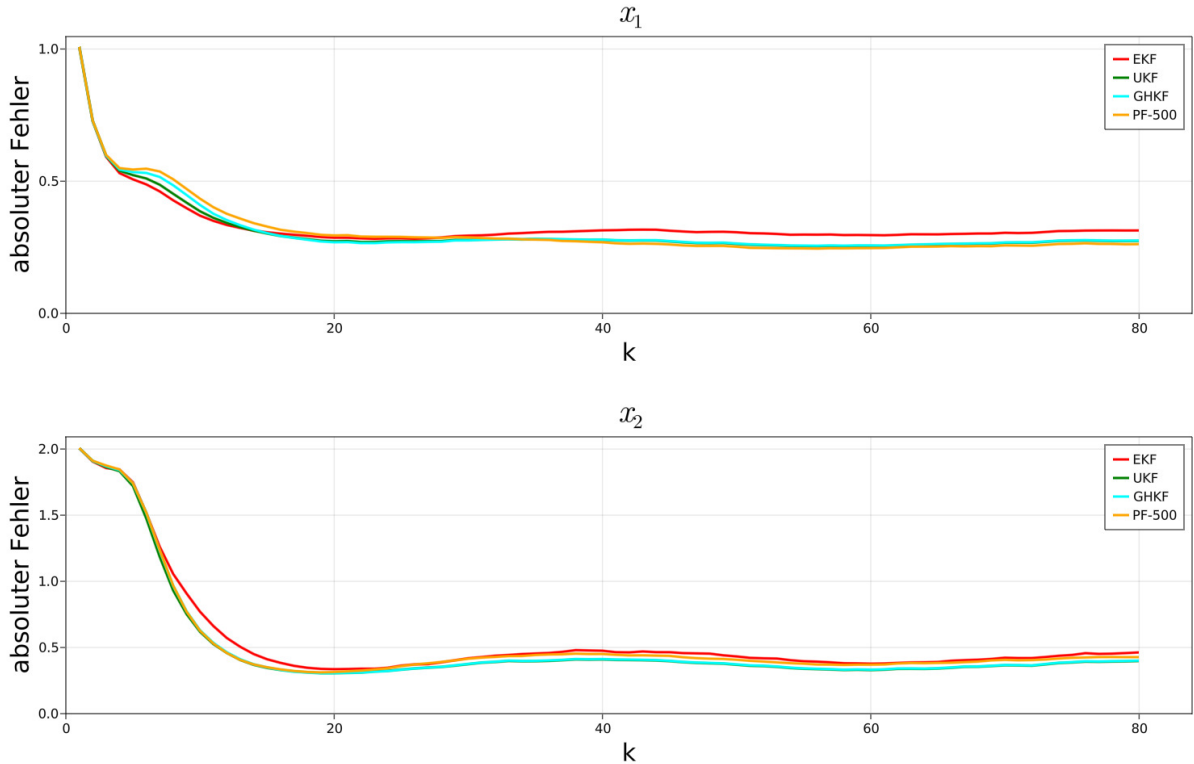


Abbildung 4.2: Durchschnittlicher Fehler der Punktschätzungen in der Initialisierungsphase

Hier können wir jedoch erkennen, dass der mittlere Fehler in den ersten Zeitschritten bei allen Filtern fast identisch ist. Für dieses Experiment wurden zusätzlich zu den Rauschprozessen der echte Anfangswert mit $x_{1,0} \sim \mathcal{U}([-2, 2])$ und $x_{1,0} \sim \mathcal{U}([-4, 4])$ zufällig generiert. Nach 10.000 Durchläufen wurde der dargestellte *root-mean-square error* (**RMSE**) der Punktschätzung von der simulierten Trajektorie berechnet.

Nun schauen wir uns die Filtergenauigkeit nach der Initialisierungsphase an. Bild 4.3 zeigt einen typischen Ausschnitt nach 1800 Zeitschritten.

Man sieht, dass die direkt beobachtete Komponente x_1 von allen Filtern gut geschätzt wird. In der zweiten Komponente fallen größere Abweichungen bei der Schätzung durch das **EKF** auf. Diese größeren Fehler treten immer an Stellen auf, an denen sich das System nicht annähernd linear verhält. Die Linearisierung der Systemfunktion führt hier also zu signifikanten Fehlern.

Tabelle 4.1 listet schließlich die mittleren Fehler der Algorithmen und die verwendete Rechenzeit für die Parameterschätzung 1.000.000 Zeitschritte.

Wie erwartet schätzt das **PF** den Zustand am besten, gefolgt von dem **UKF** und dem **GHKF**. Bemerkenswert ist, dass das Unscented Kalman-Filter dieselbe Genauigkeit wie das Gauß-Hermite Kalman-Filter aufweist, obwohl letzteres 100 Sigma-Punkte zur Schätzung verwendet. Auch die einzelnen Schätzungen liegen fast immer nah beieinander. Wäh-

KAPITEL 4. VERGLEICH DER FILTER

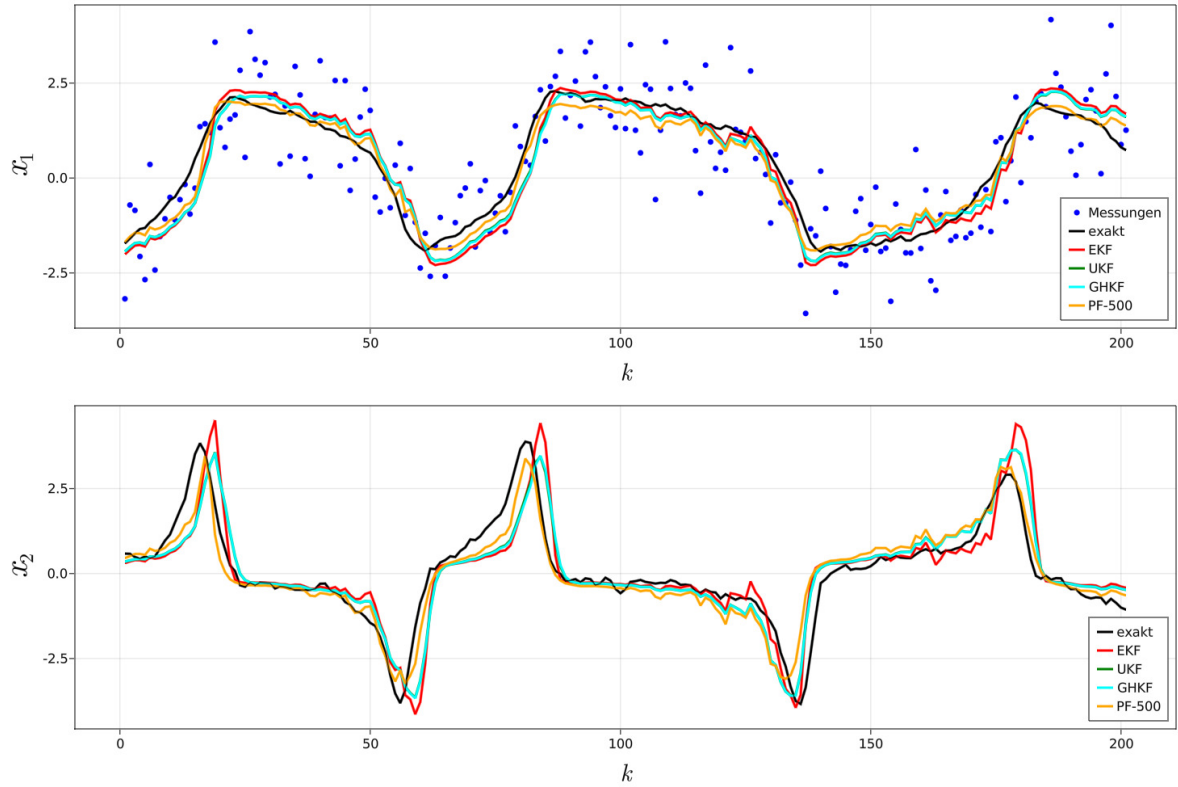


Abbildung 4.3: Repräsentativer Ausschnitt der Filtergenauigkeit

	EKF	UKF	GHKF	PF
RMSE x_1	0.3852	0.3476	0.3494	0.3343
RMSE x_2	0.6899	0.5745	0.5796	0.5978
Rechenzeit in Sekunden	2.12	14.6	195	94.3

Tabelle 4.1: Genauigkeit und Effizienz der Filter

rend insbesondere die Fehler in der zweiten Komponente relativ groß sind, konnten alle Filter der Trajektorie folgen.

Wir bemerken, dass die Rechenzeiten stark voneinander abweichen. Während das **EKF** nur knapp zwei Sekunden für eine Million Zeitschritte benötigt hat, wurden 94 Sekunden für die Berechnungen im Partikel-Filter verwendet. Außerdem benötigen die Sigma-Punkt Filter und das Partikel-Filter deutlich mehr Arbeitsspeicher, was neben dem höheren Rechenaufwand eine limitierender Faktor in eingebetteten Systemen darstellen kann. Relativierend beachten wir, dass alle Algorithmen auf nur einem CPU-Kern ausgeführt wurden, denn die Funktionsauswertung an einzelnen Punkten kann leicht parallelisiert werden. Während der Rechenaufwand für das **EKF** und das **UKF** ähnlich stark anwächst, kann das **GHKF** durch den exponentielle Anstieg der benötigten Stützstellen nur für kleine Systeme benutzt werden. Das **PF** funktioniert auch für einige große Systeme gut, theoretisch muss die Anzahl der Partikel jedoch auch exponentiell wachsen, um eine gleichbleibende Genauigkeit bei wachsendem Zustandsraum zu erzielen. Dieses Phänomen nennt man auch den Fluch der Dimensionalität [Bis06, DH03].

Für die Gauß-Hermite Integration wurde der Parameter mit $N = 10$ sehr groß gewählt. Dadurch sehen wir, dass das **UKF** sehr effizient und in dem gewählten Beispiel auch genauer als das **GHKF** ist.

Zusammenfassung

In dieser Arbeit haben wir uns zu Beginn das *Bayes-Framework* angeschaut, mit welchem wir das Filterproblem spezifiziert haben. Basierend auf dessen rekursive Lösung, dem Bayes-Filter, haben wir das Kalman-Filter für lineare Probleme mit normalverteiltem Rauschen und eine Reihe approximativer Verfahren für nichtlineare Probleme gefunden. Die ausführliche Entwicklung der Filter zeigt, wie die getroffenen Annahmen und Eigenschaften der Normalverteilung ausgenutzt wurden, um effiziente Filter zu erhalten. Außerdem dienen die Ausführungen als Leitfaden für die Entwicklung eigener Filter für ähnliche Problemstellungen. Im Gegensatz zu vielen anderen Algorithmen müssen Filter für speziellere Systeme angepasst werden. Beispielsweise wird in [EAM08] der Einfluss des Systemrauschens auf die Beobachtung berücksichtigt.

Anschließend wurde im theoretischen Vergleich gezeigt, warum das **UKF** in der Regel eine höhere Genauigkeit erzielt. Außerdem zeigte die Analyse durch die Taylorreihenentwicklung auf, wie die Sigma-Punkte und Gewichte im **UKF** gewählt wurden. Eine solche geschickte Wahl ist auch die Grundlage weiterer **SPKF**. Deshalb wurden die Berechnungen allgemeiner als für das vorgestellte Verfahren nötig ausgeführt.

Die anschließende Diskussion und die Anwendung auf Beispiele haben schließlich gezeigt, dass die Auswahl eines Filters für ein konkretes Problem ein Kompromiss zwischen der verfügbaren Rechenleistung und der geforderten Genauigkeit ist. Die theoretischen Fehler sind dabei nur ein Anhaltspunkt, sodass Filter durch Experimente und Simulationen erprobt werden müssen. Das effiziente und erprobte **EKF** mit vielen gut untersuchten Anpassungen für erhöhte Genauigkeit oder verbesserte Stabilität bleibt daher eines der meistgenutzten Filter für nichtlineare Probleme.

Das **UKF** liefert eine ähnlich effiziente Alternative mit erhöhter Genauigkeit. Der vergleichsweise geringe Rechenaufwand dieser beiden Verfahren ermöglicht ihre Anwendung auch in modernen Problemen wie dem Maschinernen [Hay04, Kapitel 2 und 7]. Für Probleme mit nicht annähernd normalverteilten Größen können das **EKF** und das **UKF** in Mixed Gaussian Filtern benutzt werden.

Dank steigender Rechenleistung können Partikel-Filter in immer größeren Problemen eingesetzt werden. Als Approximation des Bayes-Filters können **PF** oft direkt für eine Vielzahl von Modellen eingesetzt werden. Dadurch können auch Zustände in sehr komplexen Modellen einfach geschätzt werden.

Eine Gemeinsamkeit aller Verfahren ist, dass sie für konkrete Modelle analysiert und getestet werden müssen, da sich unter anderem Modellungenauigkeiten und stark nichtli-

KAPITEL 5. ZUSAMMENFASSUNG

neares Verhalten unterschiedlich auf die Algorithmen auswirken.

Die grundlegenden Ausführungen dieser Arbeit zeigen das prinzipielle Vorgehen der Entwicklung und die Funktionsweise von Filtern. Zusätzlich wurden die Verfahren anhand einfacher Beispiele demonstriert, wobei mögliche Probleme beleuchtet wurden.

Literaturverzeichnis

- [Bis06] Christopher M. Bishop. *Pattern Recognition and Machine Learning* -. Springer, Berlin, Heidelberg, 2006.
- [DH03] Fred Daum und Jim Huang. Curse of dimensionality and particle filters. In *2003 IEEE Aerospace Conference Proceedings (Cat. No. 03TH8652)*, volume 4, pages 4_1979–4_1993. IEEE, 2003.
- [EAM08] Robert J Elliott, Lakhdar Aggoun, und John B. Moore. *Hidden Markov Models - Estimation and Control*. Springer Science+Business Media, Berlin Heidelberg, 2008.
- [Hay04] Simon Haykin. *Kalman filtering and neural networks*, volume 47. John Wiley & Sons, 2004.
- [IX00] Kazufumi Ito und Kaiqi Xiong. Gaussian filters for nonlinear filtering problems. *IEEE transactions on automatic control*, 45(5):910–927, 2000.
- [JU96] Simon Julier und Jeffrey K. Uhlmann. A general method for approximating nonlinear transformations of probability distributions. Technical report, 1996.
- [JUDW95] S.J. Julier, J.K. Uhlmann, und H.F. Durrant-Whyte. A new approach for filtering nonlinear systems. In *Proceedings of 1995 American Control Conference - ACC'95*, volume 3, pages 1628–1632 vol.3, 1995.
- [JUDW00] Simon Julier, Jeffrey Uhlmann, und Hugh F Durrant-Whyte. A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Transactions on automatic control*, 45(3):477–482, 2000.
- [Kal60] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35–45, 03 1960.
- [May82] Peter S. Maybeck. *Stochastic models, estimation, and control Bd. 2*. Stochastic models, estimation, and control, Mathematics in science and engineering. New York [u.a.] : Academic Press, 1982.
- [MD17] Reiner Marchthaler und Sebastian Dingler. *Kalman-Filter: Einführung in die Zustandsschätzung und ihre Anwendung für eingebettete Systeme*. Springer Fachmedien Wiesbaden, 2017.

- [Sär13] Simo Särkkä. *Bayesian Filtering and Smoothing*, volume 3 of *Institute of Mathematical Statistics textbooks*. Cambridge University Press, 2013.
- [Sim06] Dan Simon. *Optimal State Estimation: Kalman, H_∞ , and Nonlinear Approaches*. Wiley, 2006.
- [VDM04] Rudolph Van Der Merwe. *Sigma-point Kalman filters for probabilistic inference in dynamic state-space models*. Oregon Health & Science University, 2004.