

Theorie der Sensorfusion

Public Access Version

Udo Frese

Lutz Schröder

Tom Koller

Stand: 20. März 2023

geTeXed von:

Christoph Hertzberg, Janosch Machowinski, René Wagner und
Pierre Willenbrock

Inhaltsverzeichnis

1. Ablauf des Kurses	9
1.1. 2020	9
1.2. 2022	10
2. Mathematische Grundlagen in Kurzform	11
2.1. Stochastik in \mathbb{R}	11
2.2. Stochastik in \mathbb{R}^n	12
2.3. Lineare Algebra (Matrizenrechnung)	13
3. Einführung	16
3.1. Thema der Veranstaltung	16
3.1.1. Beispiele aus der Anwendung	16
3.1.2. Lernziele	18
3.2. Einführungsbeispiel	19
3.2.1. Aussagen	19
3.3. Literatur	19
4. Wahrscheinlichkeitstheorie (in \mathbb{R})	20
4.1. Grundidee der W-Theorie	20
4.1.1. Intuition	20
4.1.2. Beispiel	21
4.1.3. Folgerungen	21
4.1.4. Bedingte Wahrscheinlichkeit	21
4.1.5. Beispiel	22
4.1.6. Unabhängige Ereignisse	22
4.2. Stetige Verteilungen auf \mathbb{R}	22
4.2.1. Wahrscheinlichkeits-Dichte	22
4.2.2. Beispiele	22
4.3. Zufallsvariablen	23
4.4. Erwartungswert	24
4.4.1. Beispiel	24
4.4.2. Etwas Integrationstheorie	24
4.4.3. Rechenregeln für Erwartungswerte	26
5. Fusion zweier Messwerte	27
5.1. Rekapitulation Wahrscheinlichkeitsrechnung	27
5.2. Z vs. z	27

5.3.	Lineare Schätzformel für zwei Messungen	28
5.3.1.	Annahmen	28
5.3.2.	Forderungen	29
5.3.3.	Generalisierung auf n Messwerte	30
5.4.	Bayes Schätzer	32
5.4.1.	x vs. X	32
5.4.2.	Beispiel	33
5.4.3.	Bayes-Schätzer für zwei Messungen	33
5.4.4.	Annahmen	34
6.	Fortsetzung Wahrscheinlichkeitstheorie	37
6.1.	Varianz	37
6.2.	Standardabweichung(Streuung)	38
6.2.1.	Eigenschaften	39
6.3.	Kovarianz	39
6.3.1.	Eigenschaften	39
6.4.	Korrelation	40
6.5.	Die Faltung	42
6.6.	Zentraler Grenzwertsatz	43
7.	Bayes-Filter	44
7.1.	Bayes-Netze	44
7.1.1.	Beispiel	44
7.2.	Problem: Zustandsschätzung	45
7.2.1.	Beispiel	45
7.2.2.	Annahmen	45
7.2.3.	Notation	46
7.3.	Algorithmus „Bayes-Filter“	46
7.3.1.	Beispiel: Ein Bayessches Spiel	47
8.	Kalman Filter	49
8.1.	Konditionierungslemma	49
8.2.	Kalman Filter	52
8.2.1.	Annahmen	52
8.2.2.	Beispiel Kohlenkarren in 1D	52
8.2.3.	Algorithmus	52
8.3.	Dynamikrauschen proportional zu Zeit oder Strecke	54
8.4.	Konsistenz	55
8.4.1.	Mahalanobis-Distanz	55
8.4.2.	Konsistenzkriterien mit verfügbarer Groundtruth	57
8.4.3.	Konsistenzkriterien ohne verfügbare Groundtruth	58
8.4.4.	Ausreißerdetektion	58

9. (Extended/Unscented) Kalman Filter	59
9.1. Nichtlineare Rechenregeln für Erwartungswerte und Varianzen	59
9.2. Extended Kalman Filter	59
9.2.1. Annahmen	59
9.2.2. Algorithmus	60
9.3. Beispiel Wasserwanne	63
9.4. σ -Punkt-Propagation/Unscented Transform	64
9.4.1. Idee	64
9.4.2. Beweisüberblick	65
9.4.3. Landau-Notation	65
9.4.4. Sigmapunktpropagation für Polynome vom Grad 3	65
9.4.5. Sigmapunktpropagation im Allgemeinfeld	66
9.5. Unscented Kalman Filter	69
9.5.1. Annahmen	69
9.5.2. Algorithmus	69
9.6. Fehlerquellen suboptimaler Filter	70
9.6.1. Approximierungsfehler	70
9.6.2. Probleme durch Inkonsistenz	73
9.6.3. Fehler in der Gaußannahme	74
10. Lineare Algebra	76
10.1. Vektor, Vektorraum	76
10.1.1. Skalarprodukt und Norm	76
10.2. Lineare Abbildung	77
10.3. Matrizen	78
10.3.1. Spezielle Matrizen	78
10.3.2. Matrix-Vektor-Multiplikation	78
10.3.3. Matrixmultiplikation	79
10.3.4. Transponierte und Inverse	80
10.3.5. Rechenregeln	82
10.4. Determinante	83
10.4.1. Die Jacobimatrix	84
11. Wahrscheinlichkeitsrechnung in \mathbb{R}^n	85
11.1. Zufallsvariable	85
11.2. Erwartungswert	85
11.2.1. Eigenschaften	85
11.3. Kovarianzmatrix	86
11.3.1. Eigenschaften	86
11.3.2. Positive Definitheit der Kovarianzmatrix	87
11.4. Mehrdimensionale Gaußverteilung	89
11.5. Mahalanobis-Distanz in \mathbb{R}^n	93

12. Mehrdimensionale Quadratische Ausgleichsrechnung	95
12.1. Lineare Ausgleichsrechnung	95
12.1.1. Annahmen	95
12.1.2. Beispiel	96
12.1.3. Zeitabhängige Probleme	96
12.1.4. Formel	96
12.2. Nichtlineare Ausgleichsrechnung als iterierte Linearisierung	98
12.2.1. Lineare Näherung (Tangente) im \mathbb{R}^n	98
12.2.2. Annahmen	98
12.2.3. Algorithmus (Gauß-Newton)	99
12.3. Nichtlineare Ausgleichsrechnung als Optimierung	99
12.3.1. A-posteriori Wahrscheinlichkeit	99
12.3.2. Konvergenz von Gauß-Newton	101
12.3.3. Softwarebibliotheken	102
13. Mehrdimensionaler Kalmanfilter	104
13.1. Kalmanfilter	104
13.1.1. Annahmen	105
13.1.2. Algorithmus	106
13.2. Beweis Kalmanfilter	106
13.3. Konsistenz des n-D Kalman Filters	108
14. Mehrdimensionaler (Extended / Unscented) Kalmanfilter	109
14.1. Nichtlineare Rechenregeln für Erwartungswert und Kovarianz	109
14.2. Extended-Kalman-Filter	109
14.2.1. Annahmen	110
14.2.2. Algorithmus	110
14.3. Variante des EKF mit nicht additivem Rauschen	110
14.3.1. Annahmen	110
14.3.2. Algorithmus	111
14.4. Herleitung EKF mit nicht additivem Rauschen	112
14.5. Sigmapunktpropagation	113
14.5.1. Berechnung der Sigmapunkte	113
14.5.2. Sigmapunktpropagation	115
14.5.3. Erwartungswert- und Kovarianzberechnung	115
14.5.4. Beweisüberblick	116
14.5.5. Sigmapunktpropagation für Polynome vom Grad 3	116
14.5.6. Sigmapunktpropagation im Allgemeinform	117
14.6. Unscented Kalman Filter	120
14.6.1. Annahmen	120
14.6.2. Algorithmus	121
14.7. Intuition bei nd-Filtern	121

15. Darstellung von Isometrien	123
15.1. Klassifizierung der Isometrien über orthonormale Matrizen	123
15.1.1. Vorbereitungen	123
15.1.2. Mazur-Ulam Theorem	124
15.1.3. Verwendung zur Darstellung der Position und Orientierung von Körpern	126
15.1.4. Verwendung als Zustand im (Extended/Unscented) Kalmanfilter	126
15.2. Rotationen	127
15.2.1. Axiomatisierung	127
15.2.2. Matrix Exponential Darstellung	129
15.3. Singularitäten im Matrix-Exponential	133
16. \boxplus-Mannigfaltigkeiten	135
16.1. Darstellung von Orientierungen im Raum	135
16.2. Motivation: UKF für 3D-Orientierungen	136
16.3. Definition	137
16.3.1. Metrik	139
16.4. Wahrscheinlichkeitsrechnung auf Mannigfaltigkeiten	140
16.4.1. Kovarianz	140
16.4.2. Erwartungswert	140
16.4.3. Berechnung des Erwartungswertes	141
16.4.4. Gaußverteilung	141
16.4.5. Mahalanobis-Distanz auf \boxplus -Mannigfaltigkeiten	142
16.4.6. Konsistenz für \boxplus -Mannigfaltigkeiten	142
16.5. Ausgleichsrechnung in \boxplus -Mannigfaltigkeiten	142
16.6. Signapunktpropagation	144
16.6.1. Generierung	144
16.6.2. Propagation	144
16.6.3. Erwartungswert und Kovarianzberechnung	144
16.6.4. Beispiel (am Modell gezeigt)	145
16.7. UKF auf \boxplus -Mannigfaltigkeiten	147
17. Beobachtbarkeit von Zuständen	149
17.1. Beobachtbarkeit	149
17.2. Schwache Beobachtbarkeit	153
17.3. Beobachtbarkeitsrangkriterium (Observability rank criteria)	156
17.3.1. Neue Systemgleichung	157
17.3.2. Lie-Ableitung	157
17.3.3. Rangkriterium mit Lie-Ableitungen	158
17.3.4. Observable Codistribution	158
18. Review der Sensorfusionsliteratur	162
18.1. Abwandlungen des Kalman Filters	162
18.1.1. Second Order EKF (SO-EKF)	162

18.1.2. Iterated Extended Kalman Filter	163
18.1.3. \boxplus -EKF	164
18.1.4. Multiplicative EKF	164
18.1.5. Error State Kalman Filter	165
18.1.6. Information Filter (IF)	166
18.1.7. Smoothing	167
18.2. Andere Filtertypen	169
18.2.1. Batch Least Squares (Quadratische Ausgleichsrechnung)	169
18.2.2. Moving Horizon Estimator (MHE)	170
18.2.3. Grid/Histogram Filter	170
18.2.4. Partikelfilter	171
18.2.5. Robust Estimation	172
18.3. Pseudomessungen	173
18.3.1. Vehicle Constraints	174
18.3.2. Topologiekarten	174
18.3.3. Zero-Velocity Update	174
18.3.4. Gebäudekarten	175
Literatur	176
A. Quizfragen und Arbeitsblätter	181
A.1. Erwartungswerte	182
A.2. Eigenschaften der 2-Messwerte-Formel	184
A.3. Generalisierung auf n -Messwerte	184
A.4. Umfang korreliert mit Fläche	187
A.5. Schnelltests einer Pflegekraft	188
A.6. Statistik aus der Lehre	189
A.7. Der landende Ballon	190
A.8. Der fortgeschritten landende Ballon	190
A.9. Neigungsschätzung im Zug	191
A.10. Anzeige der Akkulaufzeit	191
A.11. Zweimal nichtlinear konditioniert	192
A.12. Anschauliche Bedeutung von E und Cov in 2D	193
A.13. Höherdimensionale Kovarianzmatrizen	193
A.14. Mittelwertvektor und Kovarianzmatrix abgeleiteter Größen	194
A.15. Parabelfit	195
A.16. Tune in - Least Square Version	196
A.17. GPS in a nutshell	196
A.18. Der Kompass dreht sich nichtlinear im Kreis	197
A.19. Der landende Ballon 2	198
A.20. Die schnelle Waage	199
A.21. Tune in EKF Version	200
A.22. Fahrender Zug mit Bias	200
A.23. RoboCup Small Size League I	201

Inhaltsverzeichnis

A.24.RoboCup Small Size League II	201
A.25.Sigmas Cousinen vom Grade null bis drei	202
A.26.Der tapsende Inertialsensor	205
A.27.Das Auge in die Hand nehmen	206
A.28.Axiomatik anschaulich erklärt	206
A.29.Beobachtbarkeit mit einer Winkelmessung	208
A.30.Beobachtbarkeit mit einer Winkel- und Abstandsmessung	208

1. Ablauf des Kurses

1.1. 2020

Nr.	Datum	Thema	ÜZ
01	02.11	Einführung Sensorfusion (3) & Wahrscheinlichkeitstheorie in \mathbb{R} (4)	
02	4.11	Einführung Sensorfusion & Scheinkriterien	
03	9.11	Fusion zweier Messwerte: Optimaler Schätzer (5)	Ü1 \mapsto 07 ¹
04	11.11	Fragerunde & Aufgaben A1-A3	
05	16.11	Varianz und Kovarianz (6) & Bayes-Filter (7)	
06	18.11	Fragerunde & Aufgabe A4-A5 & Demo Bayes-Filter	
07	23.11	Kalman Filter (8)	Ü2 \mapsto 11
08	25.11	Besprechung Ü1 & Fragerunde & Aufgaben A6-A8	
09	30.11	Extended/Unscented Kalman Filter (9)	
10	2.12	Entfällt für Tag der Lehre	
11	7.12	Wahrscheinlichkeitsrechnung in \mathbb{R}^n (11)	Ü3 \mapsto 15
12	9.12	Besprechung Ü2 & Fragerunde & Aufgaben A9-A12	
13	14.12	Mehrdimensionale Quadratische Ausgleichsrechnung (12)	
14	16.12	Fragerunde & Aufgaben A13-A15	
15	21.12	Mehrdimensionaler Kalman Filter (13)	Ü4 \mapsto 19
	23.12	Vorlesungsfreie Zeit bis 5.1.2020	
16	6.1	Besprechung Ü3 & Fragerunde & Aufgaben A16-A18	
17	11.1	Mehrdimensionaler EKF (14)	
18	13.1	Fragerunde & Aufgabe A19-A20	
19	18.1	Mehrdimensionaler UKF (14.5) & Rotationen (15.2)	Ü5 \mapsto 23
20	20.1	Besprechung Ü4 & Fragerunde & A21-A22	
21	25.1	\boxplus -Mannigfaltigkeiten (16)	
22	27.1	Fragerunde & Aufgaben A23-A24	
23	1.2	Beobachtbarkeit von Zuständen (17)	Ü6 \mapsto 27
24	3.2	Besprechung Ü5 & Fragerunde A25-A26	
25	8.2	Beobachtbarkeit Teil II	
26	10.2	Fragerunde & A27	
27	15.2	Literaturreview Sensorfusion (In Vorbereitung)	
28	17.2	Besprechung Ü6 & Fragerunde & Prüfungsvorbereitung	

¹ Übungszettel 1 Abgabe am Termin 07

Grau hinterlegte Termine werden als Videos hochgeladen und finden nicht in Präsenz statt. Einzelne Vorlesungen können in mehrere Videos unterteilt sein.

1. Ablauf des Kurses

1.2. 2022

Nr.	Datum	Thema	Aufgaben	Ü
01	17.10	Einführung Sensorfusion (3) & Organisation	Eigen.	Ü1
02	19.10	Fusion zweier Messwerte: Optimaler Schätzer (5-5.3)	A2	
03	24.10	Bayes Schätzer (5.4)		
04	26.10	Varianz und Kovarianz (6)	A4	Ü2
	31.10	(Reformationstag)		
05	02.11	Besprechung Ü1		
06	07.11	Bayes-Filter (7)	A5	
07	09.11	Konditionierungslemma (8.1),	A6	
08	14.11	Kalman Filter, Dynamikrauschen (8.2-8.3)	A7, A8	Ü3
09	16.11	Konsistenz, Extended Kalman Filter (8.4-9.3)	A9, A10	
10	21.11	Besprechung Ü2		
11	23.11	Unscented Kalman Filter (9.4-9.5)	A11	
12	28.11	Wahrscheinlichkeitsrechnung in \mathbb{R}^n (11-11.3)	A12, A13	Ü4
13	30.11	Gauß und Mahal. in \mathbb{R}^n (11.4, 11.5)	A14	
14	05.12	Bespr. Ü3, Lin. quadr. Ausgleichsrg. in \mathbb{R}^n (12.1)	A15	
	07.12	(Tag der Lehre)		
15	12.12	Nichtlin. quadr. Ausgleichsrechnung in \mathbb{R}^n (12.2)	A16, A17	Ü5
16	14.12	Kalman Filter in \mathbb{R}^n (13)	A18-A19	
17	19.12	Besprechung Ü4	A19	
18	21.12	Extended Kalman Filter in \mathbb{R}^n (14.1-14.4) Weihnachtsferien vom 24.12.-8.1.	A20	
19	09.1	(nur Anwendungsaufgaben)	A21-A24	Ü6
20	11.1	Sigmapunktpropagation in \mathbb{R}^n (14.5)		
21	16.1	Unscented Kalman Filter (14.6); Besprechung Ü5		
22	18.1	Darstellung von Isometrien (15)		
23	23.1	\boxplus -Mannigfaltigkeiten (16-16.4.4)		Ü7
24	25.1	Quadr. Ausglrng. + UKF auf \boxplus -Mannigfaltigkeiten (16.5-16.7)	A25	
25	30.1	Besprechung Ü6	A26	
26	01.2	Literaturreview (18); mündliche Prüfungen		
27	27.02	Ferien		→

¹ Übungszettel 1 Abgabe am Termin 07

2. Mathematische Grundlagen in Kurzform

2.1. Stochastik in \mathbb{R}

Ein **Wahrscheinlichkeitsraum** ist ein Paar (Ω, p) , wobei Ω die Menge der Ergebnisse und $p(\omega)$ die Wahrscheinlichkeitsdichte eines Ergebnisses $\omega \in \Omega$ ist. Die Wahrscheinlichkeit $P(W)$ eines Ereignisses $W \subset \Omega$ ist $\int_{\omega \in W} p(\omega) d\omega$. Es muss gelten, dass $p(\omega) \geq 0$ und $P(\Omega) = 1$ sind. In den im Seminar betrachteten Fällen ist $\Omega = \mathbb{R}^n$, in diesem Abschnitt $n = 1$.

Eine **Zufallsvariable** X ist eine Funktion $\Omega \rightarrow \mathbb{R}^n$ des Ergebnisses ω . In der Notation stehen Großbuchstaben für Zufallsvariable und Kleinbuchstaben für konkrete Werte. D.h. $p(X = x)$ steht für die Wahrscheinlichkeitsdichte, dass die Zufallsvariable X einen konkreten Wert x hat. Kurz schreibt man oft $p(x)$. Für eine Zufallsvariable X definiert $(X(\Omega), x \rightarrow p(X = x))$ einen Wahrscheinlichkeitsraum für die Werte von X . Zwei Zufallsvariablen X, Y sind unabhängig, wenn $p(x, y) = p(X = x \wedge Y = y) = p(X = x)p(Y = y) \forall x, y$ ist.

Die **bedingte Wahrscheinlichkeitsdichte** dass eine Zufallsvariable X den Wert x hat, gegeben, dass eine andere Zufallsvariable Y den Wert y hat ist definiert als $p(x|y) = p(X = x|Y = y) = \frac{p(X=x \wedge Y=y)}{p(Y=y)} = \frac{p(x,y)}{p(y)}$. Mit dieser Definition ergibt sich ein bedingter Wahrscheinlichkeitsraum $(\{\omega \in \Omega | Y(\omega) = y\}, \omega \rightarrow p(\omega|Y = y))$ bestehend aus allen Ergebnissen für die $Y = y$ ist.

Der **Erwartungswert** einer Zufallsvariablen X ist $E(X) = \int_{\omega \in \Omega} \omega X(\omega) = \int_x xp(X = x)dx$. Er gibt an, welchen Wert die Zufallsgröße "im Mittel" annimmt. $E(X + Y) = E(X) + E(Y)$, $E(\lambda X) = \lambda E(X)$, $E(1) = 1$, $E(XY) = E(X)E(Y)$, falls X und Y unabhängig.

Die **Varianz** einer Zufallsvariablen ist $V(X) = E((X - E(X))^2) = E(X^2) - E(X)^2$. Sie gibt an, wie stark die Zufallsgröße um den Erwartungswert schwankt. $V(X + Y) = V(X) + V(Y)$, wenn X und Y unabhängig. $V(\lambda X) = \lambda^2 V(X)$. Hat eine Folge n unabhängiger Zufallsvariablen X_i die Varianz σ^2 , so ist die Varianz der Summe $V(\sum_{i=0}^{n-1} X_i) = n\sigma^2$ und des Mittelwertes $V(\frac{1}{n} \sum_{i=0}^{n-1} X_i) = \frac{1}{n}\sigma^2$.

Die **Kovarianz** zweier Zufallsvariablen ist $\text{Cov}(X, Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y)$. Sie gibt an, wie stark die Zufallsgrößen gekoppelt um den Erwartungswert schwanken. Die Kovarianz einer Zufallsgröße mit sich selbst ist die Varianz der Größe $\text{Cov}(X, X) = V(X)$. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$, $\text{Cov}(\lambda X, Y) = \lambda \text{Cov}(X, Y)$, $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$, $\text{Cov}(1, X) = 0$, $V(X + Y) = V(X) + 2\text{Cov}(X, Y) + V(Y)$

Die **Standardabweichung** $\sigma(X)$, englisch root of mean square (rms) ist die Quadratwurzel der Varianz. Sie hat dieselbe Einheit wie die Zufallsgröße und beschreibt damit

2. Mathematische Grundlagen in Kurzform

die "übliche" Abweichung. $\sigma(X + Y) = \sqrt{\sigma(X)^2 + \sigma(Y)^2}$ falls X und Y unabhängig, $\sigma(\lambda X) = \lambda\sigma(X)$.

Die **Korrelation** $\text{cor}(X, Y)$ ist die normalisierte Kovarianz $\text{cor}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}$. $-1 \leq \text{cor}(X, Y) \leq 1$. Sie gibt an welcher Anteil der Schwankung von X und Y gemeinsame Schwankung ist. Im Unterschied zu der Kovarianz geht die Stärke der Schwankung nicht ein.

Der mittlere quadratische Abstand eines Wertes x_0 zu einer Zufallsvariablen X ist eine quadratische Funktion in x_0 und wird durch Erwartungswert und Varianz beschrieben.

$$E((x_0 - X)^2) = V(X) + (x_0 - E(X))^2 \quad (2.1)$$

Für andere Abstandsmasse (z.B. $|\cdot|$) läßt sich der mittlere Abstand zu x_0 als Funktion von x' nicht durch endlich viele Parametern beschreiben. Aufgrund dieser Gleichung läßt sich erkennen, daß der Erwartungswert das x_0 mit dem kleinsten quadratischen Abstand ist und dieser minimale Abstand der Varianz entspricht.

Die eindimensionale **Gauß**verteilung mit Mittelwert μ und Standardabweichung σ ist definiert als:

$$\mathcal{N}(\mu, \sigma^2)(x) = p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2.2)$$

Skalare Vielfache und Summen von Gaußverteilungen sind wieder Gaußverteilungen. Ist $X \mathcal{N}(\mu_1, \sigma_1^2)$ verteilt und Y unabhängig $\mathcal{N}(\mu_2, \sigma_2^2)$ verteilt, so ist $X + Y \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ verteilt und λX ist $\mathcal{N}(\lambda\mu, (\lambda\sigma)^2)$ verteilt. Die Gaußverteilungen sind die einzige Familie von Verteilungen mit diesen Eigenschaften, endlich vielen Parametern und endlicher Varianz.

Der **zentrale Grenzwertsatz** besagt: Für eine Folge X_i unabhängiger Zufallsvariablen mit Erwartungswert μ und Standardabweichung σ , konvergiert $\frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} X_j$ gegen eine $\mathcal{N}(\mu, \sigma^2)$ Gaußverteilung.

2.2. Stochastik in \mathbb{R}^n

Der **Erwartungswert** einer Vektorwertigen n dimensionalen Zufallsvariablen X ist ein n Vektor, analog zum eindimensionalen Fall. Die i -te Komponente des Erwartungswertes ist der Erwartungswert der i -ten Komponente $(E(X))_i = E(X_i)$. Rechenregeln gelten analog zum eindimensionalen Fall: $E(X + Y) = E(X) + E(Y)$, $E(\lambda X) = \lambda E(X)$, $E(\text{const}) = \text{const}$, zusätzlich noch $E(AX) = AE(X)$ für eine Matrix A . Erwartungswerte von Untervektoren sind die entsprechenden Untervektoren des Erwartungswertes.

Der **Erwartungswert** einer Matrixwertigen $n \times n$ Zufallsvariablen X ist eine $n \times n$ Matrix analog zum eindimensionalen Fall. Die i, j Komponente des Erwartungswertes ist der Erwartungswert der i, j Komponente $(E(U))_{ij} = E(U_{ij})$. Rechenregeln gelten analog: $E(U + V) = E(U) + E(V)$, $E(\lambda U) = \lambda E(U)$, $E(\text{const}) = \text{const}$ zusätzlich noch $E(U^T) = E(U)^T$, $E(UV) = E(U)E(V)$ falls U und V unabhängig sind. Erwartungswerte von Untermatrizen sind die entsprechenden Untermatrizen des Erwartungswertes.

2.3. Lineare Algebra (Matrizenrechnung)

Die **Kovarianzmatrix** einer n dimensionalen Zufallsvariablen X ist eine $n \times n$ Matrix. Der Eintrag i, j entspricht der Kovarianz von X_i und X_j , bzw. für $i = j$ (Diagonale) der Varianz von X_i . $\text{Cov}(X)_{ij} = \text{Cov}(X_i, X_j)$. Die Matrix ist symmetrisch und positiv (semi-) definit.

$$\text{Cov}(X, Y) = E((X - E(X))(Y - E(Y))^T), \quad (2.3)$$

$$\text{Cov}(X) = \text{Cov}(X, X) = E((X - E(X))(X - E(X))^T) \quad (2.4)$$

Rechenregeln gelten analog zum eindimensionalen Fall. $\text{Cov}(\lambda X) = \lambda^2 \text{Cov}(X)$, $\text{Cov}(X + Y) = \text{Cov}(X) + \text{Cov}(Y)$, falls X und Y unabhängig, sonst $\text{Cov}(X) + \text{Cov}(Y) + \text{Cov}(X, Y) + \text{Cov}(X, Y)^T$. $\text{Cov}(AX) = A \text{Cov}(X) A^T$. Die Kovarianzmatrix eines Untervektors ist die entsprechende Untermatrix der Kovarianzmatrix.

Die **Informationsmatrix** einer Zufallsvariablen x ist die Inverse der Kovarianzmatrix $\text{Cov}(X)^{-1}$.

Die **Kovarianz** einer Zufallsvariablen X in **Richtung** n ist $\text{Cov}(n^T X) = n^T \text{Cov}(X) n$ eine quadratische Funktion in n . Daraus folgt, dass $\text{Cov}(x)$ positiv (semi-) definit ist und die Varianz in allen Richtungen gleichzeitig beschreibt.

Bei **linearisierter Fehlerfortpflanzung** wird eine Funktion f am Erwartungswert ihres Argumentes x linearisiert $f(X) \approx f(E(X)) + f'(E(X))(X - E(X))$, wobei f' die Jacobimatrix ist. Es folgt, daß $E(f(X)) \approx f(E(X))$ und $\text{Cov}(f(X)) \approx f'(E(X)) \text{Cov}(X) f'(E(X))^T$.

Die n -dimensionale Gaußverteilung wird durch die folgende Formel beschrieben:

$$\mathcal{N}(\mu, C)(x) = p(x) = \frac{1}{(2\pi)^{n/2} \sqrt{|C|}} e^{-\frac{1}{2}(x-\mu)^T C^{-1}(x-\mu)} \quad (2.5)$$

Es folgt, dass der negative Logarithmus der Wahrscheinlichkeit eine quadratische Funktion in x ist

$$-\ln(p(x)) = \text{const} + \frac{1}{2}(x - \mu)^T C^{-1}(x - \mu) \quad (2.6)$$

$$= \text{const} + x^T (C^{-1} \mu) + \frac{1}{2} x^T C^{-1} x \quad (2.7)$$

dessen Minimum der Erwartungswert μ und dessen Koeffizienten zweiter Ordnung die Inverse der Kovarianz die sogenannte Informationsmatrix ist.

2.3. Lineare Algebra (Matrizenrechnung)

Ein **Vektor** $v \in \mathbb{R}^n$ ist eine Zusammenstellung von n Zahlen (die Komponenten). Die einzelnen Komponenten werden durch Subskript Indizes bezeichnet v_i . Addition, Subtraktion und Multiplikation mit einem Skalar (\mathbb{R}) werden komponentenweise ausgeführt, mit den üblichen Rechenregeln. Das Skalarprodukt zweier Vektoren v und w ist $\sum_{i=0}^{n-1} v_i w_i$ oder in Matrixschreibweise $v^T w$. Rechenregeln ergeben sich aus den Matrixrechenregeln. Das Skalarprodukt mit sich selbst $v^T v = |v|^2$ ist geometrisch das Quadrat der Länge des

2. Mathematische Grundlagen in Kurzform

Vektors. Das Skalarprodukt zweier Vektoren $v^T w$ ist geometrisch $|v||w|\cos(\angle v, w)$ das Produkt ihrer Längen und dem Cosinus des eingeschlossenen Winkels.

Eine **lineare Abbildung** ist eine Funktion $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$, für die $f(v+w) = f(v) + f(w)$ und $f(\lambda v) = \lambda f(v)$ ist.

Eine **affine Abbildung** oder Funktion 1. Ordnung ist die Summe aus einer Konstanten und einer linearen Funktion.

Eine **Matrix** $A \in \mathbb{R}^{m \times n}$ ist eine rechteckige Zusammenstellung von Zahlen in Zeilen und Spalten (die Komponenten). Die einzelnen Komponenten werden durch Subskript Indizes bezeichnet A_{ij} . $A_{i\bullet}$ bezeichnet die i -te Zeile, $A_{\bullet j}$ die j -te Spalte. (Spalten-)Vektoren sind $n \times 1$ Matrizen. Addition, Subtraktion und Multiplikation mit einem Skalar (\mathbb{R}) werden Komponentenweise ausgeführt, mit den üblichen Rechenregeln. Das Produkt zweier Matrizen ist definiert als $(A \cdot B)_{ij} = \sum_k A_{ik} B_{kj}$, d.h. Komponente ij des Ergebnisses ist das Skalarprodukt aus Zeile $A_{i\bullet}$ und Spalte $B_{\bullet j}$. Daher muss Spaltenzahl von A und Zeilenzahl von B übereinstimmen. Matrizenmultiplikation ist nicht kommutativ (in der Regel $AB \neq BA$) aber distributiv und assoziativ. $ABC = (AB)C = A(BC)$, $(\lambda A)B = \lambda(AB)$, $A(\lambda B) = \lambda AB$, $(A+B)C = AC + BC$, $A(B+C) = AB + AC$.

Eine **Nullmatrix** 0 enthält nur 0 , $0v = 0$, $0A = 0$, $A0 = 0$.

Die $n \times n$ Einheitsmatrix I enthält überall 0 und 1 auf der Diagonale. $Iv = v$, $IA = A$, $AI = A$.

Die **Transponierte** einer Matrix ist die an der Diagonalen gespielte Matrix: $(A^T)_{ij} = A_{ji}$. Für Vektoren v, w ist $v^T w = \sum v_i w_i \in \mathbb{R}$ das Skalarprodukt (eine Zahl), während $vw^T \in \mathbb{R}^{n \times n}$ das äussere Produkt, die Matrix aller Produkte von einzel Komponenten ist $(vw^T)_{ij} = v_i w_j$. Für Matrizen A, B sind $(A+B)^T = A^T + B^T$, $(\lambda A)^T = \lambda A^T$, $(AB)^T = B^T A^T$, $A^{T-1} = A^{-1T}$.

Die **Inverse** einer A^{-1} einer $n \times n$ Matrix A ist bestimmt durch $AA^{-1} = A^{-1}A = I$. Nicht alle Matrizen sind invertierbar, wenn sie existiert ist die Inverse aber eindeutig. Regeln: $(AB)^{-1} = B^{-1}A^{-1}$, $(\lambda A)^{-1} = \frac{1}{\lambda}A^{-1}$, $A^T^{-1} = A^{-1T}$.

Eine $m \times n$ Matrix A **beschreibt eine lineare Abbildung** von \mathbb{R}^n nach \mathbb{R}^m : $f: \mathbb{R}^n \rightarrow \mathbb{R}^m: v \mapsto Av$. Da $Av = \sum_j A_{\bullet j} v_j$ bilden die Spalten von A die Bilder der Einheitsvektoren aus denen eine Linearkombination entsprechend v gebildet wird. In alternativer Sichtweise ist $(Av)_i = A_{i\bullet} v$, d.h. die i -te Komponente des Ergebnisses ist das Skalarprodukt aus v und der i -ten Zeile von A , also die Projektion von v in Richtung $A_{i\bullet}$. Für zwei lineare Funktionen $f: v \mapsto Av$ und $g: v \mapsto Bv$ wird die Hintereinanderschaltung $f \circ g: v \mapsto f(g(v)) = ABv$ durch AB beschrieben. Die Umkehrfunktion wird durch die Inverse beschrieben.

Die **Jacobimatrix** einer allgemeinen Funktion $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ an der Stelle \check{v} ist die Matrix aller partiellen Ableitungen $f'(\check{v})_{ij} = \left. \frac{\partial f(v)_i}{\partial v_j} \right|_{\check{v}}$. Auch notiert als $\left. \frac{\partial f(v)}{\partial v} \right|_{\check{v}}$. Rechenregeln für Jacobimatrizen sind analog zu denen für normale Ableitungen: $(f(v) + g(v))' = f'(v) + g'(v)$, $(\text{const})' = 0$, $(\lambda f(v))' = \lambda f'(v)$. Die Kettenregel führt zu Matrixmultiplikation: $(f(g(v)))' = f'(g(v))g'(v)$. Beispiele für Ableitung verschiedener Terme nach v : $v' = I$, $(Av)' = A$, $(v^T Av)' = v^T(A + A^T)$.

Der **Gradient** einer Funktion $f: \mathbb{R}^n \rightarrow \mathbb{R}$ ist die entsprechende $1 \times n$ Jacobimatrix (Zeilenvektor).

2.3. Lineare Algebra (Matrizenrechnung)

Eine allgemeine Funktion $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ kann in einem Punkt \check{v} **linearisiert** werden als $f(v) \approx f(\check{v}) + f'(\check{v})(v - \check{v})$. Dabei ist $f'(\check{v})$ die Jacobimatrix an der Stelle \check{v} .

Eine Matrix A ist **symmetrisch**, wenn $A^T = A$, also $A_{ij} = A_{ji}$ ist. Kovarianz und Informationsmatrix sind symmetrisch. Für symmetrische Matrizen A, B ist $A+B, A-B, \lambda A, A^T$ und A^{-1} symmetrisch. Für eine beliebige Matrix B ist $B^T A B$ symmetrisch.

Eine symmetrische Matrix $A \in \mathbb{R}^{n \times n}$ beschreibt eine **quadratische Funktion** $f: \mathbb{R}^n \rightarrow \mathbb{R}: v \mapsto v^T A v = \sum_{ij} v_i A_{ij} v_j$ (ohne konstanten oder linearen Anteil). Eine allgemeine Funktion 2. Ordnung wird durch den Ausdruck $v \mapsto v^T A v + b^T v + \gamma$ beschrieben ($A \in \mathbb{R}^{n \times n}, b \in \mathbb{R}^n, \gamma \in \mathbb{R}$). Die durch $v \mapsto A v$ beschriebene lineare Abbildung ist die Ableitung $\frac{1}{2} f'$. Für eine Zufallsvariablen x beschreibt die quadratische Funktion $v^T \text{Cov} x v$ die Varianz von $v^T x$, d.h. die Schwankung von x in Richtung v . Damit enthält die Kovarianzmatrix die Information über die Varianz in allen Richtungen. In der Gaußverteilung beschreibt der Ausdruck $(x - \mu)^T C^{-1} (x - \mu)$ die Wahrscheinlichkeit (als negativer Logarithmus) x aus der Verteilung zu ziehen.

Eine symmetrische Matrix ist **positiv definit**, wenn $v^T A v > 0 \forall v \neq 0$, d.h. wenn die beschriebene quadratische Funktion strikt positiv ist. Ist $v^T A v \geq 0 \forall v$, so ist sie positiv semi definit. Sind A, B positiv definit, so ist $A+B, \lambda A, \lambda > 0, A^{-1}$ positiv definit. Sind A, B positiv semidefinit, so ist $A+B, \lambda A, \lambda \geq 0$ positiv semidefinit.

Eine Matrix A ist **orthonormal**, wenn $A^T = A^{-1}$ oder äquivalent $A_{\bullet i} A_{\bullet j} = A_{i \bullet} A_{j \bullet} = 0$ für $i \neq j$ bzw. $= 1$ für $i = j$. Orthogonal Matrizen entsprechen linearen Abbildungen, die Vektorlängen und -winkel erhalten: $(A v)^T (A w) = v^T A^T A w = v^T A^{-1} A w = v^T w$. Sie werden oft benutzt um die Beziehung zwischen Koordinatensystemen zu beschreiben.

Ein **Eigenwert** λ mit **Eigenvektor** v einer Matrix A erfüllen $A v = \lambda v$. Es gibt (bis zu) n Eigenwerte. Die Eigenwerte der Inversen sind die Kehrwerte der Eigenwerte. Eine Matrix ist invertierbar, wenn alle Eigenwerte $\neq 0$ sind, sie ist positiv (semi-) definit, wenn alle Eigenwerte > 0 (≥ 0) sind.

Eine symmetrische Matrix A lässt sich **diagonalisieren**, d.h. zerlegen als $A = U^T D U$, wobei U eine orthogonale Matrix, D eine diagonale Matrix ist und auf der Diagonalen von D die Eigenwerte stehen. Zeile $U_{i \bullet}$ von U ist der Eigenvektor zum i -ten Eigenwert. Jede quadratische Funktion besteht in einem geeigneten Koordinatensystem (dem Eigensystem) aus unabhängigen 1-D quadratischen Funktionen der einzelnen Komponenten. Der größte und kleinste Eigenwert beschränken die Größe der dargestellten quadratische Funktion: $\lambda_{\min} v^T v \leq v^T A v \leq \lambda_{\max} v^T v$. Das Eigenwertzerlegung einer Kovarianzmatrix $\text{Cov}(x)$ beschreibt die Zufallsvariablen x als Summe von unabhängigen eindimensionalen Zufallsvariablen, in geeigneten Richtungen.

3. Einführung

3.1. Thema der Veranstaltung

Wie verknüpft man Messungen mit Fehlern zu einem möglichst genauen Ergebnis?

„Theoria combinationis observationum erroribus minimis obnoxiae“ (C. F. Gauß 1821)

3.1.1. Beispiele aus der Anwendung

- Ein Schienenfahrzeug misst seine Position mit Radar und seine Bewegung über Räder. Gesucht ist die Position.

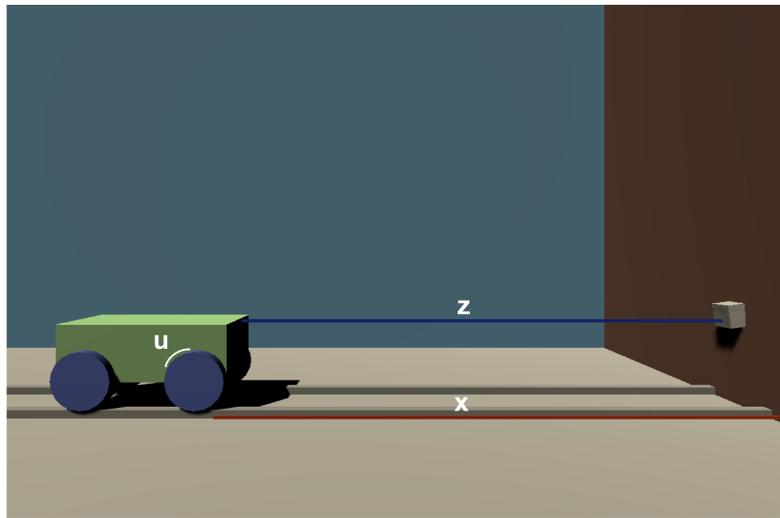


Abbildung 3.1.: Beispiel: Schienenfahrzeug

- z (Radar) ist absolut aber eher ungenau
- u (Radumdrehung) ist relativ aber eher genau
- Zustand x ändert sich über die Zeit
- Sequenz von Messungen
- nur z : ungenau, nur u : Fehler wachsend
- Im Stillstand z mitteln
- In der Fahrt erlaubt das u „so etwas wie“ über z zu mitteln

3.1. Thema der Veranstaltung

- z erlaubt den wachsenden Fehler von u zurückzusetzen
- linear, 1-dimensional
- Fahrzeug in der Ebene misst seine Bewegung über Räder und peilt den Winkel zu Funkbaken. Gesucht ist Position und Orientierung des Fahrzeugs

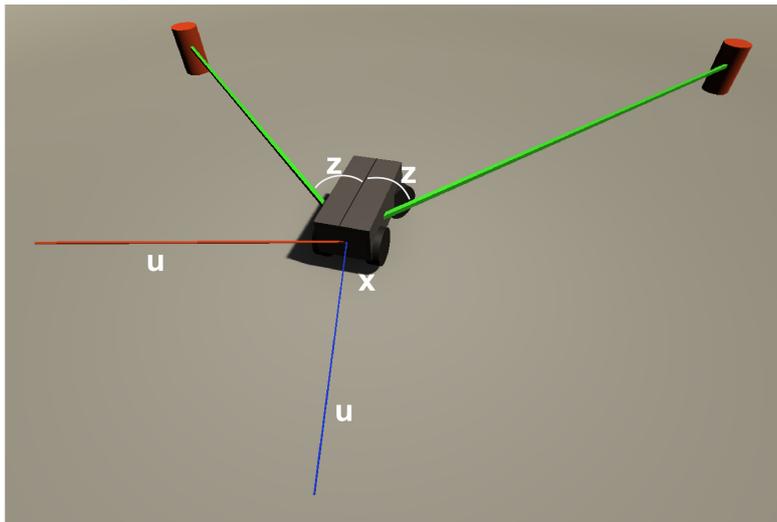


Abbildung 3.2.: Beispiel: Fahrzeug in der Ebene

- z absolut, eher ungenau, unvollständig, unzuverlässig
- u relativ, eher genau, zuverlässig
- u, z nicht linear
- u ermöglichen mehrere Messungen z zu verbinden, um den Zustand (möglichst genau) zu bestimmen
- x, u, z sind mehrdimensional
- Flugzeug misst seine Drehgeschwindigkeit und seine Beschleunigung u , sowie Winkel zu Funkbaken z . Gesucht ist die Position und die Orientierung.

3. Einführung

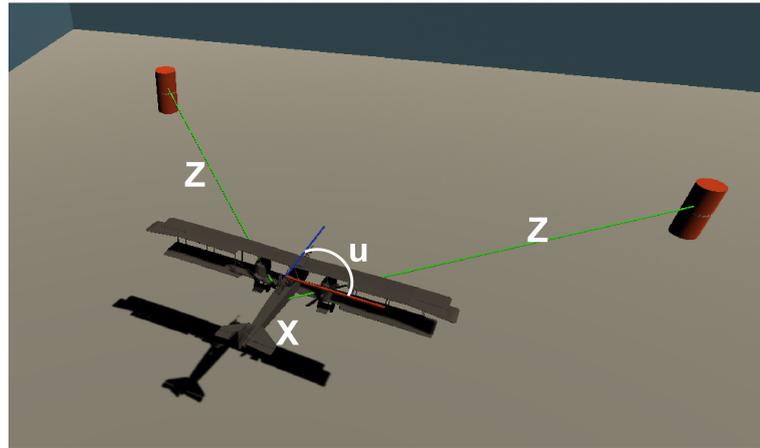


Abbildung 3.3.: Beispiel: Flugzeug

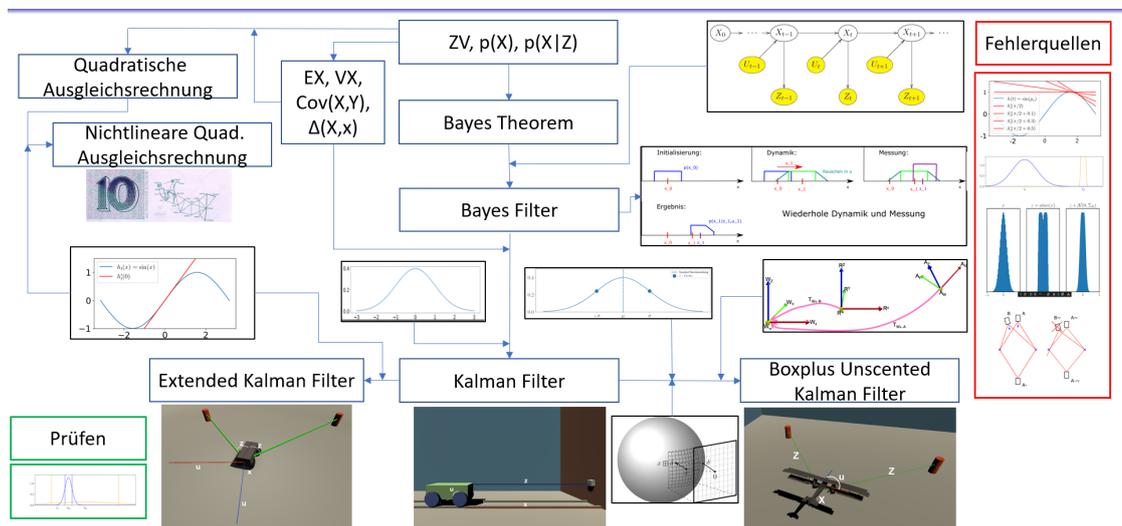


Abbildung 3.4.: Veranstaltungsübersicht

3.1.2. Lernziele

- Fehlerbehaftete Größen probabilistisch verarbeiten können
- Den (Extended, Unscented) Kalman Filter für Fusion von Messungen über die Zeit aus der Theorie der Verarbeitung unsicherer Größen heraus verstehen und beurteilen können
- Anschauliche Sensorfusionsprobleme im Formalismus des (E/U)KF modellieren und analysieren können

- Darstellung von Orientierungen im Raum durch Matrizen und deren Schätzung im UKF in seiner axiomatischen Herleitung verstehen
- Theoretische Resultate der Sensorfusion mit intuitiver Vorstellung verknüpfen können
(„Was sagt mir das?“)

3.2. Einführungsbeispiel

Eine Größe x wird von zwei Sensoren z_1 und z_2 gemessen mit Ungenauigkeit σ_1, σ_2 . Gesucht ist ein Formel für \hat{x} eine möglichst gute Schätzung.

3.2.1. Aussagen

Um eine Vorstellung für den Zusammenhang zu bekommen, fragen wir, welche Aussagen man über die Schätzung $\hat{x}(z_1, \sigma_1, z_2, \sigma_2)$ und ihre Ungenauigkeit $\sigma_{\hat{x}}$ machen kann:

$$\hat{x}(z_1, \sigma_1, z_2, \sigma_2) = \hat{x}(z_2, \sigma_2, z_1, \sigma_1) \quad (3.1)$$

$$\sigma_{\hat{x}} \leq \min(\sigma_1, \sigma_2) \quad (3.2)$$

$$\lim_{\frac{\sigma_1}{\sigma_2} \rightarrow \infty} \hat{x} = z_2 \quad (3.3)$$

$$\lim_{\frac{\sigma_1}{\sigma_2} \rightarrow \infty} \sigma(\hat{x}) = \sigma_2 \quad (3.4)$$

$$\hat{x}(z_1, 0, z_2, \sigma_2) = z_1, \text{ falls } \sigma_2 > 0 \quad (3.5)$$

$$\hat{x}(z, 0, z, 0) = z \quad (3.6)$$

$$\hat{x}(z_1, 0, z_2, 0) \text{ undefiniert, falls } z_1 \neq z_2 \quad (3.7)$$

$$\hat{x}(z_1, \sigma, z_2, \sigma) = \frac{z_1 + z_2}{2} \quad (3.8)$$

$$\text{je größer } \sigma_1 \text{ im Vergleich zu } \sigma_2, \text{ desto näher } \hat{x} \text{ an } z_2 \quad (3.9)$$

$$\min(z_1, z_2) \leq \hat{x} \leq \max(z_1, z_2) \quad (3.10)$$

3.3. Literatur

Allgemeine, weiter führende Literatur sind [50, 5].

4. Wahrscheinlichkeitstheorie (in \mathbb{R})

Quellen zu Themen dieses Kapitel (abgerufen am 22.10.18)

- <https://de.wikipedia.org/wiki/Wahrscheinlichkeitstheorie>
- <https://de.wikipedia.org/wiki/Erwartungswert>

4.1. Grundidee der W-Theorie

Definition 1. Wahrscheinlichkeitsraum nach Kolmogorow Ein Tripel (Ω, Σ, P) mit $\Sigma \subset \{A \subset \Omega\}$ heißt Wahrscheinlichkeitsraum, wenn

$$P : \Sigma \rightarrow [0 \dots 1] \quad (4.1)$$

$$P\Omega = 1 \quad (4.2)$$

$$P\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \sum_{i \in \mathbb{N}} PA_i, \quad \text{falls } i \neq j \Rightarrow A_i \cap A_j = \emptyset, \quad (4.3)$$

wobei PX links von $=$ impliziert, dass der Funktionswert existiert, also $X \in \Sigma$ ist und für alle $A \in \Sigma$ auch $\Omega - A \in \Sigma$ ist.

4.1.1. Intuition

Menge Ω von "Ergebnissen" umfasst alle möglichen Ausgänge des Zufallsexperiments. Das Zufallsexperiment besteht daraus, dass "zufällig" eins der Elemente aus Ω gewählt wird. Beispiel: Würfel: $\Omega = \{1 \dots 6\}$

Eine Menge $A \subset \Omega$ von Ergebnissen heißt Ereignis. Ein Ereignis A tritt ein, wenn das Ergebnis des Zufallsexperiments eines der Elemente von A ist. Z.B. Ereignis "gerade Zahl" $\{2, 4, 6\}$.

Die Funktion P weist einigen Ereignisse eine Zahl zwischen 0 und 1 zu, die Wahrscheinlichkeit. Z.B. $PA = \frac{|A|}{6}$ für einen Würfel.

Intuitiv bedeutet Wahrscheinlichkeit, "Grad der Gewissheit". Die Wahrscheinlichkeitstheorie schiebt dies als Interpretation außerhalb der Theorie und macht nur Aussagen über Wahrscheinlichkeiten, nicht darüber was sie bedeuten.

Die Axiome bedeuten

- Das nie eintretende Ereignis hat Wahrscheinlichkeit 0.
- Das immer eintretende Ereignis hat Wahrscheinlichkeit 1.
- Für eine Folge sich nicht gegenseitig ausschließender, also nie gleichzeitig eintretender Ereignisse ist die Wahrscheinlichkeit das eines eintritt (also der Vereinigung) die Summe der Einzelwahrscheinlichkeiten. Dies gilt auch für unendliche Folgen.

Im Würfelbeispiel ist

$$P\emptyset = \frac{|\emptyset|}{6} = 0 \quad (4.4)$$

$$P\Omega = \frac{|\{1 \dots 6\}|}{6} = \frac{6}{6} = 1 \quad (4.5)$$

$$P\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \frac{1}{6} \left| \bigcup_{i \in \mathbb{N}} A_i \right| = \frac{1}{6} \sum_{i \in \mathbb{N}} |A_i| = \sum_{i \in \mathbb{N}} P A_i \quad (4.6)$$

Σ ist die Menge von Ereignissen für die Wahrscheinlichkeiten definiert sind. In endlichen Wahrscheinlichkeitsräumen normalerweise alle Teilmengen. Für $\Omega = \mathbb{R}$ sagt der Satz von Vital, dass dies nicht geht, sondern nur allen "gutartigen" Teilmengen eine Wahrscheinlichkeit zugeordnet werden kann. Diese Problematik ignorieren wir in der Vorlesung, weil unsere Mengen alle gutartig sind.

4.1.2. Beispiel

Gezinkter Würfel: $\Omega = \{1, \dots, 6\}$, $\Sigma = \{A | A \subset \Omega\}$, $P\{1\} = \dots = P\{5\} = \frac{1}{8}$, $P\{6\} = \frac{3}{8}$, $PA = \sum_{i \in A} P\{i\}$

Dann ist z.B. $P\{1, 6\} = \frac{4}{8} = \frac{1}{2}$.

Axiom 1 und 3 gelten per Definition, für Axiom 2 muss die Summe der Wahrscheinlichkeiten $P\{1\} + \dots + P\{6\} = 1$ ergeben.

Q 1

4.1.3. Folgerungen

1. $P(A \cup B) = PA + PB - P(A \cap B)$
2. $P(\Omega - A) = 1 - PA$
3. $P(\emptyset) = 0$
4. $A \subseteq B \subseteq \Omega \implies PA \leq PB$

Der Beweis siehe Übungszettel Aufgabe 1.

Q 2

4.1.4. Bedingte Wahrscheinlichkeit

Sei B ein Ereignis mit $PB > 0$. Dann ist die bed. Wahrscheinlichkeit $P(A|B)$ die Wahrscheinlichkeit von A , wobei das Eintreten von B bereits bekannt ist.

$$P(A|B) = \frac{P(A \cap B)}{PB} \quad (4.7)$$

Durch die Konditionierung $|B$ werden nur noch die Ergebnisse betrachtet, für die B wahr ist, und unter diesen die Wahrscheinlichkeit für A bestimmt.

4. Wahrscheinlichkeitstheorie (in \mathbb{R})

4.1.5. Beispiel

Gezinkter Würfel s.o. $G = \{2, 4, 6\}$, $A = \{4, 5, 6\}$. $P(G) = \frac{5}{8}$, $P(A|G) = \frac{P\{4,6\}}{PG} = \frac{4/8}{5/8} = \frac{4}{5}$

Die Wahrscheinlichkeit, dass man mit dem gezinkten Würfel eine 4, 5 oder 6 gewürfelt hat, gegeben, dass eine gerade Zahl gewürfelt wurde, ist $\frac{4}{5}$.

Q 3

4.1.6. Unabhängige Ereignisse

Ereignisse A , B heißen *unabhängig*, wenn

$$P(A \cap B) = PA \cdot PB, \quad (4.8)$$

äquivalent (sofern $PB \neq 0$): wenn

$$PA = P(A|B) \quad (4.9)$$

(\Rightarrow : $P(A|B) = P(A \cap B)/PB = PA \cdot PB/PB = PA$, umgekehrt ähnlich).

Achtung: n Ereignisse A_1, \dots, A_n heißen *unabhängig*, wenn *alle* Teilkombinationen A_{i_1}, \dots, A_{i_k} im offensichtlichen verallgemeinerten Sinne unabhängig sind, was nicht dasselbe ist wie paarweise Unabhängigkeit!

4.2. Stetige Verteilungen auf \mathbb{R}

$\Omega = \mathbb{R}$ (typisch: Messwerte)

Wie P angeben, was ist Σ ? Problem: Tacho: $P\{100\} = 0$, deshalb kann man nicht Wahrscheinlichkeiten aller Ergebnisse angeben und die Wahrscheinlichkeiten von Ereignissen als Summe definieren. Statt dessen werden Wahrscheinlichkeitsdichten definiert, so dass sich die Wahrscheinlichkeit eines Ereignisses als Integral der Dichten über alle Ergebnisse des Ereignisses ergibt.

4.2.1. Wahrscheinlichkeits-Dichte

Definition 2. Eine *Wahrscheinlichkeitsdichte* (-funktion) ist eine integrierbare (z.B. stückweise stetige) Funktion $p: \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ mit $\int_{-\infty}^{\infty} p(x)dx = 1$.

Damit $P[a, b] = \int_a^b p(x)dx$, insbesondere: $P\{a\} = \int_a^a p(x)dx = 0$.

Allgemeiner: $PA = \int_{x \in A} p(x)dx$ für alle messbaren Teilmengen von \mathbb{R} : $A \in \Sigma := \{A \subset \mathbb{R}, A \text{ messbar}\}$.

4.2.2. Beispiele

Stetige Gleichverteilung auf $[a, b]$

$$p(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & \text{sonst} \end{cases} \quad (4.10)$$

Für $a \leq c \leq d \leq b$: $P[c, d] = \int_c^d \frac{1}{b-a} dx = \frac{d-c}{b-a}$

Gleichverteilung auf \mathbb{R} ? Nicht möglich! (Warum nicht?)

Die Normalverteilung (Gauß-Verteilung)

$\mathcal{N}(\mu, \sigma^2)$ ist gegeben durch W-Dichte

$$\mathcal{N}(\mu, \sigma^2)(x) = p(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (4.11)$$

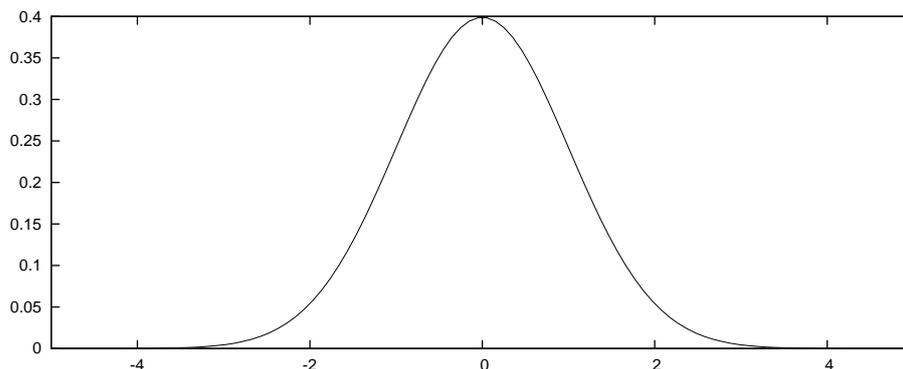


Abbildung 4.1.: Gaußverteilung mit $\sigma = 1$ (Standardnormalverteilung)

Q 4

4.3. Zufallsvariablen

Eine ZV X ist eine Funktion $X : \Omega \rightarrow \mathbb{R}$ („harmlos“, nämlich „messbar“). X liefert W-Verteilung $P(X \in \cdot)$ auf \mathbb{R} , durch

$$P(X \in A) = P\{\omega | X(\omega) \in A\} \quad (4.12)$$

Notationsvarianten wie $P(X > 1) = P(X \in (1, \infty))$.

Wir nehmen stets an, dass $P(X \in \cdot)$ gegeben ist durch W-Dichte $p(X = \cdot)$ auf \mathbb{R} ; insbesondere $P(X \in A) = \int_{x \in A} p(X = x) dx$.

ZV X, Y heißen *unabhängig*, wenn

$$\forall x, y \in \mathbb{R} : p(X = x \wedge Y = y) = p(X = x)p(Y = y). \quad (4.13)$$

In diesem Fall gilt: $P(X \in A \wedge Y \in B) = P(X \in A) \cdot P(Y \in B)$

4. Wahrscheinlichkeitstheorie (in \mathbb{R})

Beweis.

$$P(X \in A \wedge Y \in B) = \int_{x \in A} \int_{y \in B} p(X = x \wedge Y = y) dy dx \quad (4.14)$$

$$= \int_{x \in A} \int_{y \in B} p(X = x) \cdot p(Y = y) dy dx \quad (4.15)$$

$$= \int_{x \in A} p(X = x) \cdot \int_{y \in B} p(Y = y) dy dx \quad (4.16)$$

$$= \int_{x \in A} p(X = x) dx \cdot \int_{y \in B} p(Y = y) dy \quad (4.17)$$

$$= P(X \in A) \cdot P(Y \in B) \quad \square$$

Q 5

4.4. Erwartungswert

Der Erwartungswert $E(X)$ einer ZV X ist definiert als

$$E X = \int_{-\infty}^{\infty} x \cdot p(X = x) dx \quad (4.18)$$

4.4.1. Beispiel

X gleichverteilt auf $[a, b]$

$$E X = \int_{\mathbb{R}} x \cdot p(X = x) dx = \int_a^b x \cdot \frac{1}{b-a} dx \quad (4.19)$$

$$= \frac{1}{b-a} \int_a^b x dx = \frac{b+a}{2} \quad (4.20)$$

X normalverteilt nach $\mathcal{N}(\mu, \sigma^2)$

$$E X = \mu$$

Beweis siehe Übungszettel Aufgabe 2.

4.4.2. Etwas Integrationstheorie

Satz 1 (Maßtheorie). Für ZV X und Funktion $f: \mathbb{R} \rightarrow \mathbb{R}$ gilt:

$$\int_{-\infty}^{\infty} f(x) \cdot p(X = x) dx = \int_{\omega \in \Omega} f(X(\omega)) \cdot p(\omega) d\omega. \quad (4.21)$$

Allgemeiner gilt für ZVen X_1, \dots, X_n und $f: \mathbb{R}^n \rightarrow \mathbb{R}$

$$\int_{x_1=-\infty}^{\infty} \cdots \int_{x_n=-\infty}^{\infty} f(x_1, \dots, x_n) \cdot p(X_1 = x_1, \dots, X_n = x_n) dx_1 \cdots dx_n = \int_{\omega \in \Omega} f(X_1(\omega), \dots, X_n(\omega)) \cdot p(\omega) d\omega. \quad (4.22)$$

Man kann also wahlweise über \mathbb{R} bzw. \mathbb{R}^n oder über den unterliegenden W-Raum Ω integrieren. Wenn man z.B. für f die identische Funktion $\mathbb{R} \rightarrow \mathbb{R}$ wählt, erhält man eine alternative Darstellung des Erwartungswerts:

$$EX = \int_{\omega \in \Omega} X(\omega) \cdot p(\omega) d\omega \quad (4.23)$$

Statt $\int_{x_1=-\infty}^{\infty} \cdots \int_{x_n=-\infty}^{\infty}$ schreiben wir auch $\int_{x_1, \dots, x_n \in \mathbb{R}}$.

Beispiel im Diskreten

6-seitiger Würfel $X : \{1, \dots, 6\} \rightarrow \{0, 1\}, x \mapsto x \bmod 2$

ω	$p(\omega)$	$X(\omega)$
1	1/6	1
2	1/6	0
3	1/6	1
4	1/6	0
5	1/6	1
6	1/6	0

Berechnung im Bildraum $\{0, 1\}$ von X:

$$E(X) = \sum_x x \cdot P(X = x) \quad (4.24)$$

$$= 0 \cdot P(X = 0) + 1 \cdot P(X = 1) \quad (4.25)$$

$$= 0 \cdot (1/6 + 1/6 + 1/6) + 1 \cdot (1/6 + 1/6 + 1/6) \quad (4.26)$$

Berechnung im Ergebnisraum $\Omega = \{1, \dots, 6\}$:

$$E(X) = \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega) \quad (4.27)$$

$$= 1 \cdot 1/6 + 0 \cdot 1/6 + 1 \cdot 1/6 + 0 \cdot 1/6 + 1 \cdot 1/6 + 0 \cdot 1/6 \quad (4.28)$$

Im Diskreten ist der Unterschied sozusagen eine Anwendung des Distributivgesetzes. Ergebnisse mit gleichem Wert der Zufallsvariablen X werden in (4.24) zuerst zusammengefasst, d. h. ihre Wahrscheinlichkeiten addiert und dann mit dem Funktionswert multipliziert und zur Gesamtsumme addiert. In (4.27) wird jedes Ergebnis einzeln mit der Wahrscheinlichkeit multipliziert und aufsummiert.

4. Wahrscheinlichkeitstheorie (in \mathbb{R})

4.4.3. Rechenregeln für Erwartungswerte

$$E(X + Y) = \int_{\omega \in \Omega} (X + Y)(\omega) \cdot p(\omega) d\omega = \int_{\omega \in \Omega} ((X(\omega) + Y(\omega)) \cdot p(\omega) d\omega \quad (4.29)$$

$$= \left(\int_{\omega \in \Omega} X(\omega) \cdot p(\omega) d\omega \right) + \left(\int_{\omega \in \Omega} Y(\omega) \cdot p(\omega) d\omega \right) = E(X) + E(Y) \quad (4.30)$$

$$E(\lambda X) = \int_{\omega \in \Omega} (\lambda X)(\omega) \cdot p(\omega) d\omega = \int_{\omega \in \Omega} \lambda X(\omega) \cdot p(\omega) d\omega \quad (4.31)$$

$$= \lambda \int_{\omega \in \Omega} X(\omega) \cdot p(\omega) d\omega = \lambda E(X) \quad (4.32)$$

$$E(\lambda) = \int_{\omega \in \Omega} \lambda \cdot p(\omega) d\omega = \lambda \int_{\omega \in \Omega} p(\omega) d\omega = \lambda \quad (4.33)$$

X, Y **unabhängig**

$$E(XY) = \int_{\omega \in \Omega} (XY)(\omega) \cdot p(\omega) d\omega \stackrel{f:(x,y) \mapsto x \cdot y}{=} \int_{x,y \in \mathbb{R}} x \cdot y \cdot p(X = x \wedge Y = y) dx dy \quad (4.34)$$

$$= \int_{x \in \mathbb{R}} \int_{y \in \mathbb{R}} x \cdot y \cdot p(X = x) \cdot p(Y = y) dy dx \quad (4.35)$$

$$= \int_{x \in \mathbb{R}} x \cdot p(X = x) \cdot \left(\int_{y \in \mathbb{R}} y \cdot p(Y = y) dy \right) dx \quad (4.36)$$

$$= \left(\int_{x \in \mathbb{R}} x \cdot p(X = x) dx \right) \cdot \left(\int_{y \in \mathbb{R}} y \cdot p(Y = y) dy \right) = E(X) \cdot E(Y) \quad (4.37)$$

Q 6

Gegenbeispiel falls X, Y abhängig

M eine Münze mit $-1/+1$

$$E(M) = 0 \quad (4.38)$$

$$E(MM) = E(M^2) = 1 \neq E(M)^2 \quad (4.39)$$

A.1

5. Fusion zweier Messwerte

5.1. Rekapitulation Wahrscheinlichkeitsrechnung

- Kleinbuchstaben sind (meist bekannte) Zahlen
- Großbuchstaben: sind (meist unbekannte) Zufallsvariablen
- $E(X + Y) = EX + EY$, $E(\lambda X) = \lambda EX$
- X, Y unabhängig: $E(XY) = EX EY$
- $\mathcal{N}(\mu, \sigma^2)(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
- $X \sim \mathcal{N}(\mu, \sigma^2) \Rightarrow EX = \mu$

5.2. Z vs. z

Eine Variable, z.B. z ("klein- z ") steht für einen konkreten Wert, z.B. den Messwert der gerade gemessen wurde. Eine Zufallsvariable, z.B. Z ("groß- Z ") steht für eine Gesamtheit von möglichen Werten und bei welchem zufälligen Ergebnis (also auch mit welcher Wahrscheinlichkeit) diese Zufallsvariable welchen Wert hat.

Dieser Kurs nimmt je nach Kontext verschiedene Sichtweisen ein, was hier erläutert werden soll am Beispiel eines Schätzers der den Mittelwert zweier Messungen bildet (Abb. 5.1).

Das Zufallsexperiment (Würfel-Symbol) legt die Messungen fest, weil wir den Messvorgang als zufälliges Ereignis betrachten. Je nach Ergebnis des Zufallsexperimentes können dabei verschiedene Zahlen herauskommen, das ist was Zufall ausmacht (verzweigende Pfeile).

Die erste Sichtweise ist die des Algorithmus selbst (Computersymbol). Er bekommt diese konkreten Zahlen als Input und berechnet deren Mittelwert. Für ein Ergebnis des Zufallexperimentes, sind z.B. die Messungen 10 und 12 und der Algorithmus rechnet $(10 + 12)/2 = 11$. Für ein anderes Ergebnis sind die Messungen 11 und 10.3 und der Algorithmus rechnet $(11 + 10.3)/2 = 10.65$. In jedem Fall verrechnet der Algorithmus zwei konkrete Zahlen.

Nehmen wir nun die Sichtweise der oberen Figur der Abbildung, eines Algorithmusentwicklers ("Programmierer") an: Zeitlich befindet sie sich zwar eigentlich vor dem Zufallsexperiment, aber sie antizipiert beim Algorithmusentwurf, was der Algorithmus mit den konkreten Messwerten (also nach dem Zufallsexperiment) tun soll. Deshalb sieht der Algorithmus die Messungen z_1, z_2 (klein z), als konkrete Zahlen, ist aber allgemein für jede mögliche Zahl formuliert, z.B. als Formel $\hat{x}(z_1, z_2) = (z_1 + z_2)/2$. Algorithmen werden daher "mit Kleinbuchstaben" als Formeln oder Funktionen aufgeschrieben.

5. Fusion zweier Messwerte

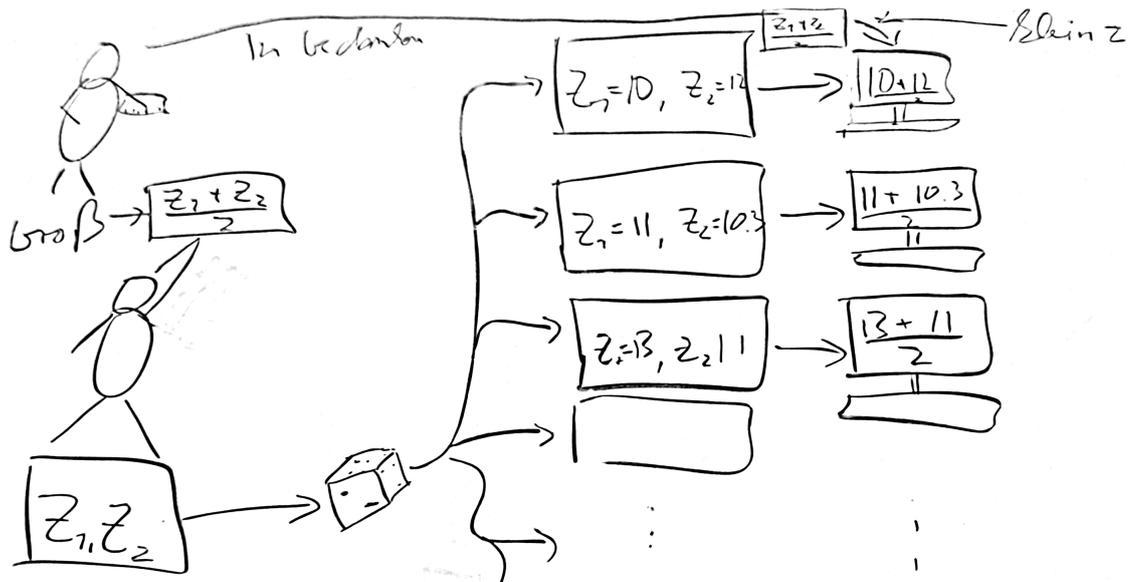


Abbildung 5.1.: Eine Zufallsvariable, z.B. eine Messung Z beinhaltet *alle möglichen* Ergebnisse der Zufallsexperimentes, wie wahrscheinlich sie sind und welchen Wert die Zufallsvariable dann hat (vgl. Text).

Nehmen wir nun die Sichtweise der unteren Figur der Abbildung, eines Algorithmens (”Wissenschaftler”) an: Sie interessiert sich dafür, wie gut der Algorithmus im Mittel funktioniert, also einem Erwartungswert über alle möglichen Ausgänge des Experiments. Daher wird die Ausgabe des Algorithmus zu einer Zufallsvariablen $\hat{X} = \hat{x}(Z_1, Z_2)$, ein deterministischer Algorithmus mit zufälligen Eingabedaten gefüttert. Man interessiert sich für die statistischen Eigenschaften von \hat{X} , analysiert also die Gesamtheit aller möglichen Ergebnisse des Zufallsexperimentes.

Zusammenfassung: Sensorfusionsalgorithmen werden später auf konkreten Zahlen als Messungen ausgeführt und daher allgemein als Algorithmus mit $\hat{x}(z_1, z_2)$ ”kleinen z ” aufgeschrieben. Für die Analyse oder Herleitung interessiert das Verhalten für alle möglichen Ergebnisse mit Wahrscheinlichkeiten und man betrachtet die Messungen Z_1, Z_2 und die daraus errechnete Schätzung $\hat{X} = \hat{x}(Z_1, Z_2)$ als Zufallsvariable mit ”großem Z ”.

5.3. Lineare Schätzformel für zwei Messungen

für eine Größe x gegeben zwei Messungen $Z_1 = z_1, Z_2 = z_2$ mit Genauigkeit σ_1, σ_2

5.3.1. Annahmen

$$E(Z_1 - x) = 0, \quad E(Z_2 - x) = 0 \quad (5.1)$$

$$E((Z_1 - x)^2) = \sigma_1^2, \quad E((Z_2 - x)^2) = \sigma_2^2 \quad (5.2)$$

Messfehler $Z_1 - x$ und $Z_2 - x$ sind unabhängig.

Q 7

Q 8

5.3.2. Forderungen

$$\hat{x} = \alpha z_1 + \beta z_2, \quad \text{also } \hat{X} = \alpha Z_1 + \beta Z_2 \quad (5.3)$$

$$E(\hat{X} - x) = 0 \quad \forall x, \quad E((\hat{X} - x)^2) \rightarrow \min \quad (5.4)$$

Das heißt, wir suchen den besten Schätzer unter allen die erwartungstreu sind (erwarteten Fehler 0) und von der Struktur eine Linearkombination der Messwerte.

Satz 2. Für zwei unabhängige erwartungstreue Messungen Z_1 , und Z_2 mit mittleren quadratischen Fehlern σ_1^2 bzw. σ_2^2 ist der optimale für alle x erwartungstreue lineare Schätzer \hat{X} mit mittlerem quadratischen Fehler $\sigma_{\hat{x}}^2$.

$$\hat{x}(z_1, \sigma_1, z_2, \sigma_2) = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} z_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} z_2, \quad \sigma_{\hat{x}}^2(\sigma_1, \sigma_2) = \frac{1}{1/\sigma_1^2 + 1/\sigma_2^2}. \quad (5.5)$$

Beweis.

$$\hat{X} = \hat{x}(Z_1, \sigma_1, Z_2, \sigma_2) \quad (5.6)$$

$$0 = E(\hat{X} - x) = E(\alpha Z_1 + \beta Z_2 - x) = \alpha E(Z_1) + \beta E(Z_2) - x \quad (5.7)$$

$$= \alpha x + \beta x - x = (\alpha + \beta - 1)x \quad (5.8)$$

$$\Rightarrow \beta = 1 - \alpha \quad \forall x = 0 \quad (5.9)$$

$$\Rightarrow \beta = 1 - \alpha, \text{ da für alle } x \text{ gültig} \quad (5.10)$$

$$\Rightarrow \hat{X} = \alpha Z_1 + (1 - \alpha) Z_2 \quad (5.11)$$

zu minimieren: $E((\hat{X} - x)^2)$

$$E((\hat{X} - x)^2) = E((\alpha Z_1 + (1 - \alpha) Z_2 - x)^2) \quad (5.12)$$

$$= E((\alpha Z_1 + (1 - \alpha) Z_2 - (\alpha + 1 - \alpha)x)^2) \quad (5.13)$$

$$= E((\alpha(Z_1 - x) + (1 - \alpha)(Z_2 - x))^2) \quad (5.14)$$

$$= E(\alpha^2(Z_1 - x)^2 + 2\alpha(1 - \alpha)(Z_1 - x)(Z_2 - x) + (1 - \alpha)^2(Z_2 - x)^2) \quad (5.15)$$

$$= \alpha^2 E((Z_1 - x)^2) + 2\alpha(1 - \alpha) E(Z_1 - x) E(Z_2 - x) + (1 - \alpha)^2 E((Z_2 - x)^2) \quad (5.16)$$

$$= \alpha^2 \sigma_1^2 + (1 - \alpha)^2 \sigma_2^2 \quad (5.17)$$

5. Fusion zweier Messwerte

Minimum bestimmen:

$$0 = \frac{d}{d\alpha} [\alpha^2 \sigma_1^2 + (1 - \alpha)^2 \sigma_2^2] = 2\alpha \sigma_1^2 + 2(1 - \alpha) \cdot (-1) \sigma_2^2 \quad (5.18)$$

$$= 2\alpha \sigma_1^2 - 2(1 - \alpha) \sigma_2^2 = 2\alpha(\sigma_1^2 + \sigma_2^2) - 2\sigma_2^2 \quad (5.19)$$

$$\Rightarrow \alpha = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \quad (5.20)$$

$$\Rightarrow \hat{X} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} Z_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} Z_2 \quad (5.21)$$

$$\sigma_{\hat{x}}^2 = E((\hat{X} - x)^2) = \alpha^2 \sigma_1^2 + (1 - \alpha)^2 \sigma_2^2, \quad \alpha = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \quad (5.22)$$

$$= \frac{\sigma_2^4 \sigma_1^2}{(\sigma_1^2 + \sigma_2^2)^2} + \frac{\sigma_1^4 \sigma_2^2}{(\sigma_1^2 + \sigma_2^2)^2} = \frac{\sigma_1^2 \sigma_2^2 (\sigma_1^2 + \sigma_2^2)}{(\sigma_1^2 + \sigma_2^2)^2} \quad (5.23)$$

$$= \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} = \frac{1}{1/\sigma_1^2 + 1/\sigma_2^2} \quad (5.24)$$

$$\Rightarrow \frac{1}{\sigma_{\hat{x}}^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \quad (5.25)$$

□

Folgerung: $\sigma_1 = \sigma_2 = \sigma \Rightarrow \sigma_{\hat{x}} = \frac{\sigma}{\sqrt{2}}$

5.3.3. Generalisierung auf n Messwerte

Wahre Größe: x

$$E(Z_i) = x$$

$Z_i - x$ sind unabhängig

$$E((Z_i - x)^2) = \sigma_i^2$$

Satz 3. Für n unabhängige erwartungstreue Messungen z_1, \dots, z_n mit mittleren quadratischen Fehlern $\sigma_1^2, \dots, \sigma_n^2$ ist der optimale erwartungstreue lineare Schätzer \hat{x} mit mittlerem quadratischen Fehler $\sigma_{\hat{x}}^2$.

$$\hat{x} = \sigma_{\hat{x}}^2 \sum_i^n \frac{z_i}{\sigma_i^2} \quad \sigma_{\hat{x}}^2 = \frac{1}{\sum_i^n \frac{1}{\sigma_i^2}} \quad (5.26)$$

Beweis.

I.A.: $n = 1$: klar

I.V.: $\hat{x}_n = \sigma_{\hat{x}_n}^2 \sum_i^n \frac{z_i}{\sigma_i^2}$

5.3. Lineare Schätzformel für zwei Messungen

I.S.:

$$\begin{aligned}\hat{x}_{n+1} &= \frac{\sigma_{n+1}^2 \cdot \hat{x}_n}{\sigma_{\hat{x}_n}^2 + \sigma_{n+1}^2} + \frac{\sigma_{\hat{x}_n}^2 \cdot z_{n+1}}{\sigma_{\hat{x}_n}^2 + \sigma_{n+1}^2} \\ &= \frac{\sigma_{n+1}^2 \cdot \sigma_{\hat{x}_n}^2}{\sigma_{n+1}^2 + \sigma_{\hat{x}_n}^2} \left(\sum_i^n \frac{z_i}{\sigma_i^2} + \frac{z_{n+1}}{\sigma_{n+1}^2} \right) \\ &= \sigma_{\hat{x}_{n+1}}^2 \sum_i^{n+1} \frac{z_i}{\sigma_i^2}\end{aligned}$$

□

5.4. Bayes Schätzer

5.4.1. x vs. X

Ein Paradoxon Sei $\sigma_1 = \sigma_2 = 1$

Ein Schätzer ist eine Funktion $\hat{x} : \mathbb{R}^2 \rightarrow \mathbb{R}$. Für welche Funktion wird der mittlere quadratische Fehler $E((\hat{x}(z_1, z_2) - x)^2)$ minimal?

Antwort: $\hat{x}(z_1, z_2) = x$, denn dann wird er 0. Das ist aber nicht die Antwort, die wir hören wollten, denn wir kennen ja x nicht. Jede Schätztheorie muss solche Artefakte verhindern. Im obigen Beispiel durch die Annahme eines linearen Zusammenhangs $\hat{x} = \alpha z_1 + \beta z_2$. Oder durch den Bayes'schen Ansatz, der nicht nur die Messfehler sondern auch die wahre Größe als zufällig betrachtet.

Bayes Modellierung X ist eine Zufallsgröße und wir kennen die Verteilung $p(Z_1 = z_1 | X = x)$ bzw. $p(Z_2 = z_2 | X = x)$. Für uns Gaußglocken im Messfehler. $Z_1 - X$ ist $\mathcal{N}(0, \sigma_1^2)$ verteilt. $Z_2 - X$ ist $\mathcal{N}(0, \sigma_2^2)$. D. h. $p(Z_1 = z_1 | X = x) = \frac{1}{\sqrt{2\pi\sigma_1}} \cdot \exp\left(-\frac{(z_1-x)^2}{2\sigma_1^2}\right)$

Wie ist $p(X = x | Z_1 = z_1 \wedge Z_2 = z_2)$?

Dieser Ansatz vermeidet die obigen Probleme. Da X eine Zufallsgröße ist und die Schätzung explizit nur von z_1 und z_2 abhängen darf, ist der wahre Werte keine konstante Funktion mehr und kann nicht der optimale (aber unbekante) Schätzer sein.

Satz 4 (Bayes). Für Zufallsvariablen X, Y gilt:

$$p(X = x | Y = y) = \frac{p(Y = y | X = x)p(X = x)}{p(Y = y)} \quad (5.27)$$

Beweis.

$$p(X = x | Y = y) \cdot p(Y = y) = p(X = x \wedge Y = y) \quad (5.28)$$

$$= p(Y = y | X = x) \cdot p(X = x) \quad (5.29)$$

Dividieren durch $p(Y = y)$ liefert die Behauptung. □

Intuition im Kontext von Messungen X ist der uns unbekante Zustand, Y die bekannte Messung. Wir fragen uns: „Was ist die Wahrscheinlichkeitsverteilung für den Zustand, gegeben, dass ich das gemessen habe, was ich gemessen habe.“ Bayes sagt dazu: „Die Wahrscheinlichkeit einen bestimmten Zustand zu haben, gegeben, dass ich das gemessen habe, was ich gemessen habe, ist gerade die Wahrscheinlichkeit das zu messen, wenn der Zustand so wäre mal der Wahrscheinlichkeit überhaupt in dem Zustand zu sein, dividiert durch die Wahrscheinlichkeit es zu messen.“

Meist betrachtet man (5.27) als Verteilung über x welches ja unbekannt ist, während Y bekannt ist. Dann ist der Nenner $p(Y = y)$ konstant und kann häufig ignoriert werden. Intuitiv: Wir haben gemessen, was wir gemessen haben; wie wahrscheinlich das war, ist uninteressant. Oder mit Sherlock Holmes:

“Wenn Du das Unmögliche ausgeschlossen hast, dann ist das, was übrig bleibt, die Wahrheit, wie unwahrscheinlich sie auch ist.”

The Adventure of the Beryl Coronet / Sherlock Holmes

5.4.2. Beispiel

Ein weißer und ein roter Würfel werden geworfen und ergeben 10 als Gesamtaugenanzahl. Wie groß ist die Wahrscheinlichkeit, dass der weiße eine 4 zeigt?

X = Augensumme zweier Würfel (weiß, rot)

Y = weißer Würfel

$$P(Y = 4|X = 10) = \frac{P(X = 10|Y = 4)P(Y = 4)}{P(X = 10)} = \frac{\frac{1}{6} \cdot \frac{1}{6}}{\frac{3}{36}} = \frac{1}{3} \quad (5.30)$$

Praktisch nützlich ist die Formel vor allen Dingen dann, wenn $p(X = 10|Y = 4)$ leichter zu formulieren/zu errechnen ist als $p(Y = 4|X = 10)$.

Q 10

5.4.3. Bayes-Schätzer für zwei Messungen

In vielen Rechnungen mit Wahrscheinlichkeitsverteilungen kann man konstante Faktoren ignorieren und statt $=$ einfach \propto schreiben und sich damit die Rechnung erheblich vereinfachen. Der Grund dafür ist der folgende:

Bei der Berechnung einer Wahrscheinlichkeitsverteilung wissen wir, dass das Integral 1 sein muss. Hat man eine sogenannte unnormierte Verteilung, deren Integral nicht 1 ist, weiß aber, dass sie proportional zu der gesuchten Verteilung ist, lässt sich die gesuchte Verteilung durch Normieren daraus gewinnen, d.h. man teilt die Verteilung durch ihr Integral.

Häufig erkennt man auch, dass die resultierende Verteilung aus einer bestimmten Klasse ist, z. B. Gaußsch, und kann die Parameter direkt ablesen, ohne formal zu normieren.

Deshalb kann man sich Rechnungen oft erheblich vereinfachen, indem man konstante Faktoren ignoriert und „proportional zu“ \propto anstatt $=$ schreibt. Da $e^{a+b} = e^a \cdot e^b$, ist konstante Faktoren zu ignorieren äquivalent dazu, konstante Summanden im Exponenten zu ignorieren.

Um die Notation zu vereinfachen, schreiben wir auch $\exp(a)$ statt e^a .

Zur Herleitung des Bayes-Schätzers verwenden wir zuerst allgemeine probabilistische Argumente (Bayes, Unabhängigkeit, Messfehler). Danach ergibt sich die konkrete rechnerische Aufgabe, das Produkt zweier Gaußverteilungen zu berechnen und zu sehen, dass es wieder eine Gaußverteilung ist und mit welchen Parametern. Diese Rechnung wird noch später benötigt, und wir ziehen sie deshalb in einen Satz vor.

Satz 5. *Das Produkt zweier Gaußverteilungen mit Parameter μ_1, σ_1^2 und μ_2, σ_2^2 ist proportional zu einer Gaußverteilungen mit Parametern analog zu Satz 2:*

$$\mathcal{N}(\mu_1, \sigma_1^2)(x) \cdot \mathcal{N}(\mu_2, \sigma_2^2)(x) \propto \mathcal{N}\left(\frac{\mu_1\sigma_2^2 + \mu_2\sigma_1^2}{\sigma_1^2 + \sigma_2^2}, \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right)(x) \quad (5.31)$$

5. Fusion zweier Messwerte

Beweis.

$$\mathcal{N}(\mu_1, \sigma_1^2)(x) \cdot \mathcal{N}(\mu_2, \sigma_2^2)(x) \quad (5.32)$$

$$\propto \exp\left(-\frac{(\mu_1 - x)^2}{2\sigma_1^2}\right) \cdot \exp\left(-\frac{(\mu_2 - x)^2}{2\sigma_2^2}\right) \quad (5.33)$$

$$= \exp\left(-\frac{(\mu_1 - x)^2}{2\sigma_1^2} - \frac{(\mu_2 - x)^2}{2\sigma_2^2}\right) \quad (5.34)$$

$$= \exp\left(-\frac{\sigma_2^2(\mu_1 - x)^2 + \sigma_1^2(\mu_2 - x)^2}{2\sigma_1^2\sigma_2^2}\right) \quad (5.35)$$

$$= \exp\left(-\frac{\sigma_2^2\mu_1^2 - 2\sigma_2^2\mu_1x + \sigma_2^2x^2 + \sigma_1^2\mu_2^2 - 2\sigma_1^2\mu_2x + \sigma_1^2x^2}{2\sigma_1^2\sigma_2^2}\right) \quad (5.36)$$

$$\propto \exp\left(-\frac{(\sigma_2^2 + \sigma_1^2)x^2 - 2(\sigma_2^2\mu_1 + \sigma_1^2\mu_2)x}{2\sigma_1^2\sigma_2^2}\right) \quad (5.37)$$

Wir wissen, dass eine quadratische Funktion im Exponenten eine Gaußverteilung bedeutet. Damit handelt es sich um eine Gaußverteilung. Wir müssen jetzt die Terme so umformen, dass sie auf die Formel $\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ passen. Da in dieser Form x^2 im Zähler einen Faktor von 1 hat, kürzen wir durch $(\sigma_2^2 + \sigma_1^2)$.

$$= \exp\left(-\frac{x^2 - 2\frac{\sigma_2^2\mu_1 + \sigma_1^2\mu_2}{\sigma_1^2 + \sigma_2^2}x}{2\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}}\right) \quad (5.38)$$

$$\propto \exp\left(-\frac{\left(x - \frac{\sigma_2^2\mu_1 + \sigma_1^2\mu_2}{\sigma_2^2 + \sigma_1^2}\right)^2}{2\frac{\sigma_1^2\sigma_2^2}{\sigma_2^2 + \sigma_1^2}}\right) \quad (5.39)$$

$$= \mathcal{N}\left(\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\mu_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}\mu_2, \frac{1}{1/\sigma_1^2 + 1/\sigma_2^2}\right)(x) \quad \square$$

Abbildung 5.2 zeigt ein grafisches Beispiel. Mit diesem rechnerischen Hintergrund ausgestattet können wir nun den Bayes-Schätzer für zwei Messwerte herleiten.

Q 11

5.4.4. Annahmen

X uniform verteilt¹

$$N_1 \text{ verteilt nach } \mathcal{N}(0, \sigma_1^2) \quad N_2 \text{ verteilt nach } \mathcal{N}(0, \sigma_2^2) \quad (5.40)$$

$$Z_1 = X + N_1 \quad Z_2 = X + N_2 \quad (5.41)$$

X, N_1, N_2 unabhängig

¹Eigentlich geht das nicht, wir rechnen hier aber so als für das gehen mit $p(X = x) = \text{const.}$

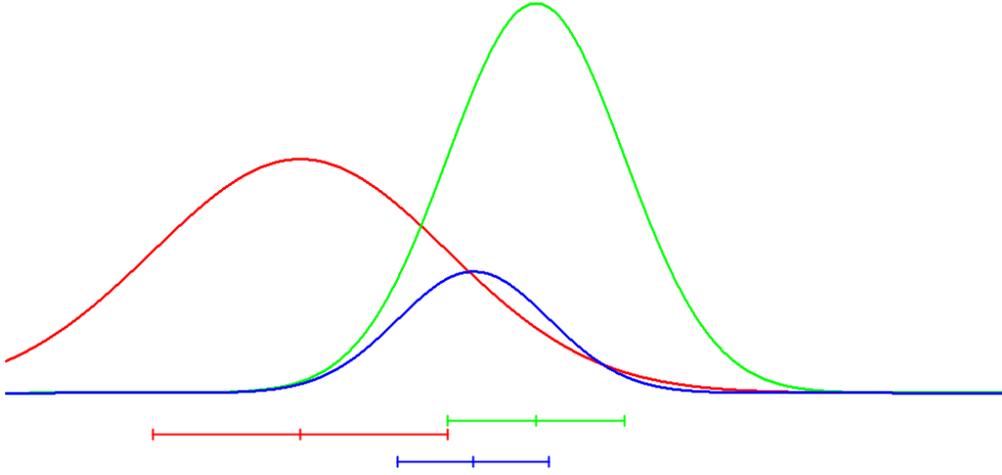


Abbildung 5.2.: Das Produkt zweier Gaußglocken (rot, grün) ist proportional zu einer Gaußglocke (blau). Die Graphen zeigen die Verteilungen, die horizontalen Linien zeigen $\mu - \sigma$, μ und $\mu + \sigma$.

Satz 6. Seien X, Z_1, Z_2, N_1, N_2 wie unter Annahmen und z_1, z_2 gegeben. Dann ist die Größe X a-posteriori ($X|Z_1 = z_1 \wedge Z_2 = z_2$) nach $\mathcal{N}(\hat{x}, \sigma_{\hat{x}}^2)$ verteilt, mit $\hat{x}, \sigma_{\hat{x}}^2$ wie in Satz 2.

$$\hat{x} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} z_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} z_2, \quad \sigma_{\hat{x}}^2 = \frac{1}{1/\sigma_1^2 + 1/\sigma_2^2}. \quad (5.42)$$

Beweis.

$$p(X = x|Z_1 = z_1 \wedge Z_2 = z_2) \stackrel{\text{Bayes}}{=} \frac{p(Z_1 = z_1 \wedge Z_2 = z_2|X = x)p(X = x)}{p(Z_1 = z_1 \wedge Z_2 = z_2)} \quad (5.43)$$

$$\propto p(Z_1 = z_1 \wedge Z_2 = z_2|X = x)p(X = x) \quad (5.44)$$

$$= p(N_1 = z_1 - x \wedge N_2 = z_2 - x|X = x)p(X = x) \quad (5.45)$$

$$= p(N_1 = z_1 - x \wedge N_2 = z_2 - x)p(X = x) \quad (5.46)$$

$$\stackrel{\text{unabh.}}{=} p(N_1 = z_1 - x)p(N_2 = z_2 - x)p(X = x) \quad (5.47)$$

$$\stackrel{\text{Ann. } p(X=x) \propto 1}{\propto} p(N_1 = z_1 - x)p(N_2 = z_2 - x) \quad (5.48)$$

$$= \mathcal{N}(0, \sigma_1^2)(z_1 - x) \cdot \mathcal{N}(0, \sigma_2^2)(z_2 - x) \quad (5.49)$$

$$= \mathcal{N}(z_1, \sigma_1^2)(x) \cdot \mathcal{N}(z_2, \sigma_2^2)(x) \quad (5.50)$$

$$\stackrel{\text{Satz 5}}{=} \mathcal{N}\left(\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} z_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} z_2, \frac{1}{1/\sigma_1^2 + 1/\sigma_2^2}\right)(x) \quad (5.51)$$

$$= \mathcal{N}(\hat{x}, \sigma_{\hat{x}}^2)(x) \quad (5.52)$$

□

5. Fusion zweier Messwerte

Es stellt sich also heraus, dass unter den Gaußschen Annahme, mit einer uniformen a-priori Verteilung $p(X)$ die a-posteriori Verteilung $p(X|z_1, z_2)$ eine Gaußverteilung ist mit denselben Parametern wie unser optimaler linearer Schätzer in [Satz 2](#). D. h. der besagte Schätzer ist optimal für Gaußverteilungen und im Allgemeinfall immer noch der beste lineare erwartungstreue Schätzer.

Q 12

6. Fortsetzung Wahrscheinlichkeitstheorie

Quellen zu Themen dieses Kapitel (abgerufen am 22.10.18)

- [https://de.wikipedia.org/wiki/Varianz_\(Stochastik\)](https://de.wikipedia.org/wiki/Varianz_(Stochastik))
- [https://de.wikipedia.org/wiki/Kovarianz_\(Stochastik\)](https://de.wikipedia.org/wiki/Kovarianz_(Stochastik))

6.1. Varianz

Definition 3.

$$V(X) = E((X - E(X))^2) \quad (6.1)$$

Die Varianz beschreibt die mittlere quadrierte Schwankungsbreite von X .

Q 13

Proposition 6.1. $V(X) = E(X^2) - E(X)^2$

Beweis.

$$V(X) = E((X - E(X))^2) = E(X^2 - 2XE(X) + E(X)^2) \quad (6.2)$$

$$= E(X^2) - E(2XE(X)) + E(E(X)^2) = E(X^2) - 2E(X)E(X) + E(X)^2 \quad (6.3)$$

$$= E(X^2) - E(X)^2 \quad \square$$

Eigenschaften:

1. $V(\lambda X) = \lambda^2 V(X)$
2. Wenn X, Y unabhängig sind, so gilt $V(X + Y) = V(X) + V(Y)$.
3. Wenn X_1, \dots, X_n unabhängig sind und $V(X_i) = \sigma^2$ für all i , so gilt

a) $V(\sum_{i=1}^n X_i) = n\sigma^2$

b) $V(\frac{1}{n} \sum_{i=1}^n X_i) = \frac{\sigma^2}{n}$

Beweis von 1.

$$V(\lambda X) = E((\lambda X)^2) - E(\lambda X)^2 = E(\lambda^2 X^2) - (\lambda E(X))^2 \quad (6.4)$$

$$= \lambda^2 E(X^2) - \lambda^2 E(X)^2 = \lambda^2 (E(X^2) - E(X)^2) \quad (6.5)$$

$$= \lambda^2 V(X) \quad \square$$

6. Fortsetzung Wahrscheinlichkeitstheorie

Beweis von 2.

$$V(X + Y) = E((X + Y)^2) - E(X + Y)^2 \quad (6.6)$$

$$= E(X^2 + 2XY + Y^2) - (E(X) + E(Y))^2 \quad (6.7)$$

$$= E(X^2) + 2E(XY) + E(Y^2) - (E(X)^2 + 2E(X)E(Y) + E(Y)^2) \quad (6.8)$$

$$= E(X^2) + 2E(X)E(Y) + E(Y^2) - E(X)^2 - 2E(X)E(Y) - E(Y)^2 \quad (6.9)$$

$$= E(X^2) + E(Y^2) - E(X)^2 - E(Y)^2 \quad (6.10)$$

$$= E(X^2) - E(X)^2 + E(Y^2) - E(Y)^2 \quad (6.11)$$

$$= V(X) + V(Y) \quad \square$$

Beweis von 3.a). direkt aus 2) ...

Bei einem Unabhängigkeitsbeweis durch Induktion wäre zu zeigen, dass die entstehenden Zwischensummen wieder unabhängig von den weiteren Summanden sind. Dazu muss man Unabhängigkeit der X_i voraussetzen in analogem Sinne wie für mehr als zwei Ereignisse, d.h. für $J \subseteq \{1, \dots, n\}$ braucht man $P(\forall j \in J, X_j = x_j) = \prod_{j \in J} P(X_j = x_j)$. Der folgende Beweis ist eine direkte Rechnung wie oben und benötigt nur paarweise Unabhängigkeit, also $p(X_i = x_i \wedge X_j = x_j) = p(X_i = x_i)p(X_j = x_j) \forall i \neq j$.

$$V(\sum_{i=1}^n X_i) = E((\sum_{i=1}^n X_i)^2) - E(\sum_{i=1}^n X_i)^2 \quad (6.12)$$

$$= E(\sum(X_i^2) + 2 \sum_{i < j} (X_i X_j)) - (\sum_{i=1}^n E(X_i))^2 \quad (6.13)$$

$$= \sum(E(X_i^2)) + 2 \sum_{i < j} (E(X_i) E(X_j)) - \quad (6.14)$$

$$\sum_{i=1}^n E(X_i)^2 - 2 \sum_{i < j} (E(X_i) E(X_j)) \quad (6.15)$$

$$= \sum(E(X_i^2) - E(X_i)^2) = \sum(V(X_i)) = \sum(\sigma^2) \quad (6.16)$$

$$= n\sigma^2. \quad (6.17)$$

□

Beweis von 3.b). aus a) und 1)

$$V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} V\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} n\sigma^2 \quad (6.18)$$

$$= \frac{\sigma^2}{n} \quad \square$$

6.2. Standardabweichung (Streuung)

$\sigma(X) = \sqrt{V(X)}$ „mittlere Abweichung“. Anders als $V(X)$ hat $\sigma(X)$ dieselbe Einheit wie X und kann mit X in ein Koordinatensystem eingetragen werden.

Q 14

6.2.1. Eigenschaften

1. $\sigma(\lambda X) = |\lambda|\sigma(X)$
2. Wenn X, Y unabhängig sind, so gilt $\sigma(X + Y) = \sqrt{\sigma(X)^2 + \sigma(Y)^2}$
3. Wenn X_i unabhängig sind und $\sigma(X_i) = \sigma$, so gilt
 - a) $\sigma(\sum X_i) = \sqrt{n}\sigma$
 - b) $\sigma(\frac{1}{n} \sum X_i) = \frac{\sigma}{\sqrt{n}}$

(Folgt alles aus den entsprechenden Eigenschaften der Varianz.)

Q 15

6.3. Kovarianz

Misst, wie stark die Schwankung von X und Y gemeinsam (verkoppelt) erfolgt.

$$\text{Cov}(X, Y) := E((X - E(X))(Y - E(Y))) \quad (6.19)$$

Proposition 6.2. $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$

Beweis. analog Varianz

□

6.3.1. Eigenschaften

1. $\text{Cov}(X, X) = V(X)$
2. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
3. $\text{Cov}(\lambda X, Y) = \lambda \text{Cov}(X, Y)$
4. $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$
5. $\text{Cov}(\lambda, X) = 0$
6. $V(X + Y) = V(X) + 2 \text{Cov}(X, Y) + V(Y)$
7. X, Y unabhängig $\Rightarrow \text{Cov}(X, Y) = 0$

Beweis von 1. klar

□

Beweis von 2. klar

□

Beweis von 3.

$$\text{Cov}(\lambda X, Y) = E(\lambda XY) - E(\lambda X)E(Y) = \lambda(E(XY) - E(X)E(Y)) = \lambda \text{Cov}(X, Y) \quad (6.20)$$

□

6. Fortsetzung Wahrscheinlichkeitstheorie

Beweis von 4.

$$\text{Cov}(X + Y, Z) = E((X + Y)Z) - E(X + Y)E(Z) \quad (6.21)$$

$$= E(XZ + YZ) - (E(X) + E(Y))E(Z) \quad (6.22)$$

$$= E(XZ) + E(YZ) - E(X)E(Z) - E(Y)E(Z) \quad (6.23)$$

$$= E(XZ) - E(X)E(Z) + E(YZ) - E(Y)E(Z) \quad (6.24)$$

$$= \text{Cov}(X, Z) + \text{Cov}(Y, Z) \quad \square$$

Beweis von 5.

$$\text{Cov}(1, Z) = E(\lambda Z) - E(\lambda)E(Z) = \lambda E(Z) - \lambda E(Z) \quad (6.25)$$

$$= 0 \quad \square$$

Beweis von 6.

$$V(X + Y) = E((X + Y)^2) - (E(X) + E(Y))^2 \quad (6.26)$$

$$= E(X^2 + 2XY + Y^2) - E(X)^2 - 2E(X)E(Y) - E(Y)^2 \quad (6.27)$$

$$= E(X^2) + 2E(XY) + E(Y^2) - E(X)^2 - 2E(X)E(Y) - E(Y)^2 \quad (6.28)$$

$$= E(X^2) - E(X)^2 + 2E(XY) - 2E(X)E(Y) + E(Y^2) - E(Y)^2 \quad (6.29)$$

$$= V(X) + 2\text{Cov}(X, Y) + V(Y) \quad \square$$

Beweis von 7.

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) \stackrel{\text{unabh.}}{=} E(X)E(Y) - E(X)E(Y) = 0 \quad (6.30)$$

\square

Q 16

6.4. Korrelation

Die Kovarianz $\text{Cov}(X, Y)$ gibt an, "wie stark" X und Y gemeinsam schwanken. Passend dazu ist $\text{Cov}(X, X) = V(X)$ die Stärke der Schwankung und durch verdoppeln einer ZV verdoppelt sich auch $\text{Cov}(X, 2Y) = 2\text{Cov}(X, Y)$. Manchmal interessiert eine „normierte Kovarianz“ die unabhängig von der Größe der Schwankung angibt, welcher "Anteil" der Schwankung gemeinsam ist.

Korrelation

Falls $\sigma(X), \sigma(Y) > 0$, ist die Korrelation der Zufallsvariablen X und Y definiert als

$$\text{cor}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)} = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)V(Y)}}. \quad (6.31)$$

Die Korrelation hat die Bedeutung, welcher Anteil der Schwankung von X und Y gemeinsam ist, wobei < 0 gegenläufige Schwankung bedeutet. Dementsprechend ist $-1 \leq \text{cor}(X, Y) \leq 1$, was im folgenden bewiesen werden soll.

Lemma 6.3 (Cauchy-Schwarz-Ungleichung).

$$E(XY)^2 \leq E(X^2) E(Y^2) \quad (6.32)$$

Zusatz: „=" gdw. $(a, b) \neq (0, 0)$ ex. mit $P(aX + bY = 0) = 1$, d.h. $aX + bY = 0$ f. ü.

Beweis. Ohne Zusatz.

$$0 \leq E((aX + bY)^2) = a^2 E(X^2) + 2ab E(XY) + b^2 E(Y^2) \quad (6.33)$$

□

Beweis. Mit Zusatz:

„dann“: Sei $aX + bY = 0$. Sei o. E. $b \neq 0$, dann $Y = -\frac{a}{b}X$, also beide Seiten $= \frac{a^2}{b^2} E(X^2)^2$

„nur dann“: Für alle $(a, b) \neq (0, 0)$ gelte $P(aX + bY = 0) < 1$. Dann:

$$0 < E((aX + bY)^2) = a^2 E(X^2) + 2ab E(XY) + b^2 E(Y^2)$$

Die Frage ist, wie wählen wir a und b ? Wir brauchen hinterher einen Term $E(X^2)E(Y^2)$. Daher wählen wir $a = \sqrt{E(Y^2)}$, $b = \sqrt{E(X^2)}$.

$$0 < E(Y^2)E(X^2) + 2\sqrt{E(Y^2)}\sqrt{E(X^2)}E(XY) + E(X^2)E(Y^2) \quad (6.34)$$

$$= 2E(X^2)E(Y^2) + 2\sqrt{E(Y^2)}\sqrt{E(X^2)}E(XY) \quad (6.35)$$

$$0 < E(XY) + \sqrt{E(Y^2)}\sqrt{E(X^2)} \quad (6.36)$$

$$-\sqrt{E(Y^2)}\sqrt{E(X^2)} < E(XY) \quad (6.37)$$

Dreht man das Vorzeichen von a um, d.h. $a = -\sqrt{E(Y^2)}$. Dann ergibt sich umgekehrt

$$0 < E(Y^2)E(X^2) - 2\sqrt{E(Y^2)}\sqrt{E(X^2)}E(XY) + E(X^2)E(Y^2) \quad (6.38)$$

$$= 2E(X^2)E(Y^2) + 2\sqrt{E(Y^2)}\sqrt{E(X^2)}E(XY) \quad (6.39)$$

$$0 < -E(XY) + \sqrt{E(Y^2)}\sqrt{E(X^2)} \quad (6.40)$$

$$E(XY) < \sqrt{E(Y^2)}\sqrt{E(X^2)} \quad (6.41)$$

6. Fortsetzung Wahrscheinlichkeitstheorie

und zusammen ergibt sich die Behauptung. □

Proposition 6.4.

$$-1 \leq \text{cor}(X, Y) \leq 1 \tag{6.42}$$

Zusatz: $\text{cor}(X, Y) = \pm 1$, gdw. linearer Zusammenhang zwischen X und Y , d.h. es ex. $(a, b, c) \neq (0, 0, 0)$ mit $P(aX + bY + c = 0) = 1$, „ $aX + bY + c = 0$ fast überall“

Beweis von Prop. 6.4.

$$(\text{Cov}(X, Y))^2 = (\text{E}((X - \text{E}X)(Y - \text{E}Y)))^2 \tag{6.43}$$

$$\stackrel{\text{C.S.}}{\leq} \text{E}((X - \text{E}(X))^2) \cdot \text{E}((Y - \text{E}(Y))^2) \tag{6.44}$$

$$= \text{V}(X) \text{V}(Y) \tag{6.45}$$

und „ $=$ “ gdw. ex. $(a, b) \neq (0, 0)$

$$a(X - \text{E}X) + b(Y - \text{E}Y) = 0 \text{ f. ü.} \tag{□}$$

6.5. Die Faltung

Für $f, g : \mathbb{R} \rightarrow \mathbb{R}$ (hinreichend harmlos) ist die Faltung $f * g : \mathbb{R} \rightarrow \mathbb{R}$ def. durch:

$$(f * g)(z) = \int_{\mathbb{R}} f(x)g(z - x)dx \tag{6.46}$$

Satz 7. Seien X, Y unabhängige ZV mit $p(X = x) = f(x)$ und $p(Y = y) = g(y)$ dann ist $p(X + Y = z) = (f * g)(z)$.

In Worten: $X + Y$ hat als Wahrscheinlichkeitsverteilung die Faltung der Verteilungen von X und Y .

Beweis. Nach unserer Definition von Wahrscheinlichkeitsdichten müssen wir zeigen, dass das Integral über die Faltungsdichte in einer (harmlosen) Teilmenge $A \subset \mathbb{R}$ die Wahrscheinlichkeit $P(X + Y \in A)$ angibt, dass die Summe $X + Y$ in A liegt. In der Tat (Ergebnis aus der Maßtheorie) reicht es, dies für den Fall zu zeigen, dass A ein Intervall $[a, b]$ ist.

$$P(X + Y \in [a, b]) \tag{6.47}$$

$$\stackrel{\text{s.o.}}{=} \iint_{x+y \in [a, b]} p(X = x \wedge Y = y) dy dx \tag{6.48}$$

$$\stackrel{X, Y \text{ unabh.}}{=} \iint_{x+y \in [a, b]} p(X = x)p(Y = y) dy dx \tag{6.49}$$

$$\stackrel{\substack{z=x+y \\ y=z-x}}{=} \int_{z=a}^b \int_{x=-\infty}^{\infty} p(X=x)p(Y=z-x)dx dz \quad (6.50)$$

$$= \int_{z=a}^b (p(X=\cdot) * p(Y=\cdot))(z) dz \quad (6.51)$$

□

Q 19

Satz 8. Sind zwei Zufallsvariablen unabhängig Gaußverteilt mit $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$ und $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$, dann ist es ihre Summe auch mit $X + Y \sim \mathcal{N}(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$.

Beweis. Nur Skizze: Anwendung der Faltung, dann ähnliche Rechnung wie bei Produkt zweier Gaußverteilungen. □

6.6. Zentraler Grenzwertsatz

„Die Summe vieler unabhängiger Zufallsvariablen ist annähernd normalverteilt.“

Formalisierung trickreich: Wörtlich genommen stimmt die Aussage nicht, weil die Summe von n ZV mit Varianz σ^2 Varianz $n\sigma^2$ hat, also nicht gegen eine Gaußverteilung konvergiert, die ja eine endliche Varianz hat. Statt dessen muss man die Summe so verschieben und skalieren, dass Erwartungswert 0 und Varianz 1 wird. Dann kann man zeigen, dass die Form der Verteilung gegen eine Einheitsgaußglocke konvergiert.

Satz 9 (ZGS). Sind $X_i, i \geq 1$, identisch verteilte unabhängige ZV mit Varianz σ^2 und $E(X_i) = \mu$, so gilt für $S_n := \sum_{i=1}^n X_i$, $S_n^* = \frac{S_n - n\mu}{\sigma\sqrt{n}}$:

$$P(S_n^* \in [a, b]) \xrightarrow[\text{gleichmäßig}]{n \rightarrow \infty} \int_a^b \mathcal{N}(0, 1)(z) dz \quad \forall a, b. \quad (6.52)$$

Ohne Beweis.

A.4

7. Bayes-Filter

7.1. Bayes-Netze

Zufallsvariablen X und Y sind unabhängig, genau dann wenn: $p(X = x|Y = y) = p(X = x)$.

Beweis: Wenn X und Y unabhängig

$$p(X = x|Y = y) = \frac{p(X = x \wedge Y = y)}{p(Y = y)} = \frac{p(X = x)p(Y = y)}{p(Y = y)} = p(X = x). \quad (7.1)$$

Umgekehrt, wenn $p(X = x|Y = y) = p(X = x)$, dann

$$p(X = x \wedge Y = y) = p(X = x|Y = y)p(Y = y) = p(X = x)p(Y = y) \quad (7.2)$$

Zufallsvariablen X und Y sind unabhängig gegeben $Z = z$, genau dann wenn:

$$p(X = x \wedge Y = y|Z = z) = p(X = x|Z = z)p(Y = y|Z = z) \quad (7.3)$$

$$\Leftrightarrow p(X = x|Y = y \wedge Z = z) = p(X = x|Z = z) \quad (7.4)$$

Derartige Abhängigkeitsstrukturen lassen sich graphisch als Bayes-Netz darstellen. Ein Bayes-Netz ist ein gerichteter Graph (V, E) , $E \subset V \times V$ ohne Kreise bei dem die Knoten $v \in V$ Zufallsvariablen X_v darstellen. Die Kanten geben intuitiv direkte Abhängigkeiten an. Eine Kante von v' nach v bedeutet, dass $X_{v'}$ X_v beeinflusst. Häufig modelliert dies Kausalität. Formal definiert sich die Wahrscheinlichkeitsverteilung der Zufallsvariablen als Produkt von je einer Verteilung pro Knoten, wobei die Verteilung eines Knoten abhängig ist von allen Elternknoten:

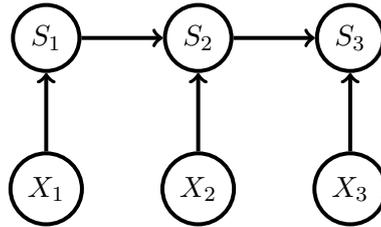
$$p(X_v = x_v \forall v \in V) = \prod_{v \in V} p(X_v = x_v | X_{v'} = x_{v'} \forall (v', v) \in E) \quad (7.5)$$

Im Knoten v wird die Verteilung $p(X_v = x_v | X_{v'} = x_{v'} \forall (v', v) \in E)$ gespeichert und die Bayesnetz-Definition (7.5) macht daraus eine Verbundverteilung aller X_v zusammen. Da die Abhängigkeitsstruktur der Verteilung im Graph repräsentiert sind können Inferenzalgorithmen "entlang dieses Graphen arbeiten".

7.1.1. Beispiel

Dreimal (unabhängig) Würfeln: X_1, X_2, X_3 .

Folgende Augensummen bilden: $S_1 = X_1, S_2 = S_1 + X_2, S_3 = S_2 + X_3$



Q 20

7.2. Problem: Zustandsschätzung

Ein System hat einen Zustand X_t der sich in Zeitschritten t ändert. Beobachtet werden Messungen $U_t = u_t$ bezüglich des Übergangs X_{t-1} auf X_t und $Z_t = z_t$ bezüglich X_t . Das große U_t bzw. Z_t sagt aus, dass die Messungen Zufallsvariablen sind, also vom zufälligen Ausgang des Experimentes abhängen. Die kleinen Buchstaben u_t bzw. z_t sagen aus, dass wir die Messungen kennen. Ihre Werte sind die Eingabe für den Algorithmus und wir befinden uns in der Situation

$$|Z_{t'} = z_{t'} \wedge U_{t'} = u_{t'} \forall t' \leq t, \quad (7.6)$$

“gegeben, dass wir gemessen haben, was wir gemessen haben.”

Gesucht ist in dieser Situation die Wahrscheinlichkeitsverteilung des aktuellen Zustandes X_t

$$p(X_t = x_t | Z_{t'} = z_{t'} \wedge U_{t'} = u_{t'} \forall t' \leq t) \quad (7.7)$$

7.2.1. Beispiel

Siehe [Abbildung 3.1](#) (Beispiel Schienenfahrzeug).

7.2.2. Annahmen

1. (Markov I) X_t ist unabhängig von allem anderen, gegeben X_{t-1}, U_t :

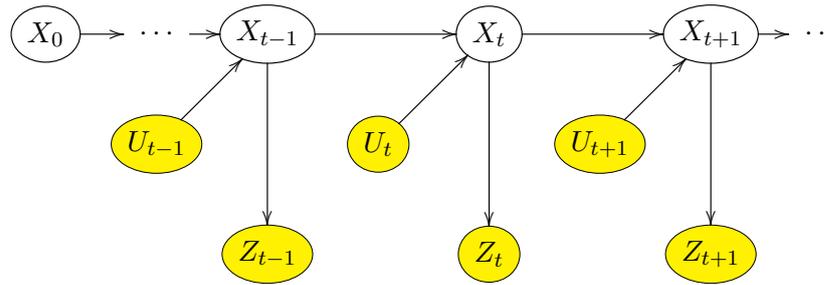
$$p(X_t = x_t | X_{t-1} = x_{t-1}, U_t = u_t, \dots) = p(X_t = x_t | X_{t-1} = x_{t-1}, U_t = u_t)$$

2. (Markov II) Die Messung Z_t ist unabhängig von allem anderen, gegeben X_t :

$$p(Z_t = z_t | X_t = x_t, \dots) = p(Z_t = z_t | X_t = x_t)$$

Das folgende Bayes-Netz zeigt diese Struktur als Graph. Die Zufallsvariablen U_t und Z_t sind hierbei gelb markiert, weil sie in einem Sensorfusionsszenario bekannt sind. Wir befinden uns also in dem Wahrscheinlichkeitsraum ”unter der Bedingung dass wir gemessen haben was wir gemessen haben.”

Q 21



7.2.3. Notation

In der Bayes-Formulierung der Zustandsschätzung mit Markov-Annahmen kommen viele Bedingungen der Form $X_t = x_t$, also dass eine bestimmte Zufallsvariable einen bestimmten Wert hat vor. Dabei kann der Wert wirklich bekannt z_t, u_t oder eine Variable im Sinne einer "für alle" Aussage oder eines Maximierungsprozesses sein. Um hier die Notation zu verkürzen vereinbaren wir, dass der Kleinbuchstabe alleine für

Großbuchstabe=Kleinbuchstabe

stehen soll. Also, z.B.

$$p(x_t | z_{1..t}, u_{1..t}) = p(X_t = x_t | Z_{t'} = z_{t'}, U_{t'} = u_{t'} \forall t' = 1 \dots t) \quad (7.8)$$

7.3. Algorithmus „Bayes-Filter“

Der Bayes Filter basiert auf der Bayes Formel und berechnet die Wahrscheinlichkeitsverteilung des Zustandes. Dazu benutzt er gegebene Messungen z_t und Zustandsübergänge u_t . Mit dem Bayes-Filter kann man Probleme der Zustandsschätzung lösen.

Der Bayes-Filter hat zwei wichtige Schritte: Der Dynamikschritt lässt die Zeit weiterticken und verarbeitet u_t . Er heißt auch Prädiktionsschritt (prediction), weil er den neuen Zustand aufgrund des alten vorhersagt und führt typischerweise zu mehr Unsicherheit. Der Messschritt verarbeitet die Messung z_t . Er heißt auch Korrekturschritt (update), weil er die Vorhersage auf Basis der Messung korrigiert.

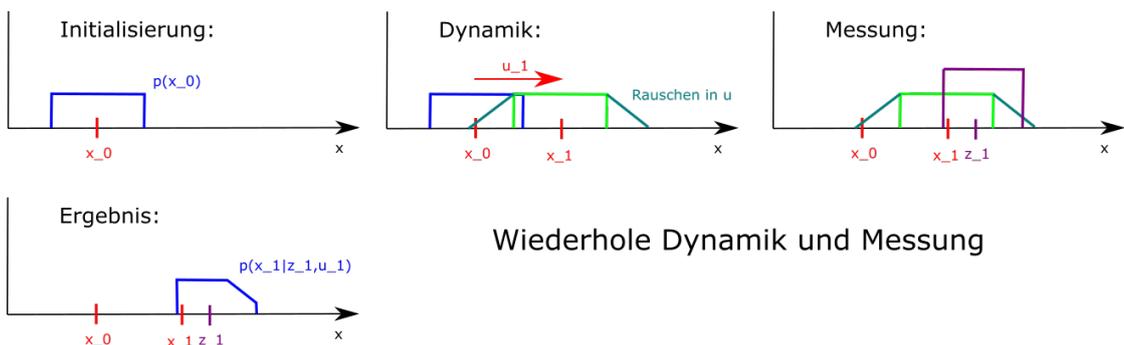


Abbildung 7.1.: Der Bayes Filter in Bildern

Notation: $p(x)$ soll heißen $p(X = x)$

Variable: Bel Wahrscheinlichkeitsverteilung Zustand (7.9)

Initialisierung: $Bel(x_0) := p(x_0) \forall x_0$ (7.10)

Dynamik(u_t): $Bel(x_t) := \int p(x_t | u_t, x_{t-1}) Bel(x_{t-1}) dx_{t-1} \forall x_t$ (7.11)

Messung(z_t): $Bel(x_t) := p(z_t | x_t) Bel(x_t) \forall x_t$ (7.12)

Anschließend normiere $Bel(x_t)$, so dass $\int Bel(x_t) = 1$.

Satz 10. Bei abwechselndem Aufruf von *Dynamik*(u_t) und *Messung*(z_t) für $t' = 1..t$ ist $Bel(x_t) = p(x_t | u_{1..t}, z_{1..t})$.

Beweis. Der Beweis erfolgt durch Induktion nach t wie häufig bei Algorithmen mit Schleifen. Im Prinzip sagt der Induktionsanfang, dass von vornherein alles richtig ist und der Induktionsschritt besagt, dass ein einzelner Schleifendurchlauf das richtige macht.

Induktionsanfang: $t = 0$ folgt aus (7.10)

Induktionsschritt: $t - 1 \rightsquigarrow t$

$$p(x_t | u_{1..t}, z_{1..t}) \stackrel{\text{Bayes}}{=} \frac{p(z_t | x_t, u_{1..t}, z_{1..t-1}) p(x_t | u_{1..t}, z_{1..t-1})}{p(z_t | u_{1..t}, z_{1..t-1})} \quad (7.13)$$

$$\stackrel{z_t \text{ const}}{\propto} p(z_t | x_t, u_{1..t}, z_{1..t-1}) p(x_t | u_{1..t}, z_{1..t-1}) \quad (7.14)$$

$$\stackrel{\text{Markov II}}{=} p(z_t | x_t) p(x_t | u_{1..t}, z_{1..t-1}) \quad (7.15)$$

$$\stackrel{\int_{-1}}{=} p(z_t | x_t) \int p(x_t, x_{t-1} | u_{1..t}, z_{1..t-1}) dx_{t-1} \quad (7.16)$$

$$= p(z_t | x_t) \int p(x_t | x_{t-1}, u_{1..t}, z_{1..t-1}) p(x_{t-1} | u_{1..t}, z_{1..t-1}) dx_{t-1} \quad (7.17)$$

$$\stackrel{\text{Markov I}}{=} p(z_t | x_t) \int p(x_t | x_{t-1}, u_t) p(x_{t-1} | u_{1..t-1}, z_{1..t-1}) dx_{t-1} \quad (7.18)$$

$$\stackrel{\text{I.A.}}{=} p(z_t | x_t) \int p(x_t | x_{t-1}, u_t) Bel(x_{t-1}) dx_{t-1} \quad (7.19)$$

$$\stackrel{\text{Algorithmus}}{\propto} Bel(x_t) \quad (7.20)$$

Obige Gleichungskette zeigt, dass $Bel(x_t) \propto p(x_t | u_{1..t}, z_{1..t})$. Sie müssen sogar gleich sein, da das Integral von Bel auf 1 normiert wird und die $p(x_t | \dots)$ als Wahrscheinlichkeitsverteilung ein Integral von 1 hat. \square

A.5

7.3.1. Beispiel: Ein Bayessches Spiel

In dem Spiel möchte ein Jäger einen Hasen fangen, hat aber nur sehr unsichere Information wo der Hase ist. Ein Bayesfilter soll ihm helfen.

7. Bayes-Filter

Es gibt 6 Gebüsch. Der Hase startet auf Gebüsch 1 und zieht zufällig in jeder Runde jeweils mit Wahrscheinlichkeit $1/6$ ein Gebüsch weiter. Dies kann der Jäger nicht beobachten.

Die Gebüsch sind magisch und in jeder Runde wispert ein zufälliges Gebüsch, ob der Hase im Verlaufe des Spiels dort schon mal war oder nicht. Welches Gebüsch wispert weiß der Jäger auch nicht.

Im 6. Gebüsch hat der Jäger ein Falle versteckt, die er mit einem Seil auslösen kann. Löst der Jäger die Falle aus und der Hase befindet sich darunter hat der Jäger gewonnen, befindet er sich nicht darunter oder entkommt er vorher, zieht also vom 6. Gebüsch ins "Ziel" hat der Hase gewonnen.

Ein Bayesfilter soll aufgrund der unsicheren Messungen (Wisperungen der Gebüsch) berechnen, wie wahrscheinlich der Hase wo ist.

- $Z_t \in \{\text{false}, \text{true}\}$ Wispert das Gebüsch "der Hase war hier" (true) oder "der Hase war nicht hier" (false)
- $U_t = ()$ Zu der Zustandsänderung gehört in diesem Beispiel keine Messung.
- $X_t \in [1, 7]$ Wo ist der Hase (7 für gewonnen)
- $p(z_t|x_t)$ modelliert, wie wahrscheinlich die Messung z_t , also Hase gesehen oder nicht gesehen eintritt, wenn der Hase im Gebüsch x_t ist.
- $p(x_{t+1}|x_t, u_t)$ modelliert die Bewegung des Hasen, d.h. das zufällige Vorziehen ($1/6$) oder stehen bleiben ($5/6$).

Ein Bayes-Filter der die Position der gejagten Figur auf dem Brett verfolgt kann als einfacher Array-Filter implementiert werden. Dabei wird in `bel[i] = p(X_t = i | z_{1..t}, u_{1..t})` gespeichert und mit jedem Mess/Dynamikschritt aktualisiert.

Verwendet man als Taktik für den Jäger, die Falle zu aktivieren, wenn Feld 6 der wahrscheinlichste Aufenthaltsort des Hasen ist, so gewinnt der Bayes-Schätzer ca. 37% (126/340) der Spiele. Zum Vergleich: Die sensorlose Strategie am Zeitpunkt mit der größten Wahrscheinlichkeit ($t = 5 \cdot 6 = 30$) die Falle zu aktivieren hat nur in 19% der Fälle Erfolg.

Demo: bayessimulation.py

Q 22

8. Kalman Filter

Die ursprüngliche Idee war es, den Kalman Filter herzuleiten indem in den Bayes Filter konkrete Gaußverteilungen eingesetzt werden. Das ist aber sehr rechenaufwändig. Von daher leiten wir im Folgenden den Kalman Filter direkt her und nutzen dabei schon bewiesene Resultate.

8.1. Konditionierungslemma

Definition 4. Zwei Zufallsvariablen $X, Z : \Omega \rightarrow \mathbb{R}$ heißen *gemeinsam Gaußverteilt*, wenn

$$p(X = x \wedge Z = z) \propto \exp \left(-\frac{1}{2} \begin{pmatrix} x - EX \\ z - EZ \end{pmatrix}^T \begin{pmatrix} VX & \text{Cov}(X, Z) \\ \text{Cov}(X, Z) & VZ \end{pmatrix}^{-1} \begin{pmatrix} x - EX \\ z - EZ \end{pmatrix} \right) \quad (8.1)$$

Man sieht im Exponenten stehen Terme $\propto x^2$, $\propto xz$ und $\propto z^2$ die ihre Koeffizienten aus der Inversen der angegebenen Matrix nehmen. Der Erwartungswert verschiebt das ganze.

Diese Definition ist eine Spezialisierung der Definition für mehrdimensionale Gaußverteilungen, die wir in Kapitel 11 einführen werden. Sie greift auf die Konzepte Vektor und Matrix zurück, die wir ausführlich in Kapitel 10 diskutieren. Dieser Vorgriff ist kurz. Er wird benötigt, weil wir über eine gemeinsame Verteilung von X und Z sprechen wollen um $p(X = x \wedge Z = z)$ auszurechnen. Dies wollen wir in einer Art tun, die strukturell genauso ist wie der analoge Beweis für mehrdimensionale X und Z in Kapitel 13. Daher dieser methodische Vorgriff.

Q 23

8. Kalman Filter

Bemerkung 8.1. Sind Zufallsvariablen $X, Z : \Omega \rightarrow \mathbb{R}$ Gaußverteilt und unabhängig, dann sind sie auch gemeinsam Gaußverteilt.

Beweis.

$$p(X = x \wedge Z = z) = p(X = x)p(Z = z) \quad (8.2)$$

$$\propto \exp\left(-\frac{1}{2} \frac{(x - EX)^2}{VX}\right) \exp\left(-\frac{1}{2} \frac{(z - EZ)^2}{VZ}\right) \quad (8.3)$$

$$= \exp\left(-\frac{1}{2} \begin{pmatrix} x - EX \\ z - EZ \end{pmatrix}^T \begin{pmatrix} VX^{-1} & 0 \\ 0 & VZ^{-1} \end{pmatrix} \begin{pmatrix} x - EX \\ z - EZ \end{pmatrix}\right) \quad (8.4)$$

$$= \exp\left(-\frac{1}{2} \begin{pmatrix} x - EX \\ z - EZ \end{pmatrix}^T \begin{pmatrix} VX & 0 \\ 0 & VZ \end{pmatrix}^{-1} \begin{pmatrix} x - EX \\ z - EZ \end{pmatrix}\right) \quad (8.5)$$

□

Bemerkung 8.2. Sind Zufallsvariablen $X, \delta : \Omega \rightarrow \mathbb{R}$ Gaußverteilt und unabhängig und $c, h \in \mathbb{R}$, dann sind X und $Z = cX + h + \delta$ gemeinsam Gaußverteilt.

Wir beweisen Bemerkung 8.2 hier nicht, weil sie mit unserem jetzigen Handwerkszeug sehr umständlich zu beweisen ist. Mit den Mitteln aus Kapitel 11 folgt sie unmittelbar. Inhaltlich ist das Argument, dass nach Bemerkung 8.1 X und δ gemeinsam Gaußsch sind und dies erhalten bleibt, wenn man lineare Operationen, wie $(x, \delta) \mapsto cX + h + \delta$ ausführt.

Q 24

Lemma 8.3 (Konditionierungslemma). *Sind X und Z gemeinsam Gaußverteilt und $z \in \mathbb{R}$, dann ist $X|Z = z$ (d.h. unter der Bedingung $Z = z$) auch Gaußverteilt und*

$$E(X|Z = z) = EX + \frac{\text{Cov}(X, Z)}{VZ}(z - EZ) \quad (8.6)$$

$$V(X|Z = z) = VX - \frac{\text{Cov}(X, Z)^2}{VZ}. \quad (8.7)$$

Beweis. Betrachten wir die Verteilung von X unter der Bedingung $Z = z$. Wir werden die Formel aus (8.1) nehmen, den Wert von z fixieren, die bekannte Formel für die Inverse einer 2×2 -Matrix anwenden und feststellen, dass das Ergebnis eine Gaußverteilung in x ist mit fester Varianz und von z abhängigem Erwartungswert. Diese beiden Größen lesen wir dann an der Formel ab.

$$p(X = x|Z = z) = \frac{p(X = x \wedge Z = z)}{p(Z = z)} \quad (8.8)$$

$$\propto p(X = x \wedge Z = z) \quad (8.9)$$

$$= \exp\left(-\frac{1}{2} \begin{pmatrix} x - EX \\ z - EZ \end{pmatrix}^T \begin{pmatrix} VX & \text{Cov}(X, Z) \\ \text{Cov}(X, Z) & VZ \end{pmatrix}^{-1} \begin{pmatrix} x - EX \\ z - EZ \end{pmatrix}\right) \quad (8.10)$$

Jetzt wenden wir die bekannte Formel für die Inverse einer 2×2 -Matrix an,

$$= \exp \left(-\frac{1}{2} \frac{1}{VXVZ - \text{Cov}(X, Z)^2} \begin{pmatrix} x - EX \\ z - EZ \end{pmatrix}^T \begin{pmatrix} VZ & -\text{Cov}(X, Z) \\ -\text{Cov}(X, Z) & VX \end{pmatrix} \begin{pmatrix} x - EX \\ z - EZ \end{pmatrix} \right) \quad (8.11)$$

Wir lösen das Vektor-Matrix-Vektor-Produkt auf:

$$= \exp \left(-\frac{1}{2} \frac{1}{VXVZ - \text{Cov}(X, Z)^2} \left(VZ(x - EX)^2 - 2 \text{Cov}(X, Z)(x - EX)(z - EZ) + VX(z - EZ)^2 \right) \right) \quad (8.12)$$

Das ganze ist eine Verteilung in x . z ist konstant. Daher sortieren wir nach $(x - EX)^2$ und lassen konstante Summanden im Exponenten weg ($(z - EZ)$ ist konstant). Wir ziehen den Faktor vor x aus der Klammer, weil wir auf die bekannte Gaußverteilung hinaus wollen.

$$\propto \exp \left(-\frac{1}{2} \frac{VZ}{VXVZ - \text{Cov}(X, Z)^2} \left((x - EX)^2 - 2 \frac{\text{Cov}(X, Z)}{VZ} (x - EX)(z - EZ) \right) \right)$$

und ergänzen quadratisch

$$\propto \exp \left(-\frac{1}{2} \frac{VZ}{VXVZ - \text{Cov}(X, Z)^2} \left((x - EX) - \frac{\text{Cov}(X, Z)}{VZ} (z - EZ) \right)^2 \right) \quad (8.13)$$

$$= \exp \left(-\frac{1}{2} \frac{\overbrace{\left(x - \left(EX + \frac{\text{Cov}(X, Z)}{VZ} (z - EZ) \right) \right)}^{E(X|Z=z)}}{\underbrace{VX - \frac{\text{Cov}(X, Z)^2}{VZ}}_{V(X|Z=z)}} \right) \quad (8.14)$$

□

Dieses Lemma wird uns im Kalman-Filter die Formel für den Messschritt liefern, weil es erlaubt, eine Messung $Z = z$ "hinter den Konditionsstrich" zu bringen, d.h. die Verteilung von X nach Eintreffen einer Messung $Z = z$ zu aktualisieren.

A.6

8.2. Kalman Filter

Der Kalman-Filter ist ein Bayesfilter mit Gaußverteilungen und linearen Modellen.

8.2.1. Annahmen

1. X_0 ist $\mathcal{N}(\mu_0, \sigma_0^2)$ verteilt
2. $X_t = a_t X_{t-1} + b_t U_t + g_t + \varepsilon_t$; $a_t, b_t, g_t \in \mathbb{R}$, $\varepsilon_t \sim \mathcal{N}(0, \sigma_{\varepsilon t}^2)$
3. $Z_t = c_t X_t + h_t + \delta_t$; $c_t, h_t \in \mathbb{R}$, $\delta_t \sim \mathcal{N}(0, \sigma_{\delta t}^2)$
4. X_0 und alle ε_t, δ_t unabhängig

a_t, b_t, g_t, c_t, h_t sind bekannt und kommen in der Regel aus der Problemdefinition, sind meistens also schon zur Programmierzeit fest. Die Zufallsvariablen ε_t und δ_t sind die unbekanntenen Messfehler. Der Zustand X_t ist natürlich auch unbekannt und soll geschätzt werden. Die Messungen u_t und z_t sind bekannt, weil sie gemessen werden. Allerdings erst zur Laufzeit.

Vergleiche [50, Tabelle 3.1].

8.2.2. Beispiel Kohlenkarren in 1D

Vergleiche Abb. 3.1.

Ein Kohlenkarren bewegt sich auf einer Schiene, wir wollen wissen, wo er sich befindet. Ein Radar misst die Funkwellenlaufzeit zum Karren und zurück aber nicht sehr genau. Ein Radumdrehungssensor misst die Umdrehung des Rades als Winkel.

Diese Situation soll in einem KF modelliert werden.

- X_t ist die wahre Position des Karren (in m)
- Z_t ist die gemessene Signallaufzeit des Radars (in s)
- U_t ist der Winkel um den sich das Rad seit $t - 1$ gedreht hat
- Dynamikmodell: $X_t = X_{t-1} + rU_t + \varepsilon_t$, r Radradius, $\varepsilon_t \sim \mathcal{N}(0, \sigma_{\varepsilon t}^2)$ verteilt, $\sigma_{\varepsilon t}^2$ aus Datenblatt oder gemäß Kap. 8.3, $a_t = 1, b_t = r, g_t = 0$
- Messmodell: $Z_t = \frac{2}{c}X_t + \delta_t$, c Lichtgeschwindigkeit, $c_t = \frac{2}{c}, h_t = 0, \delta_t \sim \mathcal{N}(0, \sigma_{\delta t}^2)$, $\sigma_{\delta t}$ aus Datenblatt (z.B. $10ns$)
- Initialisierung (anfangs sicher): $\mu_0 = x_0, \sigma_0 = 0$ (oder klein)
- Initialisierung (anfangs unsicher): $\mu_0 = \frac{x_{\min} + x_{\max}}{2}, \sigma_0^2 = \frac{(x_{\max} - x_{\min})^2}{12}$ (Gauß-Näherung an Uniformverteilung $\mathcal{U}(x_{\min}, x_{\max})$)

8.2.3. Algorithmus

Der Kalman-Filter hat wie der Bayes-Filter zwei wichtige Schritte: Der Dynamikschritt lässt die Zeit weiterticken und verarbeitet u_t . Er heißt auch Prädiktionsschritt (prediction), weil er den neuen Zustand aufgrund des alten vorhersagt und führt typischerweise

zu mehr Unsicherheit. Der Messschritt verarbeitet die Messung z_t . Er heißt auch Korrekturschritt (update), weil er die Vorhersage auf Grund von der Messung korrigiert.

$$\text{Variablen:} \quad \mu_t, \sigma_t^2 \quad (8.15)$$

$$\text{Initialisierung:} \quad \mu_0 = E(X_0), \sigma_0^2 = V(X_0) \quad (8.16)$$

$$\text{Dynamik}(U_t = u_t): \quad \bar{\mu}_t = a_t \mu_{t-1} + b_t u_t + g_t \quad (8.17)$$

$$\bar{\sigma}_t^2 = a_t^2 \sigma_{t-1}^2 + \sigma_{\varepsilon t}^2 \quad (8.18)$$

$$\text{Messung}(Z_t = z_t): \quad k_t = \frac{\bar{\sigma}_t^2 c_t}{\bar{\sigma}_t^2 c_t^2 + \sigma_{\delta t}^2} \quad (8.19)$$

$$\mu_t = \bar{\mu}_t + k_t(z_t - (c_t \bar{\mu}_t + h_t)) \quad (8.20)$$

$$\sigma_t^2 = \bar{\sigma}_t^2 - k_t c_t \bar{\sigma}_t^2 \quad (8.21)$$

Der Kalman Filter repräsentiert die aktuelle Verteilung von X , gegeben alle schon gemachten Messungen, als eine Gaußverteilung mit Mittelwert μ_t und Varianz σ_t^2 . Diese beiden Größen sind Variablen im Algorithmus und werden bei jeder Messung verändert. Die Größen haben Indizes, um im Beweis auf die Werte zu verschiedenen Zeitpunkten Bezug nehmen zu können. Die Variablen im Programm müssen jeweils nur den aktuellen Wert speichern. Die Variablen mit Balken benennen jeweils Größen vor Integration einer Messung aber nach dem Dynamikschritt, die Größen ohne Balken die nach Integration der Messung. Um die Notation zu vereinfachen, wird der Algorithmus für den Fall abwechselnder Dynamik- und Messschritte beschrieben. Es können aber auch Dynamik- und Messschritte in beliebiger Reihenfolge erfolgen.

Q 25

Satz 11. *Nach Ausführung der Kalman Filter Formeln ist*

$$X_t | u_{1..t}, z_{1..t-1} \sim \mathcal{N}(\bar{\mu}_t, \bar{\sigma}_t^2), \quad \text{falls } t > 0 \quad (8.22)$$

$$X_t | u_{1..t}, z_{1..t} \sim \mathcal{N}(\mu_t, \sigma_t^2) \quad (8.23)$$

verteilt.

Beweis. Der Beweis rechnet im Wesentlichen mit den bekannten Gesetzen nach, dass die Dynamikformeln die Veränderung der Verteilung von X_{t-1} nach X_t im Dynamikschritt beschreiben. Die resultierende Verteilung ist konditioniert unter den Dynamikmessungen $u_{1..t}$ aber nur den Messungen $z_{1..t-1}$. Dann wird das Konditionierungslemma mit $Z = Z_t$ angewendet, um die aktuelle Messung hinter den Konditionierungsstrich zu bringen und die Verteilung entsprechend anzupassen.

Formal durch Induktion nach t .

I.A.: $t = 0$ folgt aus Annahme 1.

8. Kalman Filter

I.S.: $t - 1 \rightsquigarrow t$

$$X_{t-1} | u_{1...t-1}, z_{1...t-1} \sim \mathcal{N}(\mu_{t-1}, \sigma_{t-1}^2) \quad (8.24)$$

$$a_t X_{t-1} | u_{1...t-1}, z_{1...t-1} \sim \mathcal{N}(a_t \mu_{t-1}, a_t^2 \sigma_{t-1}^2) \quad (8.25)$$

$$b_t u_t + g_t | u_{1...t-1}, z_{1...t-1} \text{ ist konstant} \quad (8.26)$$

$$\varepsilon_t \sim \mathcal{N}(0, \sigma_{\varepsilon t}^2) \quad (8.27)$$

$$X_t | u_{1...t}, z_{1...t-1} \sim \mathcal{N}(a_t \mu_{t-1} + b_t u_t + g_t, a_t^2 \sigma_{t-1}^2 + \sigma_{\varepsilon t}^2) = \mathcal{N}(\bar{\mu}_t, \bar{\sigma}_t^2) \quad (8.28)$$

Wir wollen jetzt das Konditionierungslemma auf $X | u_{1...t}, z_{1...t-1}$ mit $Z = Z_t$ anwenden. Dazu benötigen wir $EX, EZ, VX, VZ, \text{Cov}(X, Z)$ unter der Vorbedingung $|u_{1...t}, z_{1...t-1}$ (in folgenden Gleichungen als *):

$$E(X|*) = \bar{\mu}_t, \quad (8.29)$$

$$E(Z|*) = c_t \bar{\mu}_t + h_t, \quad (8.30)$$

$$V(X|*) = \bar{\sigma}_t^2, \quad (8.31)$$

$$V(Z|*) = V(c_t X_t + h_t + \delta_t) = \bar{\sigma}_t^2 c_t^2 + \sigma_{\delta t}^2 \quad (8.32)$$

$$\text{Cov}(X, Z|*) = c_t \text{Cov}(X_t, X_t) = \bar{\sigma}_t^2 c_t \quad (8.33)$$

Wir setzen ein (alle E, V und Cov unter $|*$):

$$X_t | u_{1...t}, z_{1...t} = X_t | z_t, * \quad (8.34)$$

$$= \mathcal{N}\left(EX + \frac{\text{Cov}(X, Z)}{VZ}(z_t - EZ), VX - \frac{\text{Cov}(X, Z)^2}{VZ}\right) \quad (8.35)$$

$$= \mathcal{N}(\bar{\mu}_t + k_t(z_t - (c_t \bar{\mu}_t + h_t)), \bar{\sigma}_t^2 - k_t c_t \bar{\sigma}_t^2) = \mathcal{N}(\mu_t, \sigma_t^2) \quad (8.36)$$

$$k_t = \frac{\text{Cov}(X_t, Z_t)}{VZ_t} = \frac{\bar{\sigma}_t^2 c_t}{\bar{\sigma}_t^2 c_t^2 + \sigma_{\delta t}^2} \quad \square$$

A.7 + A.8

8.3. Dynamikrauschen proportional zu Zeit oder Strecke

Der Kalman-Filter modelliert ein bestimmtes Rauschen pro Messung, insbesondere pro Dynamik Messung U_t als $\sigma_{\varepsilon t}^2$. Das ist bei vielen Sensoren eine sinnvolle Art, Unsicherheit anzugeben.

Bei manchen Sensoren, besonders solchen die eine Änderungsrate messen (z.B. Gyrometer, Accelerometer) gibt man aber oft eine Rauschdichte σ_ε an. Eigenschaft dieser Sensoren ist, dass sich innerhalb einer Zeit von T im aufintegrierten Signal eine Varianz von $T\sigma_\varepsilon^2$ aufintegriert, egal in wie viele Zeitschritte die Zeitspanne T zerlegt wird.

Technischer Grund dafür ist, dass diese Sensoren ihr kontinuierliches Signal mitteln, so dass bei weniger Messungen die einzelne Messung weniger rauscht und alle zusammen immer genauso viel Information tragen. Man erkennt diese Angaben daran, dass die Einheit von σ_ε die Einheit der Messung Z durch \sqrt{Hz} bzw. die Einheit der aufintegrierten Messung mal \sqrt{Hz} ist.

In so einem Fall setzt man

$$\sigma_{\epsilon t}^2 = \sigma_{\epsilon}^2 \Delta T, \quad (8.37)$$

an, wobei ΔT die physikalische Zeitdauer eines Schrittes ist (wir hatten für die vereinfachte Schreibweise im Skript Zeitschritte angenommen, dass die Zeit t die Subskript zu den verschiedenen Größen ist in Schritten gemessen wird).

Analog gibt es Sensoren die nicht "pro Zeit" sondern pro "aufakkumulierter Änderung" rauschen. Z.B. ist es bei einem Raddrehensensor eine sinnvolle Modellierung, dass das Rauschen in einer Varianz proportional zur zurückgelegten Strecke ist, unabhängig davon in wie vielen Schritten diese Strecke zerlegt wurde.

In so einem Fall setzt man

$$\sigma_{\epsilon t}^2 = \sigma_{\epsilon}^2 |u_t|, \quad (8.38)$$

an, wobei die Abhängigkeit des Rauschens von u_t formal eine gewisse Näherung ist. Die Einheit von σ_{ϵ} wäre dann, z.B. bei einem Odometer \sqrt{m} , ein Wert von 1 hieße, dass bei $1m$ gefahrener Strecke Rauschen mit Standardabweichung $1m$ entsteht.

8.4. Konsistenz

Der Begriff der Konsistenz soll Auskunft darüber geben, ob ein Schätzer/Filter wie der Kalman Filter richtig funktioniert. Da der Kalman Filter das Ergebnis eines Zufallsprozesses schätzt, ist es nie möglich exakt die Wahrheit zu schätzen (man kennt sie meist auch gar nicht, um damit zu vergleichen). Das macht es schwierig zu bewerten, ob der Schätzer richtig angewendet wurde, richtig implementiert wurde oder überhaupt geeignet ist, den Zustand des Systems zu schätzen. In diesem Abschnitt geht es also darum, zu bewerten, wann der Filter korrekt ist.

8.4.1. Mahalanobis-Distanz

Schätzt man den Zustand eines Systems, wird man immer etwas falsch liegen. Das liegt in der Natur der Zufallsprozesse, die man versucht zu beobachten. Die wirkliche Wahrheit wird man nie herausfinden. Wichtig ist aber, dass die Schätzung sich nicht zu weit vom wahren Wert entfernt. Schauen wir uns ein Beispiel an (Abbildung 8.1). Der wahre Wert x_t wird von 2 Schätzern \hat{x}_{t1} und \hat{x}_{t2} geschätzt. Welcher Schätzer ist besser? Auf den ersten Blick ist \hat{x}_{t2} deutlich dichter an der Wahrheit und damit der bessere Schätzer.

Schätzer wie der Kalman Filter oder der Bayes Filter schätzen allerdings eine Wahrscheinlichkeitsverteilung, da ein einzelner Wert den Zufall nicht gut abbildet. Nehmen wir also an, dass die beiden Schätzer eine Gaußverteilung vom wahren Wert schätzen wie in Abbildung 8.2.

\hat{x}_{t1} hat eine hohe Unsicherheit/Varianz geschätzt, d.h. der Schätzer „weiß“, dass er gerade keine genaue Aussage über den wahren Wert treffen kann. \hat{x}_{t2} hat hingegen eine geringe Unsicherheit und „glaubt“ den wahren Wert gut zu kennen. Schau auf die Wahrscheinlichkeitsdichte der beiden Schätzer für den wahren Wert. Diese ist für \hat{x}_{t1} höher

8. Kalman Filter

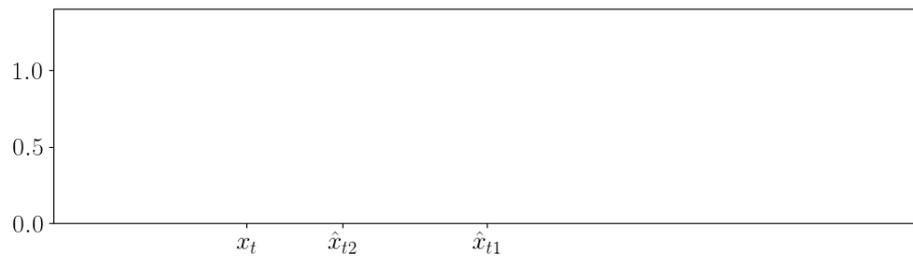


Abbildung 8.1.: Schätzfehler zweier Schätzer

obwohl der Erwartungswert der Verteilung weiter weg ist. Das heißt, in der Schätzung von \hat{x}_{t1} ist der wahre Wert plausibler zu erklären.

Insgesamt ist der Fehler von \hat{x}_{t1} zwar dennoch größer, aber dadurch dass \hat{x}_{t1} eine deutlich höhere Varianz hat passt der wahre Wert besser zur Verteilung von \hat{x}_{t1} .

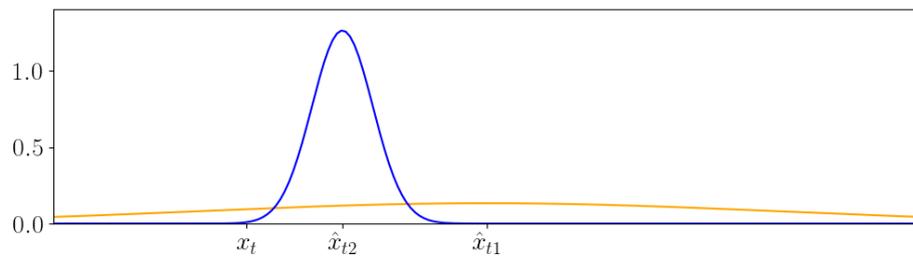


Abbildung 8.2.: Schätzfehler mit Gauß

Trägt man die Standardabweichungen der Gaußlocken auf (Abbildung 8.3) sieht man es sehr deutlich. Der wahre Wert ist weniger als eine Standardabweichung von Erwartungswert von \hat{x}_{t1} entfernt, aber mehr als 2 Standardabweichungen von \hat{x}_{t2} .

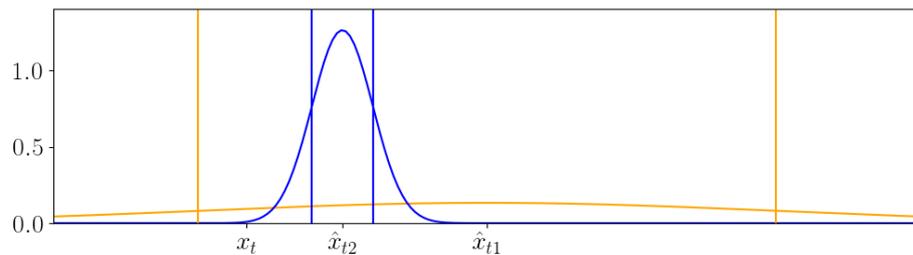


Abbildung 8.3.: Schätzfehler mit Gauß und 1σ -Intervall

Das Konzept, den Fehler eines Schätzers nicht nur durch den absoluten Abstand sondern auch mit der geschätzten Varianz zu bewerten lässt sich mit der Mahalanobis-Distanz formalisieren. Sie ist ein Maß dafür, wie plausibel ein Messwert/wahrer Wert zu einer Verteilung passt.

Definition 5. Mahalanobis-Distanz in 1D Sei $X : \Omega \mapsto \mathbb{R}$ eine Zufallsvariable mit $E(X) = \hat{x}$ und $V(X) = \sigma^2$, dann ist die Mahalanobis-Distanz von X zu einer Realisierung x definiert als:

$$\Delta(x, X) = \Delta(x, \hat{x}, \sigma^2) = \frac{|x - \hat{x}|}{\sigma}$$

Die Mahalanobis-Distanz ist einheitenlos, weil sie die Distanz mit der Standardabweichung vergleicht. Eine Mahalanobis-Distanz von 1 bedeutet als immer 1σ weit vom Mittelwert entfernt. Bei einer Gaußverteilung gibt es deshalb auch feste Prozentzahlen für die Wahrscheinlichkeit, dass die Mahalanobis-Distanz $\leq 1, 2, 3$ ist, nämlich ca. 68%, 95% und 99,7%. In der Praxis wird oft die quadrierte Mahalanobis-Distanz verwendet, weil diese im mehrdimensionalen Fall leichter zu handhaben ist:

$$\Delta^2(x, \hat{x}, \sigma^2) = \frac{(x - \hat{x})^2}{\sigma^2} \quad (8.39)$$

Satz 12. Der Erwartungswert der quadrierten 1D-Mahalanobis-Distanz ist 1.

$$E(\Delta^2(X, X)) = 1. \quad (8.40)$$

Beweis.

$$E(\Delta^2(X, X)) = E\left(\frac{(x - E X)^2}{V X}\right) = \frac{E((x - E X)^2)}{V X} = \frac{V X}{V X} = 1 \quad (8.41)$$

□

Noch genauer, aber ohne Beweis, kann man sagen, dass die quadrierte 1D-Mahalanobis-Distanz einer χ^2 -Verteilung mit 1 Freiheitsgrad folgt.

8.4.2. Konsistenzkriterien mit verfügbarer Groundtruth

Hat man Groundtruth x_t verfügbar, sei es durch einen weiteren Sensor oder sei es in Simulation, kann man die Konsistenz der vom Filter berechneten Zustandsverteilung untersuchen. Dazu kann man das z.B. das $\mu \pm 2\sigma$ Intervall und die Groundtruth plotten und visuell beurteilen, ob die Groundtruth ca. 95% der Zeit darin liegt (Varianz stimmt) und ungefähr mittig (Erwartungstreue). Formeller kann man den Normalized Estimation Error Squared (NEES) ausrechnen:

$$\frac{1}{T} \sum_{t=1}^T \Delta^2(x_t, \mu_t, \sigma_t^2) \quad (8.42)$$

Der Erwartungswert ist 1, deshalb sollte das Mittel über die Zeit nah an 1 sein. Ein formeller statistischer Test ist nicht möglich, weil die Zustände und damit die Δ^2 abhängig sind. Wirkliche Fehler im Modell zeigen sich aber oft in großen Abweichungen, wo der Mittelwert deutlich über 10 ist. Schlägt dieser Test fehl, kann man untersuchen,

8. Kalman Filter

ob der Filter einen Bias hat, indem man die mittlere Abweichung zwischen Groundtruth und Schätzung (inkl. Vorzeichen) berechnet.

$$\frac{1}{T} \sum_{t=1}^T x_t - \mu_t \quad (8.43)$$

8.4.3. Konsistenzkriterien ohne verfügbare Groundtruth

Ist keine Groundtruth verfügbar, kann man statt dessen die Konsistenz der vorhergesagten Verteilung der Messungen überprüfen, zu diesen hat man ja in Gestalt der Messung selbst „Groundtruth“. Man berechnet dazu die normalized innovation squared (NIS):

$$\frac{1}{T} \sum_{t=1}^T \Delta^2(z_t, E(Z_t|u_{1..t}, z_{1..t-1}), V(Z_t|u_{1..t}, z_{1..t-1})) \quad (8.44)$$

$$= \frac{1}{T} \sum_{t=1}^T \Delta^2(z_t, c_t \bar{\mu}_t + h_t, \bar{\sigma}_t^2 c_t^2 + \sigma_{\delta t}^2) \quad (8.45)$$

Die Formeln stammen dabei aus dem Beweis zum Messschritt des Kalman-Filters.

Dieser Term ist im Erwartungswert 1 und deshalb auch im statistischen Mittel über die Zeit ungefähr 1. Die Situation hier ist sogar besser, als beim Zustand. Die sogenannten Innovations $z_t - E(Z_t|...)$ sind nämlich unabhängig über die Zeit. Dies ist eine nicht offensichtliche Eigenschaft des Kalman-Filters. Sie rührt anschaulich daher, dass alle Information aus der Messung eines Zeitpunktes in die Zustandsverteilung integriert ist. Daher kürzt sie sich in der Innovation der nächsten Messung raus. Deshalb sind fundierte statistische Tests möglich, die wir aber hier nicht diskutieren wollen.

Da die Innovations unabhängig sind, kann man sich die Innovations über die Zeit plotten und visuell auf Unabhängigkeit untersuchen. Insbesondere überkonfidente Filter, also wenn die Varianz im Filter kleiner als in Wirklichkeit ist, zeigen sich in aufeinanderfolgenden Innovations die in die selbe Richtung gehen. Der Filter lässt sich nur schwer von etwas überzeugen. Dies kann man leicht erkennen.

8.4.4. Ausreißerdetektion

Die Mahalanobis-Distanz der Innovation $\Delta^2(z_t, c_t \bar{\mu}_t + h_t, \bar{\sigma}_t^2 c_t^2 + \sigma_{\delta t}^2)$ kann man auch für eine einzelne Messung im Betrieb ausrechnen. Wirkliche Ausreißer haben oft sehr hohe Werte, die für gaußverteilte Messfehler kaum vorkommen. Diese Messungen können dann ignoriert werden. Ein typischer Schwellwert ist dabei 9, was 3σ und 0.3% fälschlich verworfenen Messungen entspricht. Diese Maßnahme löst nicht alle Probleme, aber ist ein einfacher Schutz gegen Ausreißer.

9. (Extended/Unscented) Kalman Filter

9.1. Nichtlineare Rechenregeln für Erwartungswerte und Varianzen

Wir kennen schon Rechenregeln um für eine Zufallsvariable X und eine lineare Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = ax + b$ aus Erwartungswert EX und Varianz VX den Erwartungswert $E(f(X))$ und die Varianz $V(f(X))$ zu berechnen, nämlich

$$E(f(X)) = E(aX + b) = aEX + b \quad (9.1)$$

$$V(f(x)) = V(ax + b) = a^2VX \quad (9.2)$$

Für den Fall, dass f nichtlinear ist gibt es keine allgemeingültige exakte Lösung. Man kann aber die Funktion f im Erwartungswert linear nähern und daraus eine Näherung von $E(f(X))$ und $V(f(X))$ bestimmen:

$$f(x) \approx f(EX) + (x - EX)f'(EX) \quad (9.3)$$

$$\begin{aligned} E(f(X)) &\approx E(f(EX) + (X - EX)f'(EX)) \\ &= f(EX) + (EX - EX)f'(EX) = f(EX) \end{aligned} \quad (9.4)$$

$$\begin{aligned} V(f(X)) &\approx V(f(EX) + (X - EX)f'(EX)) \\ &= V(Xf'(EX)) = (f'(EX))^2 VX \end{aligned} \quad (9.5)$$

$$\begin{aligned} \text{Cov}(f(X), X) &\approx \text{Cov}(f(EX) + (X - EX)f'(EX), X) \\ &= f'(EX) \text{Cov}(X, X) = f'(EX)V(X) \end{aligned} \quad (9.6)$$

Dieses Vorgehen ist in Abbildungen ?? - ?? entnommen aus [50] gezeigt.

9.2. Extended Kalman Filter

9.2.1. Annahmen

1. X_0 ist ungefähr $\mathcal{N}(\mu_0, \sigma_0^2)$
2. $X_t = g_t(X_{t-1}, u_t) + \varepsilon_t$; $\varepsilon_t \sim \mathcal{N}(0, \sigma_{\varepsilon_t}^2)$, g differenzierbar
3. $Z_t = h_t(X_t) + \delta_t$; $\delta_t \sim \mathcal{N}(0, \sigma_{\delta_t}^2)$, h differenzierbar

Sorry, this version does not include copyrighted Material.

Abbildung 9.1.: Abbildung aus [50, Fig. 3.3].

9. (Extended/Unscented) Kalman Filter

Sorry, this version does not include copyrighted Material.

Abbildung 9.2.: Abbildung aus [50, Fig. 3.4].

Sorry, this version does not include copyrighted Material.

Abbildung 9.3.: Abbildung aus [50, Fig. 3.5].

4. X_0 und alle ε_t, δ_t sind unabhängig.

Die Annahmen für den Kalman Filter sind ein Spezialfall für den EKF bei dem die Funktionen linear sind $g(x, u) = ax + bu + g$, $h(x) = cx + h$.

9.2.2. Algorithmus

Der Extended Kalman Filter [50][Tabelle 3.3] wendet die Gleichungen des Kalmanfilters an, indem er in jedem Schritt die nichtlinearen Funktionen g und h durch Linearisierungen \bar{g} und \bar{h} , also Tangenten am aktuellen Erwartungswert ersetzt. Dabei ist für den Dynamikschritt μ_{t-1} der aktuelle Erwartungswert und für den Messschritt ist es $\bar{\mu}_t$ (vgl. Kap. 8.2.3).

$$\bar{g}_t(x, u) = g_t(\mu_{t-1}, u) + (x - \mu_{t-1})g'_t(\mu_{t-1}, u) \quad (9.7)$$

$$\bar{h}_t(x) = h_t(\bar{\mu}_t) + (x - \bar{\mu}_t)h'_t(\bar{\mu}_t), \quad (9.8)$$

$$\bar{g}_t(x, u) \approx g_t(x, u) \quad \bar{h}_t(x, u) \approx h_t(x, u) \quad (9.9)$$

Die Funktionen \bar{g} und \bar{h} sind linear und können daher in den KF Algorithmus in Sektion 8.2.3 eingesetzt werden. Das Ergebnis ist der Extended Kalman Filter, ein näherungsweise Algorithmus für nichtlineare Dynamik- und Messfunktionen.

Initialisierung: μ_0, σ_0^2 nach 1.

Dynamik: (u_t)

$$\bar{\mu}_t = g_t(\mu_{t-1}, u_t) \quad (9.10)$$

$$\bar{\sigma}_t^2 = a_t^2 \cdot \sigma_{t-1}^2 + \sigma_{\varepsilon t}^2, \quad a_t = g'_t(\mu_{t-1}, u_t) \quad (9.11)$$

Messung (z_t)

$$k_t = \frac{\bar{\sigma}_t^2 c_t}{\bar{\sigma}_t^2 c_t^2 + \sigma_{\delta t}^2}, \quad c_t = h'_t(\bar{\mu}_t) \quad (9.12)$$

$$\mu_t = \bar{\mu}_t + k_t(z_t - h_t(\bar{\mu}_t)) \quad (9.13)$$

$$\sigma_t^2 = \bar{\sigma}_t^2 - k_t c_t \bar{\sigma}_t^2 \quad (9.14)$$

Der KF ist ein exakter Algorithmus. Dementsprechend haben wir bewiesen, dass die Verteilung von $X|z_{1...t}, u_{1...t}$ *exakt* so ist, wie der Algorithmus sie berechnet.

Der EKF ist ein Naherungsalgorithmus, bei dem man kaum mathematisch harte Aussagen ber das Ergebnis beweisen kann. Insbesondere ist "" in (9.4), (9.5) und (9.6) kein mathematisch exakt definierter Operator (jedenfalls wie wir ihn benutzen). Er sagt vielmehr dass die linke Seite eine sinnvoll begrndete Naherung der rechten Seite ist. Diese Rechnungen sind so "richtig", wie die nichtlineare Funktion im Bereich der Unsicherheit hinreichend linear ist. Das obliegt dem Entwickler zu beurteilen.

In diesem Sinne ist das folgende kein Satz, sondern eine Herleitung als begrndeter Weg an eine sinnvolle Naherung zu kommen.

Herleitung 1. *Nach Ausfhrung der Extended Kalman Filter Formeln ist*

$$X_t|u_{1\dots t}, z_{1\dots t-1} \text{ungefhr} \sim \mathcal{N}(\bar{\mu}_t, \bar{\sigma}_t^2), \quad \text{falls } t > 0 \quad (9.15)$$

$$X_t|u_{1\dots t}, z_{1\dots t} \text{ungefhr} \sim \mathcal{N}(\mu_t, \sigma_t^2) \quad (9.16)$$

verteilt.

Induktion nach t .

I.A.: $t = 0$ folgt aus Annahme 1.

I.S.: $t - 1 \rightsquigarrow t$

$$X_{t-1}|u_{1\dots t-1}, z_{1\dots t-1} \sim \mathcal{N}(\mu_{t-1}, \sigma_{t-1}^2) \quad (9.17)$$

$$g_t(X_{t-1}, u_t)|u_{1\dots t-1}, z_{1\dots t-1} \sim \mathcal{N}(g_t(\mu_{t-1}, u_t), g_t'(\mu_{t-1}, u_t)^2 \sigma_{t-1}^2) \quad (9.18)$$

$$\varepsilon_t \sim \mathcal{N}(0, \sigma_{\varepsilon t}^2) \quad (9.19)$$

$$X_t|u_{1\dots t}, z_{1\dots t-1} \sim \mathcal{N}(g_t(\mu_{t-1}, u_t), g_t'(\mu_{t-1}, u_t)^2 \sigma_{t-1}^2 + \sigma_{\varepsilon t}^2) \quad (9.20)$$

$$= \mathcal{N}(\bar{\mu}_t, \bar{\sigma}_t^2) \quad (9.21)$$

Wir wollen jetzt das Konditionierungslemma auf $X|u_{1\dots t}, z_{1\dots t-1}$ mit $Z = Z_t$ anwenden. Dazu benotigen wir $EX, EZ, VX, VZ, \text{Cov}(X, Z)$ unter der Vorbedingung $|u_{1\dots t}, z_{1\dots t-1}$ (in folgenden Gleichungen als $*$):

$$E(X|*) = \bar{\mu}_t, \quad (9.22)$$

$$E(Z|*) = h(\bar{\mu}_t) \quad (9.23)$$

$$V(X|*) = \bar{\sigma}_t^2, \quad (9.24)$$

$$V(Z|*) = V(h_t(X_t) + \delta_t) = h'(\bar{\mu}_t)^2 \bar{\sigma}_t^2 + \sigma_{\delta t}^2 \quad (9.25)$$

$$\text{Cov}(X, Z|*) = \text{Cov}(X, h_t(X_t) + \delta_t) \quad (9.26)$$

$$= h'(\bar{\mu}_t) \bar{\sigma}_t^2 + 0 \quad (9.27)$$

Wir setzen ein (alle E, V und Cov unter $|*$):

$$X_t|u_{1\dots t}, z_{1\dots t} = X_t|u_{1\dots t}, z_{1\dots t-1}, z_t \quad (9.28)$$

$$= \mathcal{N}\left(EX + \frac{\text{Cov}(X, Z)}{VZ}(z_t - EZ), VX - \frac{\text{Cov}(X, Z)^2}{VZ}\right) \quad (9.29)$$

9. (Extended/Unscented) Kalman Filter

$$= \mathcal{N}(\bar{\mu}_t + k_t(z_t - h_t(\bar{\mu}_t)), \bar{\sigma}_t^2 - k_t h'(\bar{\mu}_t) \bar{\sigma}_t^2) = \mathcal{N}(\mu_t, \sigma_t^2) \quad (9.30)$$

$$k_t = \frac{\text{Cov}(X_t, Z_t)}{V Z_t} = \frac{\bar{\sigma}_t^2 h'(\bar{\mu}_t)}{\bar{\sigma}_t^2 h'(\bar{\mu}_t)^2 + \sigma_{\delta t}^2} \quad (9.31)$$

[50][Abb. 3.4]

9.3. Beispiel Wasserwanne

Gesucht ist das Wasservolumen in einer Badewanne. Diese wird über einen Zulauf kontinuierlich, aber unregelmäßig befüllt. Im Mittel fließen jede Minute $0.01 m^3$ zu, die Standardabweichung ist aber mit $0.005 m^3$ recht hoch. Ein Schwimmer misst den Wasserstand, aber fehlerbehaftet, u.a. wegen Wellen in der Wanne (Höhe $0.05 m$). Außerdem misst ein Durchflussmesser im Ablauf, wie viel Wasser die Wanne verlässt. Für den Durchflussmesser haben wir in einem Test bei Abfluss von $1 m^3$ einen Fehler von $0.01 m^3$ gemessen. Die Wanne hat eine komplizierte Form, der Füllstand beim Volumen v ist als eine Funktion $f(v)$ gegeben.

Dieser Vorgang soll als EKF modelliert werden.

- X_t Volumen in der Wanne [m^3]
- Z_t Füllstandshöhe [m]
- $h_t(x) = f(x)$ Füllstand aus Volumen berechnet
- $\sigma_\delta = 0.05 m$ Messrauschen und Wellenhöhe
- U_t Wasserablauf seit $t - 1$, [m^3]
- $g(x, u) = x + u + \frac{0.01 m^3}{60 s} \delta T$ Änderung des Volumens durch Abfluss und mittleren Zufluss
- $\sigma_{\varepsilon}^2 = \sigma_{\varepsilon+}^2 + \sigma_{\varepsilon-}^2$, Die Unsicherheit in der Dynamik ergibt sich aus einer Unsicherheit beim Zufluss und einer Unsicherheit beim Abfluss
- $\sigma_{\varepsilon+}^2 = \frac{(0.005 m^3)^2}{60 s} \delta T$ Zuflussvarianz proportional zur Zeit, so dass sich in $60 s$ Unsicherheit $0.005 m^3$ ergibt.
- $\sigma_{\varepsilon-}^2 = \frac{(0.01 m^3)^2}{1 m^3} |u|$ Abflussvarianz proportional zur Menge Wasser die abfloss, so dass sich bei Abfluss von $1 m^3$ eine Unsicherheit von $0.01 m^3$ ergibt.

Was wäre, wenn wir ohne Schwimmer messen? Die Abflussmenge hängt vom Druck, d.h. von der Füllstandshöhe ab. Man kann grob die Durchflussmessung sowohl für U_t verwenden, weil sie die Änderung des Volumens misst, als auch für Z_t , weil sie wegen des funktionalen Zusammenhangs indirekt die Füllhöhe misst. Diese Doppelverwendung widerspricht formal der Unabhängigkeitsforderung, ist aber in diesem Sonderfall okay.

- Z_t Wasserablauf seit $t - 1$ [m]
- $h(x) = \gamma f(x)$ unter der Annahme, dass der Abfluss proportional zum Druck ist. Der Faktor müsste vorher bestimmt werden.

A.9-A.10

9.4. σ -Punkt-Propagation/Unscented Transform

9.4.1. Idee

Die Näherungsformeln aus Kap. 9.1 propagieren $E X$ und $V X$ durch eine nichtlineare Funktion f mit Hilfe von Funktionswert und Ableitung an der Stelle $E X$. Dies ist der Versuch eine Gaußglocke, durch die nichtlineare Funktion zu propagieren. Diese Approximation ist je nach Nicht-Linearität der Funktion f unzureichend.

Eine anderer Ansatz ist die unscented transform/ σ -Punkt-Propagation nach [46]. Diese basiert auf 2 Prinzipien:

1. Es ist einfach einen einzelnen Punkt durch eine nicht-lineare Funktion zu propagieren (einfacher als eine ganze PDF)
2. Es ist möglich, einen Satz von Punkten in der PDF zu finden, die die PDF gut repräsentieren (Berechnung von $E(X)$ und $V(X)$)

Diese beiden Prinzipien kombiniert man folgendermaßen: Wir wählen eine Menge \mathcal{X} von σ -Punkten aus der PDF von X , sodass $E \mathcal{X} = E X$ und $V \mathcal{X} = V X$. Die σ -Punkte werden einzeln durch die nichtlineare Funktion f propagiert. Der Erwartungswert und die Varianz der so entstehenden Verteilung sind eine gute Näherung an die Propagation der gesamten Verteilung X . Bei richtiger Wahl, der σ -Punkte ist diese Approximation korrekt bis zur 3. Ordnung in σ statt nur bis zur 1. Ordnung wie der EKF. Dies wollen wir im folgenden zeigen.

Als σ - Punkte von $\mathcal{N}(\mu, \sigma)$ wählen wir (Zu sehen in 9.4):

$$\mathcal{X} = \{\mu + \sqrt{2}\sigma, \mu, \mu, \mu - \sqrt{2}\sigma\} \quad (9.32)$$

Dass die Erwartungswert μ zweimal vorkommt ist, dient dabei Kompatibilität mit der Literatur¹.

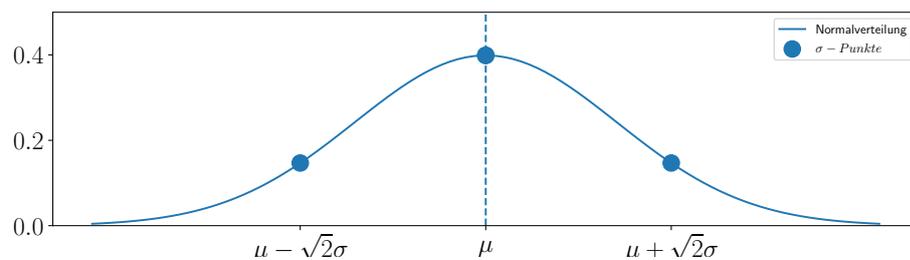


Abbildung 9.4.: Sigma-Punkte der Normalverteilung.

In der Statistik wird oft ein Intervall $[\mu \pm \sigma]$ als „Stellvertreter“ für die Verteilung von X betrachtet. Z.B. wird es oft in Diagramme eingezeichnet und für Gaußverteilungen ist auch bekannt, dass X mit einer Wahrscheinlichkeit von $\approx 68.6\%$ in diesem Intervall liegt. Es soll aber zusätzlich ein dritter Signmapunkt μ verwendet werden. Erst dadurch können die Signmapunkte erfassen, wie nichtlinear eine Funktion ist.

Sorry, this version does not include copyrighted Material.

Abbildung 9.5.: Abbildung aus [50, Fig. 3.7].

Sorry, this version does not include copyrighted Material.

Abbildung 9.6.: Abbildung aus [50, Fig. 3.8].

Abbildungen 9.5 und 9.6 entnommen aus [50] illustrieren das Vorgehen und vergleichen es mit der Propagation über Ableitungen.

Wir wollen nun zeigen, dass diese Approximation für den Erwartungswert 3. Ordnung korrekt ist.

9.4.2. Beweisüberblick

Im Überblick funktioniert der Beweis wie folgt: Wir zeigen zuerst, dass für ein Polynom 3. Ordnung die Sigmapunktpropagation den exakten Erwartungswert $E(f(X))$ liefert. Im Allgemeinfall nutzen wir eine Taylorapproximation 3. Ordnung, dessen Glieder 0-3 exakt sind und dessen Restglied 4. Ordnung einen Fehler produziert, der aber $o(\sigma^3)$ ist. D.h. wenn $\sigma \rightarrow 0$ geht, fällt der Fehler schneller als σ^3 . Diese Aussage gilt dann auch für die Varianz, die ja als Erwartungswert $E((f(X) - E(f(X)))^2)$ definiert ist. Die Varianz aus der Sigmapunktpropagation ist aber selbst für Polynome der Ordnung 2 nicht korrekt, weil dadurch der Term im Erwartungswert ein Polynom der Ordnung 4 wird.

9.4.3. Landau-Notation

Die Rechnungen lassen sich mit der Landau-Notation, dem „O-Kalkül“ leichter aufschreiben [56]. Dieses ist aus der Laufzeitanalyse von Algorithmen bekannt. Für die Landau-Notation muss man einen Grenzwertprozess festlegen (bei Algorithmen meist $n \rightarrow \infty$). In unserem Fall ist das $\sigma \rightarrow 0$, d.h. wir betrachten ein immer kleiner werdendes σ . Für kleiner werdende σ wird der Fehler kleiner und die Landau-Notation notiert, wie schnell.

Dabei steht hier $o(\sigma^3)$ für eine Funktion h , die schneller fällt, als σ^3 , d.h. für die

$$\lim_{\sigma \rightarrow 0} \frac{h(\sigma)}{\sigma^3} = 0 \quad (9.33)$$

ist.

9.4.4. Sigmapunktpropagation für Polynome vom Grad 3

Satz 13. Sei f Polynom dritter Ordnung und $X \sim N(\mu, \sigma)$ verteilt. Seien $\mathcal{X} = (\mu + \sqrt{2}\sigma^2, \mu, \mu - \sqrt{2}\sigma^2)$ die dazugehörigen Sigmapunkte. Dann ist $E(f(\mathcal{X})) = E(f(X))$, die Sigmapunktpropagation also korrekt.

¹vgl. Erläuterung für den n -D Fall

9. (Extended/Unscented) Kalman Filter

Beweis. Sowohl $E(f(\mathcal{X}))$, als auch $E(f(X))$ sind linear in f . Summen oder skalare Vielfache in f können also herausgezogen werden. Deshalb reicht es, wenn wir die Gleichheit auf einer Basis der Polynome vom Grad 3 prüfen. Zweckmäßigerweise nehmen wir dafür die $(x - \mu)^i, i = 0 \dots 3$.

- $(x - \mu)^0$: $E(f(X)) = E(1) = 1$ und die Sigmapunkte sind auch alle 1 und damit auch ihr Mittelwert $E(f(\mathcal{X})) = E(1) = 1$.
- $(x - \mu)^1$: ist anti-symmetrisch um μ und die Gaußverteilung symmetrisch um μ . Deshalb ist $E(f(X)) = 0$, weil sich jeweils die anti-symmetrischen Punkte gegeneinander herauskürzen. Das selbe, also $E(f(\mathcal{X})) = 0$ gilt für die Sigmapunktpropagation, weil die Sigmapunkte symmetrisch um μ liegen.
- $(x - \mu)^2$: $E((X - \mu)^2) = VX = \sigma^2$. Den Erwartungswert der Sigmapunkte berechnen wir als:

$$E(f(\mathcal{X})) = \frac{1}{4} \sum_{j=1}^4 f(\mathcal{X}^{[j]}) = \frac{1}{4} \sum_{j=1}^4 (\mathcal{X}^{[j]} - \mu)^2 \quad (9.34)$$

$$= \frac{1}{4} \left((\sqrt{2\sigma^2})^2 + 0^2 + 0^2 + (-\sqrt{2\sigma^2})^2 \right) \quad (9.35)$$

$$= \frac{1}{4} (2\sigma^2 + 2\sigma^2) = \sigma^2 \quad (9.36)$$

In diesem Term liegt der Unterschied zur Berechnung des Erwartungswertes durch Linearisierung, welcher nur den $(x - \mu)^0$ Term richtig behandelt.

- $(x - \mu)^3$: Ist antisymmetrisch und daher $E(f(\mathcal{X})) = E(f(X)) = 0$.

□

9.4.5. Sigmapunktpropagation im Allgemeinform

Wir wollen nun für f eine allgemeine dreimal stetig differenzierbare Funktionen betrachten und zeigen, dass der Fehler der Sigmapunktpropagation dann $o(\sigma^3)$, also dritter Ordnung ist.

Wir machen eine Taylorapproximation 3. Ordnung von f sowohl für $E(f(X))$, also auch $E(f(\mathcal{X}))$. Auf dem entstehenden Polynom dritter Ordnung ist die Sigmapunktpropagation exakt und beide stimmen überein, wir müssen nur noch für beide Restglieder zeigen, dass sie $o(\sigma^3)$ sind.

Lemma 9.1. Sei $r = o(|x - \mu|^3)$ für $x \rightarrow \mu$ und r „hinreichend harmlos“. Sei außerdem $X \sim \mathcal{N}(\mu, \sigma^2)$ mit $\sigma^2 \rightarrow 0$. Dann ist $E(r(X)) = o(\sigma^3)$.

Beweis. Die Idee ist, $x = \mu + \lambda\sigma$ mit $\lambda \sim \mathcal{N}(0, 1)$ zu setzen, während $\sigma \rightarrow 0$ geht. Für ein festes λ ist das ein Punkt der an qualitativ der selben Stelle der Gaußverteilung bleibt und mit kleiner werdendem σ näher zu μ rückt. Dieser Ansatz ist leichter, als ein festes

x zu betrachten über dem die Gaußverteilung immer kleiner wird.

$$\lim_{\sigma \rightarrow 0} \frac{\mathbb{E}(r(X))}{\sigma^3} = \lim_{\sigma \rightarrow 0} \int \frac{r(\mu + \lambda\sigma)}{\sigma^3} \mathcal{N}(0, 1)(\lambda) d\lambda \quad (9.37)$$

$$= \int \lim_{\sigma \rightarrow 0} \frac{r(\mu + \lambda\sigma)}{\sigma^3} \mathcal{N}(0, 1)(\lambda) d\lambda \quad (9.38)$$

Diese Vertauschung von Limes und Integral setzt voraus, dass der Integrand „hinreichend harmlos“ ist. Ich möchte hier nicht die genauen Bedingungen diskutieren, da wir in der ganzen Veranstaltung davon ausgehen, dass auftretenden Funktionen integrierbar sind, was eine ähnliche Annahme ist.

Der Bruch konvergiert gegen 0, weil $r(x) = o(|x - \mu|^3)$ und $|x - \mu|^3 = \sigma^3$ ist.

$$= \int 0 \mathcal{N}(0, 1)(\lambda) d\lambda = 0 \quad (9.39)$$

.

□

Wir machen jetzt eine ähnliche Analyse für das Verhalten des Restglieds in der Sigmapunktpropagation.

Lemma 9.2. Sei $r = o(|x - \mu|^3)$ für $x \rightarrow \mu$. Sei außerdem $X \sim \mathcal{N}(\mu, \sigma^2)$ mit $\sigma^2 \rightarrow 0$ und $\mathcal{X} = (\mu + \sqrt{2\sigma^2}, \mu, \mu, \mu - \sqrt{2\sigma^2})$. Dann ist $\mathbb{E}(r(\mathcal{X})) = o(\sigma^3)$.

Beweis.

$$\mathbb{E}(r(\mathcal{X})) = \frac{1}{4} \sum_{j=1}^4 r(\mathcal{X}^j) \quad (9.40)$$

$$= \frac{1}{4} \left(r(\mu + \sqrt{2\sigma^2}) + r(\mu) + r(\mu) + r(\mu - \sqrt{2\sigma^2}) \right) \quad (9.41)$$

Wenn $\sigma \rightarrow 0$ konvergiert, konvergieren die Argumente von r gegen μ und zwar mit einem $|x - \mu|$ Wert von $\sqrt{2}\sigma$ bzw. 0. Damit kann man die $o(|x - \mu|^3)$ Zusicherung für r einsetzen.

$$= \frac{1}{4} \left(o((\sqrt{2}\sigma)^3) + 0 + 0 + o((\sqrt{2}\sigma)^3) \right) = o(\sigma^3) \quad (9.42)$$

□

Satz 14. Gegeben sei eine dreimal stetig differenzierbare Funktion $f : \mathbb{R} \mapsto \mathbb{R}$ und eine Gaußverteilung $X = \mathcal{N}(\mu, \sigma)$. Sei $\mathcal{X} = (\mu + \sqrt{2\sigma^2}, \mu, \mu, \mu - \sqrt{2\sigma^2})$, dann ist $\mathbb{E}(f(\mathcal{X}))$ eine 3. Ordnung korrekte Approximation von $\mathbb{E}(f(X))$, d.h. $\mathbb{E}(f(\mathcal{X})) - \mathbb{E}(f(X)) = o(\sigma^3)$ für $\sigma \rightarrow 0$.

Beweis. Sei p das Taylorpolynom 3. Ordnung entwickelt um μ und r das dazugehörige Restglied. Damit ist

$$f(x) = p(x) + r(x), \quad r(x) = o(|x - \mu|^3) \quad (9.43)$$

$$\mathbb{E}(f(\mathcal{X})) - \mathbb{E}(f(X)) = \mathbb{E}(p(\mathcal{X})) - \mathbb{E}(p(X)) + \mathbb{E}(r(\mathcal{X})) - \mathbb{E}(r(X)) \quad (9.44)$$

9. (Extended/Unscented) Kalman Filter

Da die Sigmapunktpropagation auf Polynomen dritter Ordnung exakt ist (Satz 13), kürzen sich die ersten beiden Terme raus. Die zweiten beiden Terme sind $o(\sigma^3)$ nach Lemma 9.2 und 9.1.

$$= 0 + 0 + o(\sigma^3) + o(\sigma^3) = o(\sigma^3) \quad (9.45)$$

□

Die Varianz lässt sich ebenfalls aus der Verteilung der Sigmapunkte zurückgewinnen.

Satz 15. Gegeben sei eine dreimal stetig differenzierbare Funktion $f : \mathbb{R} \mapsto \mathbb{R}$ und eine Gaußverteilung $X = \mathcal{N}(\mu, \sigma^2)$. Sei $\mathcal{X} = (\mu + \sqrt{2\sigma^2}, \mu, \mu - \sqrt{2\sigma^2})$, dann ist $V(f(\mathcal{X}))$ eine 3. Ordnung korrekte Approximation von $V(f(X))$, d.h. $V(f(\mathcal{X})) - V(f(X)) = o(\sigma^3)$ für $\sigma \rightarrow 0$.

Beweis. Zum Beweis ziehen wir Satz 14 und die Definition der Varianz als Erwartungswert $V(Y) = E((Y - EY)^2)$ heran. Zwei Fehler entstehen, einer im äußeren Erwartungswert, einer im inneren, beide zeigen wir als $o(\sigma^3)$, so dass der Gesamtfehler auch $o(\sigma^3)$ ist.

$$V(f(\mathcal{X})) = E((f(\mathcal{X}) - E(f(\mathcal{X})))^2) \quad (9.46)$$

Betrachten wir das Argument im Erwartungswert als Funktion $h(x) = (f(x) - E(f(\mathcal{X})))^2$ und wenden Satz 14 an.

$$= E((f(X) - E(f(\mathcal{X})))^2) + o(\sigma^3) \quad (9.47)$$

Aus dem ersten Übungszettel ist bekannt, dass $E((X - \alpha)^2) = VX + (\alpha - EX)^2$, daher können wir das $E(f(\mathcal{X}))$ durch $E(f(X))$ mit einem zusätzlichen Fehlerterm ersetzen.

$$= V(f(X)) + (E(f(\mathcal{X})) - EX)^2 + o(\sigma^3) \quad (9.48)$$

$$= V(f(X)) + O(\sigma^3)^2 + o(\sigma^3) = V(f(X)) + o(\sigma^3) \quad (9.49)$$

□

Es ist bemerkenswert, dass der Fehler zwar auch bei der Varianz $o(\sigma^3)$ ist, die Varianz aber schon für Polynome der Ordnung 2 inexakt wird. Der Grund dafür ist, dass ein quadratisches Polynom als f zu einem Polynom 4. Ordnung im Erwartungswert der Varianzformel führt. Ein konkretes Gegenbeispiel ist X^2 für $X \sim \mathcal{N}(0, 1)$. Dieses hat laut Sigmapunktpropagation $E(X^2) = 1$ (richtig) und $V(X^2) = 1$. Richtig wäre aber $V(X^2) = E(X^4) - E(X^2)^2 = 3 - 1 = 2$ nach [57, Momente].

Es ist denkbar die σ -Punkte anders zu wählen. Je nach Wahl der Sigmapunkte ist die Approximation X-Order Korrekt. Die präsentierte Sigmapunktpropagation ist der Spezialfall für eine Dimension und lässt sich später auf N-Dimensionen generalisieren. In [46] wurde als Sigmapunkte für 1D nur $(\mu + \sqrt{\sigma^2}, \mu - \sqrt{\sigma^2})$ verwendet. Das hat den

Nachteil, dass die Sigmapunkte die Nichtlinearität einer Funktion gar nicht merken. Dieser Nachteil tritt aber nur in 1D auf und war daher wohl nicht im Fokus des Artikels. Der Vorschlag hier im Skript entspringt der Idee, eine zweite „Dummydimension“ hinzuzufügen, die die Funktionen ignorieren. Zwei Sigmapunkte variieren dann in der zweiten Dimension und sind in der ersten μ . Diese Konstruktion ist auch der Grund für den doppelten Sigmapunkt μ .

9.5. Unscented Kalman Filter

Der Unscented Kalman Filter funktioniert analog zum Extended Kalman Filter, nutzt aber die Sigmapunktpropagation anstelle von Ableitungen um Erwartungswert und Varianz von X durch das Dynamik- oder Messmodell zu propagieren.

9.5.1. Annahmen

1. X_0 ist ungefähr $\mathcal{N}(\mu_0, \sigma_0^2)$
2. $X_t = g_t(X_{t-1}, u_t) + \varepsilon_t$; $\varepsilon_t \sim \mathcal{N}(0, \sigma_{\varepsilon t}^2)$, g stetig
3. $Z_t = h_t(X_t) + \delta_t$; $\delta_t \sim \mathcal{N}(0, \sigma_{\delta t}^2)$, h stetig
4. X_0 und alle ε_t, δ_t sind unabhängig.

9.5.2. Algorithmus

Initialisierung: μ_0, σ_0^2 nach 1.

Dynamik: (u_t)

$$\mathcal{X}_{t-1} = \begin{pmatrix} \mu_{t-1} & \mu_{t-1} & \mu_{t-1} + \sqrt{2\sigma_{t-1}^2} & \mu_{t-1} - \sqrt{2\sigma_{t-1}^2} \end{pmatrix} \quad (9.50)$$

$$\bar{\mathcal{X}}_t^* = g(\mathcal{X}_{t-1}, u_t) \quad (9.51)$$

$$\bar{\mu}_t = \frac{1}{4} \sum_{i=1}^2 \bar{\mathcal{X}}_t^{*[i]} \quad (9.52)$$

$$\bar{\sigma}_t^2 = \frac{1}{4} \sum_{i=1}^2 (\bar{\mathcal{X}}_t^{*[i]} - \bar{\mu}_t)^2 + \sigma_{\varepsilon t} \quad (9.53)$$

9. (Extended/Unscented) Kalman Filter

Messung (z_t)

$$\bar{\mathcal{X}}_t = \left(\bar{\mu}_{t-1} \quad \bar{\mu}_{t-1} \quad \bar{\mu}_t + \sqrt{2\bar{\sigma}_t^2} \quad \bar{\mu}_t - \sqrt{2\bar{\sigma}_t^2} \right) \quad (9.54)$$

$$\bar{\mathcal{Z}}_t = h(\bar{\mathcal{X}}_t) \quad (9.55)$$

$$\hat{z}_t = \frac{1}{4} \sum_{i=1}^4 \mathcal{Z}_t^{[i]} \quad (9.56)$$

$$\sigma_t^{2z} = \frac{1}{4} \sum_{i=1}^4 (\bar{\mathcal{Z}}_t^{[i]} - \hat{z}_t)^2 + \sigma_{\delta t}^2 \quad (9.57)$$

$$\sigma_t^{2zx} = \frac{1}{4} \sum_{i=1}^4 (\bar{\mathcal{Z}}_t^{[i]} - \hat{z}_t)(\bar{\mathcal{X}}_t^{[i]} - \mu_t) \quad (9.58)$$

$$k_t = \frac{\bar{\sigma}_t^{2zx}}{\bar{\sigma}_t^{2z}} \quad (9.59)$$

$$\mu_t = \bar{\mu}_t + k_t(z_t - h_t(\bar{\mu}_t)) \quad (9.60)$$

$$\sigma_t^2 = \bar{\sigma}_t^2 - k_t \bar{\sigma}_t^{2zx} \quad (9.61)$$

Wie der EKF ist auch der UKF ein approximativer Algorithmus über den man keine harten mathematischen Aussagen beweisen kann. Auf eine Herleitung als sinnvolle Approximation verzichten wir hier, da aus den obigen Formeln direkt die Sigmapunkt-propagation erkennbar ist.

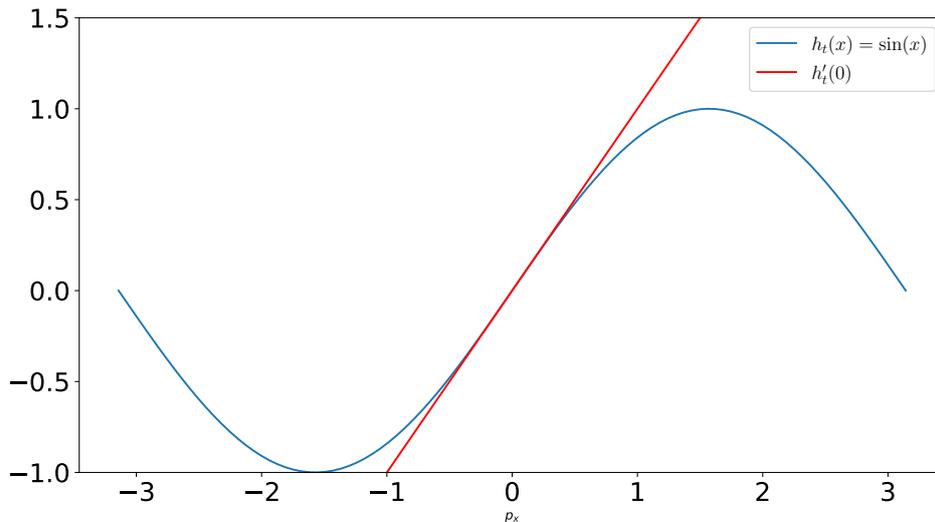
9.6. Fehlerquellen suboptimaler Filter

Während der Kalman Filter optimal die Wahrscheinlichkeitsverteilung schätzt, sind EKF und UKF suboptimal. Sie schätzen zwar eine Wahrscheinlichkeitsverteilung, aber diese kann beliebig falsch sein, d.h. die Filter können inkonsistent werden. Wir schauen uns in diesem Kapitel die vermeintlich größte Fehlerquelle des E- und UKF an: Approximierungsfehler.

9.6.1. Approximierungsfehler

Filter für nichtlineare Systeme wie der EKF und UKF approximieren das System, um diese allgemein behandeln zu können. Die Wahrscheinlichkeitsverteilung des approximierten Systems sieht allerdings anders aus, als die des Originalsystems. Daher sorgt dieser Schritt schon dafür, dass die Filter suboptimal sind. Wie verfälscht das Ergebnis durch die Approximation ist, hängt davon ab, inwieweit das System der Approximation entspricht. Im EKF wird beispielsweise eine lineare Approximation benutzt. Daher ist der EKF optimal, wenn das System linear ist (Er wird wieder zum normalen KF). Bei Systemen die sich lokal linear verhalten, ist der EKF ebenfalls dicht am optimalen Ergebnis. Das schauen wir uns einmal an einem Beispiel an.

Wir versuchen mit einem EKF den Zustand X zu schätzen. Als Messung haben wir $Z_t = \sin(x)$. Die Messfunktion h_t ist in Abbildung 9.7 gezeigt. Gehen wir davon aus,

Abbildung 9.7.: Eine Messung $h(t)$ mit Linearisierung am wahren Wert)

das genau jetzt $x = 0$. Die rote Linie zeigt das approximierte Modell, das der EKF für die Schätzung benutzt. Die rote Linie approximiert die Sinus-Funktion in einem großen Bereich sehr gut. Daher ist davon auszugehen, dass der Linearisierungsfehler (ein linearer Approximierungsfehler) klein ist.

Allerdings sollten wir noch berücksichtigen, dass der EKF einen Schätzfehler haben kann. Das heißt, dass obwohl der wahre Zustand bei $x = 0$ liegt, der EKF leicht daneben linearisiert. Dementsprechend müssen wir davon ausgehen, dass die Taylor-Approximierung am falschen Arbeitspunkt durchgeführt wird. Schauen wir uns in Abbildung 9.8 an, wie sich das auswirkt.

Durch die fehlerhafte Schätzung des EKF wird die Sinusfunktion nicht am wahren Wert linearisiert, wodurch die Linearisierung mehr von der Funktion abweicht. Allerdings ist der Fehler dadurch ebenfalls gering, da die Sinusfunktion um 0 herum beinahe linear ist. Wir sehen, dass die verschiedenen Linearisierungen nur geringfügig auseinander fächern. Dementsprechend kann man weiterhin einen geringen Fehler des EKF durch Linearisierung erwarten.

Anders sieht die ganze Situation aus, wenn der wahre Wert bei $x = \frac{\pi}{2}$ liegt. Schauen wir wieder auf die verschiedenen möglichen Approximierungen in Abbildung 9.9:

Die verschiedenen Approximierungen fächern stark auseinander. Sie bilden die Messfunktion nur über einen sehr kleinen Bereich gut ab. Daher müssen wir davon ausgehen, dass der EKF um $x = \frac{\pi}{2}$ herum starke Linearisierungsfehler aufweist.

Ein Linearisierungsfehler heißt im Endeffekt, dass der Wert $c_t = h_t^*(\bar{\mu}_t)$ im Messschritt falsch berechnet wird. Dadurch wird der Kalman Gain k_t verfälscht und die Innovation wird falsch gewichtet. Der EKF über- oder unterkorrigiert seine bisherige Schätzung. Dies führt zu Inkonsistenz, da die Varianz im Filter ebenfalls zu stark oder zu schwach

9. (Extended/Unscented) Kalman Filter

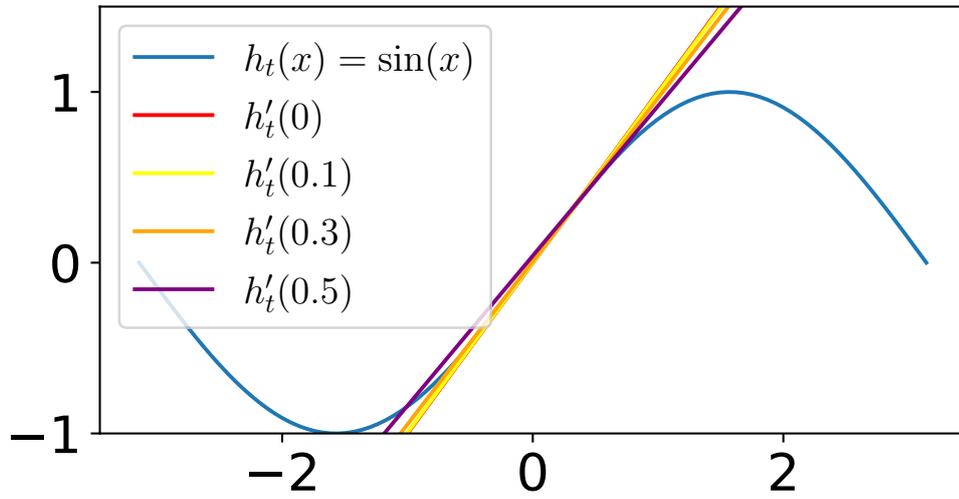


Abbildung 9.8.: Eine Messung $h(t)$ mit möglichen Linearisierungen)

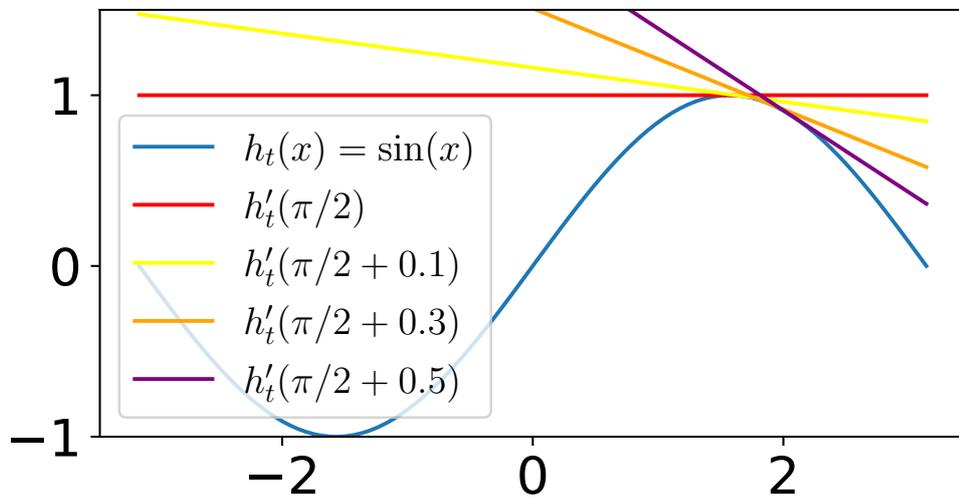


Abbildung 9.9.: Eine Messung $h(t)$ mit möglichen Linearisierungen)

angepasst wird.

Auch wenn der UKF eine Approximation 3. Grades durchführt, tritt dasselbe Problem auf. Die Approximierung passt nicht überall gut zum Messmodell, wenn dieses eine höhere Ordnung als 3 hat. Der UKF hat ebenfalls kleine Fehler in der Schätzung und approximiert daher mit Bezug auf einen falschen Arbeitspunkt. Auch hier wird der Kalman Gain falsch berechnet.

Trotz dieser Fehler können EKF und UKF zu annähernd konsistenten Schätzungen kommen. Dies ist sogar der Normalfall. Dies hat 2 (und wahrscheinlich mehr) Gründe: Zum einen, sind die Messfunktionen in weiten Teilen annähernd linear, wie es bei der Sinuskurve um $x = 0$ der Fall war. Zum anderen, hängt der Einfluss der Approximierungsfehler an der Innovation. Der Kalman Gain wird bei der Korrektur im Messschritt mit der Innovation $z - h(x)$ multipliziert. Wenn die Innovation gering ist, geht daher auch der Approximierungsfehler nur geringfügig in die Schätzung ein. Das heißt, solange der Filter die Messung gut vorhersagt, sind die Approximierungsfehler gering. „gut“ ist hier wieder nicht näher bestimmbar. Man kann keine klare Grenze ziehen, ab welcher Innovation die Approximierungsfehler Überhand nehmen.

Solange die Schätzung also dicht am wahren Wert liegt, sind die Approximierungen tauglich. Aber: Wenn der Filter eine schlechte Schätzung hat, funktioniert die Approximierung nicht mehr richtig. Der Filter wird inkonsistent.

9.6.2. Probleme durch Inkonsistenz

Ist ein E/UKF inkonsistent, über- oder unterschätzt er den Fehler des eigenen Schätzwertes. Dadurch wird die Schätzung falsch korrigiert. Wir führen uns das anhand der Abbildung 9.10 vor Augen:

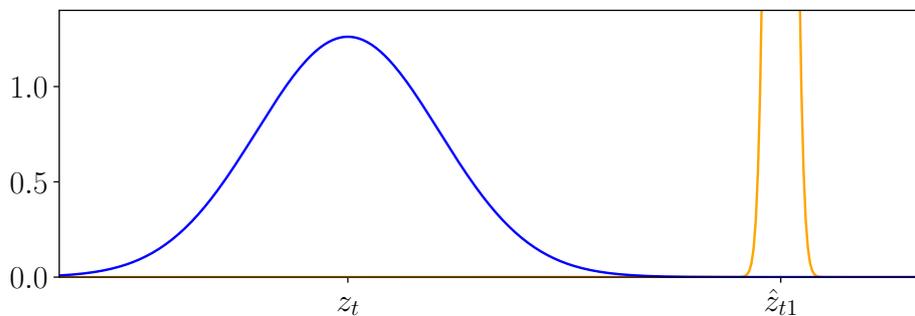


Abbildung 9.10.: Eine Messung z_t und die Erwartete Messung \hat{z}_t mit inkonsistent geschätzter Varianz.

Die erwartete Messung \hat{z}_t liegt in einem extrem unwahrscheinlichen Bereich der Verteilung von z_t und umgekehrt. Das liegt daran, dass der Fehler vom inkonsistenten Schätzer stark unterschätzt wird. Da die Varianz der Messung sehr viel höher ist, als die Varianz der erwarteten Messung, wird die erwartete Messung (und damit auch der Zustand) im Messschritt nur geringfügig geändert. Der Schätzer „glaubt“, die Messung sei eine Fehl-

9. (Extended/Unscented) Kalman Filter

messung, da sie nicht zur Erwartung passt. Im EKF und UKF äußert sich dieser Glaube darin, dass der Kalman Gain klein wird.

Daher korrigiert der Schätzer seinen Erwartungswert nur sehr langsam in Richtung der tatsächlichen Messung. Der Schätzer braucht also einige Messungen, bis er sich von der aufgebauten Inkonsistenz erholen kann. Zusätzlich sorgt die falsche Erwartung für Approximierungsfehler, wodurch sich die Schätzung auch weiter verschlechtern kann.

9.6.3. Fehler in der Gaußannahme

Eine Kernannahme des (E/U)KF ist das alle Verteilungen Gaußverteilt sind. Diese Annahme ist für 2 Teile der Filter zwingend nötig. Zum einen basiert das Konditionierungslemma des Messschrittes auf der Gauß-Annahme. Zum anderen repräsentieren („speichern“) die Filter die Verteilung nur mit Erwartungswert und Varianz. Andere Verteilungen können damit nicht dargestellt werden.

Die meisten Rauschphänomene kann man plausibel als gaußverteilt annehmen. Lineare Modelle ändern daran auch nichts. Die lineare Transformation einer Gaußverteilung ist wieder eine Gaußverteilung. Wie sieht das aber bei nichtlinearen Modellen aus? Wir wollen uns das Beispiel in Abbildung 9.11 ansehen.

Wir schauen uns die Wahrscheinlichkeitsverteilung einer Messfunktion $z = \arctan(x)$ an. Da dies nicht analytisch möglich ist, bedienen wir uns einer numerischen Approximation um die Wahrscheinlichkeitsverteilung darzustellen. Das heißt, wir schneiden den Zustandsraum in gleich große Abschnitte und geben die jeweilige Wahrscheinlichkeit dieses Abstandes an. Der Zustand x ist gaußverteilt und das entsprechende Histogramm sieht man im linken Bild. In der Mitte sieht man, wie die Messung z verteilt ist, wenn x gaußverteilt ist. Es ist sofort sichtbar, dass diese Verteilung nicht mehr Gauß ist.

Das die Verteilung von z nicht mehr Gauß ist widerspricht der Gaußannahme im EKF und führt daher zu Approximierungsproblemen. Der Filter stellt die Verteilung von z als Gaußverteilung mit Erwartungswert und Varianz dar. Selbst wenn Erwartungswert und Varianz korrekt berechnet wären, würde man damit die Verteilung vereinfachen. Die Vereinfachung hätte in diesem Beispiel einen Erwartungswert von 0 und würde damit verschleiern, dass Messungen um die $-0,5$ und $0,5$ wahrscheinlicher sind als Messungen von 0.

Nichtlineare Funktionen führen also dazu, dass die Gaußannahme nicht mehr stimmt. Eine weitere Fehlerquelle der Algorithmen für nichtlineare Modelle.

Schauen wir uns aber noch das rechte Bild an. In diesem Bild ist das Messrauschen des Sensors hinzugefügt. Dadurch gleicht die Verteilung wieder mehr einer Gaußverteilung. Dies äußert sich im Zentralen Grenzwertsatz (Kapitel 6.6). Dieser sagt aus, dass die Summe vieler Zufallsvariablen unabhängig von der Verteilung der einzelnen ZVs annähernd Gaußverteilt ist.

Dieser Umstand sorgt dafür, dass dieser Approximierungsfehler häufig nur einen geringen Einfluss auf das Messergebnis hat und die Linearisierungsfehler deutlich überwiegen.

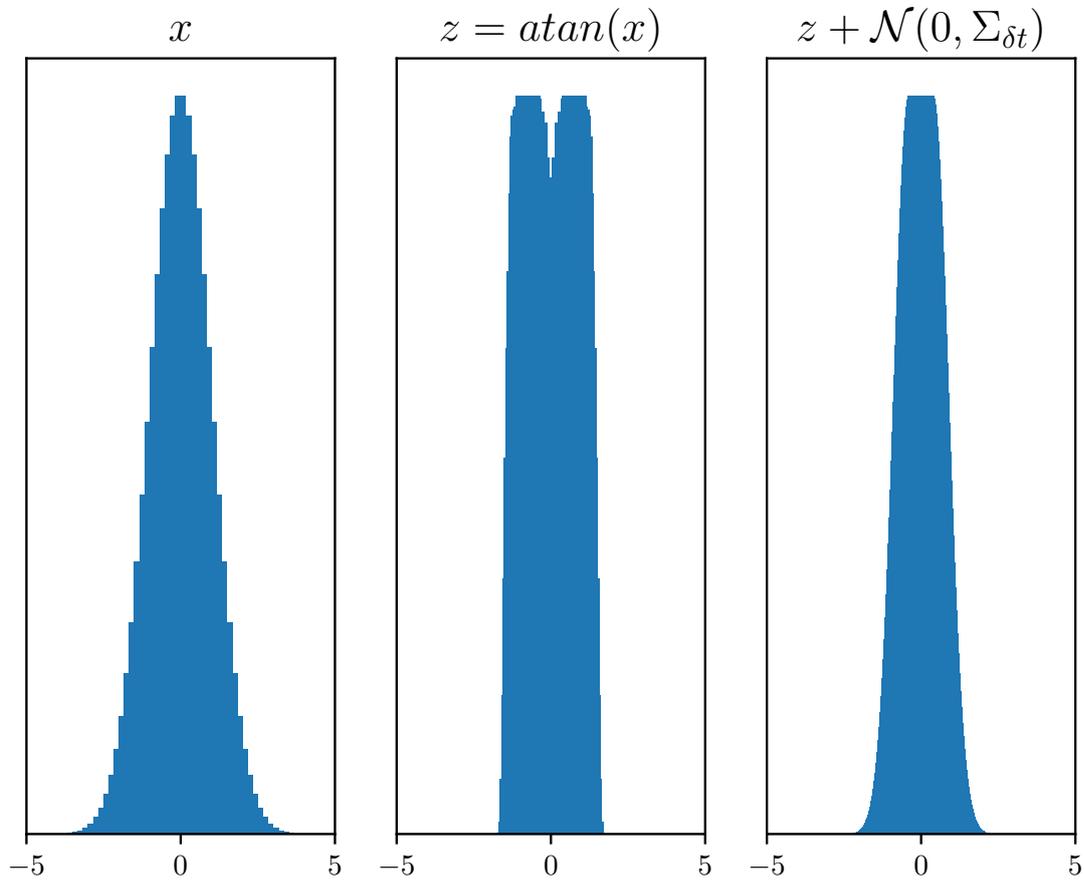


Abbildung 9.11.: Wie sieht die Messung in einem EKF wirklich aus? In der Graphik sind 3 Wahrscheinlichkeitsverteilungen numerisch approximiert dargestellt. Das heißt, die Graphik zeigt die Verteilungen in Histogrammen. Links ist die aktuelle Verteilung des Zustandes x gezeigt. Diese ist eine richtige Gaußverteilung. In der Mitte sieht man, wie eine Messung $z = \arctan(x)$ verteilt ist. Rechts ist die mittlere Verteilung plus Messrauschen gezeigt.

10. Lineare Algebra

10.1. Vektor, Vektorraum

Ein Vektor ist ein Element v von \mathbb{R}^n , $n \in \mathbb{N}$ aufgefasst nach Bedarf als

Zeilenvektor (v_1, \dots, v_n)

Spaltenvektor $\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$

Vektoren werden komponentenweise addiert und mit Skalaren $\lambda \in \mathbb{R}$ multipliziert:

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} + \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} v_1+w_1 \\ \vdots \\ v_n+w_n \end{pmatrix} \quad \lambda \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} \lambda v_1 \\ \vdots \\ \lambda v_n \end{pmatrix} \quad (10.1)$$

Spezielle Vektoren

Nullvektor $0 = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$

Einheitsvektoren $e_j = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \text{ (} j\text{te Stelle)} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$

Es gelten die offensichtlichen Rechenregeln, d. h. $(\mathbb{R}^n, +, \cdot, 0)$ ist ein \mathbb{R} -Vektorraum.

10.1.1. Skalarprodukt und Norm

Für $v, w \in \mathbb{R}^n$ setze

$$\langle v, w \rangle = v^\top w = \sum_{i=1}^n v_i w_i \quad \text{Skalarprodukt} \quad (10.2)$$

$$\|v\| = \sqrt{\langle v, v \rangle} = \sqrt{\sum_{i=1}^n v_i^2} \quad \text{Norm} \quad (10.3)$$

Geometrische Interpretation

$\|v\|$ Länge von v , $\langle v, w \rangle = \|v\| \cdot \|w\| \cdot \cos \angle(v, w)$.

Insbesondere $v \perp w \Leftrightarrow \langle v, w \rangle = 0$

10.3. Matrizen

Eine Matrix $A \in \mathbb{R}^{m \times n}$ besteht aus Zahlen $A_{ij} \in \mathbb{R}$

$$A = \begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \vdots \\ A_{m1} & \cdots & A_{mn} \end{pmatrix} \quad (10.10)$$

Ein Spaltenvektor ist eine $\mathbb{R}^{n \times 1}$ -Matrix. Ein Zeilenvektor ist eine $\mathbb{R}^{1 \times n}$ -Matrix.

Eine $m \times n$ -Matrix A besteht aus m Zeilenvektoren $A_{i\bullet}$ bzw. aus n Spaltenvektoren $A_{\bullet j}$.

Matrizen der gleichen Form werden komponentenweise addiert und mit Skalaren multipliziert; es gelten wieder die Vektorraum-Gesetze.

10.3.1. Spezielle Matrizen

Nullmatrix 0 ($0_{ij} = 0$)

Einheitsmatrix I ($I_{\bullet j} = e_j$)

10.3.2. Matrix-Vektor-Multiplikation

Definition 7. Für $A \in \mathbb{R}^{m \times n}$ und $v \in \mathbb{R}^n$ ist das Matrix-Vektor-Produkt Av ein Vektor $\in \mathbb{R}^m$, nämlich

$$(Av)_i = \sum_j A_{ij}v_j. \quad (10.11)$$

Definition 8. Eine lineare Abbildung $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ induziert eine Matrix $A^f \in \mathbb{R}^{m \times n}$ durch $(A^f)_{\bullet j} = f(e_j)$, $j = 1, \dots, n$.

Definition 9. Eine $m \times n$ -Matrix A induziert eine lineare Abbildung

$$f_A : \mathbb{R}^n \rightarrow \mathbb{R}^m, v \mapsto Av \quad (10.12)$$

(also $(Av)_i = \sum_{j=1}^n A_{ij}v_j = A_{i\bullet}v$)

Satz 16. Diese Zuordnung linearer Abbildungen zu Matrizen und umgekehrt sind gegenseitig invers.

$$f_{A^f} = f \quad A^{f_A} = A \quad (10.13)$$

Insbesondere stellt die von linearen Abbildung induzierte Matrix tatsächlich diese Abbildung dar.

Anschaulich behauptet dieses Theorem (10.8) in Matrix-Schreibweise:

$$\begin{pmatrix} f(v)_1 \\ f(v)_2 \\ \vdots \\ f(v)_m \end{pmatrix} = \begin{pmatrix} f(e_1)_1 & f(e_2)_1 & \dots & f(e_n)_1 \\ f(e_1)_2 & f(e_2)_2 & \dots & f(e_n)_2 \\ \vdots & \vdots & \dots & \vdots \\ f(e_1)_m & f(e_2)_m & \dots & f(e_n)_m \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \tag{10.14}$$

Beweis.

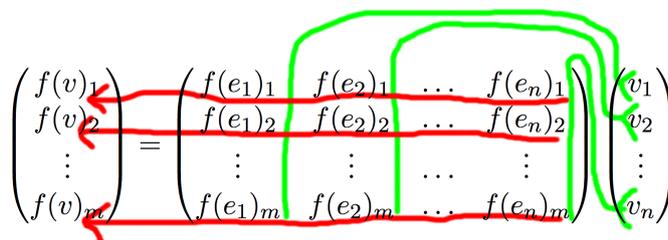
$$(A^f \cdot v)_i = \sum_{j=1}^n f(e_j)_i v_j = \left(\sum_{j=1}^n v_j f(e_j) \right)_i \stackrel{f \text{ lin.}}{=} f \left(\underbrace{\sum_{j=1}^n v_j e_j}_v \right)_i = f(v)_i \tag{10.15}$$

$$\Rightarrow f_{A^f} = f \tag{10.16}$$

$$f_A(e_j) = A \cdot e_j = A_{\bullet j} \tag{10.17}$$

$$\Rightarrow A^{f_A} = A \quad \square$$

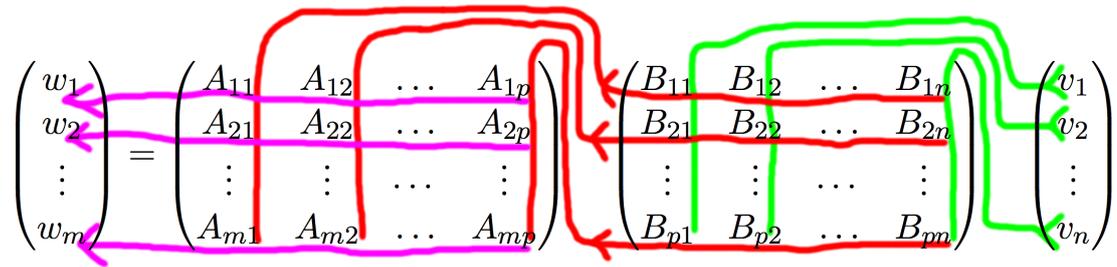
D. h. $m \times n$ -Matrixen „sind“ lineare Abbildungen $\mathbb{R}^n \rightarrow \mathbb{R}^m$. Dabei korrespondieren bei einer Matrix Spalten zu Eingaben die von „rechts und über oben“ in die Matrix hineinkommen (grüne Pfeile) und Zeilen zu Ausgaben die die Matrix nach links weitergibt (rote Pfeile). Dies erklärt auch die Bedingung, dass der Spaltenraum zur Eingabe (Vektor rechts von der Matrix) und der Zeilenraum zur Ausgabe (Ergebnisvektor) passen muss.



10.3.3. Matrixmultiplikation

Wir können nun lineare Abbildungen durch Matrizen darstellen und mit Hilfe der MV-Multiplikation auf Vektor-Argumente anwenden. Wir wollen nun Matrizen, also Abbildungen verknüpfen. Die wichtigste Verknüpfung ist die Matrixmultiplikation die der Abbildungskomposition entspricht.

Die Ausgangssituation ist dabei, dass ein Vektor nacheinander zwei lineare Abbildungen durchläuft, die durch Matrizen B und A dargestellt sind $w = A(Bv)$:



Wir möchten die Gesamtabbildung, welche wieder linear ist durch eine Matrix darstellen und diese aus A und B ausrechnen. Es wird sich im folgenden herausstellen, dass dies das Matrixprodukt AB ist.

Definition 10. Für $A \in \mathbb{R}^{m \times p}, B \in \mathbb{R}^{p \times n}$ ist $AB \in \mathbb{R}^{m \times n}$ definiert durch

$$(AB)_{ij} = \sum_{k=1}^p A_{ik} B_{kj} \quad (10.18)$$

Nicht alle Matrizen A und B lassen sich miteinander multiplizieren: Die Zeilenzahl von B muss der Spaltenzahl von A entsprechen. Das ist aus der Funktionskompositionssichtweise logisch, denn die Ausgabe der ersten Abbildung (Zeilen von B) muss Eingabe der zweiten Abbildung (Spalten von A) sein. Die Ergebnismatrix stellt dann eine Abbildung von den Eingaben von B zu den Ausgaben von A dar und hat daher die Zeilenzahl von A und die Spaltenzahl von B .

Dieser Sachverhalt wird durch folgendes Theorem bewiesen:

Theorem 10.2. *Matrix-Multiplikation entspricht Komposition der dargestellten Funktionen:*

$$f_{AB} = f_A \circ f_B : \mathbb{R}^n \xrightarrow{f_B} \mathbb{R}^p \xrightarrow{f_A} \mathbb{R}^m \quad (10.19)$$

Beweis.

$$\begin{aligned} f_A(f_B(v))_i &= \sum_{k=1}^p A_{ik} \cdot f_B(v)_k = \sum_{k=1}^p A_{ik} \cdot \sum_{j=1}^n B_{kj} v_j = \sum_{k=1}^p \sum_{j=1}^n A_{ik} \cdot B_{kj} v_j \quad (10.20) \\ &= \sum_{j=1}^n \left(\sum_{k=1}^p A_{ik} B_{kj} \right) v_j = \sum_{j=1}^n (AB)_{ij} v_j = f_{AB}(v)_i \quad \square \end{aligned}$$

Folgerung: Matrix-Multiplikation ist assoziativ: $(AB)C = A(BC)$

Ferner $f_1 = \text{id}$ (bzw. $A^{\text{id}} = I$), also $IA = A = AI$

10.3.4. Transponierte und Inverse

Transponierte

$$(A^\top)_{ij} = A_{ji}, \quad A \in \mathbb{R}^{m \times n} \Rightarrow A^\top \in \mathbb{R}^{n \times m} \quad (10.21)$$

Umkehrfunktion linearer Funktionen

Wir wollen rechnerisch eine durch eine Matrix dargestellte lineare Funktion umkehren. Dafür müssen wir zuerst einmal nachprüfen, dass die Umkehrung wieder linear ist also überhaupt durch eine Matrix dargestellt werden kann:

Satz 17. Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ linear und bijektiv (eindeutig). Dann ist f^{-1} auch linear.

Beweis.

$$f^{-1}(v + w) = f^{-1}(f(f^{-1}(v)) + f(f^{-1}(w))) \quad (10.22)$$

$$= f^{-1}(f(f^{-1}(v) + f^{-1}(w))) = f^{-1}(v) + f^{-1}(w) \quad (10.23)$$

$$f^{-1}(\lambda v) = f^{-1}(f(\lambda f^{-1}(v))) \quad (10.24)$$

$$= f^{-1}(f(\lambda f^{-1}(v))) = \lambda f^{-1}(v) \quad (10.25)$$

□

Inverse

$B \in \mathbb{R}^{n \times n}$ heißt Inverse von $A \in \mathbb{R}^{n \times n}$, falls $AB = I$ und $BA = I$.

Lemma 10.3. Wenn eine Inverse existiert ist sie eindeutig.

Beweis. Seien B, B' zwei Inverse von A . Dann ist $B' = BAB' = B$. □

Man schreibt daher $B = A^{-1}$; es gilt auch $A^{-1}A = I$. Für quadratische Matrizen ist $BA = I \Leftrightarrow AB = I$, das ist ein Satz aus der linearen Algebra, den wir hier nicht beweisen.

Satz 18. Umkehrung einer bijektiven Funktion entspricht dem Invertieren einer Matrix.

$$f_{A^{-1}} = (f_A)^{-1} \quad (10.26)$$

Beweis. Ist die durch A dargestellte Funktion f_A bijektiv, so gibt es eine lineare Umkehrfunktion (Satz 17), die wieder durch eine Matrix $B = A_{f_A^{-1}}$ dargestellt wird.

Diese Matrix B muss A^{-1} sein, weil

$$id = f_A \circ f_B = f_{AB} \Rightarrow I = AB \quad (10.27)$$

$$id = f_B \circ f_A = f_{BA} \Rightarrow I = BA \quad (10.28)$$

$$\Rightarrow B = A^{-1} \quad (10.29)$$

□

10.3.5. Rechenregeln

$$(A + B) + C = A + B + C \quad A + B + C = A + (B + C) \quad (10.30)$$

$$A + B = B + A \quad \lambda A + \lambda B = \lambda(A + B) \quad (10.31)$$

$$(AB)C = ABC \quad = A(BC) \quad (10.32)$$

$$(\lambda A)B = \lambda AB \quad A(\lambda B) = \lambda AB \quad (10.33)$$

$$(-A)B = -AB \quad A(-B) = -AB \quad (10.34)$$

$$\text{i.A. } AB \neq BA \quad (10.35)$$

$$(A + B)C = AC + BC, \quad C(A + B) = CA + CB \quad (10.36)$$

$$(A + B)^\top = A^\top + B^\top \quad (A - B)^\top = A^\top - B^\top \quad (10.37)$$

$$(\lambda A)^\top = \lambda A^\top \quad (10.38)$$

$$(AB)^\top = B^\top A^\top, \quad \text{_}^\top \text{ ist linear} \quad (10.39)$$

$$(AB)^{-1} = B^{-1}A^{-1}, \quad (\lambda A)^{-1} = \frac{1}{\lambda} \cdot A^{-1}, \quad (10.40)$$

$$(A^{-1})^{-1} = A, \quad (A^\top)^\top = A, \quad (10.41)$$

$$(A^\top)^{-1} = (A^{-1})^\top \quad (10.42)$$

$$AA^{-1} = A^{-1}A = I \quad (10.43)$$

In Worten: Addition und Multiplikation verhalten sich normal, aber die Reihenfolge muss gewahrt bleiben. Invertierung kann nur in Produkte hereingezogen werden, dabei kehrt sich die Reihenfolge um. Skalare können aus Summen, Produkten und invers aus Inversen herausgezogen werden. Transposition kann in alle Operationen (+, -, *, Inverse) hineingezogen werden, bei der Multiplikation kehrt sich die Reihenfolge um.

Satz 18 bedeutet, dass wir bijektive lineare Funktionen in ihrer Matrixdarstellung invertieren können, er liefert aber noch keinen Algorithmus dafür. Praktisch verwendet man meist LU bzw. Choleskyzerlegung und Spezialalgorithmen. Ein Divide-and-Konquer-Ansatz ergibt sich aus dem Blockmatrixinversionslemma, dass wir für das mehrdimensionale Konditionierungslemma brauchen und deshalb hier festhalten:

Bemerkung 10.4. Setzt man Matrizen blockweise zusammen, gelten analoge Rechenregeln für die Blöcke wie bei einzelnen Zahlen. Insbesondere:

$$\begin{pmatrix} x \\ z \end{pmatrix}^\top \begin{pmatrix} P & Q \\ R & S \end{pmatrix} \begin{pmatrix} x \\ z \end{pmatrix} = \begin{pmatrix} x \\ z \end{pmatrix}^\top \begin{pmatrix} Px + Qz \\ Rx + Sz \end{pmatrix} = x^\top Px + x^\top Qz + z^\top Rx + z^\top Sz. \quad (10.44)$$

Ist $\begin{pmatrix} P & Q \\ R & S \end{pmatrix}$ symmetrisch, also $P = P^\top, R = Q^\top, S = S^\top$, dann gilt vereinfacht sich obiger Ausdruck zu

$$= x^\top Px + 2x^\top Qz + z^\top Sz. \quad (10.45)$$

Lemma 10.5 (Block-Matrix-Inversions-Lemma). *Es ist:*

$$\begin{pmatrix} P & Q \\ R & S \end{pmatrix}^{-1} = \begin{pmatrix} \tilde{P} & \tilde{Q} \\ \tilde{R} & \tilde{S} \end{pmatrix} \quad (10.46)$$

Mit:

$$\tilde{P} = (P - QS^{-1}R)^{-1} \quad \text{Schur-Komplement} \quad (10.47)$$

$$\tilde{Q} = -\tilde{P}(QS^{-1}) \quad (10.48)$$

$$\tilde{R} = -(S^{-1}R)\tilde{P} \quad (10.49)$$

$$\tilde{S} = S^{-1} + (S^{-1}R)\tilde{P}(QS^{-1}). \quad (10.50)$$

Dies gilt, sofern die Inversen existieren. Die Umkehrung gilt im allgemeinen nicht (Bsp.: $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$). Für SPD Matrizen sind allerdings S und $P - QS^{-1}R$ auch SPD und invertierbar.

Beweis. Nachrechnen, dass $\begin{pmatrix} \tilde{P} & \tilde{Q} \\ \tilde{R} & \tilde{S} \end{pmatrix} \begin{pmatrix} P & Q \\ R & S \end{pmatrix} = I \dots$ □

10.4. Determinante

Die Determinante ist eine Funktion $\det : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ die Matrizen Zahlen zuordnet. Anschaulich entspricht die Zahl dem Faktor mit dem sich das Volumen von Punktmenge verändert, wenn sie mit der Matrix A malgenommen, also der linearen Abbildung f_A unterzogen werden:

$$\text{Volumen}(\underbrace{AM}_{\{Am|m \in M\}}) = |\det A| \text{Volumen}(M) \quad (10.51)$$

Ist $\det(A) < 0$ wird die Punktmenge M dabei gespiegelt, also rechts und links vertauscht.

Wichtige Eigenschaften:

$$\det(AB) = \det(A) \det(B)$$

$$\det(A^{-1}) = (\det(A))^{-1}$$

$$\det(A^T) = \det(A)$$

Ferner gilt: Eine Matrix A ist *invertierbar*, d.h. besitzt eine Inverse A^{-1} , genau dann, wenn $\det(A) \neq 0$. Das ist äquivalent zu der $\text{rang}(A) < n$ (Die Matrix hat keinen vollen Rang oder auch sie ist singulär).

Weiterhin erhalten einige Operationen auf Spalten oder Zeilen den Rang der Matrix und damit die Invertierbarkeit (siehe Tabelle 10.1).

Tabelle 10.1.: Entnommen aus (Enthält weitere Eigenschaften): https://www.tu-ilmeneu.de/fileadmin/media/num/neundorf/Dokumente/Lehre/hm/Determinante_Rechenregeln.pdf

Die Matrix B entstehe aus der Matrix A durch	Einfluss auf Determinante
Vertauschung zweier Zeilen oder Spalten von A	$\det(B) = -\det(A)$
Multiplikation einer Spalte/Zeile mit $\lambda \in \mathbb{R}$	$\det(B) = \lambda \det(A)$
Addition des λ -fachen einer Spalte/Zeile zu einer anderen	$\det(B) = \det(A)$

10.4.1. Die Jacobimatrix

Für $f : \mathbb{R}^m \rightarrow \mathbb{R}^n$ ist die Jacobimatrix eine $m \times n$ Matrix, definiert als:

$$f'(v) = \frac{\partial f(v)}{\partial v} = \left(\frac{\partial f(v)_i}{\partial v_j} \right)_{ij} \quad (10.52)$$

Damit lässt sich f an der Stelle x_0 nähern als

$$f(x) \approx \underbrace{f(x_0)}_{\mathbb{R}^m} + \underbrace{f'(x_0)}_{\mathbb{R}^{m \times n}} \underbrace{(x - x_0)}_{\mathbb{R}^n} \quad (10.53)$$

11. Wahrscheinlichkeitsrechnung in \mathbb{R}^n

Ziel dieses Kapitels ist es, die Begriffe Zufallsvariable, Erwartungswert, Varianz und Gauß-Verteilung auf \mathbb{R}^n zu generalisieren.

11.1. Zufallsvariable

$$X : \Omega \rightarrow \mathbb{R}^n, U : \Omega \rightarrow \mathbb{R}^{m \times n}$$

11.2. Erwartungswert

Die Definition des Erwartungswertes lässt sich direkt vom eindimensionalen Fall übertragen. Dabei ist die integrierte Größe einfach ein Vektor bzw. eine Matrix an Stelle einer Zahl.

$$E(X) = \int_{\omega \in \Omega} X(\omega)p(\omega)d\omega \quad (11.1)$$

$$= \int_{x \in \mathbb{R}^n} xp(X = x)dx \quad (11.2)$$

$$E(U) = \int_{\omega \in \Omega} U(\omega)p(\omega)d\omega \quad (11.3)$$

$$= \int_{u \in \mathbb{R}^{n \times m}} up(U = u)du \quad (11.4)$$

11.2.1. Eigenschaften

Da Integration, Summe und Produkt mit $\cdot p(\dots)$ komponentenweise geschehen, ist der Erwartungswert eines Vektors/einer Matrix einfach der Vektor/die Matrix der Erwartungswerte der Komponenten.

$$E(X)_i = E(X_i) \quad E(U)_{i,j} = E(U_{i,j}) \quad (11.5)$$

Beweis.

$$E(U)_{i,j} = \left(\int_{\omega \in \Omega} U(\omega)p(\omega)d\omega \right)_{i,j} \quad (11.6)$$

$$= \int_{\omega \in \Omega} U(\omega)_{i,j}p(\omega)d\omega \quad (11.7)$$

$$= E(U_{i,j}) \quad \square$$

11. Wahrscheinlichkeitsrechnung in \mathbb{R}^n

Weitere Eigenschaften sind (Beweis: Übungsaufgabe):

1. $E(U + V) = E(U) + E(V)$ für $U, V : \Omega \rightarrow \mathbb{R}^{m \times n}$
2. $E(\lambda U) = \lambda E(U)$ für $U : \Omega \rightarrow \mathbb{R}^{m \times n}, \lambda \in \mathbb{R}$
3. $E(A) = A$ für $A \in \mathbb{R}^{m \times n}$
4. $E(U^\top) = E(U)^\top$ für $U : \Omega \rightarrow \mathbb{R}^{m \times n}$
5. $E(UV) = E(U)E(V)$ für $U : \Omega \rightarrow \mathbb{R}^{m \times n}, V : \Omega \rightarrow \mathbb{R}^{n \times p}$ unabhängig, insbesondere $E(AV) = AEV$ und $E(UB) = EUB$ für $A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{n \times p}$

11.3. Kovarianzmatrix

Seien $X : \Omega \rightarrow \mathbb{R}^n, Y : \Omega \rightarrow \mathbb{R}^m$ Zufallsvariablen. Bei der Definition der Kovarianzmatrix muss gegenüber der Varianz eine Generalisierung des $(\dots)^2$ gefunden werden. Wir definieren:

$$\text{Cov}(X) = E((X - E(X))(X - E(X))^\top) \quad (11.8)$$

$$\text{Cov}(X, Y) = E((X - E(X))(Y - E(Y))^\top) \quad (11.9)$$

Dadurch wird $\text{Cov}(X)$ zu einer $n \times n$ und $\text{Cov}(X, Y)$ zu einer $n \times m$ Matrix. Auf den ersten Blick mag unklar sein, warum man die Transposition nicht an den ersten Faktor herandefiniert. In diesem Fall wäre das Ergebnis eine Zahl. Die folgenden Eigenschaften zeigen, dass die gewählte Definition sinnvoll ist. Die Eigenschaften der alternativen Definition werden in einer Übungsaufgabe analysiert.

11.3.1. Eigenschaften

Die Notation $\text{Cov}(X)$ ist eigentlich nur eine Schreibvereinfachung für $\text{Cov}(X) = \text{Cov}(X, X)$. Die Kovarianzmatrix $\text{Cov}(X, Y)$ besteht aus einer Matrixanordnung der Kovarianzen der Komponenten von X und Y :

$$\text{Cov}(X, Y)_{i,j} = \text{Cov}(X_i, Y_j). \quad (11.10)$$

Beweis. Wir machen zunächst die Beobachtung für $u \in \mathbb{R}^n, v \in \mathbb{R}^m$:

$$(uv^\top)_{ij} = \sum_{k=1}^1 u_{i,k}v_{j,k} = u_i v_j \quad (11.11)$$

Das heißt: Der Ausdruck uv^\top bildet die Matrix aller möglichen Produkte von Komponenten, das sogenannte äußere Produkt zweier Vektoren. Im Unterschied dazu bildet der Ausdruck $u^\top v$ das Skalarprodukt, d. h. die Summe der Produkte entsprechender Komponenten auch inneres Produkt genannt. Damit folgt:

$$\text{Cov}(X, Y)_{i,j} = E(((X - E(X))(Y - E(Y))^\top)_{i,j}) \quad (11.12)$$

$$= E((X_i - E(X_i))(Y_j - E(Y_j))) \quad (11.13)$$

$$= \text{Cov}(X_i, Y_j) \quad \square$$

Q 26
A.12
A.13

Weitere Eigenschaften sind (Beweis: Übungsaufgabe):

1. $\text{Cov}(X, Y) = E(XY^\top) - E(X)E(Y)^\top$ für $X : \Omega \rightarrow \mathbb{R}^n, Y : \Omega \rightarrow \mathbb{R}^m$
2. $\text{Cov}(X) = E(XX^\top) - E(X)E(X)^\top$ für $X : \Omega \rightarrow \mathbb{R}^n$
3. $\text{Cov}(\lambda X, \mu Y) = \lambda\mu \text{Cov}(X, Y)$ für $\lambda, \mu \in \mathbb{R}, X : \Omega \rightarrow \mathbb{R}^n, Y : \Omega \rightarrow \mathbb{R}^m$
4. $\text{Cov}(\lambda X) = \lambda^2 \text{Cov}(X)$ für $\lambda \in \mathbb{R}, X : \Omega \rightarrow \mathbb{R}^n$
5. $\text{Cov}(X, Y) = \text{Cov}(Y, X)^\top$ für $X : \Omega \rightarrow \mathbb{R}^n, Y : \Omega \rightarrow \mathbb{R}^m$
6. $\text{Cov}(X) = \text{Cov}(X)^\top$ für $X : \Omega \rightarrow \mathbb{R}^n$
7. $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$ für $X, Y : \Omega \rightarrow \mathbb{R}^n, Z : \Omega \rightarrow \mathbb{R}^m$
8. $\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$ für $X : \Omega \rightarrow \mathbb{R}^n, Y, Z : \Omega \rightarrow \mathbb{R}^m$
9. $\text{Cov}(X + Y) = \text{Cov}(X) + \text{Cov}(Y) + \text{Cov}(X, Y) + \text{Cov}(X, Y)^\top$ für $X, Y : \Omega \rightarrow \mathbb{R}^n$
10. $\text{Cov}(X, Y) = 0$ für $X : \Omega \rightarrow \mathbb{R}^n, Y : \Omega \rightarrow \mathbb{R}^m$ unabhängig
11. $\text{Cov}(X + Y) = \text{Cov}(X) + \text{Cov}(Y)$ für $X, Y : \Omega \rightarrow \mathbb{R}^n$ unabhängig
12. $\text{Cov}(AX, BY) = A \text{Cov}(X, Y) B^\top$ für $X : \Omega \rightarrow \mathbb{R}^n, Y : \Omega \rightarrow \mathbb{R}^m, A \in \mathbb{R}^{p \times n}, B \in \mathbb{R}^{q \times m}$
13. $\text{Cov}(AX) = A \text{Cov}(X) A^\top$ für $X : \Omega \rightarrow \mathbb{R}^n, A \in \mathbb{R}^{p \times n}$

Die Beweise lassen sich meist durch Rückgriff auf die entsprechenden eindimensionalen Eigenschaften und die Komponenteneigenschaft nachweisen, alternativ auch über direktes Rechnen mit der Matrixdefinition oder Eigenschaft 1 bzw. 2. Der Beweis für die Aussagen über $\text{Cov}(X)$ ergibt sich durch Einsetzen in die entsprechenden Aussagen über $\text{Cov}(X, Y)$. Besonders interessant sind Eigenschaften 12 und 13. Sie sagen aus, dass die Kovarianzmatrix alle Information enthält, die man braucht, um die Kovarianzen linearer Abbilder der Zufallsvariablen zu berechnen. Solange man im Linearen bleibt, und das tut die Theorie hinter dem Kalman-Filter etc., enthält die Kovarianzmatrix also alle Information, die man benötigt.

11.3.2. Positive Definitheit der Kovarianzmatrix

Lemma 11.1. $\text{Cov}(X)$ induziert eine quadratische Funktion $r \mapsto r^\top \text{Cov}(X)r$ die jeder Richtung r die Varianz von X "in Richtung r " zuordnet. Formal: $r^\top \text{Cov}(X)r = V(r^\top x)$, wobei für $\|r\| = 1$ die Größe $r^\top x$ als Skalarprodukt gerade die Projektion in Richtung r ist.

Beweis.

$$V(r^\top X) = \text{Cov}(r^\top X) = r^\top \text{Cov}(X)r^{\top\top} = r^\top \text{Cov}(X)r \quad \square$$

Das heißt, für jede Linearkombination der Komponenten von X ergibt sich die Varianz als $r^\top \text{Cov}(X)r$, wenn die Koeffizienten in r stehen.

Im Eindimensionalen war die Varianz ≥ 0 . Auch diese Eigenschaft generalisiert sich auf mehrere Dimensionen. Dabei ist der entsprechende Begriff, dass die Kovarianzmatrix positiv definit ist.

Q 27

11. Wahrscheinlichkeitsrechnung in \mathbb{R}^n

Definition 11. Eine symmetrische Matrix $A \in \mathbb{R}^{n \times n}$, $A^\top = A$ heißt *positiv semidefinit*, wenn $r^\top A r \geq 0, \forall r$, und *positiv definit*, wenn außerdem $r^\top A r > 0 \forall r \neq 0$.

D.h. eine Matrix ist positiv semidefinit, wenn ihre induzierte quadratische Funktion f nichtnegativ ist, und positiv definit, wenn außerdem der Nullvektor die einzige Nullstelle von f ist („positiver“ geht es nicht, da der Nullvektor Nullstelle jeder rein quadratischen Funktion ist).

Lemma 11.2. Eine symmetrisch positiv definite Matrix A hat eine symmetrisch positiv definite Inverse A^{-1} .

Beweis. Für $r \neq 0$ ist $Ar \neq 0$, weil sonst $r^\top A r = 0$ wäre. Damit ist A invertierbar. D.h. $A^{-1}r \neq 0$. Durch einsetzen erhalten wir:

$$0 < (A^{-1}r)^\top A(A^{-1}r) = r^\top A^{-1}AA^{-1}r = r^\top A^{-1}r \quad (11.14)$$

□

Korollar 11.3. $\text{Cov}(X)$ ist symmetrisch und positiv semidefinit.

Symmetrie gilt nach Eigenschaft 6 oben, positive Semidefinitheit folgt aus Nichtnegativität der Varianz und Lemma 11.1.

Zusammenfassend kann man die Kovarianzmatrix aus 3 verschiedenen Blickwinkeln sehen. Die Kovarianzmatrix ist

- eine als Erwartungswert eines äußeren Produktes definierte Größe
- eine systematische Zusammenstellung der paarweisen Kovarianzen der Komponenten
- eine Funktion, die jede Linearkombinationen der Komponenten ihre Varianz zuordnet.

Der erste Punkt ermöglicht, mit der Definition durch Anwenden von Matrix- und $E(\dots)$ -Regeln zu rechnen. Zudem garantiert er, dass derselbe geometrische Vektor unabhängig vom gewählten Koordinatensystem dieselbe Kovarianz beibehält.

Der zweite Punkt zeigt, dass in der Kovarianzmatrix die Information enthalten ist, die sich aus einer eindimensionalen Sichtweise auf die Vektoren ergeben würde. Außerdem ermöglicht er, einzelne Zahlen einer Matrix herauszupicken und zu verstehen, ohne sich die ganze Matrix anzuschauen.

Der letzte Punkt sagt aus, dass die Informationsmatrix die Unsicherheit in allen Richtungen repräsentiert, nicht nur in den Koordinatenrichtungen. Die Aussage ist auch insofern bemerkenswert, da es ja unendlich viele Richtungen gibt und man für die Varianz in allen Richtungen a priori unendlich viele Zahlen bräuchte. Die dritte Aussage besagt, dass die n^2 vielen Zahlen der Kovarianzmatrix schon genug sind.

11.4. Mehrdimensionale Gaußverteilung

Wir wollen jetzt die Gaußverteilung auf mehrere Dimensionen generalisieren (11.1). Analog zum eindimensionalen Fall gibt es zu jedem Mittelwert und jeder positiv definiten Kovarianzmatrix genau eine Gaußverteilung. Wir wollen die aus der Literatur bekannte Formel für diese Verteilung in zwei Schritten herleiten.

Zuerst definieren wir uns eine Std.-Gaußverteilung in n Dimensionen einfach als Vektor von n unabhängigen Std.-Gaußverteilungen in einer Dimension. Dann generieren wir daraus mit einer affinen Abbildung Gaußverteilungen mit beliebigem Erwartungswert und beliebiger Kovarianz.

Definition 12. $X : \Omega \rightarrow \mathbb{R}^n$ ist standard-Gaußverteilt, falls die X_i unabhängig und standard-Gaußverteilt sind.

Auffällig an der Standard-Gaußverteilung ist, dass sie rund ist, also in alle Richtungen gleich aussieht. Das ist nicht selbstverständlich, weil wir sie ja aus einer Gaußverteilung pro Dimension zusammengebaut haben. Man könnte also erwarten, dass man diese „Herkunft“ sieht und die Achsen als Richtung besonders sind. Dies ist z.B. der Fall, wenn man zwei 1D Uniformverteilungen zusammenbaut: Die resultierende Verteilung ist ein achsenparalleler Quader. Ein Begründung für diese Besonderheit liefert das folgende Lemma, das die Dichte der Standardgaußverteilung herleitet.

Lemma 11.4. Die Dichte der \mathbb{R}^n Standard-Gaußverteilung ist:

$$p(X = x) = \frac{1}{\sqrt{2\pi}^n} \exp\left(-\frac{1}{2}x^\top \mathbf{I}x\right) = \frac{1}{\sqrt{2\pi}^n} \exp\left(-\frac{1}{2}\sum_i x_i^2\right) \quad (11.15)$$

Beweis.

$$p(X = x) \stackrel{\text{unabhängig}}{=} \prod_{i=1}^n p(X_i = x_i) \quad (11.16)$$

$$= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x_i^2\right) \quad (11.17)$$

$$= \frac{1}{\sqrt{2\pi}^n} \exp\left(-\frac{1}{2}\sum_i x_i^2\right) \quad (11.18)$$

$$= \frac{1}{\sqrt{2\pi}^n} \exp\left(-\frac{1}{2}x^\top \mathbf{I}x\right) \quad \square$$

Man sieht, dass $p(X = x)$ von nur von $\sum_i x_i^2 = \|x\|^2$ abhängt. Daher ist die Standardgaußverteilung rund.

Lemma 11.5. Die Standard-Gaußverteilung X hat $E(X) = 0$ und $\text{Cov}(X) = \mathbf{I}$.

11. Wahrscheinlichkeitsrechnung in \mathbb{R}^n

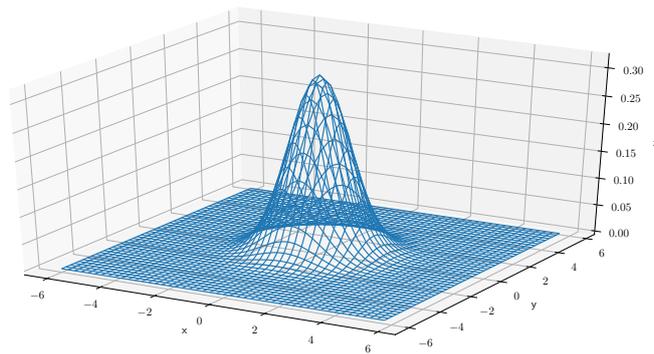


Abbildung 11.1.: PDF der 2D-Standard-Gaußverteilung

Beweis.

$$E(X)_i = E(X_i) = 0 \quad (11.19)$$

$$\text{Cov}(X)_{i,j} = \text{Cov}(X_i, Y_j) = \begin{cases} 0, & i \neq j, \text{ weil unabhängig} \\ 1, & i = j, \text{ weil Std.Gauß} \end{cases} \quad (11.20)$$

$$\Rightarrow \text{Cov}(X) = I \quad \square$$

Q 28

Das folgende Lemma ergibt sich aus der Transformationsregel für Integrale. Als Beispiel für seine Aussage: Halbiere ich eine 1-D Gaußsche Zufallsvariable, so staucht sich die Gaußglocke nicht nur um den Faktor 2, sie muss auch entsprechend höher werden, weil die Werte entsprechend „dichter“ liegen.

Lemma 11.6. Seien $X : \Omega \rightarrow \mathbb{R}^n$, $A \in \mathbb{R}^{n \times n}$ invertierbar, $b \in \mathbb{R}^n$ und $Y = AX + b$ dann

$$p(Y = y) = \frac{p(X = A^{-1}(y - b))}{|\det A|} \quad (11.21)$$

Beweis. Nach der Transformationsregel der Integralrechnung (man erinnere sich, dass A die Jacobimatrix von $x \mapsto Ax + b$ ist): per Substitution $y = Ax - b$, damit $x = A^{-1}(y - b)$ und $dx = dy/|\det A|$, gilt für jedes Ereignis B

$$\begin{aligned} P(Y \in B) &= P(X \in A^{-1}[B - b]) \\ &= \int_{x \in A^{-1}[B - b]} p(X = x) dx \\ &= \int_{y \in B} \frac{p(X = A^{-1}(y - b))}{|\det A|} dy \end{aligned}$$

□

Ein weiteres Lemma, die Cholesky Zerlegung, wird normalerweise in der Numerik bewiesen. Es sagt aus, dass es eine auf Matrizen generalisierte Quadratwurzel gibt. Diese Matrix lässt sich effizient berechnen mit der sogenannten Cholesky Zerlegung.

Lemma 11.7 (Cholesky). *Sei $\Sigma \in \mathbb{R}^{n \times n}$, symmetrisch positiv definit, dann gibt es $L \in \mathbb{R}^{n \times n}$, so dass $LL^\top = \Sigma$.*

Q 29

Mit diesen Hilfsmitteln zeigen wir, dass es zu jedem Erwartungswert und jeder pos. definiten Kovarianzmatrix eine Gaußlocke gibt, und leiten deren Formel her.

Satz 19. *Zu jedem $\mu \in \mathbb{R}^n$ und jedem symmetrisch positiv semi-definiten $\Sigma \in \mathbb{R}^{n \times n}$ mit Cholesky-Zerlegung $\Sigma = LL^\top$ erhält man per*

$$X = LY + \mu$$

ein affines Bild X der Standard-Gaußverteilung Y mit Erwartungswert μ , Kovarianz Σ und Dichte

$$p(X = x) = \frac{1}{\sqrt{2\pi^n} \sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(x - \mu)^\top \Sigma^{-1}(x - \mu)\right) \quad (11.22)$$

Beweis.

$$E(X) = E(LY + \mu) = LE(Y) + \mu \quad (11.23)$$

$$= \mu \quad (11.24)$$

$$\text{Cov}(X) = \text{Cov}(LY + \mu) = \text{Cov}(LY) + \text{Cov}(\mu) \quad (11.25)$$

$$= \text{Cov}(LY) = L \text{Cov}(Y) L^\top = LL^\top \quad (11.26)$$

$$= \Sigma \quad (11.27)$$

Damit stimmen Erwartungswert und Varianz. Nachzurechnen ist noch, ob die Verteilung die oben angegebene Form hat:

$$p(X = x) = \frac{p(Y = y)}{|\det L|} \quad \text{wobei: } y = L^{-1}(x - \mu) \quad (11.28)$$

$$= \frac{1}{\sqrt{2\pi^n} \det L} \exp\left(-\frac{1}{2}y^\top I y\right) \quad (11.29)$$

da, $\det \Sigma = \det(LL^\top) = \det L^2$

$$= \frac{1}{\sqrt{2\pi^n} \sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}y^\top I y\right) \quad (11.30)$$

$$= \frac{1}{\sqrt{2\pi^n} \sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(L^{-1}(x - \mu))^\top I(L^{-1}(x - \mu))\right) \quad (11.31)$$

$$= \frac{1}{\sqrt{2\pi^n} \sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(x - \mu)^\top L^{-\top} L^{-1}(x - \mu)\right) \quad (11.32)$$

$$= \frac{1}{\sqrt{2\pi^n} \sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(x - \mu)^\top (LL^\top)^{-1}(x - \mu)\right) \quad (11.33)$$

$$= \frac{1}{\sqrt{2\pi^n} \sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(x - \mu)^\top \Sigma^{-1}(x - \mu)\right) \quad \square$$

11. Wahrscheinlichkeitsrechnung in \mathbb{R}^n

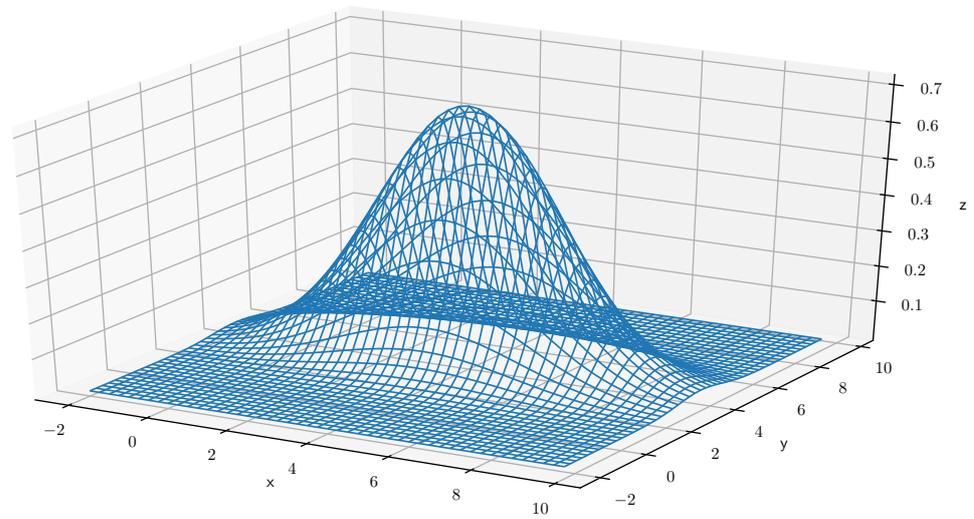


Abbildung 11.2.: Gaußverteilungen in 2D mit unabhängigen X_i

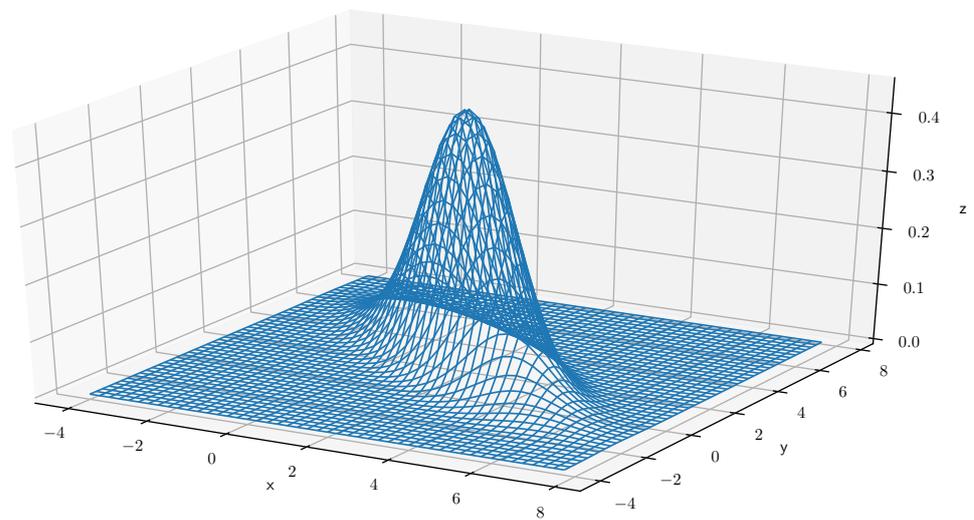


Abbildung 11.3.: Gaußverteilungen in 2D mit abhängigen X_i

11.5. Mahalonobis-Distanz in \mathbb{R}^n

Die Mahalonobis-Distanz, ist ein Distanzmaß, womit sich bestimmen lässt, wie plausibel ein Wert zu einer Wahrscheinlichkeitsverteilung passt. Im Kern, normiert die Mahalonobis-Distanz den Abstand vom getesteten Wert und dem Erwartungswert der Verteilung mit der Standardabweichung der Verteilung. Damit gibt sie die Differenz als Vielfaches der Standardabweichung wieder.

Auch die Mahalonobis-Distanz lässt sich auf mehrere Dimensionen generalisieren.

Definition 13. Sei $X : \Omega \mapsto \mathbb{R}^n$ eine Zufallsvariable mit $E(X) = \mu$ und $\text{Cov}(X) = \Sigma$, dann ist die Mahalonobis-Distanz von X zu einer Realisierung x bestimmt durch:

$$\Delta(x, X) = \sqrt{(x - \mu)^T \Sigma^{-1} (x - \mu)}$$

Im mehrdimensionalen kann man nicht mehr vom Vielfachen der Standardabweichung sprechen. Allerdings ist die Interpretation ähnlich. Für eine Standardgaußverteilung ist $\Sigma = I$ und $\Delta(x, X) = \|x - EX\|$ einfach die euklidische Distanz zum Erwartungswert 0. Eine allgemeine Gaußverteilung haben wir ja definiert in dem wir die Standardgaußverteilung verschoben und mit L linear transformiert haben. Transformiert man die euklidische Distanz mit, so erhält man die Mahalonobisdistanz.

Lemma 11.8. Sei $X : \Omega \mapsto \mathbb{R}^n$ mit $EX = \mu$ und $\text{Cov} X = \Sigma = LL^T$. Dann ist

$$\Delta(x, X) = \|L^{-1}(x - \mu)\|. \quad (11.34)$$

Beweis. Wir setzen $y = L^{-1}(x - \mu)$, so dass $x = Ly + \mu$ ist.

$$\Delta(x, X)^2 = (x - \mu)^T \Sigma^{-1} (x - \mu) \quad (11.35)$$

$$= (Ly + \mu - \mu)^T \Sigma^{-1} (Ly + \mu - \mu) \quad (11.36)$$

$$= y^T L^T (LL^T)^{-1} Ly = y^T L^T L^{-T} L^{-1} Ly = y^T y = \|y\|^2 \quad (11.37)$$

□

Man sieht allerdings, dass $\Delta(x, X)$ dem Exponenten einer Gaußverteilung sehr ähnlich ist. Man bezeichnet die Mahalonobisdistanz auch als Logarithmus der Gaußdichte. Das ist mathematisch allerdings nicht ganz richtig, weil es den Faktor $\frac{1}{2}$ und den Vorfaktor ignoriert.

Im eindimensionalen ist für eine Gaußverteilung $E(\Delta(X, X)^2) = 1$ (siehe 8.4). Im mehrdimensionalen ist es eins pro Dimension:

Lemma 11.9. Erwartungswert der Mahalonobisdistanz bei Gaußverteilungen Sei $X : \Omega \mapsto \mathbb{R}^n$ eine Gaußverteilte Zufallsvariable mit $E(X) = \mu$ und $\text{Cov}(X) = \Sigma$, dann ist der Erwartungswert der quadratischen Mahalonobisdistanz aller Elemente in X gegeben durch:

$$E(\Delta(X, X)^2) = E((X - \mu)^T \Sigma^{-1} (X - \mu)) = n$$

11. Wahrscheinlichkeitsrechnung in \mathbb{R}^n

Beweis. Sei Y die Standardgaußverteilung in \mathbb{R}^n , so dass $X = LY + \mu$ mit $LL^T = \Sigma$ ist. Dann ist:

$$E(\Delta^2(X, X)^2) = E(\|Y\|^2) = E\left(\sum_{i=1}^n Y_i^2\right) = \sum_{i=1}^n E(Y_i^2) = n \cdot 1 \quad (11.38)$$

□

Der Erwartungswert entspricht also der Anzahl der Dimensionen n . Eine Erklärung dafür liefert der Blick auf die Standardgaußverteilung in \mathbb{R}^n . In jeder Dimension der Verteilung ist der erwartete Fehler 1. Da die Dimensionen unabhängig sind, summieren sich die Fehler auf, d.h. im Durchschnitt ist ein Element der Verteilung 1 in jeder Dimension vom Erwartungswert entfernt. Der Gesamtabstand zum Erwartungswert ergibt sich dann als die Norm zu \sqrt{n} , was der Mahalanobis-Distanz entspricht. Da die Mahalanobis-Distanz eine Normierung des Abstandes auf die Kovarianz durchführt, bleibt diese Eigenschaft auch in allgemeinen mehrdimensionalen Gaußverteilungen erhalten.

12. Mehrdimensionale Quadratische Ausgleichsrechnung

12.1. Lineare Ausgleichsrechnung

Die lineare quadratische Ausgleichsrechnung ist das n -dimensionale Pendant zur Zwei-Messungs Formel (**Satz 6**) in einer Dimension. Anders als dort brauchen wir aber formal nur eine Messung zu betrachten, da sich mehrere vektorielle Messungen zu einem größeren Vektor zusammenstapeln lassen. Die Frage ist also, was ist die Verteilung des gesuchten Zustandes X wenn ich eine Messung Z beobachtet habe und Z linear plus Rauschen von X abhängt. Der Zustand ist erst mal nicht von der Zeit abhängig, man spricht in dem Kontext häufiger von Parametern als vom Zustand.

Diese Formel wird sehr häufig genutzt. Sie ist die Grundlage z.B. für Ausgleichsgeraden oder -kurven und Kalibrierung von Messgeräten. Historisch wurde sie von C. F. Gauß zur Vermessung erfunden.

Zur Vorbereitung benötigen wir das folgende Lemma. Es besagt, dass ein Exponential einer quadratischen Funktion immer proportional zu einer Gaußverteilung ist und gibt Erwartungswert und Kovarianz an.

Lemma 12.1. *Ist $X : \Omega \rightarrow \mathbb{R}^n$ verteilt als Exponential einer quadratischen Funktion $p(X = x) \propto \exp\left(-\frac{1}{2}(x^\top Ax + x^\top b + \gamma)\right)$ mit $A \in \mathbb{R}^{n \times n}$ sym. pos. def., $b \in \mathbb{R}^n$, $\gamma \in \mathbb{R}$, so ist X Gaußsch mit*

$$E(X) = \frac{-A^{-1}b}{2} \qquad \text{Cov}(X) = A^{-1} \qquad (12.1)$$

Beweis. Positiv definite Matrizen sind invertierbar.

$$\begin{aligned} & \exp\left(-\frac{1}{2}\left(x - \frac{-A^{-1}b}{2}\right)^\top A \left(x - \frac{-A^{-1}b}{2}\right)\right) \\ & \propto \exp\left(-\frac{1}{2}(x^\top Ax + b^\top A^{-1}Ax/2 + x^\top AA^{-1}b/2)\right) \\ & = \exp\left(-\frac{1}{2}(x^\top Ax + x^\top b)\right) \qquad \square \end{aligned} \qquad (12.2)$$

Q 30

12.1.1. Annahmen

1. Zustand $X : \Omega \rightarrow \mathbb{R}^n$ uniform verteilt, gesucht

12. Mehrdimensionale Quadratische Ausgleichsrechnung

2. $Z = AX + b + \varepsilon$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $\varepsilon : \Omega \rightarrow \mathbb{R}^m \sim \mathcal{N}(0, \Sigma_\varepsilon)$, Z gemessen als z
3. X und ε unabhängig

A bildet ab von Zustand nach Messung. Jede Zeile in A bzw. b entspricht einer Messung, jede Spalte von A einem unbekanntem Zustand (man sagt meist Parameter). Typischerweise ist $m > n$, weil nur bei mehr Messungen als zu schätzenden Freiheitsgraden sich eine Redundanz der Messungen ergibt, durch die die quadratische Ausgleichsrechnung die Genauigkeit verbessern kann.

12.1.2. Beispiel

Eine Wägeelektronik liefert eine Spannung abhängig von der gewogenen Masse, wobei der Zusammenhang linear ist. Nullpunkt und Faktor dieses linearen Zusammenhangs sollen kalibriert werden. Dazu werden Gewichte von $m_1 = 5$, $m_2 = 10$, $m_3 = 15$ und $m_4 = 20$ kg auf die Waage gelegt und der jeweilige Messwert abgelesen ($Z_i = z_i$ in Volt). Die Messungen sind unabhängig und rauschen um $\sigma = 10mV$.

Der Zustand (bzw. die gesuchten Parameter) $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ besteht aus Nullpunkt (X_1) und Faktor (X_2),

$$Z_i = X_1 + m_i X_2 + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, 0.01^2) \quad (12.3)$$

$$A = \begin{pmatrix} 1 & 5 \\ 1 & 10 \\ 1 & 15 \\ 1 & 20 \end{pmatrix}, \quad b = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \Sigma_\varepsilon = \begin{pmatrix} 0.01^2 & & & \\ & 0.01^2 & & \\ & & 0.01^2 & \\ & & & 0.01^2 \end{pmatrix} \quad (12.4)$$

A und b drücken zusammen aus, dass Nullpunkt X_1 plus Masse mal *Faktor* X_2 die Spannung ist die die Waage liefern sollte. Die Diagonalstruktur von Σ_ε drückt Unabhängigkeit der Messungen aus.

Es hängt das Ergebnis nicht vom absoluten Wert des Rauschens ab, sondern nur vom Verhältnis der Komponenten, daher kann man auch 1 statt 0.01^2 verwenden, was aussagt, dass sie gleich sind.

Q 31
A.15

12.1.3. Zeitabhängige Probleme

Man kann aber auch zeitanhängige Probleme mit quadratischer Ausgleichsrechnung behandeln, indem verschiedene Komponenten von X sich auf verschiedene Zeitpunkte beziehen. Z.B. kann man bei einem Filter-Problem $X = [X_{T_{\text{start}}}, \dots, X_{T_{\text{end}}}]$ setzen, also die Zustände X_t für das Zeitfenster $t = T_{\text{start}}, \dots, T_{\text{end}}$ schätzen. Hat das Zeitfenster immer eine feste Dauer und endet auf die Gegenwart, so spricht man von einem „Sliding Window“ Ansatz.

12.1.4. Formel

Die lineare Quadratische Ausgleichsrechnung ist mehr eine Formel als ein Algorithmus.

$$\hat{x} = (A^\top \Sigma_\varepsilon^{-1} A)^{-1} A^\top \Sigma_\varepsilon^{-1} (z - b) \quad (12.5)$$

$$\Sigma_{\hat{x}} = (A^\top \Sigma_\varepsilon^{-1} A)^{-1} \quad (12.6)$$

Satz 20. *Unter den obigen Annahmen ist $X|Z = z$ Gaußsch mit*

$$E(X|Z = z) = (A^\top \Sigma_\varepsilon^{-1} A)^{-1} A^\top \Sigma_\varepsilon^{-1} (z - b) \quad (12.7)$$

$$\text{Cov}(X|Z = z) = (A^\top \Sigma_\varepsilon^{-1} A)^{-1}, \quad (12.8)$$

sofern A n linear unabhängige Spalten hat (Rank n).

Wäre A invertierbar, würde sich (12.7) zu $E(X|Z = z) = A^{-1}(z - b)$ reduzieren, d. h. es würde der eine eindeutige Wert für X bestimmt, bei dem alle Messungen z exakt passen. Folgerung: Nur wenn $m > n$, also Redundanz in den Messungen vorhanden ist, hilft die Ausgleichsrechnung und dann lässt sich die Inverse in (12.7) nicht in die Klammer hineinziehen.

Hat A Rank $< n$, sind also Spalten von A linear abhängig, dann gibt es einen Vektor $v \in \mathbb{R}^n$ mit $Av = 0$. Änderungen von x um v beeinflussen (die Verteilung von) Z nicht, man sagt dieser Freiheitsgrad ist nicht beobachtbar. Als Folge wird $A^\top \Sigma_\varepsilon^{-1} A$ nicht invertierbar und die a-posteriori Verteilung ist nicht Gaußsch.

Beweis.

$$p(X = x|Z = z) \stackrel{\text{Bayes}}{=} \frac{p(Z = z|X = x)p(X = x)}{p(Z = z)} \quad (12.9)$$

$$\stackrel{X \text{ uniform}}{\propto} p(Z = z|X = x) = p(\varepsilon = z - (Ax + b)|X = x) \quad (12.10)$$

$$\stackrel{\varepsilon, X \text{ unabh.}}{=} p(\varepsilon = z - b - Ax) \quad (12.11)$$

$$\propto \exp\left(-\frac{1}{2}((z - b) - Ax)^\top \Sigma_\varepsilon^{-1}((z - b) - Ax)\right) \quad (12.12)$$

$$\propto \exp\left(-\frac{1}{2}\left(x^\top A^\top \Sigma_\varepsilon^{-1} Ax - x^\top A^\top \Sigma_\varepsilon^{-1}(z - b) - (z - b)^\top \Sigma_\varepsilon^{-1} Ax\right)\right) \quad (12.13)$$

$$= \exp\left(-\frac{1}{2}\left(x^\top A^\top \Sigma_\varepsilon^{-1} Ax - x^\top A^\top \Sigma_\varepsilon^{-1}(z - b) \cdot (2)\right)\right) \quad (12.14)$$

$$= \exp\left(-\frac{1}{2}\left(x^\top \underbrace{A^\top \Sigma_\varepsilon^{-1} A}_{A \text{ Lemma 12.1}} x + x^\top \underbrace{A^\top \Sigma_\varepsilon^{-1}(b - z)}_{b \text{ Lemma 12.1}} \cdot (2)\right)\right) \quad (12.15)$$

$A^\top \Sigma_\varepsilon^{-1} A$ ist positiv definit, weil

$$x^\top (A^\top \Sigma_\varepsilon^{-1} A)x = (Ax)^\top \Sigma_\varepsilon^{-1} (Ax) > 0. \quad (12.16)$$

12. Mehrdimensionale Quadratische Ausgleichsrechnung

Das > 0 gilt, weil Σ_ε positiv definit ist und Ax wegen Rank n nicht 0 sein kann. Damit ist nach Lemma 12.1

$$\Rightarrow X|Z = z \text{ Gaußsch} \quad (12.17)$$

$$E(X|Z = z) = (A^\top \Sigma_\varepsilon^{-1} A)^{-1} A^\top \Sigma_\varepsilon^{-1} (z - b) \quad (12.18)$$

$$\text{Cov}(X|Z = z) = (A^\top \Sigma_\varepsilon^{-1} A)^{-1} \quad \square$$

12.2. Nichtlineare Ausgleichsrechnung als iterierte Linearisierung

12.2.1. Lineare Näherung (Tangente) im \mathbb{R}^n

Lemma 12.2. Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ stetig partiell diffbar und $x_0 \in \mathbb{R}^n$. Dann ist

$$\bar{f}(x) = f(x_0) + \left. \frac{\partial f(x)}{\partial x} \right|_{x=x_0} \cdot (x - x_0) \quad (12.19)$$

die lineare Näherung an f in x_0 , d.h.

$$\|f(x) - \bar{f}(x)\| = o(\|x - x_0\|) \quad (12.20)$$

Beweis. Normalerweise in der Analysis bewiesen als "Stetig partiell diffbar impliziert total diffbar". \square

12.2.2. Annahmen

Die nichtlineare Ausgleichsrechnung erweitert die Annahmen von oben auf einen nicht-linearen Zusammenhang zwischen Zustand und Messung.

1. X uniform verteilt $X : \Omega \rightarrow \mathbb{R}^n$, gesucht
2. $Z = f(X) + \varepsilon$, $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ stetig diffbar, $\varepsilon : \Omega \rightarrow \mathbb{R}^m \sim \mathcal{N}(0, \Sigma_\varepsilon)$, gemessen als z
3. X und ε unabhängig

Idee: f linearisieren:

$$\bar{f}(x) = f(x_0) + \left. \frac{\partial f(x)}{\partial x} \right|_{x=x_0} \cdot (x - x_0) \quad (12.21)$$

$$= Ax + b \quad (12.22)$$

$$A = \left. \frac{\partial f(x)}{\partial x} \right|_{x=x_0} \quad (12.23)$$

$$b = f(x_0) - \left. \frac{\partial f(x)}{\partial x} \right|_{x=x_0} x_0 = f(x_0) - Ax_0 \quad (12.24)$$

$$f(x) \approx \bar{f}(x) \quad (12.25)$$

12.3. Nichtlineare Ausgleichsrechnung als Optimierung

$$E(X|Z = z) \approx (A^\top \Sigma_\varepsilon^{-1} A)^{-1} A^\top \Sigma_\varepsilon^{-1} (z - b) \quad (12.26)$$

$$= (A^\top \Sigma_\varepsilon^{-1} A)^{-1} A^\top \Sigma_\varepsilon^{-1} (z - f(x_0) + Ax_0) \quad (12.27)$$

$$= (A^\top \Sigma_\varepsilon^{-1} A)^{-1} A^\top \Sigma_\varepsilon^{-1} (z - f(x_0)) + (A^\top \Sigma_\varepsilon^{-1} A)^{-1} (A^\top \Sigma_\varepsilon^{-1} A) x_0 \quad (12.28)$$

$$= (A^\top \Sigma_\varepsilon^{-1} A)^{-1} A^\top \Sigma_\varepsilon^{-1} (z - f(x_0)) + I * x_0 \quad (12.29)$$

$$\text{Cov}(X|Z = z) \approx (A^\top \Sigma_\varepsilon^{-1} A)^{-1} \quad (12.30)$$

Die Linearisierung und damit die Schätzung ist nur eine Näherung. Wie gut sie ist hängt davon ab, wie linear sich f im Bereich zwischen dem gewählten Linearisierungspunkt und der Schätzung verhält. Man kann dieses Ergebnis erheblich verbessern, indem man iteriert, d. h. die letzte Schätzung als neuen Linearisierungspunkt nimmt und erneut durchrechnet.

12.2.3. Algorithmus (Gauß-Newton)

Startwert μ_0 wählen

Wiederhole bis μ_i konvergiert:

$$A_i := \left. \frac{\partial f(x)}{\partial x} \right|_{x=\mu_{i-1}} \quad (12.31)$$

$$\mu_{i+1} := \mu_i + (A_i^\top \Sigma_\varepsilon^{-1} A_i)^{-1} A_i^\top \Sigma_\varepsilon^{-1} (z - f(\mu_i)) \quad (12.32)$$

$$\text{danach } \Sigma = (A^\top \Sigma_\varepsilon^{-1} A)^{-1} \quad (12.33)$$

A.16

A.17

12.3. Nichtlineare Ausgleichsrechnung als Optimierung

Der Gauß-Newton Algorithmus ist das am häufigsten verwendete Verfahren für z.B. Kalibrierungsprobleme. Erst einmal ist es nur eine plausible Näherung. Man kann allerdings eine stärkere Aussage beweisen, nämlich, dass das Verfahren gegen eines lokales Maximum der a-posteriori Wahrscheinlichkeit $p(X = x|Z = z)$ konvergiert.

12.3.1. A-posteriori Wahrscheinlichkeit

Wir wollen die a-posteriori Verteilung $p(X = x|Z = z)$ explizit hinschreiben:

Satz 21. *Unter den obigen Annahmen ist*

$$p(X = x|Z = z) \propto \exp\left(-\frac{1}{2}(z - f(x))^\top \Sigma_\varepsilon^{-1} (z - f(x))\right) \quad (12.34)$$

12. Mehrdimensionale Quadratische Ausgleichsrechnung

Beweis.

$$p(X = x|Z = z) = \frac{p(Z = z|X = x)p(X = x)}{p(Z = z)} \propto p(Z = z|X = x) \quad (12.35)$$

$$= p(\varepsilon = z - f(x)) \propto \exp\left(-\frac{1}{2}(z - f(x))^\top \Sigma_\varepsilon^{-1}(z - f(x))\right) \quad (12.36)$$

□

Lemma 12.3. *Es gelten die obigen Annahmen und seien zusätzlich die ε_i unabhängig und $\sigma(\varepsilon_i) = \sigma_i$. Dann ist die a-posteriori Wahrscheinlichkeit*

$$p(X = x|Z = z) \propto \exp\left(-\frac{1}{2} \sum_i \left(\frac{z_i - f_i(x)}{\sigma_i^2}\right)^2\right) \quad (12.37)$$

Beweis.

$$\Sigma_\varepsilon = \text{Cov}(\varepsilon)_{ij} = \begin{cases} 0 & i \neq j \\ \sigma_i^2 & i = j \end{cases} \quad (12.38)$$

$$p(X = x|Z = z) \propto \exp\left(-\frac{1}{2}(z - f(x))^\top \Sigma_\varepsilon^{-1}(z - f(x))\right) \quad (12.39)$$

$$\propto \exp\left(-\frac{1}{2} \sum_i (z - f(x))_i \sigma_i^{-2} (z - f(x))_i\right) \quad (12.40)$$

$$= \exp\left(-\frac{1}{2} \sum_i \left(\frac{z_i - f_i(x)}{\sigma_i^2}\right)^2\right) \quad (12.41)$$

□

Satz 22 (Quadratische Ausgleichsrechnung). *Es gelten die Annahmen wie in Lemma 12.3. Dann ist die Maximum-A-Posteriori-Schätzung das Minimum des mittleren quadratischen Fehlers*

$$\arg \max_x p(X = x|Z = z) = \arg \min_x \sum_i \left(\frac{z_i - f_i(x)}{\sigma_i^2}\right)^2 \quad (12.42)$$

Beweis.

$$\arg \max_x p(X = x|Z = z) = \arg \max_x \exp\left(-\frac{1}{2} \sum_i \left(\frac{z_i - f_i(x)}{\sigma_i^2}\right)^2\right) \quad (12.43)$$

$$= \arg \max_x -\frac{1}{2} \sum_i \left(\frac{z_i - f_i(x)}{\sigma_i^2}\right)^2 \quad (12.44)$$

$$= \arg \min_x \sum_i \left(\frac{z_i - f_i(x)}{\sigma_i^2}\right)^2 \quad (12.45)$$

□

Intuitiv ist $z_i - f_i(x)$ das sogenannte Residuum, also der Unterschied, was man beim Zustand x hätte messen sollen und was man gemessen hat. Das quadrierte Residuum ist ein Maß für den Fehler, welches immer positiv ist, positive und negative Residuen gleich bewertet und für größere Residuen steigt. Dividiert durch σ_i^2 wird, wie bei der quadrierten Mahalanobis-Distanz das Residuum im Verhältnis zur Messunsicherheit gesetzt, so dass ein Residuum von $\pm\sigma_i$ immer zu 1 wird. Der einzelne Summand bewertet also, wie plausibel der Messwert z_i ist, wenn $X = x$ wäre. Summiert ist es die Gesamtplausibilität und die gilt es zu optimieren.

Deshalb ist eine allgemeine Herangehensweise an NL quadratische Ausgleichsrechnung einen Optimierungsalgorithmus auf die obige Loss-Funktion anzuwenden. Der Gauß-Newton-Algorithmus ist ein solcher, auch wenn wir ihn anders als iterierte Linearisierung hergeleitet haben und das etabliert das folgende Lemma.

12.3.2. Konvergenz von Gauß-Newton

Lemma 12.4. Wenn (12.32) konvergiert, so gegen einen kritischen Punkt (Maximum, Minimum, Sattelpunkt) der a-posteriori Wahrscheinlichkeitsdicht $p(X = x|Z = z)$.

Beweis. Wir betrachten zuerst, was es bedeutet, dass (12.32) konvergiert:

$$\mu = \lim_i \mu_i \tag{12.46}$$

$$A = \lim_i \left. \frac{\partial f(x)}{\partial x} \right|_{x=\mu_i} = \left. \frac{\partial f(x)}{\partial x} \right|_{x=\mu}. \tag{12.47}$$

Letzteres gilt wegen stetiger Diffbarkeit von f . Bildet man von beiden Seiten von (12.32) den Grenzwert, so folgt

$$0 = (A^\top \Sigma_\epsilon^{-1} A)^{-1} A^\top \Sigma_\epsilon^{-1} (z - f(\mu)) \tag{12.48}$$

$$0 = A^\top \Sigma_\epsilon^{-1} (z - f(\mu)) \tag{12.49}$$

$$0 = (A_{\bullet j})^\top \Sigma_\epsilon^{-1} (z - f(\mu)) \quad \forall j \tag{12.50}$$

$$0 = \frac{\partial f(x)}{\partial x_j}^\top \Sigma_\epsilon^{-1} (z - f(\mu)) \quad \forall j \tag{12.51}$$

Nun betrachten wir die Ableitung der Wahrscheinlichkeit nach x_j .

$$\frac{\partial p(X = x|Z = z)}{\partial x_j} \tag{12.52}$$

$$\propto \frac{\partial}{\partial x_j} \exp \left(-\frac{1}{2} (z - f(x))^\top \Sigma_\epsilon^{-1} (z - f(x)) \right) \tag{12.53}$$

12. Mehrdimensionale Quadratische Ausgleichsrechnung

$$= \exp(\dots) \left(-\frac{1}{2}\right) \left(-(z - f(x))^T \Sigma_\varepsilon^{-1} \frac{\partial f(x)}{\partial x_j} - \frac{\partial f(x)^T}{\partial x_j} \Sigma_\varepsilon^{-1} (z - f(x)) \right) \quad (12.54)$$

$$= \exp(\dots) \left(\frac{\partial f(x)^T}{\partial x_j} \Sigma_\varepsilon^{-1} (z - f(x)) \right) \quad (12.55)$$

Ist $x = \mu$, also im Grenzwert, ergibt sich 0. □

Diese hart bewiesene Aussage alleine betrachtet ist noch nicht sehr aufschlussreich, will man doch gegen ein Maximum konvergieren. Folgendes Argument sagt, dass dies meistens der Fall ist: Wäre f linear, ergäbe sich in einem Schritt das einzige Maximum (Lemma 20). Ist f moderat nicht-linear und startet man nah genug am Maximum wird zwar die Funktion etwas verzerrt, aber aus dem Maximum wird nicht ein Sattelpunkt oder Minimum, dazu müsste sich die Funktion schon sehr ändern. Deshalb konvergiert der Algorithmus von einem guten Startwert typischerweise zum Maximum.

12.3.3. Softwarebibliotheken

Ceres

Es gibt noch verbesserte Varianten des Algorithmus, z. B. den Levenbergh-Marquardt Algorithmus [37, Kap. 15.5]. Dieser Algorithmus wird sehr weit genutzt zur Kalibrierung oder zur Vermessung, wenn Schätzfehler größer sind als der Linearisierungsbereich der Messfunktionen. Eine Implementierung findet sich z. B. in der GNU-Scientific Library. Praktisch besonders gut geeignet ist die von Google entwickelte C++-Bibliothek Ceres [2]. Diese ermöglicht die Messfunktion als Code hinzuschreiben, bildet automatisch Ableitungen unterstützt die Formulierung eines Modells für komplexe Situationen und nutzt eine häufig vorhandene Struktur namens "Dünnbesetztheit" aus.

scipy

Eine nicht ganz so mächtige Python Alternative ist `scipy.optimize.least_squares`. Diese erlaubt, die Modell in Python zu definieren und verwendet numerische Ableitungen für den Levenbergh-Marquardt-Algorithmus. Da alle Modellfunktionen in Python ausgewertet werden und das für die numerischen Ableitungen vielfach ist die Software recht langsam und nur für kleine Probleme (ca. $n < 10, m < 200$) geeignet.

Vergleiche dazu auch "Anwendungen der Bildverarbeitung (14)". Wichtiger Tipp: Option `jac='3-point'` setzen!

Tensorflow Probability

Das Tensorflow Probability Paket [20] (Python) unterstützt Schätzaufgaben in probabilistischen Modellen. Eigentlich ist das genau das Thema der Veranstaltung, allerdings ist gerade quadratische Ausgleichsrechnung nicht direkt implementiert. Es gibt allerdings in der Bibliothke mehrere Optimierungsverfahren, von denen sich zum Beispiel BFGS

12.3. Nichtlineare Ausgleichsrechnung als Optimierung

(oder LBFGS für große Probleme) gut nutzen lässt, um direkt (12.42) zu lösen. Man profitiert von zwei sehr nützlichen Features von Tensorflow. Zum einen bildet Tensorflow automatisch die für den Optimierungsalgorithmus nötigen Ableitungen. Das spart Arbeit und Fehlersuche. Zum anderen compiliert Tensorflow die Python geschriebenen Modellfunktionen f_i , so dass sie viel schneller ausgewertet werden, als wenn man die gleichen Funktionen mit `scipy` verwendet. Das reduziert die Rechenzeit erheblich und macht auch die Verarbeitung größerer Probleme möglich.

13. Mehrdimensionaler Kalmanfilter

13.1. Kalmanfilter

In dieser Sektion wollen wir den n -dimensionalen Kalman Filter herleiten. Im Unterschied zur allgemeinen quadratischen Ausgleichsrechnung modelliert der Kalman Filter den Wechsel von Zuständen über die Zeit. Um eine neue Messung zu integrieren, muss er jeweils nur diese Messungen sowie die berechneten Erwartungswert und Kovarianz des Zustandes betrachten. Er muss nicht die ganze Vergangenheit aller Messungen durchlaufen.

Analog wie im 1-D KF ergibt sich der Dynamikschritt aus den Rechenregeln für Erwartungswerte und Kovarianzen die wir schon auf n-D verallgemeinert haben. Für den Messschritt benötigen wir ein Lemma, "das eine Messung hinter den Konditionierungsstrich bringt", d.h. das aus Erwartungswert und Kovarianz vor der Messung, Erwartungswert und Kovarianz nach der Messung ausrechnet.

Wir benötigen dafür das Blockmatrix-Inversionslemma 10.5, das hier noch einmal hingeschrieben ist.

$$\begin{pmatrix} P & Q \\ R & S \end{pmatrix}^{-1} = \begin{pmatrix} \tilde{P} & \tilde{Q} \\ \tilde{R} & \tilde{S} \end{pmatrix} \quad (13.1)$$

Mit:

$$\tilde{P} = (P - QS^{-1}R)^{-1} \quad \text{Schur-Komplement} \quad (13.2)$$

$$\tilde{Q} = -\tilde{P}(QS^{-1}) \quad (13.3)$$

$$\tilde{R} = -(S^{-1}R)\tilde{P} \quad (13.4)$$

$$\tilde{S} = S^{-1} + (S^{-1}R)\tilde{P}(QS^{-1}). \quad (13.5)$$

BM-Inv-Lemma
als Folie

Die n-D Version dieses Konditionierungslemmas folgt nun:

Lemma 13.1 (Konditionierungslemma). *Seien $X : \Omega \rightarrow \mathbb{R}^n, Z : \Omega \rightarrow \mathbb{R}^m$, so dass $\begin{pmatrix} X \\ Z \end{pmatrix}$ Gauß-verteilt, dann ist $X|Z = z$ Gaußsch mit*

$$E(X|Z = z) = E X + \text{Cov}(X, Z) \text{Cov}(Z)^{-1}(z - E Z) \quad (13.6)$$

$$\text{Cov}(X|Z = z) = \text{Cov} X - \text{Cov}(X, Z) \text{Cov}(Z)^{-1} \text{Cov}(X, Z)^T \quad (13.7)$$

Beweis. Reduktion auf $E X = E Z = 0$. Zuerst den Spezialfall

$$p(X = x | Z = z) = \frac{p(X = x \wedge Z = z)}{p(Z = z)} \quad (13.8)$$

$$\propto p\left(\begin{pmatrix} X \\ Z \end{pmatrix} = \begin{pmatrix} x \\ z \end{pmatrix}\right) \quad (13.9)$$

$$\propto \exp\left(-\frac{1}{2} \begin{pmatrix} x \\ z \end{pmatrix}^\top \begin{pmatrix} \text{Cov}(X) & \text{Cov}(X, Z) \\ \text{Cov}(X, Z)^\top & \text{Cov}(Z) \end{pmatrix}^{-1} \begin{pmatrix} x \\ z \end{pmatrix}\right) \quad (13.10)$$

Nach BM-Inversionslemma: mit $P = \text{Cov}(X)$, $Q = \text{Cov}(X, Z)$, $S = \text{Cov}(Z)$:

$$= \exp\left(-\frac{1}{2} \begin{pmatrix} x \\ z \end{pmatrix}^\top \begin{pmatrix} \tilde{P} & \tilde{Q} \\ \tilde{Q}^\top & \tilde{S} \end{pmatrix} \begin{pmatrix} x \\ z \end{pmatrix}\right) \quad (13.11)$$

$$\propto \exp\left(-\frac{1}{2} \left(x^\top \tilde{P} x + x^\top (2\tilde{Q}z)\right)\right) \quad (13.12)$$

also nach Lemma 12.1 Gaußsch mit $A = \tilde{P}$, $b = 2\tilde{Q}z$

$$E(X|Z = z) = -A^{-1}b/2 = -\tilde{P}^{-1}\tilde{Q}z = \tilde{P}^{-1}\tilde{P}QS^{-1}z \quad (13.13)$$

$$= QS^{-1}z = \text{Cov}(X, Z) \text{Cov}(Z)^{-1}z \quad (13.14)$$

$$\text{Cov}(X|Z = z) = A^{-1} = \tilde{P}^{-1} = P - QS^{-1}R \quad (13.15)$$

$$= \text{Cov}(X) - \text{Cov}(X, Z) \text{Cov}(Z)^{-1} \text{Cov}(X, Z)^\top \quad (13.16)$$

Allgemeinfall: Setze $X' = X - E X$, $Z' = Z - E Z$, $z' = z - E Z$

$$E(X|Z = z) = E(X'|Z' = z') + E X = E X + \text{Cov}(X, Z) \text{Cov}(Z)^{-1}(z - E Z) \quad (13.17)$$

$$\text{Cov}(X|Z = z) = \text{Cov}(X'|Z' = z') = \text{Cov}(X) - \text{Cov}(X, Z) \text{Cov}(Z)^{-1} \text{Cov}(X, Z)^\top \square$$

Q 32

13.1.1. Annahmen

- $X_t : \Omega \rightarrow \mathbb{R}^n$ Zustand zur Zeit t
- $U_t : \Omega \rightarrow \mathbb{R}^p$ Zustandsübergangsmessung $t - 1 \rightarrow t$
- $Z_t : \Omega \rightarrow \mathbb{R}^m$ Messung

1. $X_0 \sim \mathcal{N}(\mu_0, \Sigma_0)$

2. $X_t = A_t X_{t-1} + B_t U_t + g_t + \varepsilon_t$; $A_t \in \mathbb{R}^{n \times n}$, $B_t \in \mathbb{R}^{n \times p}$, $g_t \in \mathbb{R}^n$, $\varepsilon_t \sim \mathcal{N}(0, \Sigma_{\varepsilon t})$

3. $Z_t = C_t X_t + h_t + \delta_t$; $C_t \in \mathbb{R}^{m \times n}$, $h_t \in \mathbb{R}^m$, $\delta_t \sim \mathcal{N}(0, \Sigma_{\delta t})$

4. Alle $X_0, \varepsilon_t, \delta_t$ unabhängig

Man sieht, dass die Struktur analog zu den Annahmen des eindimensionalen KF in Kapitel 8.2.1 ist. Jede Größe (X_{t-1}, X_t, U_t) geht mit einem gewissen Faktor (A_t, B_t, C_t) in die jeweilige Formel für X_t bzw. Z_t ein; es gibt konstante Summanden (g_t, h_t) und Zufallsvariablen für das Rauschen die gaußverteilt mit Erwartungswert 0 und bekannter

13. Mehrdimensionaler Kalmanfilter

Kovarianz sind. Im 1D-Fall waren dies alles Zahlen. Die Faktoren A_t, B_t, C_t werden zu Matrizen, weil sie angeben, mit welchem Faktor jede Komponente von X_{t-1}, X_t bzw. U_t in jede Komponente von X_t bzw. Z_t geht. Die konstanten Summanden g_t und h_t werden zu Vektoren, weil sie für jede Komponente einen Wert angeben. Die Rauschzufallsvariablen ϵ_t, δ_t werden vektoriell, beschrieben durch eine symmetrische Kovarianzmatrix $(\Sigma_{\epsilon t}, \Sigma_{\delta t})$.

Das Geschehen in diesem Modell spielt sich in drei "Räumen" ab, dem Zustandsraum (X_t), dem Messraum (Z_t) und dem Raum der Zustandsübergangs- (ZÜ) Messungen (U_t). Die Matrizen definieren lineare Abbildungen zwischen diesen Räumen. Es ist hilfreich, sich zu vergegenwärtigen, welche von wo nach wo abbildet:

Matrix (bzw. Vektor)	von (Spalten)	nach (Zeilen)
A_t	Zustand	Zustand
B_t	ZÜ-Messung	Zustand
C_t	Zustand	Messung
Σ_t	Zustand	Zustand
$\Sigma_{\epsilon t}$	Zustand	Zustand
$\Sigma_{\delta t}$	Messung	Messung
g_t		Zustand
h_t		Messung

Q 33

13.1.2. Algorithmus

Variablen: $\mu_t \in \mathbb{R}^n, \Sigma_t \in \mathbb{R}^{n \times n}$ (13.18)

Initialisierung: $\mu_0 = E(X_0), \Sigma_0 = \text{Cov}(X_0)$ (13.19)

Dynamik($U_t = u_t$): $\bar{\mu}_t = A_t \mu_{t-1} + B_t u_t + g_t$ (13.20)

$$\bar{\Sigma}_t = A_t \Sigma_{t-1} A_t^\top + \Sigma_{\epsilon t} \quad (13.21)$$

Messung($Z_t = z_t$): $K_t = \bar{\Sigma}_t C_t^\top (C_t \bar{\Sigma}_t C_t^\top + \Sigma_{\delta t})^{-1}$ (13.22)

$$\mu_t = \bar{\mu}_t + K_t (z_t - (C_t \bar{\mu}_t + h_t)) \quad (13.23)$$

$$\Sigma_t = \bar{\Sigma}_t - K_t C_t \bar{\Sigma}_t \quad (13.24)$$

Q 34

13.2. Beweis Kalmanfilter

Um die Ausdrücke im Beweis nicht zu lang werden zu lassen, benutzen wir wieder die verkürzte Schreibweise $|z_t$ anstelle von $|Z_t = z_t$, usw.

Satz 23. Nach Ausführung der KF Formel ist

$$X_t | u_{1..t}, z_{1..t-1} \sim \mathcal{N}(\bar{\mu}_t, \bar{\Sigma}_t) \quad \forall t > 0 \quad (13.25)$$

$$X_t | u_{1..t}, z_{1..t} \sim \mathcal{N}(\mu_t, \Sigma_t) \quad \forall t \geq 0 \quad (13.26)$$

Der Beweis besteht im Wesentlichen daraus, nachzurechnen, dass die Kalman Filter Formeln „richtig“ sind. Für den Dynamikschritt ist das eine unmittelbare Anwendung der Rechenregeln. Für den Messschritt ergibt sich die Aussage aus dem Konditionierungslemma.

Beweis. Beweis durch Induktion nach t .

I.A. $t = 0$ aus Annahme (1)

I.S. $t - 1 \rightarrow t$

Dynamikschritt: Propagiere X_{t-1} durch Annahme (2) um X_t zu erhalten.

$$X_{t-1}|u_{1..t-1}, z_{1..t-1} \sim \mathcal{N}(\mu_{t-1}, \Sigma_{t-1}) \quad (13.27)$$

$$A_t X_{t-1}|u_{1..t-1}, z_{1..t-1} \sim \mathcal{N}\left(A_t \mu_{t-1}, A_t \Sigma_{t-1} A_t^\top\right) \quad (13.28)$$

$$\varepsilon_t \sim \mathcal{N}(0, \Sigma_{\varepsilon t}) \quad (13.29)$$

$$X_t = A_t X_{t-1} + B_t U_t + g_t + \varepsilon_t|u_{1..t}, z_{1..t-1} \quad (13.30)$$

$$\sim \mathcal{N}\left(A_t \mu_{t-1} + B_t u_t + g_t, A_t \Sigma_{t-1} A_t^\top + \Sigma_{\varepsilon t}\right) \quad (13.31)$$

$$\sim \mathcal{N}(\bar{\mu}_t, \bar{\Sigma}_t) \quad (13.32)$$

Messschritt: Wende das Konditionierungslemma an und berechne die benötigten Kovarianzen nach den Rechenregeln aus Annahme (3): Alles ist konditioniert unter $|u_{1..t}, z_{1..t-1}$, was wir mit $|*$ abkürzen, mit $Z = Z_t$, $z = z_t$, $X = X_t$. Damit erlaubt das Konditionierungslemma eine weitere Messung z_t hinter den Konditionierungsstrich zu bringen. Es folgt

$$X_t|u_{1..t}, z_{1..t} = X_t|*, z_t \text{ Gaußsch} \quad (13.33)$$

Um die Terme für Erwartungswert und Kovarianz zusammensetzen, berechnen wir zuerst die im Lemma auftretenden Kovarianzausdrücke aus den im Algorithmus auftretenden Variablen.

$$\mathbb{E}(Z_t|*) = \mathbb{E}(C_t X_t + h_t + \delta_t|*) \quad (13.34)$$

$$= C_t \bar{\mu}_t + h_t \quad (13.35)$$

$$\text{Cov}(Z_t|*) = \text{Cov}(C_t X_t + h_t + \delta_t|*) \quad (13.36)$$

$$= \text{Cov}(C_t X_t|*) + \text{Cov}(\delta_t|*) \quad (13.37)$$

$$= C_t \text{Cov}(X_t|*) C_t^\top + \Sigma_{\delta t} \quad (13.38)$$

$$= C_t \bar{\Sigma}_t C_t^\top + \Sigma_{\delta t} \quad (13.39)$$

$$\text{Cov}(X_t, Z_t|*) = \text{Cov}(X_t, C_t X_t + h_t + \delta_t|*) \quad (13.40)$$

$$= \text{Cov}(X_t, C_t X_t|*) \quad (13.41)$$

$$= \text{Cov}(X_t, X_t|*) C_t^\top \quad (13.42)$$

13. Mehrdimensionaler Kalmanfilter

$$= \bar{\Sigma}_t C_t^\top \quad (13.43)$$

$$\text{Cov}(X_t, Z_t|*) \text{Cov}(Z_t|*)^{-1} = \bar{\Sigma}_t C_t^\top (C_t \bar{\Sigma}_t C_t^\top + \Sigma_{\delta t})^{-1} \quad (13.44)$$

$$= K_t \quad (13.45)$$

Nun setzen wir nach dem Konditionierungslemma zusammen:

$$\text{E}(X_t|u_{1..t}, z_{1..t}) = \text{E}(X_t|*, z_t) \quad (13.46)$$

$$= \text{E}(X_t|*) + \text{Cov}(X_t, Z_t|*) \text{Cov}(Z_t|*)^{-1} (z_t - \text{E}(Z_t|*)) \quad (13.47)$$

$$= \bar{\mu}_t + K_t (z_t - (C_t \bar{\mu}_t + h_t)) = \mu_t \quad (13.48)$$

$$\text{Cov}(X_t|u_{1..t}, z_{1..t}) = \text{Cov}(X_t|*, z_t) \quad (13.49)$$

$$= \text{Cov}(X_t|*) - \text{Cov}(X_t, Z_t|*) \text{Cov}(Z_t|*)^{-1} \text{Cov}(X_t, Z_t|*)^\top \quad (13.50)$$

$$= \bar{\Sigma}_t - K_t C_t \bar{\Sigma}_t = \Sigma_t \quad \square$$

A.19

13.3. Konsistenz des n-D Kalman Filters

Die Konsistenzkriterien ändern sich im Vergleich zum eindimensionalen nur geringfügig:

1. Erwartungstreue Schätzung des wahren Wertes $\text{E}(X - \hat{x}) = 0$
2. Übereinstimmung von geschätzter und wahrer Varianz $\text{E}(\Delta(X, \hat{X})^2) = n$
3. Innovations ($z - \text{E}(Z|\bar{\mu}_t)$) ebenfalls erwartungstreu und übereinstimmend
4. Innovations sind unabhängig und Gaußverteilt

Lediglich der Sollwert von $\text{E}(\Delta(X, \hat{X})^2)$ ist nicht mehr 1 sondern n (Was im 1D Fall wieder 1 ist). Als optimaler Filter ist der n-D Kalman Filter immer konsistent, vorausgesetzt das System ist korrekt modelliert. Zum testen, nutzt man die selben Verfahren wie beim 1D Kalman Filter (Siehe 8.4).

14. Mehrdimensionaler (Extended / Unscented) Kalmanfilter

14.1. Nichtlineare Rechenregeln für Erwartungswert und Kovarianz

Sind Mess- oder Dynamikmodell nichtlinear, müssen Erwartungswertvektor und Kovarianzmatrix durch eine nichtlineare Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ propagiert werden. Hier gibt es wie im 1D Fall keine exakte Lösung, aber eine Näherung basierend auf der Linearisierung am Erwartungswert (10.53):

$$X : \Omega \rightarrow \mathbb{R}^n \quad f : \mathbb{R}^n \rightarrow \mathbb{R}^m \quad (14.1)$$

$$f(x) \approx \underbrace{f(EX)}_{\in \mathbb{R}^m} + \underbrace{f'(EX)}_{\in \mathbb{R}^{m \times n}} \underbrace{(x - EX)}_{\in \mathbb{R}^n} \quad (14.2)$$

$$\mathbb{E}(f(X)) \approx \mathbb{E}(f(EX) + f'(EX)(X - EX)) = f(EX) \quad (14.3)$$

$$\text{Cov}(f(X)) \approx \text{Cov}(f(EX) + f'(EX)(X - EX)) \quad (14.4)$$

$$= f'(EX) \text{Cov}(X - EX) f'(EX)^T \quad (14.5)$$

$$= f'(EX) \text{Cov}(X) f'(EX)^T \quad (14.6)$$

$$\text{Cov}(X, f(X)) \approx \text{Cov}(X, f(EX) + f'(EX)(X - EX)) \quad (14.7)$$

$$= \text{Cov}(X, f'(EX)X) = \text{Cov}(X) f'(EX)^T \quad (14.8)$$

14.2. Extended-Kalman-Filter

Der Extended Kalman Filter ersetzt wie im eindimensionalen die linearen Funktionen für Zustandsübergang und Messung durch differenzierbare Funktionen. Diese Funktionen werden dann am jeweils aktuellen Schätzwert linearisiert. Der EKF iteriert nicht und man kann deshalb über das Ergebnis keine Aussage machen, außer dass es eine Näherung an das exakte Ergebnis ist. Der Algorithmus ergibt sich, in dem man beim Propagieren von Erwartungswert und Kovarianz durchs Dynamik bzw. Messmodell nicht die exakten Rechenregeln wie im linearen Fall (KF) benutzt, sondern die linearisierten Näherungsregeln aus (14.3) und (14.6). Diese verwenden den Funktionswert und dessen Ableitung (Jacobi-Matrix) am Erwartungswert. Es ergeben sich ähnliche Formeln wie im linearen Fall, bei denen die Matrizen A_t , B_t und C_t nicht vorgegeben werden, sondern als Jacobi-Matrix von g nach x und u bzw. h nach x bestimmt werden.

14. Mehrdimensionaler (Extended / Unscented) Kalmanfilter

14.2.1. Annahmen

1. X_0 ist ungefähr $\mathcal{N}(\mu_0, \Sigma_0)$
2. $X_t = g_t(X_{t-1}, U_t) + \varepsilon_t$; $g_t : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}^n$, diffbar, ε_t ist ungefähr $\mathcal{N}(0, \Sigma_{\varepsilon t})$
3. $Z_t = h_t(X_t) + \delta_t$; $h_t : \mathbb{R}^n \rightarrow \mathbb{R}^m$, diffbar, δ_t ist ungefähr $\mathcal{N}(0, \Sigma_{\delta t})$
4. Alle ε_t, δ_t sind unabhängig.

Q 35

14.2.2. Algorithmus

$$\text{Dynamik}(U_t = u_t): \quad \bar{\mu}_t = g(\mu_{t-1}, u_t) \quad (14.9)$$

$$\bar{\Sigma}_t = A_t \Sigma_{t-1} A_t^\top + \Sigma_{\varepsilon t}, \quad A_t = g^{(x)}(\mu_{t-1}, u) \quad (14.10)$$

$$\text{Messung}(Z_t = z_t): \quad C_t = h'(\bar{\mu}_t) \quad (14.11)$$

$$K_t = \bar{\Sigma}_t C_t^\top (C_t \bar{\Sigma}_t C_t^\top + \Sigma_{\delta t})^{-1} \quad (14.12)$$

$$\mu_t = \bar{\mu}_t + K_t(z_t - h(\bar{\mu}_t)) \quad (14.13)$$

$$\Sigma_t = \bar{\Sigma}_t - K_t C_t \bar{\Sigma}_t \quad (14.14)$$

A.20

14.3. Variante des EKF mit nicht additivem Rauschen

Die bisherigen Annahmen gehen davon aus, dass exakt das bekannte u in die Dynamikfunktion g eingeht und zusätzliches Rauschen im Zustand addiert wird. Das macht Sinn, z. B. für die Position eines Fahrzeugs und Störungen wie Wind, Wellen, oder ähnliches. Oft will man aber auch ausdrücken, dass die Zustandsübergangsmessung u fehlerhaft ist analog zu Annahme (3) für Z . Zudem kann man physikalische Modellparameter (z.B. Federkonstanten, Reifenumfang) oder zusätzliche Einflussgrößen (z.B. Seitenwind) nie exakt bestimmen oder diese sind schwanken ebenfalls (z.B. mit Temperatur, Luftdruck). Aus dieser Unsicherheit ergibt sich dann eine Unsicherheit im Zustand, aber wir benötigen eine Formel um diese auszurechnen. Das folgende Modell trägt dem Rechnung und erlaubt dem Rauschvektor ε_t an beliebigen Stellen des Modells anzusetzen. Dementsprechend darf der Rauschvektor auch eine beliebige Dimension q haben.

14.3.1. Annahmen

- $X_t : \Omega \mapsto \mathbb{R}^n$ Zustand zur Zeit t
- $U_t : \Omega \mapsto \mathbb{R}^p$ Übergangsmessung von $t - 1 \mapsto t$
- $Z_t : \Omega \mapsto \mathbb{R}^m$ Messung zur Zeit t

$$1. X_0 \approx \mathcal{N}(\mu_0, \Sigma_0)$$

$$2. X_t = g_t(X_{t-1}, U_t, \varepsilon_t); g \text{ diffbar}; \varepsilon_t : \Omega \rightarrow \mathbb{R}^q; \varepsilon_t \approx \mathcal{N}(0, \Sigma_{\varepsilon t}); \Sigma_{\varepsilon t} \in \mathbb{R}^{q \times q}, \text{SPSD}$$

$$3. Z_t = h(X_t) + \delta_t; \delta_t \approx \mathcal{N}(0, \Sigma_{\delta t}), h \text{ diffbar}$$

4. Alle ε_t, δ_t unabhängig

Man mag sich fragen, wozu man das Rauschen als Eingabe in die Modellfunktion sieht, schließlich ist es ja unbekannt. In der Tat wird die Funktion vom späteren Algorithmus immer am Erwartungswert ausgewertet, also für das Rauschen bei $\epsilon = 0$. Allerdings ist die Funktion ja nicht nur zum praktischen Auswerten da, sondern ein formales Modell und die Ableitung nach ϵ erlaubt dadurch systematisch zu ermitteln, wie stark sich das Rauschen auf den Zustand auswirkt bei komplizierteren Zusammenhängen als einfachem addieren. Diese Möglichkeit motiviert den Ansatz.

Diese Annahmen sind eine Verallgemeinerung der bisherigen. Mit

$$g(x, u, \epsilon) = g^{\text{alt}}(x, u) + \epsilon \quad (14.15)$$

modelliert man das bekannte additive Rauschen im Zustand. Mit

$$g(x, u, \epsilon) = g^{\text{alt}}(x, u + \epsilon) \quad (14.16)$$

modelliert man Rauschen in der Zustandsübergangsmessung. Um Rauschen in Modellparametern zu modellieren muss man die Komponenten von ϵ auf die Modellparameter in den Domänenformeln des Modells addieren.

Es bleibt allerdings die Einschränkung, dass die Rauschvektoren verschiedener Zeitpunkte unabhängig angenommen werden. Besonders für schwankende Modellparameter ist das oft nur eine sehr grobe Näherung, alternativ können die Modellparameter in den Zustand aufgenommen werden.

Um mehrere dieser Effekte gleichzeitig zu modellieren, muss man den Rauschvektor ϵ größer machen und unterschiedliche Teile an den verschiedenen Stellen verwenden.

14.3.2. Algorithmus

Im Algorithmus ändert sich nur der Dynamikschritt. Es kommt ein Term hinzu, der die Unsicherheit addiert, die sich aus $\Sigma_{\epsilon t}$ durch g ergibt. Bei der Berechnung der Modelle wird der Rauschvektor auf 0 gesetzt.

$$\text{Dynamik}(U_t = u_t): \quad \bar{\mu}_t = g(\mu_{t-1}, u_t, 0) \quad (14.17)$$

$$\bar{\Sigma}_t = A_t \Sigma_{t-1} A_t^\top + B_t \Sigma_{\epsilon t} B_t^\top \quad (14.18)$$

$$A_t = g^{(x)}(\mu_{t-1}, u_t, 0), \quad B_t = g^{(\epsilon)}(\mu_{t-1}, u_t, 0) \quad (14.19)$$

$$\text{Messung}(Z_t = z_t): \quad C_t = h^{(x)}(\bar{\mu}_t) \quad (14.20)$$

$$K_t = \bar{\Sigma}_t C_t^\top (C_t \bar{\Sigma}_t C_t^\top + \Sigma_{\delta t})^{-1} \quad (14.21)$$

$$\mu_t = \bar{\mu}_t + K_t (z_t - h(\bar{\mu}_t)) \quad (14.22)$$

$$\Sigma_t = \bar{\Sigma}_t - K_t C_t \bar{\Sigma}_t \quad (14.23)$$

Vergleichen wir den Algorithmus mit dem gewöhnlichen EKF. Wenn g das Rauschen ϵ einfach addiert, ist $g^{(\epsilon)}(\dots) = I$ und $\Sigma_{\epsilon t}$ wird wie im gewöhnlichen EKF addiert. Insofern sind beide Algorithmen kompatibel. Ein weiterer häufiger Fall ist, dass ϵ als

Messrauschen zu u addiert wird. Dann ist $g^{(\epsilon)}(\dots) = g^{(u)}(\dots)$, so dass die Ableitung des neuen Zustands nach u bestimmt, wie stark sich das Messrauschen den Zustand beeinflusst. Auch das macht Sinn.

14.4. Herleitung EKF mit nicht additivem Rauschen

Es folgt eine Herleitung des EKF mit nicht additivem Rauschen. In der Struktur folgt sie den Beweisen für den KF und EKF in 1-D: Es werden mit den Rechenregeln für Erwartungswert und Kovarianz und den Annahmen aus $E(X_{t-1})$ und $\text{Cov}(X_{t-1})$ die Größen $E(X_t)$, $E(Z_t)$, $\text{Cov}(X_t)$, $\text{Cov}(Z_t)$ und $\text{Cov}(X_t, Z_t)$ ausgerechnet. Zusätzlich wird im Messschritt das Konditionierungslemma angewendet.

Im Fall des EKF bedarf es dabei den Rechenregeln aus Sektion 14.1 für Propagation durch nichtlineare Funktionen. Weil diese Formeln nur näherungsweise gelten und keine "harte" Aussage machen, ist die ganze Herleitung des EKF kein Beweis sondern nur eine "Argumentation dass der Algorithmus eine sinnvolle Näherung ist". Hergeleitet wird die Aussage, dass nach Ausführung der KF Formel gilt:

$$X_t|u_{1..t}, z_{1..t-1} \approx \mathcal{N}(\bar{\mu}_t, \bar{\Sigma}_t) \forall t > 0 \quad (14.24)$$

$$X_t|u_{1..t}, z_{1..t} \approx \mathcal{N}(\mu_{t-1}, \Sigma_t) \forall t \geq 0 \quad (14.25)$$

Wie üblich ergibt sich der Induktionsanfang aus Annahme 1. Und wir zeigen den Induktionsschritt $t - 1 \rightarrow t$. Alle Ausdrücke stehen unter der Vorbedingung $* = u_{1..t}, z_{1..t-1}$.

$$E(X_t|*) = E(g_t(X_{t-1}, U_t, \epsilon_t)) \quad (14.26)$$

$$= E \left(g_t \left(\begin{array}{c} X_{t-1} \\ u_t \\ \epsilon_t \end{array} \middle| * \right) \right) \approx g \left(\begin{array}{c} E(X_{t-1}|*) \\ u_t \\ E \epsilon_t \end{array} \right) = g(\mu_{t-1}, u_t, 0) = \bar{\mu}_t \quad (14.27)$$

$$\text{Cov}(X_t|*) = \text{Cov}(g_t(X_{t-1}, U_t, \epsilon_t)|*) \quad (14.28)$$

$$= \text{Cov} \left(g_t \left(\begin{array}{c} X_{t-1} \\ u_t \\ \epsilon_t \end{array} \middle| * \right) \right) \quad (14.29)$$

$$\approx g'_t \left(\begin{array}{c} E X_{t-1} \\ u_t \\ E \epsilon_t \end{array} \right) \text{Cov} \left(\left(\begin{array}{c} X_t \\ u_t \\ \epsilon_t \end{array} \middle| * \right) \right) g'_t \left(\begin{array}{c} E X_{t-1} \\ u_t \\ E \epsilon_t \end{array} \right)^T \quad (14.30)$$

$$= g'_t \left(\begin{array}{c} \mu_{t-1} \\ u_t \\ 0 \end{array} \right) \left(\begin{array}{cc} \Sigma_{t-1} & \\ & 0 \\ & & \Sigma_{\epsilon t} \end{array} \right) g'_t \left(\begin{array}{c} \mu_{t-1} \\ u_t \\ 0 \end{array} \right)^T \quad (14.31)$$

$$g'_t \left(\begin{array}{c} \mu_{t-1} \\ u_t \\ 0 \end{array} \right) = (g^{(x)}(\mu_{t-1}, u_t, 0) \quad g^{(u)}(\mu_{t-1}, u_t, 0) \quad g^{(\epsilon)}(\mu_{t-1}, u_t, 0)) \\ = g^{(x)}(\mu_{t-1}, u_t) \Sigma_{t-1} g^{(x)}(\mu_{t-1}, u_t)^T + g^{(\epsilon)}(\mu_{t-1}, u_t) \Sigma_{\epsilon t} g^{(\epsilon)}(\mu_{t-1}, u_t)^T \quad (14.32)$$

$$= A_t \Sigma_{t-1} A_t^T + B_t \Sigma_{\epsilon t} B_t^T = \bar{\Sigma}_t \quad (14.33)$$

$$E(Z_t|*) = E(h_t(X_t) + \delta_t|*) \approx h_t(E X) \quad (14.34)$$

$$\text{Cov}(Z_t|*) = \text{Cov}(h_t(X_t) + \delta_t|*) = \text{Cov}(h_t(X_t)|*) + \text{Cov}(\delta_t) \quad (14.35)$$

$$= h_t'(E X) \text{Cov}(X|*) h_t'(E X)^T + \text{Cov}(\delta_t) = C_t \bar{\Sigma}_t C_t^T + \Sigma_{\delta t} \quad (14.36)$$

$$\text{Cov}(X_t, Z_t|*) = \text{Cov}(X_t, h_t(X_t) + \delta_t|*) = \text{Cov}(X_t, h_t(X_t)|*) + \text{Cov}(X_t, \delta_t|*) \quad (14.37)$$

$$= \text{Cov}(X_t|*) h_t'(E X)^T = \Sigma_t C_t^T \quad (14.38)$$

Diese Formeln sind die selben wie beim KF, bis auf dass C_t statt vorgegeben als Jakobimatrix von h_t bestimmt wird. Der Rest der Herleitung ist die Anwendung des Konditionierungslemmas und läuft analog.

A.21

A.23

A.24

14.5. Sigmapunktpropagation

Analog zum 1D-Fall betrachten wir hier die Möglichkeit Erwartungswert und Kovarianz durch eine nichtlineare Funktion zu propagieren, indem man die Funktion auf sogenannten Sigmapunkten als Stellvertreter für die ganze Verteilung auswertet und aus der Statistik der Ergebnisse Erwartungswert und Kovarianz des Ergebnisses schätzt. Die Idee stammt aus [1].

14.5.1. Berechnung der Sigmapunkte

In 1D sind die Sigmapunkte einer Verteilung $\mathcal{N}(\mu, \sigma^2)$ die Punkte $(\mu, \mu, \mu \pm \sqrt{2}\sigma)$, d.h. der Erwartungswert und ein Vielfaches der Standardabweichung in beide Richtungen. Wie lässt sich diese Idee nun auf mehrere Dimensionen generalisieren? Wir wollen gedanklich zwei Schritten machen, indem wir zuerst sinnvolle Sigmapunkte für eine n-D Standardnormalverteilung definieren und dann das Konzept auf allgemeine Normalverteilungen erweitern.

Für die n-D Standard-Gaußglocke $\mathcal{N}(0, I)$ erscheint es sinnvoll, den Erwartungswert als Sigmapunkt beizubehalten und in jeder Dimension ein Vielfaches der Standardabweichung in beide Richtungen zu gehen.

$$\mathcal{X} = \left(0 \quad 0 \quad \lambda e_i \Big|_{i=1}^n \quad -\lambda e_i \Big|_{i=1}^n \right). \quad (14.39)$$

Die Schreibweise bedeutet, dass \mathcal{X} eine Liste von Vektoren ist. Die einzelnen Sigmapunkte sind dann $\mathcal{X}^{[i]}$, $i = 1 \dots |\mathcal{X}|$.

Die Frage ist nun, welches Vielfaches λ ? Betrachten wir die Sigmapunkte aus der Sicht einer festen Dimension, z.B. der ersten. Im 1-D Fall hatten wir zwei Sigmapunkte ohne

14. Mehrdimensionaler (Extended / Unscented) Kalmanfilter

Abweichung und zwei Sigmapunkte mit Abweichung. Der Faktor $\sqrt{2}$ führte dazu, dass im Mittel die quadrierte Abweichung, wie gewünscht 1 war. Im n -D Fall gibt es zwei Sigmapunkte mit Abweichung in die feste Dimension. Dem stehen aber $2n$ Sigmapunkte ohne Abweichung gegenüber. Zwei, weil sie überhaupt keine Abweichung haben und der Rest $2(n-1)$ weil die Abweichung in eine senkrechte Richtung zur betrachteten Dimension geht. Dementsprechend muss das Vielfache $\lambda = \sqrt{1+n}$ sein, damit im Mittel die quadrierte Abweichung 1 ist.¹

$$\mathcal{X} = \begin{pmatrix} 0 & 0 & \sqrt{1+ne_i} \Big|_{i=1}^n & -\sqrt{1+ne_i} \Big|_{i=1}^n \end{pmatrix}. \quad (14.40)$$

Für die Generalisierung auf allgemeine Gaußlocken erinnern wir uns an Satz 19 der sagt, dass man $\mathcal{N}(\mu, \Sigma)$ aus $\mathcal{N}(0, I)$ erhält, indem man die lineare Abbildung $x \mapsto \mu + Lx$ anwendet, wobei L die Choleskyzerlegung $LL^T = \Sigma$ von Σ ist.

Dementsprechend erhält man sinnvolle Sigmapunkte von $\mathcal{N}(\mu, \Sigma)$, indem man auf (14.40) die Sigmapunkte von $\mathcal{N}(0, I)$ die selbe lineare Abbildung anwendet:

$$\mathcal{X} = \mu + L \begin{pmatrix} 0 & 0 & \sqrt{1+ne_i} \Big|_{i=1}^n & -\sqrt{1+ne_i} \Big|_{i=1}^n \end{pmatrix} \quad (14.41)$$

$$= \begin{pmatrix} \mu & \mu & \mu + \sqrt{1+nLe_i} \Big|_{i=1}^n & \mu - \sqrt{1+nLe_i} \Big|_{i=1}^n \end{pmatrix} \quad (14.42)$$

$$= \begin{pmatrix} \mu & \mu & \mu + \sqrt{1+nL_{\bullet i}} \Big|_{i=1}^n & \mu - \sqrt{1+nL_{\bullet i}} \Big|_{i=1}^n \end{pmatrix} \quad (14.43)$$

Die Choleskyzerlegung von Σ ist in gewisser Hinsicht ein Pendant zur Quadratwurzel in 1-D. Deshalb schreibt man [50] die Matrix L auch symbolisch als $L = \sqrt{\Sigma}$. Mit dieser nicht ganz sauberen Notation kann man kurz schreiben:

$$\mathcal{X} = \begin{pmatrix} \mu & \mu & \mu + \sqrt{1+n\sqrt{\Sigma_{\bullet i}}} \Big|_{i=1}^n & \mu - \sqrt{1+n\sqrt{\Sigma_{\bullet i}}} \Big|_{i=1}^n \end{pmatrix} \quad (14.44)$$

$$= \begin{pmatrix} \mu & \mu & \mu + \sqrt{(n+1)\Sigma} & \mu - \sqrt{(n+1)\Sigma} \end{pmatrix}. \quad (14.45)$$

$$= \begin{pmatrix} 2 \times & \mu \pm \sqrt{(n+1)\Sigma} \end{pmatrix} \quad (14.46)$$

Q 36

Es sollte noch erwähnt werden, dass es verschiedene L mit $LL^T = \Sigma$ gibt. Die Choleskyzerlegung ist nur eine davon. Und diese L ergeben verschiedene aber gleichwertige Sigmapunkte.

Auch „sinnvoll“ gewählte Sigmapunkte sind nicht immer optimal. Um in n Dimensionen ebenfalls eine 3. Ordnung korrekte Approximation von Erwartungswert und Kovarianz zu erhalten muss die Kovarianz bei der Bildung der Sigmapunkte mit dem Faktor $n+1$ skaliert werden.

¹[1] definiert $\mathcal{X} = (\pm\sqrt{ne_i} \Big|_{i=1}^n)$ Dies hat den Nachteil, dass in 1D mit nur 2 Sigmapunkten die Nichtlinearität der Funktion von den Sigmapunkten nicht wahrgenommen wird. Z.B. hat mit dieser Definition in A.11 auch die Sigmapunktpropagation eine a-posteriori Varianz von 0. Deshalb haben wir im Skript die zwei Sigmapunkte bei μ hinzugefügt und die Skalierung entsprechend auf $\sqrt{n+1}$ angepasst. Man kann diese Version auch aus der Originalversion herleiten, indem man eine Dummydimension zufügt, die von der betrachteten Funktion ignoriert wird. Die zwei Sigmapunkte die diese Dimension aufspannen wirken dann effektiv wie das doppelte μ und n erhöht sich um 1.

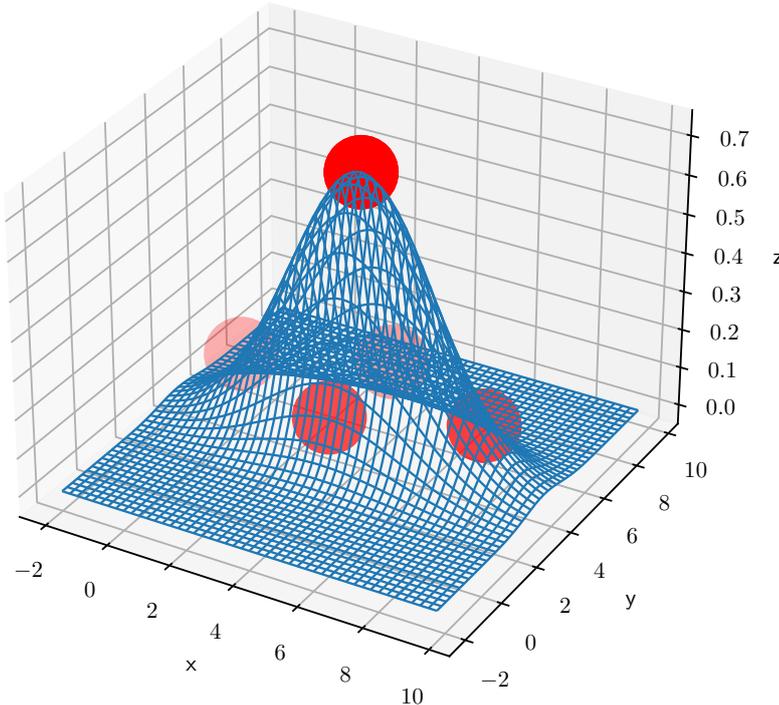


Abbildung 14.1.: Sigma-Punkte in 2-D

14.5.2. Sigmapunktpropagation

Die Sigma-Punkte \mathcal{X} die die Verteilung von X repräsentieren werden nun durch die entsprechende Modellfunktion f propagiert (Dynamikfunktion g oder Messfunktion h). Dies notieren wir als

$$\bar{\mathcal{X}} = f(\mathcal{X}) \quad \Leftrightarrow \quad \bar{\mathcal{X}}^{[i]} = f(\mathcal{X}^{[i]}) \quad (14.47)$$

14.5.3. Erwartungswert- und Kovarianzberechnung

Aus den transformierten Sigmapunkten $\bar{\mathcal{X}}$ werden dann der Erwartungswert $E(f(X))$, die Kovarianz $\text{Cov}(f(X))$ und die Kovarianz $\text{Cov}(f(X), X)$ näherungsweise zurückgewonnen.

$$E(f(X)) \approx E(f(\mathcal{X})) = \bar{\mu} = \frac{1}{|\mathcal{X}|} \sum_k \bar{\mathcal{X}}^{[k]} \quad (14.48)$$

$$\text{Cov}(f(X)) \approx \text{Cov}(f(\mathcal{X})) = \bar{\Sigma} = \frac{1}{|\mathcal{X}|} \sum_k (\bar{\mathcal{X}}^{[k]} - \bar{\mu})(\bar{\mathcal{X}}^{[k]} - \bar{\mu})^T \quad (14.49)$$

$$\text{Cov}(f(X), X) \approx \text{Cov}(f(\mathcal{X}), \mathcal{X}) = \bar{\Sigma}_{fx} = \frac{1}{|\mathcal{X}|} \sum_k (\bar{\mathcal{X}}^{[k]} - \bar{\mu})(\mathcal{X}^{[k]} - \mu)^T \quad (14.50)$$

14. Mehrdimensionaler (Extended / Unscented) Kalmanfilter

Gleichung (14.48) ist einfach der Mittelwert der Sigmapunkte.

Gleichungen (14.49) und (14.50) entsprechen der Definition $\text{Cov}(f(X)) = E((f(X) - E(f(X)))(f(X) - E(f(X)))^T)$ wobei der $E(\dots)$ als Mittelwert der Sigmapunkte approximiert wird.

Anders als beim EKF, wo die Funktion nur in einer Umgebung des Mittelwertes ausgewertet wird, wird beim UKF die Funktion also in dem Bereich, wo der größte Anteil der Verteilung liegt, ausgewertet. Dadurch liegt die Berechnung mit Sigma-Punkt-Propagation meist näher am exakten Wert als die Berechnung über die Jacobi-Matrix. Außerdem werden keine Ableitungen benötigt. Im Ausgleich dafür, wertet der UKF die Modelle häufiger aus und ist daher etwas langsamer. Der Effekt ist stärker mit höherer Dimension des Zustandes.

14.5.4. Beweisüberblick

Wir wollen nun zeigen, dass auch in n -D die Sigmapunktpropagation eine 3. Ordnung korrekte Näherung an den Erwartungswert bzw. die Kovarianzmatrix bildet.

Der Beweis ist wie in 1-D strukturiert, wird durch die mehreren Dimensionen aber etwas komplizierter. Wir zeigen zuerst, dass die Sigmapunktpropagation für Polynome vom Grad 3 den exakten Erwartungswert liefert. Wegen Linearität sowohl vom Erwartungswert als auch von der Sigmapunktpropagation reicht es wieder aus, die auf μ zentrierten Monome vom Grad 0 bis 3 zu betrachten. Da es sich um ein Polynom von n Variablen handelt, können die Monome aber unterschiedliche Komponenten verknüpfen. Z.B. bei Dimension 2 sind zu betrachtende Monome der Grade 0-3 folgende $1; x - \mu_x, y - \mu_y; (x - \mu_x)^2, (x - \mu_x)(y - \mu_y), (y - \mu_y)^2; (x - \mu_x)^3, (x - \mu_x)^2(y - \mu_y), (x - \mu_x)(y - \mu_y)^2, \text{ und } (y - \mu_y)^3$. Wieder wird Grad 0 offensichtlich sein und die Grade 1 und 3 aus Antisymmetrie folgen. In Grad 2 steckt wieder der Kern, insbesondere der richtige Faktor λ .

Wir erweitern dann die Aussage auf den Allgemeinfall differenzierbarer Funktionen, indem wir die Funktion als Polynom vom Grad 3 plus Restglied $o(\|x - \mu\|^3)$ darstellen. Wir zeigen sowohl dass der exakte Erwartungswert des Restglied, als auch die Sigmapunktpropagation des Restglieds $o(\|x - \mu\|^3)$ ist. Damit ist der Fehler der Sigmapunktpropagation $o(\|x - \mu\|^3)$.

In n -D ist nicht nur das Argument der betrachteten Funktion vektorwertig, es ist auch der Erwartungswert ein Vektor oder wie bei der Kovarianz sogar eine Matrix. Dies generalisiert sich aber einfach elementweise.

14.5.5. Sigmapunktpropagation für Polynome vom Grad 3

Satz 24. Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}$ Polynom dritter Ordnung und $X \sim N(\mu, \Sigma)$ verteilt. Seien $\mathcal{X} = \begin{pmatrix} 2 \times \\ \mu \\ \mu \pm \sqrt{(n+1)\Sigma} \end{pmatrix}$ die dazugehörigen Sigmapunkte. Dann ist $E(f(\mathcal{X})) = E(f(X))$, die Sigmapunktpropagation also korrekt.

A.25

Beweis. Wir setzen $L = \sqrt{\Sigma}$. Sowohl $E(f(\mathcal{X}))$, als auch $E(f(X))$ sind linear in f . Summen oder skalare Vielfache in f können also herausgezogen werden. Deshalb reicht es, wenn wir die Gleichheit auf einer Basis der Polynome vom Grad 3 prüfen.

Zweckmäßigerweise nehmen wir die auf μ zentrierten Monome, die jetzt auch verschiedene Komponenten von X miteinander malnehmen können:

- Grad 0, $f(x) = 1$: $E(f(X)) = E(1) = 1$ und die Sigmapunkte sind auch alle 1 und damit auch ihr Mittelwert $E(f(\mathcal{X})) = E(1) = 1$.
- Grad 1, $f(x) = (x - \mu)_i$: Die Funktion ist antisymmetrisch um μ , während die Gaußverteilung symmetrisch um μ ist. Deshalb ist $E(f(X)) = 0$, weil sich jeweils die anti-symmetrischen Punkte gegeneinander herauskürzen. Das selbe, also $E(f(\mathcal{X})) = 0$ gilt für die Sigmapunktpropagation, weil die Sigmapunkte symmetrisch um μ liegen.
- Grad 2, $f(x) = (x - \mu)_i(x - \mu)_j$: Den Erwartungswert der Sigmapunkte berechnen wir als:

$$E(f(\mathcal{X})) = \frac{1}{2n+2} \sum_{k=1}^{2n+2} f(\mathcal{X}^{[k]}) \quad (14.51)$$

$$= \frac{1}{2n+2} \sum_{k=1}^{2n+2} (\mathcal{X}^{[k]} - \mu)_i (\mathcal{X}^{[k]} - \mu)_j \quad (14.52)$$

Betrachten wir die Summe genauer. Die μ Sigmapunkte fallen heraus, weil die Klammer 0 ist.

$$= \frac{1}{2n+2} \sum_{k=1}^n ((\sqrt{n+1}L_{\bullet k})_i (\sqrt{n+1}L_{\bullet k})_j + (-\sqrt{n+1}L_{\bullet k})_i (-\sqrt{n+1}L_{\bullet k})_j)$$

Die $+\sqrt{\dots}$ und $-\sqrt{\dots}$ Sigmapunkte haben den selben Wert, denn $-1 \cdot -1 = 1$. Die $\sqrt{n+1}$ zieht sich aus beiden Klammern heraus und vor die Summe als $(n+1)$.

$$= \frac{2(n+1)}{2n+2} \sum_{k=1}^n L_{ik} L_{jk} = \sum_{k=1}^n L_{ik} L_{kj}^T = (LL^T)_{ij} \quad (14.53)$$

$$= \Sigma_{ij} = \text{Cov}(X)_{ij} = E((X - \mu)_i (X - \mu)_j) = E(f(X)) \quad (14.54)$$

- Grad 3, $f(x) = (x - \mu)_i(x - \mu)_j(x - \mu)_k$ ist antisymmetrisch um μ und daher $E(f(\mathcal{X})) = E(f(X)) = 0$

□

14.5.6. Sigmapunktpropagation im Allgemeinfall

Wir betrachten analog zum 1D Fall nun für f eine dreimal stetig differenzierbare Funktion und wollen zeigen, dass der Fehler der Sigmapunktpropagation dritter Ordnung ist. Im 1D Fall hieß dies $o(\sigma^3)$, im n -D Fall gibt es aber statt eines Sigmas eine ganze

14. Mehrdimensionaler (Extended / Unscented) Kalmanfilter

Kovarianzmatrix. Dritter Ordnung heißt in dem Fall $o(\|L\|_2^3) = o(\|\Sigma\|_2^{3/2})$. Die sogenannte 2-Norm misst den größten Singulärwert, was in diesem Fall die Wurzel des größten Eigenwertes von Σ , also die größte Standardabweichung ist.

Wir zeigen analog zum 1D-Fall, dass sowohl der Erwartungswert $E(f(X))$, als auch die Sigmapunktpropagation $E(f(\mathcal{X})) = o(\|L\|_2^3)$ ist, wenn die Funktion $f = o(\|x - \mu\|^3)$ ist.

Lemma 14.1. *Sei $r = o(\|x - \mu\|^3)$ für $x \rightarrow \mu$ und $r : \mathbb{R}^n \rightarrow \mathbb{R}$ „hinreichend harmlos“. Sei außerdem $X \sim \mathcal{N}(\mu, \Sigma)$ mit $\Sigma \rightarrow 0$. Dann ist $E(r(X)) = o(\|\Sigma\|_2^{3/2})$.*

Beweis. Analog zu $x = \mu + \lambda\sigma$ im 1D-Fall setzen wir $x = \mu + Lv$, $v \sim \mathcal{N}(0, I_n)$, so dass man ein festes v betrachten kann, während $L \rightarrow 0$ geht.

$$\lim_{\Sigma \rightarrow 0} \frac{E(r(X))}{\|\Sigma\|_2^{3/2}} = \lim_{L \rightarrow 0} \int \frac{r(\mu + Lv)}{\|\Sigma\|_2^{3/2}} \mathcal{N}(0, I_n)(v) dv \quad (14.55)$$

$$= \int \lim_{\Sigma \rightarrow 0} \frac{r(\mu + Lv)}{\|\Sigma\|_2^{3/2}} \mathcal{N}(0, I_n)(v) dv \quad (14.56)$$

Wie im 1D-Fall setzt diese Vertauschung von Limes und Integral voraus, dass der Integrand „hinreichend harmlos“ ist.

Das Argument von r geht gegen μ . Wir setzen $r = o(\|x - \mu\|^3)$ ein:

$$= \int \lim_{\Sigma \rightarrow 0} \frac{o(\|Lv\|^3)}{\|\Sigma\|_2^{3/2}} \mathcal{N}(0, I_n)(v) dv \quad (14.57)$$

$$= \int \lim_{L \rightarrow 0} \frac{o(\|L\|_2^3 \|v\|^3)}{\|L\|_2^3} \mathcal{N}(0, I_n)(v) dv \quad (14.58)$$

$$= \int \lim_{L \rightarrow 0} o(\|v\|^3) \mathcal{N}(0, I_n)(v) dv \quad (14.59)$$

v ist in jedem der Limes über die integriert wird eine Konstante, also $o(\|v\|^3) = o(1)$ und $o(1)$ geht gegen 0.

$$= \int 0 \mathcal{N}(0, I_n)(v) dv = 0 \quad (14.60)$$

□

Wie in 1D machen wir eine ähnliche Analyse für die Sigmapunktpropagation des Restgliedes.

Lemma 14.2. *Sei $r = o(\|x - \mu\|^3)$ für $x \rightarrow \mu$. Sei außerdem $X \sim \mathcal{N}(\mu, \Sigma)$ mit $\Sigma \rightarrow 0$ und $\mathcal{X} = \begin{pmatrix} 2 \times \\ \mu & \mu \pm \sqrt{(n+1)\Sigma} \end{pmatrix}$ die dazugehörigen Sigmapunkte. Dann ist $E(r(\mathcal{X})) = o(\|\Sigma\|_2^{3/2})$.*

Beweis.

$$\mathbb{E}(r(\mathcal{X})) = \frac{1}{2n+2} \sum_{k=1}^{2n+2} r(\mathcal{X}^k) \quad (14.61)$$

Wenn $\Sigma \rightarrow 0$, konvergieren die Sigmapunkte, also alle Argumente von r gegen μ und zwar mit

$$\|x - \mu\| \leq \sqrt{n+1} \|L\|_2 = \sqrt{n+1} \|\Sigma\|_2^{1/2} \quad (14.62)$$

Damit kann man die $o(\|x - \mu\|^3)$ Zusicherung einsetzen:

$$\mathbb{E}(r(\mathcal{X})) = \frac{1}{2n+2} \sum_{k=1}^{2n+2} o((\sqrt{n+1} \|\Sigma\|_2^{1/2})^3) = o(\|\Sigma\|_2^{3/2}) \quad (14.63)$$

□

Satz 25. Gegeben sei eine dreimal stetig differenzierbare Funktion $f : \mathbb{R}^n \mapsto \mathbb{R}$ und eine Gaußverteilung $X = \mathcal{N}(\mu, \Sigma)$. Seien $\mathcal{X} = \left(\begin{smallmatrix} 2 \times \\ \mu \end{smallmatrix} \quad \mu \pm \sqrt{(n+1)\Sigma} \right)$ die Sigmapunkte, dann ist $\mathbb{E}(f(\mathcal{X}))$ eine 3. Ordnung korrekte Approximation von $\mathbb{E}(f(X))$, d.h. $\mathbb{E}(f(\mathcal{X})) - \mathbb{E}(f(X)) = o(\|\Sigma\|_2^{3/2})$ für $\Sigma \rightarrow 0$.

Beweis. Sei p das Taylorpolynom 3. Ordnung entwickelt um μ und r das dazugehörige Restglied. Damit ist

$$f(x) = p(x) + r(x), \quad r(x) = o(\|x - \mu\|^3) \quad (14.64)$$

$$\mathbb{E}(f(\mathcal{X})) - \mathbb{E}(f(X)) = (\mathbb{E}(p(\mathcal{X})) - \mathbb{E}(p(X))) + \mathbb{E}(r(\mathcal{X})) - \mathbb{E}(r(X)) \quad (14.65)$$

Da die Sigmapunktpropagation auf Polynomen dritter Ordnung exakt ist (Satz 24), kürzen sich die Klammer raus. Die zweiten beiden Terme sind $o(\|\Sigma\|_2^{3/2})$ nach Lemma 14.2 und 14.1.

$$= 0 + o(\|\Sigma\|_2^{3/2}) + o(\|\Sigma\|_2^{3/2}) = o(\|\Sigma\|_2^{3/2}) \quad (14.66)$$

□

In n -D müssen wir den Satz noch für vektor- und matrixwertige Funktionen generalisieren. Dies funktioniert einfach komponentenweise.

Satz 26. Gegeben sei eine dreimal stetig differenzierbare Funktion $f : \mathbb{R}^n \mapsto \mathbb{R}^{m \times p}$ und eine Gaußverteilung $X = \mathcal{N}(\mu, \Sigma)$. Seien $\mathcal{X} = \left(\begin{smallmatrix} 2 \times \\ \mu \end{smallmatrix} \quad \mu \pm \sqrt{(n+1)\Sigma} \right)$ die Sigmapunkte, dann ist $\mathbb{E}(f(\mathcal{X}))$ eine 3. Ordnung korrekte Approximation von $\mathbb{E}(f(X))$, d.h. $\mathbb{E}(f(\mathcal{X})) - \mathbb{E}(f(X)) = o(\|\Sigma\|_2^{3/2})$ für $\Sigma \rightarrow 0$.

Beweis. Die Aussage gilt für jede Komponente des Erwartungswertes $\in \mathbb{R}^{m \times p}$ nach Satz 25. Damit gilt sie auch für die Matrix als Ganzes. □

14. Mehrdimensionaler (Extended / Unscented) Kalmanfilter

Ebenso wie im 1D Fall die Varianz lässt sich auch die Kovarianzmatrix über die Sigmapunkte approximieren, weil sie als Erwartungswert definiert ist.

Satz 27. Gegeben sei eine dreimal stetig differenzierbare Funktion $f : \mathbb{R}^n \mapsto \mathbb{R}^{m \times p}$ und eine Gaußverteilung $X = \mathcal{N}(\mu, \Sigma)$. Seien $\mathcal{X} = \begin{pmatrix} \mu \\ \mu \pm \sqrt{(n+1)\Sigma} \end{pmatrix}$ die Sigmapunkte, dann ist $\text{Cov}(f(\mathcal{X}))$ eine 3. Ordnung korrekte Approximation von $\text{Cov}(f(X))$, d.h. $\text{Cov}(f(\mathcal{X})) - \text{Cov}(f(X)) = o(\|\Sigma\|_2^{3/2})$ für $\Sigma \rightarrow 0$.

Beweis. Der Beweis ist analog zu Satz 15. Wir nutzen die Definition der Kovarianzmatrix als Erwartungswert $\text{Cov}(Y) = \mathbb{E}((Y - \mathbb{E}Y)(Y - \mathbb{E}Y)^T)$ und zeigen, dass sowohl der Fehler im äußeren als auch im inneren Erwartungswert $o(\|\Sigma\|_2^{3/2})$ ist.

$$\text{Cov}(f(\mathcal{X})) = \mathbb{E}((f(\mathcal{X}) - \mathbb{E}(f(\mathcal{X}))(f(\mathcal{X}) - \mathbb{E}(f(\mathcal{X})))^T) \quad (14.67)$$

Betrachten wir das Argument im Erwartungswert als Funktion $h(x) = (f(x) - \mathbb{E}(f(\mathcal{X}))(f(x) - \mathbb{E}(f(\mathcal{X})))^T$ und wenden Satz 26 an.

$$= \mathbb{E}((f(X) - \mathbb{E}(f(\mathcal{X}))(f(X) - \mathbb{E}(f(\mathcal{X})))^T) + o(\|\Sigma\|_2^{3/2}) \quad (14.68)$$

Analog zum 1D-Fall ist $\mathbb{E}((X - \bar{x})(X - \bar{x})^T) = \text{Cov}(X) + (\mathbb{E}X - \bar{x})(\mathbb{E}X - \bar{x})^T$. Daher können wir das $\mathbb{E}(f(\mathcal{X}))$ durch $\mathbb{E}(f(X))$ mit einem zusätzlichen Fehlerterm ersetzen.

$$= \text{Cov}(f(X)) + (\mathbb{E}(f(\mathcal{X})) - \mathbb{E}(f(X)))(\mathbb{E}(f(\mathcal{X})) - \mathbb{E}(f(X)))^T + o(\|\Sigma\|_2^{3/2}) \quad (14.69)$$

$$= \text{Cov}(f(X)) + o(\|\Sigma\|_2^{3/2})^2 + o(\|\Sigma\|_2^{3/2}) \quad (14.70)$$

$$= \text{Cov}(f(X)) + o(\|\Sigma\|_2^{3/2}) \quad (14.71)$$

Achtung: Wie in 1D ist auch hier die Approximation selbst für Polynome vom Grad 2 nicht mehr exakt, weil durch die Quadrierung ein Polynom vom Grad 4 entsteht. \square

14.6. Unscented Kalman Filter

Der Unscented Kalman Filter verwendet die oben besprochene Sigmapunktpropagation um im Dynamikschritt Erwartungswert und Kovarianz des neuen Zustands und im Messschritt, Erwartungswert und Kovarianz von Z auszurechnen. Letztere wird dann wie beim KF/EKF ins Konditionierungslemma eingesetzt.

14.6.1. Annahmen

Wie beim EKF.

1. X_0 ist ungefähr $\mathcal{N}(\mu_0, \Sigma_0)$
2. $X_t = g_t(X_{t-1}, U_t) + \varepsilon_t$; $g_t : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}^n$, ε_t ist ungefähr $\mathcal{N}(0, \Sigma_{\varepsilon t})$
3. $Z_t = h_t(X_t) + \delta_t$; $h_t : \mathbb{R}^n \rightarrow \mathbb{R}^m$, δ_t ist ungefähr $\mathcal{N}(0, \Sigma_{\delta t})$
4. Alle ε_t, δ_t sind unabhängig.

14.6.2. Algorithmus

Dynamik(u_t):

$$\mathcal{X}_{t-1} = \begin{pmatrix} 2 \times & \\ \mu_{t-1} & \mu_{t-1} \pm \sqrt{(n+1)\Sigma_{t-1}} \end{pmatrix} \quad (14.72)$$

$$\bar{\mathcal{X}}_t^* = g(\mathcal{X}_{t-1}, u_t) \quad (14.73)$$

$$\bar{\mu}_t = \frac{1}{|\mathcal{X}|} \sum_i \bar{\mathcal{X}}_t^{*[i]} \quad (14.74)$$

$$\bar{\Sigma}_t = \frac{1}{|\mathcal{X}|} \sum_i (\bar{\mathcal{X}}_t^{*[i]} - \bar{\mu}_t)(\bar{\mathcal{X}}_t^{*[i]} - \bar{\mu}_t)^\top + \Sigma_{\epsilon t} \quad (14.75)$$

alternativ kann auch nicht-additives Rauschen verwendet werden, indem man die Sigmapunkte im Raum (x, ϵ) für die Verteilung $\mathcal{N}((\mu_{t-1}), (\Sigma_{t-1}))$ definiert.

Messung (z_t):

$$\bar{\mathcal{X}}_t = \begin{pmatrix} 2 \times & \\ \bar{\mu}_t & \bar{\mu}_t \pm \sqrt{(n+1)\bar{\Sigma}_t} \end{pmatrix} \quad (14.76)$$

$$\bar{\mathcal{Z}}_t = h(\bar{\mathcal{X}}_t) \quad (14.77)$$

$$\hat{z}_t = \frac{1}{|\mathcal{X}|} \sum_i \bar{\mathcal{Z}}_t^{[i]} \quad (14.78)$$

$$\Sigma_t^z = \frac{1}{|\mathcal{X}|} \sum_i (\bar{\mathcal{Z}}_t^{[i]} - \hat{z}_t)(\bar{\mathcal{Z}}_t^{[i]} - \hat{z}_t)^\top + \Sigma_{\delta t} \quad (14.79)$$

$$\bar{\Sigma}_t^{x,z} = \frac{1}{|\mathcal{X}|} \sum_i (\bar{\mathcal{X}}_t^{[i]} - \bar{\mu}_t)(\bar{\mathcal{Z}}_t^{[i]} - \hat{z}_t)^\top \quad (14.80)$$

$$K_t = \bar{\Sigma}_t^{x,z} (\Sigma_t^z)^{-1} \quad (14.81)$$

$$\mu_t = \bar{\mu}_t + K_t(z_t - \hat{z}_t) \quad (14.82)$$

$$\Sigma_t = \bar{\Sigma}_t - K_t(\bar{\Sigma}_t^{x,z})^\top \quad (\text{oder: } \bar{\Sigma}_t - K_t \bar{\Sigma}_t^z K_t^\top) \quad (14.83)$$

14.7. Intuition bei nd-Filtern

Eine besondere Rolle kommt bei mehrdimensionalen Zuständen der Kovarianz bei. Die Kovarianzmatrix gibt abseits der Hauptdiagonalen die Kovarianzen verschiedener Komponenten des Zustandes untereinander an. Diese NebenkompONENTEN geben also an, wie viel der Unsicherheit einer Komponente an der Unsicherheit einer anderen Komponente liegt. Dadurch ist es möglich, Zustände genauer zu schätzen, die nicht direkt gemessen werden.

Ein Beispiel: Wir betrachten den fortgeschrittenen landenden Ballon (AA.8). Die Höhe und Sinkgeschwindigkeit des Ballons werden geschätzt. Dazu wird die Beschleunigung

14. Mehrdimensionaler (Extended / Unscented) Kalmanfilter

und die Höhe gemessen. Das heißt, es gibt keine direkte Messung der Sinkgeschwindigkeit. Dennoch kann die Sinkgeschwindigkeit genauer geschätzt werden, als einfach nur die Beschleunigung aufzuintegrieren.

Dadurch dass die Sinkgeschwindigkeit im Dynamik-Schritt auf die Höhe gerechnet wird, haben die Höhe und die Sinkgeschwindigkeit eine positive Kovarianz. Wenn im Messschritt die Höhe gemessen wird, sorgt die Kovarianz im Zustand nun dafür, dass auch die Schätzung der Sinkgeschwindigkeit angepasst wird. Der Filter „verst“also, dass die Sinkgeschwindigkeit anders gewesen sein muss, um zu der gemessenden Höhe zu gelangen. Ist die gemessene Höhe größer als erwartet, muss die Sinkgeschwindigkeit geringer gewesen sein und umgekehrt. Man kann es auch folgendermaßen betrachten: Dem Filter steht die Information aus allen bisherigen Messungen zur Verfügung (repräsentiert über die Wahrscheinlichkeitsverteilung vom Zustand). Mit je 2 aufeinanderfolgenden Messungen der Höhe, lässt sich die Sinkgeschwindigkeit durch Differenzierung bestimmen. Daher sind mehr Informationen über die Sinkgeschwindigkeit vorhanden, als das Aufintegrieren der Beschleunigung.

Allgemein heißt das, dass die \mathbb{R}^n -Filter Zusammenhänge die im Dynamikmodell zwischen den Zuständen entstehen mit der Kovarianzmatrix repräsentieren. Mithilfe dieser Zusammenhänge, werden im Messschritt auch Zustände korrigiert, die nicht an der Messung beteiligt sind, also nicht direkt ins Messmodell eingehen. Diese Eigenschaft ist sehr wünschenswert, da es einem erspart, weitere Informationen aus der Messung händisch zu extrahieren. Zudem wird der Dynamikschritt und damit die prediction der Filter genauer, durch die genauere Schätzung aller Zustände.

15. Darstellung von Isometrien

Langfristig wollen wir im folgenden die Darstellung der Pose (Orientierung und Position) eines Körpers im Raum schätzen. Dabei wird sich herausstellen, dass diese Größe eine besondere mathematische Struktur hat (Mannigfaltigkeit) die berücksichtigt werden muss.

In diesem Kapitel wird diese besondere mathematische Struktur analysiert. Die Pose eines Körpers im Raum lässt sich darstellen durch eine Funktion $f : \mathbb{R}^3 \rightarrow \mathbb{R}^3$, die jeden Punkt in Körperkoordinaten die Weltkoordinaten des selben Punktes zuordnet. Damit definiert diese Funktion die Abbildung zwischen Körper- und Weltkoordinaten und damit indirekt die Pose des Körpers. Die Funktion erhält die Länge, weil natürlich zwei Punkte in Körper- wie Weltkoordinaten den gleichen Abstand haben.

15.1. Klassifizierung der Isometrien über orthonormale Matrizen

15.1.1. Vorbereitungen

Wir starten daher mit der Analyse der mathematischen Struktur (Klassifizierung) solcher sogenannter Isometrien.

Definition 14. Eine Funktion $f : \mathbb{R}^m \rightarrow \mathbb{R}^n$ ist eine Isometrie, g.d.w.

$$|f(x) - f(y)| = |x - y| \quad \forall x, y \in \mathbb{R}^m. \quad (15.1)$$

Q 37

Lemma 15.1. Jede Isometrie $f : \mathbb{R}^m \rightarrow \mathbb{R}^n$ ist stetig.

Beweis. Sei $x_i \in \mathbb{R}^m$ mit $x_i \rightarrow x$. Zu zeigen ist, dass $f(x_i) \rightarrow f(x)$. Da $x_i \rightarrow x$, konvergiert $x_i - x \rightarrow 0$ und $|x_i - x| \rightarrow 0$. Da f eine Isometrie ist, $|f(x_i) - f(x)| = |x_i - x| \rightarrow 0$ und damit $f(x_i) \rightarrow f(x)$ und f ist stetig. \square

Lemma 15.2. Wenn eine stetige Funktion $f : \mathbb{R}^m \rightarrow \mathbb{R}^n$, $f(a+b) = f(a) + f(b)$ erfüllt, dann ist sie linear.

Beweis. Für Linearität ist noch zu zeigen $f(\lambda a) = \lambda f(a)$ für $\lambda \in \mathbb{R}$. Wir zeigen dies zuerst für natürliche λ , dann für Kehrwerte natürlicher λ , dann für Brüche und zuletzt für reelle Zahlen:

15. Darstellung von Isometrien

$f(\lambda a) = \lambda f(a) \forall \lambda \in \mathcal{N}$ folgt durch Induktion mit

$$f(\lambda a) = f((\lambda - 1)a + a) = f((\lambda - 1)a) + f(a) = (\lambda - 1)f(a) + f(a) = \lambda f(a) \quad (15.2)$$

Nun zu den Kehrwerten natürlicher Zahlen: Hier nutzen wir obiges Resultat und drehen es um:

$$f(a) = f\left(\frac{\lambda}{\lambda}a\right) = \lambda f\left(\frac{1}{\lambda}a\right) \quad (15.3)$$

$$f\left(\frac{1}{\lambda}a\right) = \frac{1}{\lambda}f(a) \quad (15.4)$$

Für gebrochene $\lambda = \frac{p}{q}$ werden beide Resultate kombiniert.

$$f(\lambda a) = f\left(\frac{p}{q}a\right) = pf\left(\frac{1}{q}a\right) = \frac{p}{q}f(a) = \lambda f(a) \quad (15.5)$$

Das Gesamtergebn ergibt sich dann aus Stetigkeit, weil jedes reelle λ als Grenzwert einer Folge von rationalen λ_i dargestellt werden kann:

$$f(\lambda a) = f\left(\lim_i \lambda_i a\right) = \lim_i f(\lambda_i a) = \lim_i \lambda_i f(a) = \lambda f(a) \quad (15.6)$$

□

Lemma 15.3. *Erhält eine stetige Funktion $f : \mathbb{R}^m \rightarrow \mathbb{R}^n$ Mittelpunkte, d.h. $f\left(\frac{a+b}{2}\right) = \frac{f(a)+f(b)}{2}$, so ist sie affin, d.h. $f(x) = Ax + b$ für $A \in \mathbb{R}^{n \times m}, b \in \mathbb{R}^n$.*

Beweis. Wir zeigen, dass $g : x \mapsto f(x) - f(0)$ linear ist, dann folgt die Aussage mit $b = f(0)$ und A der Matrix die g repräsentiert. Dafür ist nach Lemma 15.2 zu zeigen, dass $g(a+b) = g(a) + g(b)$.

$$g(a) = g\left(\frac{2a+0}{2}\right) = \frac{g(2a) + g(0)}{2} = \frac{g(2a)}{2} \quad (15.7)$$

$$\Rightarrow g(2a) = 2g(a) \quad (15.8)$$

$$g(a+b) = g\left(2\frac{a+b}{2}\right) = 2g\left(\frac{a+b}{2}\right) = 2\frac{g(a) + g(b)}{2} = g(a) + g(b). \quad (15.9)$$

□

15.1.2. Mazur-Ulam Theorem

Theorem 15.4 (Mazur-Ulam Theorem [53], Spezialfall). *Jede Isometrie $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ist affin.*

$$f(x) = Qx + t, Q \in \mathbb{R}^{n \times n}, t \in \mathbb{R}^n, \quad (15.10)$$

wobei Q orthonormal ist ($Q^T Q = I$).

Die Beweisidee ist folgende: In der euklidischen Norm bzw. Metrik ist die Mitte zwischen zwei Punkten a und b eindeutig bestimmt, als der einzige Punkt bei dem die Summe der Entfernungen zu a und b gleich der Entfernung von a zu b ist.

Da Isometrien Entfernungen erhalten, erhalten sie wegen dieser Eigenschaften auch Mittelpunkte und sind damit affin. Formal beweist sich dies als:

Beweis. Als Isometrie ist f nach Lemma 15.1 stetig und nach Lemma 15.3 müssen wir nur noch zeigen, dass f Mittelpunkte erhält.

Seien dazu $a, b \in \mathbb{R}^n$.

$$\left| f(a) - f\left(\frac{a+b}{2}\right) \right| = \left| a - \frac{a+b}{2} \right| = \frac{|a-b|}{2} = \frac{|f(a) - f(b)|}{2} \quad (15.11)$$

Analog gilt

$$\left| f(b) - f\left(\frac{a+b}{2}\right) \right| = \frac{|f(a) - f(b)|}{2} \quad (15.12)$$

Es gibt nur einen Punkt, der genau halb so weit von $f(a)$ und $f(b)$ entfernt ist, wie die beiden voneinander. Das ist die Mitte $\frac{f(a)+f(b)}{2}$. Also ist

$$f\left(\frac{a+b}{2}\right) = \frac{f(a) + f(b)}{2} \quad (15.13)$$

und nach Lemma 15.3 f affin von der Form $f(x) = Qx + b$.

Es bleibt noch zu zeigen, dass $Q^T Q = I$, also Q orthonormal ist. Der Beweis basiert auf der Beobachtung, dass

$$|a+b|^2 - |a-b|^2 = (a+b)^2 - (a-b)^2 = 4a \cdot b, \quad (15.14)$$

sich also das Skalarprodukt $a \cdot b$ aus den Normen von $a+b$ und $a-b$ rekonstruieren lässt, die beide von f erhalten werden. Im Detail

$$|f(e_i + e_j) - f(0)|^2 - |f(e_i - e_j) - f(0)|^2 = |e_i + e_j|^2 - |e_i - e_j|^2 = 4e_i \cdot e_j \quad (15.15)$$

$$= \begin{cases} 4 & i = j \\ 0 & \text{sonst} \end{cases} \quad (15.16)$$

$$|f(e_i + e_j) - f(0)|^2 - |f(e_i - e_j) - f(0)|^2 = |Q(e_i + e_j)|^2 - |Q(e_i - e_j)|^2 \quad (15.17)$$

$$= |Q_{\bullet i} + Q_{\bullet j}|^2 - |Q_{\bullet i} - Q_{\bullet j}|^2 \quad (15.18)$$

$$= 4Q_{\bullet i} \cdot Q_{\bullet j} = 4(Q^T Q)_{ij} \quad (15.19)$$

Also ist $Q^T Q = I$. □

Man sieht, dass Q 6 Zwangsbedingungen erfüllen muss: Alle 3 Spalten müssen Länge 1 haben und Spalte 1-2, 2-3, 1-3 müssen orthogonal zueinander sein. Anschaulich hat man bei 9 Zahlen nur 3 Freiheitsgrade und 6 Zwangsbedingungen.

15. Darstellung von Isometrien

Unter den orthonormalen Matrizen gibt es wegen $|Q|^2 = |Q^T||Q| = |Q^T Q| = |I| = 1$ nur solche mit Determinante 1 und -1 . Letztere spiegeln eine Achse (z.B. $\begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$). Sie können als physische Bewegungen nicht auftreten: ist $Q : [0 \dots 1] \rightarrow \mathbb{R}^{3 \times 3}$ eine stetige Bewegung mit $Q(0) = I$ und $Q(t)^T Q(t) = I$, so ist $|Q(t)|$ anfangs $+1$ und immer ± 1 . Da es stetig ist kann es nicht nach -1 springen. Anschaulich drehen die $|Q| = -1$ Isometrien die Händigkeit um: Aus einem Rechtssystem wird ein Linkssystem.

Diese Beobachtung gibt Anlass für die Definition der Speziellen Orthogonalen Gruppe:

Definition 15.

$$SO(n) = \{Q \in \mathbb{R}^{n \times n} | Q^T Q = I, |Q| = 1\} \quad (15.20)$$

15.1.3. Verwendung zur Darstellung der Position und Orientierung von Körpern

Die Darstellung von Posen im Raum durch orthonormale Matrizen ist weit verbreitet in Computergrafik, Robotik, Computational Geometry und Bildverarbeitung. Dazu definiert man Koordinatensysteme (z.B. A, B) die fest an einem Körper oder der Welt sind und betrachtet die Isometrie $f_{B \leftarrow A}$, die einen Punkt von z.B. A -Koordinaten nach B -Koordinaten umrechnet, also mit A -Koordinaten eines Punktes als Eingabe die B -Koordinaten des selben Punktes ausrechnet. Nach dem Mazur-Ulam Theorem, lässt sich $f(x)_{B \leftarrow A}$ als $f(x)_{B \leftarrow A} = Qx + t$ darstellen. Zur weiteren Vereinfachung wird die Matrix Q und der Vektor t zu einer homogenen Matrix

$$\begin{pmatrix} f(x) \\ 1 \end{pmatrix} = \begin{pmatrix} Q & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ 1 \end{pmatrix} = T_{W \leftarrow B} x \quad (15.21)$$

zusammengefasst und diese als $T_{W \leftarrow B}$ entsprechend den Koordinatensystemen zwischen denen sie abbildet benannt.

Diese Matrizen werden im Computer gespeichert. Sie lassen sich nach den folgenden Regeln verknüpfen um damit zu rechnen:

$$T_{A \leftarrow B} = T_{B \leftarrow A}^{-1} \quad (15.22)$$

$$T_{C \leftarrow A} = T_{C \leftarrow B} T_{B \leftarrow A}, \quad (15.23)$$

Wobei die Konsistenz der Rechnung mit der "Pfeilschreibweise" von rechts nach links abgelesen werden kann. Ein Beispiel findet sich in Abbildung 15.1. Die selben Rechenregeln gelten auch für die Matrizen Q die die Orientierung darstellen ohne Translation t .

Q 39

15.1.4. Verwendung als Zustand im (Extended/Unscented) Kalmanfilter

Für die Nutzung im Zustand eines Kalmanfilters zur Schätzung von Orientierungen ergibt sich aber ein Problem: Die Bedingung $Q^T Q = I$ muss erfüllt bleiben und davon

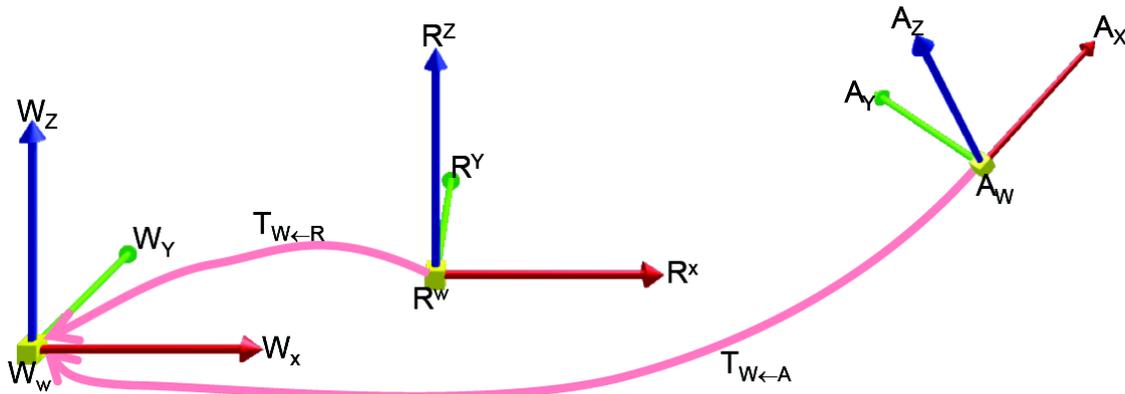


Abbildung 15.1.: Die Pose zweier Körper A (“Aircraft”) und R “Runway” wird im Weltkoordinatensystem W durch zwei homogene Matrizen $T_{W \leftarrow A}$ und $T_{W \leftarrow R}$ dargestellt. Durch Verknüpfung und Invertierung erhält man die Pose des Flugzeugs A relativ zur Landebahn R : $T_{R \leftarrow A} = T_{W \leftarrow R}^{-1} T_{W \leftarrow A}$. Durch die “Pfeilschreibweise” lässt sich die Konsistenz des Ausdrucks überprüfen.

weiss der Filter nichts. Man kann also nicht einfach die 9 Zahlen aus Q in den Zustand eines Kalmanfilters tun, sondern die Struktur von Q muss berücksichtigt werden.

Der Vollständigkeit halber sollte noch erwähnt werden, dass man eine Rotation auch als Quaternion (also 4 Zahlen) darstellen kann, deren Quadratsumme 1 sein muss. Das macht die Repräsentation etwas kompakter hat aber den gleichen Nachteil, dass der Filter eine Bedingung berücksichtigen muss.

Wir analysieren nun eine spezielle Klasse von Isometrien, nämlich Rotationen. Das Endergebnis ist, dass in 3D (!) jede Isometrie durch eine Verschiebung und eine Rotation um eine Achse definiert werden kann. Dies liefert auch eine alternative Art, Posen im Raum in einem Kalmanfilter zu schätzen, die aber auch Probleme zeigt.

15.2. Rotationen

15.2.1. Axiomatisierung

Die besondere Form der Isometrie ist die Rotation um eine vorgegebene Achse im Raum. Wir definieren diese Rotation axiomatisch über.

Definition 16. $\text{Rot}(v, \alpha, p)$ für $v \in \mathbb{R}^3$, $|v| = 1$, $\alpha \in \mathbb{R}$, $p \in \mathbb{R}^3$ (dreht p um v mit Winkel α) heißt Rotation, wenn

1. $\text{Rot}(v, 0, p) = p$
2. $\frac{\partial \text{Rot}(v, \alpha, p)}{\partial \alpha} = v \times \text{Rot}(v, \alpha, p)$

Die zweite Bedingung ergibt sich daraus, dass sich ein rotierender Punkt tangential am Kreis bewegt. Damit steht seine Bewegungsrichtung senkrecht sowohl auf dem

15. Darstellung von Isometrien

Radius (also dem Punkt selbst) als auch der Achse. Außerdem steigt die Geschwindigkeit proportional mit der Entfernung zur Achse. Dies alles wird durch den \times Ausdruck ausgedrückt.

Es stellt sich bei dieser Axiomatisierung der Rotation die Frage, ob es nicht ein bisschen übertrieben ist auf Differentialgleichungen zurückzugreifen. So kann man nämlich nicht unabhängig sagen, wie eine einzelne Rotationsmatrix zu einem festen α aussieht, sondern die Abhängigkeit von α muss als Funktion betrachtet werden.

Allerdings ist bemerkenswert, dass die Definition ohne \sin oder \cos auskommt, d. h. die Bedeutung von Winkeln in der Abhängigkeit von α wird implizit mitdefiniert. Das scheint bei einer Definition die jeweils für ein festes α gilt schwierig. Außerdem zeigt sich im Folgenden, dass die Definition sich gut zum Beweisen eignet. Dazu benutzen wir den Existenz und Eindeutigkeitsatz für Differentialgleichungen, der unter bestimmten Bedingungen (die hier erfüllt sind) zusichert dass es eine und genau eine Lösung gibt.

Bemerkung 15.5. Rot existiert und ist eindeutig. Wegen Existenz und Eindeutigkeitsatz für Differentialgleichungen.

Durch diesen Satz können wir generell folgende Beweisstrategie verwenden. Wir haben einen Ausdruck von dem wir zeigen wollen, dass er $\text{Rot}(v, \alpha, p)$ ist. Wir nennen ihn zuerst $\text{Rot}'(v, \alpha, p)$. Dann zeigen wir, dass auch $\text{Rot}'(v, \alpha, p)$ die Axiome (1) und (2) erfüllt. Nach dem Eindeutigkeitsatz für DGL muss dann $\text{Rot}' = \text{Rot}$ sein.

Wir wollen uns im folgenden die Struktur von Rotationen klarer machen. Dazu zeigen wir zuerst, dass die Rotation linear im rotierten Punkt ist.

Lemma 15.6. $\text{Rot}(v, \alpha, p)$ ist linear in p , d. h. es lässt sich durch eine Matrix $\text{Rot}(v, \alpha)$ darstellen.

$$\text{Rot}(v, \alpha, p) = \underbrace{[\text{Rot}(v, \alpha, e_1), \text{Rot}(v, \alpha, e_2), \text{Rot}(v, \alpha, e_3)]}_{\text{Rot}(v, \alpha)} \cdot p \quad (15.24)$$

Beweis. $\text{Rot}'(v, \alpha, p)$ wie in (15.24) und prüfe die beiden Rotationsaxiome.

$$1.: \quad \text{Rot}'(v, 0, p) = \text{Rot}(v, 0)p = [\text{Rot}(v, 0, e_1), \text{Rot}(v, 0, e_2), \text{Rot}(v, 0, e_3)]p \quad (15.25)$$

$$= [e_1, e_2, e_3]p = \text{I}p = p \quad (15.26)$$

$$2.: \quad \frac{\partial \text{Rot}'(v, \alpha, p)}{\partial \alpha} = \frac{\partial (\text{Rot}(v, \alpha)p)}{\partial \alpha} = \sum_{i=1}^3 \frac{\partial \text{Rot}(v, \alpha, e_i)}{\partial \alpha} p_i \quad (15.27)$$

$$= \sum_{i=1}^3 (v \times \text{Rot}(v, \alpha, e_i)) p_i = v \times \sum_{i=1}^3 (\text{Rot}(v, \alpha) e_i p_i) \quad (15.28)$$

$$= v \times \left(\text{Rot}(v, \alpha) \sum_{i=1}^3 (e_i p_i) \right) = v \times (\text{Rot}(v, \alpha)p) \quad (15.29)$$

$$= v \times \text{Rot}'(v, \alpha, p) \quad (15.30)$$

Damit folgt $\text{Rot}' = \text{Rot}$ und damit (15.24) aus dem Eindeutigkeitsatz für DGL. \square

Die Matrix $\text{Rot}(v, \alpha) \in SO(3)$ (ohne Beweis).

Die Definition $\text{Rot}(v, \alpha)$ betrachtet die Achse (v , Vektor der Länge 1) und den Winkel (α) getrennt. Es ist naheliegend, die Informationen zusammenzuführen und den Winkel, um den gedreht werden soll, in der Länge des Vektors zu codieren. Diese Darstellung stellt eine Rotation durch einen einzelnen Vektor dar.

15.2.2. Matrix Exponential Darstellung

Definition 17 (Matrix Exponential).

$$\text{Rot}(v) = \text{Rot}(v/|v|, |v|), \quad \text{Rot}(0) = I \quad (15.31)$$

Drehung um v mit Winkel $|v|$.

Anschaulich könnte man die Darstellung auch als “Skalierte Achse” bezeichnen, weil die Achse als Vektor um den Winkel skaliert ist.

Ein Problem bei der Definition ist, dass man nicht mehr zwischen negiertem Winkel und negierter Achse unterscheiden kann. Da ist aber auch nicht notwendig, weil beides einfach zur invertierten Drehung führt. Ein schwierigerer Problemfall stellt $v = 0$ dar, also keine Drehung. Dann ist $v/|v|$ undefiniert. Andererseits macht es auch keinen Unterschied, ob man nicht um die X-Achse oder nicht um die Y-Achse dreht. Anschaulich macht die Definition also Sinn.

Der Name “Matrix Exponential” für $\text{Rot}(v)$ rührt daher, dass $\text{Rot}(v)$ gleich einem Ausdruck mit der Matrix Exponentialfunktion ist, wie wir gleich beweisen werden. Dadurch ist auch gezeigt, dass $\text{Rot}(0)$ kein Sonderfall ist, also in der Tat die Definition auch für 0 sinnvoll ist.

Definition 18. Für eine quadratische Matrix $A \in \mathbb{R}^{n \times n}$ ist das Matrixexponential durch die übliche Potenzreihe der Exponentialfunktion definiert in die statt einer Zahl eine Matrix eingesetzt wird.

$$\exp A = \sum_{i=0}^{\infty} \frac{1}{i!} A^i \quad (15.32)$$

In dem Matrix-Exponential Ausdruck für $\text{Rot}(v)$ spielt das Kreuzprodukt aus Axiom 2 eine Rolle und um es einer Matrix-Exponentialfunktion zu übergeben, muss es als Matrix dargestellt werden.

Definition 19. Für einen Vektor $v \in \mathbb{R}^3$ ist

$$[v]_{\times} = \begin{pmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{pmatrix} \quad (15.33)$$

15. Darstellung von Isometrien

die Matrix, die “Kreuzprodukt mit v von links” darstellt:

$$[v]_{\times} p = v \times p \quad (15.34)$$

Das folgende Lemma leitet den Matrix-Exponential-Ausdruck für $\text{Rot}(v)$ her:

Lemma 15.7.

$$\text{Rot}(v, \alpha, p) = \exp(\alpha[v]_{\times}) p = \sum_{i=0}^{\infty} \frac{1}{i!} \alpha^i [v]_{\times}^i p \quad (15.35)$$

$$\text{Rot}(v, \alpha) = \exp(\alpha[v]_{\times}) = \sum_{i=0}^{\infty} \frac{1}{i!} \alpha^i [v]_{\times}^i \quad (15.36)$$

$$\text{Rot}(v) = \exp[v]_{\times} = \sum_{i=0}^{\infty} \frac{1}{i!} [v]_{\times}^i \quad (15.37)$$

Beweis. Definiere

$$\text{Rot}'(v, \alpha, p) = \exp(\alpha[v]_{\times}) p = \sum_{i=0}^{\infty} \frac{1}{i!} \alpha^i [v]_{\times}^i p \quad (15.38)$$

und zeige, dass $\text{Rot}'(v, \alpha, p)$ die Rotationsaxiome erfüllt.

$$\text{Rot}'(v, 0, p) = Ip = p \quad (15.39)$$

$$\frac{\partial \text{Rot}'(v, \alpha, p)}{\partial \alpha} = \frac{\partial}{\partial \alpha} \sum_{i=0}^{\infty} \frac{1}{i!} \alpha^i [v]_{\times}^i p \quad (15.40)$$

$$= \sum_{i=1}^{\infty} \frac{1}{i!} i \alpha^{i-1} [v]_{\times}^i p \quad (15.41)$$

$$= [v]_{\times} \sum_{i=1}^{\infty} \frac{1}{(i-1)!} \alpha^{i-1} [v]_{\times}^{i-1} p \quad (15.42)$$

$$= [v]_{\times} \text{Rot}'(v, \alpha, p) \quad (15.43)$$

$$= v \times \text{Rot}'(v, \alpha, p) \quad (15.44)$$

Damit folgt (15.35) aus dem Existenz und Eindeutigkeitsatz. (15.36) folgt direkt daraus, dass (15.35) die Form “Matrix mal p ” hat. (15.37) folgt durch Einsetzen der Definition (15.31):

$$\text{Rot}(v) = \text{Rot}\left(\frac{v}{|v|}, |v|\right) = \exp\left(\widehat{\left|\frac{v}{|v|}\right.}\right) = \exp([v]_{\times}) \quad (15.45)$$

□

Dieses Result ist zum ersten mal eine konkrete Formel für die vorher ja nur axiomatisch definierte Rotation, wenn auch als unendliche Potenzreihe. Das nächste Lemma

gibt eine “noch konkretere” Formel mit relem \sin, \cos an, wie sie auch üblicherweise in Programmen implementiert wird:

Lemma 15.8 (Rodriguez-Formel).

$$\text{Rot}(v, \alpha) = I + s[v]_{\times} + c[v]_{\times}^2, \quad (15.46)$$

$$= \begin{pmatrix} 1 + cx^2 - c & cxy - sz & cxz + sy \\ cxy + sz & 1 + cy^2 - c & cyz - sx \\ cxz - sy & cyz + sx & 1 + cz^2 - c \end{pmatrix}, \quad (15.47)$$

$$\text{mit } v = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad c = 1 - \cos \alpha, \quad s = \sin \alpha \quad (15.48)$$

Beweis.

Wir beweisen die Formel, indem wir die Axiome überprüfen:

$$v = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad (15.49)$$

1. Axiom: Für $\alpha = 0$ gilt:

$$c = 0, \quad s = 0 \quad (15.50)$$

$$\Rightarrow \text{Rot}(v, 0) * p = I * p = p, \quad (15.51)$$

2. Axiom: Wir nutzen, dass $\text{Rot}(v, \alpha, p)$ linear ist.

$$\frac{\partial c}{\partial \alpha} = s, \quad \frac{\partial s}{\partial \alpha} = 1 - c, \quad \frac{\partial x}{\partial \alpha} = 0, \quad \frac{\partial y}{\partial \alpha} = 0, \quad \frac{\partial z}{\partial \alpha} = 0, \quad (15.52)$$

$$\Rightarrow \frac{\partial \text{Rot}(v, \alpha, p)}{\partial \alpha} = \frac{\partial \text{Rot}(v, \alpha) * p}{\partial \alpha} = ((1 - c)[v]_{\times} + s[v]_{\times}^2) * p \quad (15.53)$$

$$[v]_{\times} * \text{Rot}(v, \alpha, p) = [v]_{\times} * \text{Rot}(v, \alpha) * p = ([v]_{\times} + s[v]_{\times}^2 + c[v]_{\times}^3) * p \quad (15.54)$$

Berechnen wir noch den Ausdruck $[v]_{\times}^3$:

$$[v]_{\times}^3 = \begin{pmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{pmatrix} * \begin{pmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{pmatrix} * \begin{pmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{pmatrix} \quad (15.55)$$

15. Darstellung von Isometrien

$$= \begin{pmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{pmatrix} * \begin{pmatrix} -v_2^2 - v_3^2 & v_1 v_2 & v_1 v_3 \\ v_1 v_2 & -v_1^2 - v_3^2 & v_2 v_3 \\ v_1 v_3 & v_2 v_3 & -v_1^2 - v_2^2 \end{pmatrix} \quad (15.56)$$

$$= \begin{pmatrix} 0 & v_3(v_1^2 + v_2^2 + v_3^2) & -v_2(v_1^2 + v_2^2 + v_3^2) \\ -v_3(v_1^2 + v_2^2 + v_3^2) & 0 & v_1(v_1^2 + v_2^2 + v_3^2) \\ v_2(v_1^2 + v_2^2 + v_3^2) & -v_1(v_1^2 + v_2^2 + v_3^2) & 0 \end{pmatrix} \quad (15.57)$$

$$(v_1^2 + v_2^2 + v_3^2) = \|v\|^2 = 1 \quad (15.58)$$

$$\Rightarrow [v]_{\times}^3 = \begin{pmatrix} 0 & v_3 & -v_2 \\ -v_3 & 0 & v_1 \\ v_2 & -v_1 & 0 \end{pmatrix} = -[v]_{\times} \quad (15.59)$$

$$\Rightarrow [v]_{\times} * \text{Rot}(v, \alpha, p) = ([v]_{\times} + s[v]_{\times}^2 - c[v]_{\times}) * p = ((1 - c)[v]_{\times} + s[v]_{\times}^2) * p \quad (15.60)$$

Damit ist auch Axiom 2 erfüllt.

$$(15.61)$$

□

Mit der Rodriguez-Formel können wir jede Rotation um eine Achse als Matrix schreiben. Ein Theorem aus der linearen Algebra (Klassifikation für orthonormalen Matrizen) liefert, dass in 3D auch jedes $Q \in SO(3)$ als $\text{Rot}(v)$ dargestellt werden kann, also letztlich jede Händigkeitserhaltende Isometrie als eine Translation plus eine Rotation um eine Achse. In höheren Dimensionen gilt das nicht mehr (Übung).

Theorem 15.9 (Eulers Rotationstheorem). *Sei $Q \in SO(3)$, dann gibt es ein $v \in \mathbb{R}^3$, so dass $Qv = v$ ist.*

Beweis.

$$|Q - I| = |(Q - I)^T| = |Q^T - I| = |Q^T(I - Q)| \quad (15.62)$$

$$= |I - Q| = (-1)^3 |Q - I| = -|Q - I| \quad (15.63)$$

Es folgt $|Q - I| = 0$ und es gibt ein v mit $(Q - I)v = 0$, also $Qv = v$. □

Theorem 15.10. *Sei $Q \in SO(3)$, dann gibt es ein $v \in \mathbb{R}^3$, so dass $Q = \text{Rot}(v)$ ist.*

Beweis. Nur Idee: Nach Eulers Rotationstheorem gibt es einen Vektor v mit $Qv = v$ der die Achse definiert, weil er unverändert bleibt. Senkrecht dazu hat nach dem selben Satz Q die Form einer 2D-Drehmatrix $\begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}$ von der man den Winkel α übernimmt. □

Damit können wir auch die Umkehrfunktion der Rotation, also sozusagen die Arcus-Rotation definieren. Theorem 15.10 zeigt, dass sich jede $SO(3)$ Orientierung durch einen 3D Vektor des Matrix Exponential darstellen kann. $SO(3)$ hat also, wie die Zahl der 6 Nebenbedingungen schon vermuten ließ, genau 3 Freiheitsgrade. Für Drehungen um π allerdings ist das Ergebnis nicht eindeutig, man kann sich um v und $-v$ drehen, also sozusagen 180° in die eine oder in die andere Richtung.

Definition 20. $\text{aRot}(Q) = v \Leftrightarrow \text{Rot}(v) = Q \wedge |v| < \pi$

15.3. Singularitäten im Matrix-Exponential

Das Resultat 15.10 legt nahe, den Vektor v zur Darstellung einer Orientierung im Raum in den Zustand des Kalmanfilters zu übernehmen. Da jeder Wert von v eine gültige Orientierung $\text{Rot}(v)$ liefert, muss der Filter keine Nebenbedingung berücksichtigen.

Q 41

Leider ist auch dies keine Lösung, weil diese Art der Parametrierung sogenannte Singularitäten hat, also Parameterwerte bei denen eine kleine Änderung der Orientierung eine große Änderung der Parameter erfordert. Der Kalmanfilter sieht die möglichen Änderungen der Orientierung nur über die Parameter und kann so einer Änderung dann nicht folgen.

Anschaulich startet man mit der Erkenntnis, dass die ganze Kugel $|v| = 2\pi$ von Rot auf $\text{Rot}(v) = I$ abgebildet wird, weil eine volle Drehung um gleich welche Achse steht wieder am Anfang endet. Dabei sind für v auf dieser Kugel 2 der drei Freiheitsgrade von v ohne Effekt und der eine übriggebliebener (Verlängerung/Verkürzung von v) kann nicht die drei Freiheitsgrade abdecken, die die Orientierung tatsächlich hat.

Um dies formal zu zeigen betrachten wir die Ableitung von $\text{Rot}(v)$ bei $v = (2\pi, 0, 0)^T$ also einer kompletten Drehung um die X Achse. Der Freiheitsgrad des Verlängerns von v führt zu einer Drehung um X, nach der obigen Überlegung müssten also die Drehung um Y und Z fehlen. Wir betrachten daher $\text{Rot}(v)_{21}$, also die Welt-Y-Koordinate der Körper-X-Achse. Nach unseren Überlegungen sollten die Parameter v sich (differentiell) nicht auf $\text{Rot}(v)_{21}$ auswirken, die Ableitung von $\text{Rot}(v)_{21}$ nach v also 0 sein.

$$\frac{\partial \text{Rot}(x, y, z)_{21}}{\partial(x, y, z)} \Big|_{x=2\pi, y=0, z=0} \tag{15.64}$$

$$= \frac{\partial \text{Rot} \left(\frac{(x, y, z)}{\sqrt{x^2 + y^2 + z^2}}, \sqrt{x^2 + y^2 + z^2} \right)}{\partial(x, y, z)} \Big|_{x=2\pi, y=0, z=0} \tag{15.65}$$

15. Darstellung von Isometrien

$$= \frac{\partial \text{Rot}(x, y, z, \alpha)_{21}}{\partial(x, y, z, \alpha)} \Big|_{x=1, y=0, z=0, \alpha=2\pi} \frac{\partial \left(\frac{(x, y, z)}{\sqrt{x^2 + y^2 + z^2}}, \sqrt{x^2 + y^2 + z^2} \right)}{\partial(x, y, z)} \Big|_{x=2\pi, y=0, z=0} \quad (15.66)$$

$$= \frac{\partial((1 - \cos \alpha)xy + \sin \alpha z)}{\partial(x, y, z, \alpha)} \Big|_{x=1, y=0, z=0, \alpha=2\pi} \begin{pmatrix} * & * & * \\ * & * & * \\ * & * & * \end{pmatrix}, \quad (15.67)$$

$$= ((1 - \cos \alpha)y \quad (1 - \cos \alpha)x \quad \sin \alpha \quad \sin \alpha xy + \cos \alpha z) \Big|_{x=1, y=0, z=0, \alpha=2\pi} \begin{pmatrix} * & * & * \\ * & * & * \\ * & * & * \end{pmatrix} \quad (15.68)$$

$$= (0 \quad 0 \quad 0 \quad 0) \begin{pmatrix} * & * & * \\ * & * & * \\ * & * & * \end{pmatrix} \quad (15.69)$$

$$= (0 \quad 0 \quad 0) \quad (15.70)$$

Es kann also bei $v = (2\pi, 0, 0)^T$ keine "differentielle" Änderung von v die Welt-Y-Koordinate der Körper-X-Achse differentiell verändern.

Q 42

Zumindest für Filter, die wie der (Extended/Unscented) Kalmanfilter differenzierbare Funktionen voraussetzen, weil sie linearisieren, funktioniert also diese Parametrisierung nicht.

In der Tat gilt sogar (ohne Beweis)

Theorem 15.11. *Es gibt keine singularitätenfreie Minimalparametrisierung der 3D Orientierungen, d.h. eine Funktion $f : \mathbb{R}^3 \rightarrow SO(3)$, bei der $\frac{\partial f(v)}{\partial v} \in \mathbb{R}^{9 \times 3}$ überall vollen Rank (3) hat.*

Es gibt also keine Möglichkeit durch geschickte Parameterwahl 3D Orientierungen transparent in den (E/U)KF wie er ist zu integrieren, wir müssen überlegen, wie man den (E/U)KF erweitert, so dass er mit Strukturen, wie $SO(3)$ klar kommt.

Q 43

16. \boxplus -Mannigfaltigkeiten

Wir wollen in diesem Kapitel Sensorfusion betrachten, bei der die gesuchte Größe X die Orientierung eines Objektes im Raum beinhaltet.

16.1. Darstellung von Orientierungen im Raum

Wir benötigen dazu folgende Resultate aus dem Kapitel 15 die hier noch mal aufgeführt sind, falls das Kapitel nicht behandelt wurde.

- Orientierungen im 3-D Raum werden als Rotationsmatrizen $R \in \mathbb{R}^{3 \times 3}$ dargestellt.
- Rotationsmatrizen sind orthonormal ($R^T R = I$) mit Determinante $\det R = 1$.
- Die Menge dieser Rotationsmatrizen heißt

$$SO(n) = \{R \mid R \in \mathbb{R}^{n \times n}, R^T R = I, \det R = 1\} \quad (16.1)$$

- Rotationsmatrizen bestehen aus 9 Zahlen, haben aber nur 3 Freiheitsgrade, weil der Rest durch die 6 Bedingungen in $R^T R = I$ festgelegt ist.
- Es gibt eine Funktion $\text{Rot} : \mathbb{R}^3 \rightarrow SO(3)$ (Matrix-Exponential), wobei $\text{Rot}(v)$ eine Drehung um v mit Winkel $|v|$ ist.
- Rot ist stetig und surjektiv (jede Rotation lässt sich darstellen),
- Es gibt eine teilweise Umkehrfunktion $\text{aRot} : SO(3) \rightarrow \mathbb{R}^3$ mit

$$\text{Rot}(\text{aRot}(R)) = R, \quad (16.2)$$

$$\text{aRot}(\text{Rot}(v)) = v \text{ für } |v| < \pi. \quad (16.3)$$

- Es gibt sogenannte "Singularitäten" v , an denen $\frac{\partial \text{Rot}(v)}{\partial v}$ Rank < 3 hat.
- An diesen Stellen lassen sich kleine Änderungen in $\text{Rot}(v)$ nicht mit kleinen Änderungen in v erreichen. Dies ist analog zum Länge/Breite System der Geodäsie, wo bei einer Bewegung über den Pol die Länge springt.
- Derartige Singularitäten haben alle Funktionen $\mathbb{R}^3 \rightarrow SO(3)$

Soll ein Sensorfusionsalgorithmus so eine Orientierung R schätzen, kann man nicht einfach die 9 Zahlen von R als Unbekannte in X sehen, weil der Algorithmus von den Nebenbedingungen die eine $SO(3)$ -Matrix erfüllen muss nichts weiß. Außerdem erschie-
ne so ein Vorgehen auch wenig effizient, weil man 9 Unbekannte für 3 Freiheitsgrade benötigt.

16. \boxplus -Mannigfaltigkeiten

Sorry, this version does not include copyrighted Material.

Abbildung 16.1.: Der \boxplus -Operator bildet einen Teil des Zustandsraums (hier Sphäre) in der Umgebung eines Zustandes x in den \mathbb{R}^n (hier Ebene) bijektiv ab. Dadurch können Sensorfusionsalgorithmen darauf operieren ohne die topologische Struktur des Zustandsraums explizit zu berücksichtigen. Für einen Referenzzustand x und eine vektorielle Änderung δ ist der geänderte Zustand $x \boxplus \delta$. (auf Folie). Copyright [19].

Deshalb erscheint es naheliegend, die Rotationsmatrix über Parameter darzustellen, z.B. als $R = \text{Rot}(v)$ oder auch über sogenannte Euler-Winkel. Dann versagt der Algorithmus aber an den Singularitäten, weil er Änderungen von R nur durch Änderungen von v erreichen kann und das an diesen Stellen nicht geht.

In diesem Kapitel wollen wir einen systematischen Ansatz herleiten, Strukturen wie $SO(3)$ in Sensorfusionsalgorithmen zu behandeln.

16.2. Motivation: UKF für 3D-Orientierungen

Die Idee ist, solche Zustandsräume, wie z.B. $SO(3)$ zu behandeln, indem man im UKF die Mittelwerte als Zustand (also im Beispiel als 3×3 Matrix) aber deren Kovarianzen als über 3 Parameter beschriebene *Änderungen* darstellt. Es wird also die Orientierung selbst als Matrix und nur kleine Änderungen der Orientierung über Parameter dargestellt. Formal führt das zu allgemeinen Zuständen, so genannten Mannigfaltigkeiten, die aber über Operationen verfügen müssen um kleine Änderungen als Vektor zu parametrisieren.

Dieser Zustandsraum ist für den UKF eine Blackbox. Der UKF greift nur zu über: a) die Dynamik- und Messfunktionen, die ja spezifisch für jedes Messmodell und damit auch für jeden Zustandsraum ist, und b) über einen \boxplus -Operator (Abb. 16.1) der eine kleine als Vektor dargestellte Änderung auf einen Zustand draufrechnet und einen \boxminus -Operator der kleine Änderungen im Zustand als Vektor darstellt.

Alle generischen Formeln des UKF (Sigma-Punkt Propagation, Kalman Gain, Zustandsaktualisierung) stützen sich auf diese vektorielle Darstellung der Änderung relativ zu einem Referenzwert (meist der Mittelwert) ab. Die Eleganz des Ansatzes liegt darin, dass man meistens einfach nur das reguläre $+$ durch \boxplus und $-$ durch \boxminus ersetzen muss.

Wir diskutieren hier überblicksartig die Theorie dieser sogenannten \boxplus -Mannigfaltigkeiten, eine ausführlichere Diskussion findet sich in dem Papier [19].

Christoph Hertzberg, René Wagner, Udo Frese, Lutz Schröder: *Integrating Generic Sensor Fusion Algorithms with Sound State Representations through Encapsulation of Manifolds*, Information Fusion (2012).

16.3. Definition

Definition 21 (\boxplus -Mannigfaltigkeit). Sei $\mathcal{S} \subset \mathbb{R}^s$ ausgestattet mit zwei Operatoren

$$\boxplus_{\mathcal{S}} : \mathcal{S} \times \mathbb{R}^n \rightarrow \mathcal{S}, \quad (16.4)$$

$$\boxminus_{\mathcal{S}} : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}^n. \quad (16.5)$$

Sei außerdem $V \subset \mathbb{R}^n$ eine offene Umgebung von 0 und \boxplus stetig differenzierbar im rechten Argument und \boxminus im linken. Dann ist $(\mathcal{S}, \boxplus, \boxminus, V)$ eine \boxplus -Mannigfaltigkeit wenn folgende Axiome gelten:

$$(x \boxplus 0 = x) \quad (16.6a)$$

$$\forall y \in \mathcal{S} : x \boxplus (y \boxminus x) = y \quad (16.6b)$$

$$\forall \delta \in V : (x \boxplus \delta) \boxminus x = \delta \quad (16.6c)$$

$$\forall \delta_1, \delta_2 \in \mathbb{R}^n : \|(x \boxplus \delta_1) \boxminus (x \boxplus \delta_2)\| \leq \|\delta_1 - \delta_2\|. \quad (16.6d)$$

Das Axiom (16.6a) macht $\delta = 0$ zum neutralen Element von \boxplus . Es ist in Klammern gesetzt, weil es in der ursprünglichen Veröffentlichung gefordert wurde, es sich aber aus (16.6d) und (16.6b) folgern lässt. Axiome (16.6b) und (16.6c) machen \boxplus surjektiv und innerhalb von V injektiv (Links- und Rechtsinverses) und definieren, dass \boxminus die Umkehrung von \boxplus ist. $y \boxminus x$ liefert den Vektor der \boxplus addiert auf x y ergibt. V ist zweckmässigerweise der Bereich in dem \boxplus eindeutig ist.

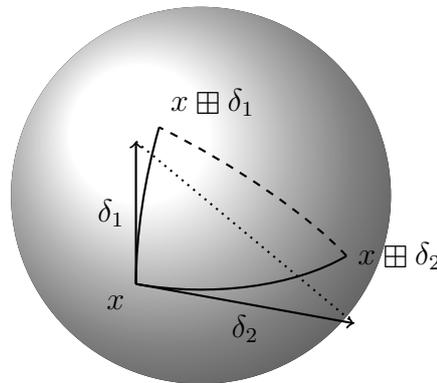


Abbildung 16.2.: Der Abstand zwischen $\mu \boxplus \delta_1$ und $\mu \boxplus \delta_2$ (gestrichelt) ist kleiner als die Distanz zwischen δ_1 und δ_2 in der durch \boxplus zugeordneten Karte um μ (gepunktet). (auf Folie)

Axiom (16.6d) ist schwerer zu verstehen. Es sagt aus, dass die Distanz zwischen $\mu \boxplus \delta_1$ und $\mu \boxplus \delta_2$ in der von \boxplus um μ definierten Karte mindestens so groß ist wie die von \boxminus induzierte Distanz (Abb 16.2). Intuitiv bedeutet das, dass der Effekt einer Änderung durch \boxplus um 0 am größten ist, also nicht für große δ überproportional wächst. Diese Bedingung ist nötig um weiter unten einen Erwartungswert definieren zu können. Die

16. \boxplus -Mannigfaltigkeiten

Gleichung ist mehr als eine Dreiecksgleichung. Letztere hätte man, wenn auf der rechten Seite $\|\delta_1\| + \|\delta_2\|$ stünde.

Außerdem impliziert Axiom (16.6d), dass 0 das neutrale Element von \boxplus ist:

Lemma 16.1. $x \boxplus 0 = x$

Beweis. Die Beweisidee ist die Beobachtung, dass für $\delta_1 = \delta_2$ (16.6d) aussagt, dass \boxminus von gleichen Elementen 0 ist. Formal:

$$x \stackrel{(b)}{=} x \boxplus (x \boxminus x) \quad (16.7)$$

$$\|x \boxminus x\| = \|(x \boxplus (x \boxminus x)) \boxminus (x \boxplus (x \boxminus x))\| \stackrel{(d)}{\leq} \|(x \boxminus x) - (x \boxminus x)\| = 0 \quad (16.8)$$

$$x \boxplus 0 = x \boxplus (x \boxminus x) = x \quad (16.9)$$

□

Bei Vektorräumen ist einfach $\boxplus = +, \boxminus = -$ und alles bleibt beim alten. Bei Rotationen gibt es z. B. zwei Möglichkeiten die im Folgenden diskutiert werden. Im Allgemeinen muss für jede Mannigfaltigkeit eine spezielle Parametrisierung gewählt und als \boxplus, \boxminus definiert werden.

Lemma 16.2. $(\mathbb{R}^n, +, -, \mathbb{R}^n)$, also \mathbb{R}^n mit dem üblichen Operationen ist eine \boxplus -Mannigfaltigkeit.

Beweis. (16.6b), (16.6c) und (16.6d) gelten alle, weil x sich herauskürzt. □

Eine einfache Parametrisierung besteht darin, die relative Orientierung als Matrix Exponential darzustellen.

Definition 22. Parametrisierung von $\mathcal{S} = SO(3)$ durch Relativrotationen in lokalen Koordinaten.

$$x \boxplus \delta = x \cdot \text{Rot}(\delta) \quad (16.10)$$

$$y \boxminus x = \text{aRot}(x^{-1}y) \quad (16.11)$$

Mit dieser Parametrisierung werden die Orientierungen zu einer \boxplus -Mannigfaltigkeit.

Lemma 16.3. Mit Definition 22 wird $(SO(3), \boxplus, \boxminus, \{v \mid |v| < \pi\})$ zu einer \boxplus -Mannigfaltigkeit.

Beweis. Wir rechnen die Axiome (16.6b) und (16.6c) nach:

$$x \boxplus (y \boxminus x) = y \text{Rot}(\text{aRot}(x^{-1}y)) = x(x^{-1}y) = x \quad (16.12)$$

$$(x \boxplus \delta) \boxminus x = \text{aRot}(x^{-1}x \cdot \text{Rot}(\delta)) = \text{aRot}(\text{Rot}(\delta)) \stackrel{|\delta| \leq \pi}{=} \delta \quad (16.13)$$

Der Beweis für das vierte Axiom ist komplizierter und in [19]. □

In realen Anwendungen hat man meist zusammengesetzte Zustandsräume, z.B. wenn bei einer IMU eine Position (\mathbb{R}^3) eine Geschwindigkeit (\mathbb{R}^3) und eine Orientierung ($SO(3)$) zu schätzen ist. Das folgende Lemma zeigt, wie man sich aus \boxplus -Mannigfaltigkeiten für die Teile eine \boxplus -Mannigfaltigkeit für das Ganze zusammenbauen kann. Die Grundidee ist dabei komponentenweise vorzugehen.

Lemma 16.4. Sind $\mathcal{S} = (S, \boxplus_S, \boxminus_S, V_S)$ und $\mathcal{T} = (T, \boxplus_T, \boxminus_T, V_T)$ \boxplus -Mannigfaltigkeit, so lässt sich daraus eine \boxplus -Mannigfaltigkeit für Paare aus S und T bilden:

$$\mathcal{M} = \mathcal{S} \times \mathcal{T} = (\mathcal{S} \times \mathcal{T}, \boxplus_M, \boxminus_M, V_S \times V_T) \quad (16.14)$$

$$\begin{pmatrix} s \\ t \end{pmatrix} \boxplus_M \begin{pmatrix} \delta_s \\ \delta_t \end{pmatrix} = \begin{pmatrix} s \boxplus_S \delta_s \\ t \boxplus_T \delta_t \end{pmatrix}, \quad \begin{pmatrix} s_2 \\ t_2 \end{pmatrix} \boxminus_M \begin{pmatrix} s_1 \\ t_1 \end{pmatrix} = \begin{pmatrix} s_2 \boxminus_S s_1 \\ t_2 \boxminus_T t_1 \end{pmatrix} \quad (16.15)$$

16.3.1. Metrik

Der \boxminus Operator berechnet den Unterschied zwischen zwei Elementen von \mathcal{S} als Vektor. Dessen Länge ist dann ein Maß für den Abstand. Wir zeigen, dass es die Axiome einer Metrik erfüllt.

Definition 23. Sei X eine Menge. Eine Funktion $d : X \times X \rightarrow \mathbb{R}$ heisst Metrik, wenn

$$d(x, y) > 0 \quad \forall x \neq y \quad (16.16)$$

$$d(x, y) = d(y, x) \quad (16.17)$$

$$d(x, z) \leq d(x, y) + d(y, z) \quad (16.18)$$

Lemma 16.5. Der Operator \boxminus definiert eine Metrik d auf \mathcal{S} durch

$$d_{\mathcal{S}}(x, y) := \|y \boxminus x\|. \quad (16.19)$$

Beweis. Die erste Bedingung folgt aus $x \boxplus 0 = x$, weil damit für alle $x \neq y$ $y \boxminus x \neq 0$ ist. Symmetrie folgt aus (16.6d):

$$\begin{aligned} d(x, y) &= \|y \boxminus x\| = \|(y \boxplus 0) \boxminus (y \boxplus (x \boxminus y))\| \\ &\leq \|0 - (x \boxminus y)\| = \|x \boxminus y\| = d(y, x) \end{aligned} \quad (16.20)$$

und anders herum, was Gleichheit impliziert. Die Dreiecksungleichung folgt ebenfalls aus (16.6d):

$$d(x, z) = \|z \boxminus x\| \quad (16.21)$$

$$= \|(y \boxplus (z \boxminus y)) \boxminus (y \boxplus (x \boxminus y))\| \quad (16.22)$$

$$\leq \|(z \boxminus y) - (x \boxminus y)\| \quad (16.23)$$

$$\leq \|(x \boxminus y)\| + \|(z \boxminus y)\| \quad (16.24)$$

$$= d(x, y) + d(y, z) \quad \square$$

Bei den betrachteten Orientierungen in 3D hat die Metrik $d(x, y) = \|x \boxminus y\|$ die Bedeutung eines Winkels. Sie ist der kleinste Winkel mit dem man x auf y drehen kann.

16.4. Wahrscheinlichkeitsrechnung auf Mannigfaltigkeiten

Wir generalisieren jetzt unsere probabilistischen Standardbegriffe – Erwartungswert, Kovarianz und Gaußverteilung – mit Hilfe von \boxplus und \boxminus auf allgemeine Mannigfaltigkeiten. Es muss bemerkt werden, dass diese Begriffe leider nicht mehr die mathematische Strenge besitzen, wie ihre Vektorraum Pendant, Aussagen kann man teilweise nur in Näherung machen. Allgemein kann man aber sagen, dass der dominante Fehler eines Filters eh darin liegt, dass die Funktionen linearisiert werden. Im Verhältnis dazu sind die Probleme dieser Definitionen eher klein.

Definition 24. Sei $(\mathcal{S}, \boxplus, \boxminus, V)$ eine \boxplus -Mannigfaltigkeit und $X, Y : \Omega \rightarrow \mathcal{S}$ Zufallsvariablen in \mathcal{S} . Dann ist definiert.

$$E(X) = \arg \min_{x \in \mathcal{S}} E(\|X \boxminus x\|^2) \quad (16.25)$$

$$\text{Cov}(X, Y) = E((X \boxminus E X)(Y \boxminus E Y)^\top) \quad (16.26)$$

$$\mathcal{N}(\mu, \Sigma) = \mu \boxplus \mathcal{N}(0, \Sigma) \quad (16.27)$$

16.4.1. Kovarianz

Die Definition der \mathcal{S} -Kovarianz (16.26) ist einfach. Sie ist eh definiert über die Differenz zum Mittelwert ins äußere Produkt erhoben. Diese Differenz wird einfach über \boxminus statt – errechnet. Wie im \mathbb{R}^n Fall ist die Kovarianz eine Matrix von Zahlen $\mathbb{R}^{n \times n}$.

Die Intuition aus der Metrik hilft bei der Interpretation von Kovarianzen, weil mit

$$\text{tr Cov}(X) = \sum_i E((X \boxminus EX)_i^2) = E(\|X \boxminus EX\|^2) = E(\|d(X, EX)\|^2) \quad (16.28)$$

$\sqrt{\text{tr Cov}(X)}$ die Unsicherheit von X im Sinne der Interpretation durch d ist. Bei einer Drehung z.B. gibt $\sqrt{\text{tr Cov}(X)}$ den mittleren Drehfehler als Winkel an.

16.4.2. Erwartungswert

Schwieriger ist die Definition des \mathcal{S} -Erwartungswertes, weil \mathcal{S} i. A. keine Möglichkeit bietet, Elemente „aufzusummieren“. Statt dessen verwenden wir eine Eigenschaft, die in Vektorräumen äquivalent ist und auch auf \boxplus -Systemen ausdrückbar ist. In Vektorräumen gilt

$$EX = \arg \min_x E(\|X - x\|^2) \quad (16.29)$$

wie in Übungszettel 1 Aufgabe 3 gezeigt. Der Erwartungswert EX ist gleichzeitig auch der Wert mit der geringsten erwarteten quadratischen Distanz zu X . Der Distanzbegriff lässt sich auch auf \boxplus -Systemen definieren als $\|x \boxminus y\|$. Daraus ergibt sich (16.25) als Definition für den Erwartungswert auf \boxplus -Systemen. Der \mathcal{S} -Erwartungswert ist damit ein Element von \mathcal{S} .

16.4.3. Berechnung des Erwartungswertes

Da mit (16.25) $E(X)$ nicht explizit definiert ist, stellt sich die Frage, wie man ihn berechnet. Eine naheliegende Idee ist, wenn man zu EX will, die erwartete \boxminus Differenz mit \boxplus zu addieren. Letztere ist ein Vektor, so dass man von ihr den Erwartungswert bilden kann. Bei einem Vektorraum wäre man in einem Schritt fertig, bei einer \boxplus -Mannigfaltigkeit muss man von einem Startwert μ_0 aus iterieren.

$$\mu_{k+1} = \mu_k \boxplus E(X \boxminus \mu_k) \quad (16.30)$$

$E(X \boxminus \mu_k)$ ist ein Vektorerwartungswert also konzeptionell kein Problem, bei der praktischen Durchführung stellt sich natürlich noch die Frage, wie er berechnet wird. Wir zeigen jetzt, dass diese Iteration $E(\|X \boxminus \mu_k\|^2)$ minimiert. Wir zeigen formal, dass jeder Schritt des Algorithmus sofern er denn etwas verändert den zu minimierenden Wert $E(\|X \boxminus \mu_k\|^2)$ verringert. Daraus folgt, dass dieser Wert konvergiert aber leider nicht direkt, dass auch die μ_k konvergieren.

Dafür folgern wir zuerst die folgende Formel aus (16.6d):

$$\|y \boxminus (x \boxplus \delta)\| = \|x \boxplus (y \boxminus x) \boxminus (x \boxplus \delta)\| \leq \|(y \boxminus x) - \delta\| \quad (16.31)$$

Bei einem Vektorraum gäbe es vektorielle Gleichheit $y - (x + \delta) = y - x - \delta$, bei einer \boxplus -Mannigfaltigkeit wird der resultierende Vektor höchstens kürzer. Daraus folgt die Ungleichung:

$$E(\|X \boxminus \mu_{k+1}\|^2) = E(\|X \boxminus (\mu_k \boxplus E(X \boxminus \mu_k))\|^2) \quad (16.32)$$

$$\stackrel{(16.31)}{\leq} E(\|(X \boxminus \mu_k) - E(X \boxminus \mu_k)\|^2) \quad (16.33)$$

$$= E(\|(X \boxminus \mu_k)\|^2 - 2(X \boxminus \mu_k)^T E(X \boxminus \mu_k) + \|E(X \boxminus \mu_k)\|^2) \quad (16.34)$$

$$= E(\|(X \boxminus \mu_k)\|^2) - \|E(X \boxminus \mu_k)\|^2 \quad (16.35)$$

$$= E(\|(X \boxminus \mu_k)\|^2) - \|\mu_{k+1} \boxminus \mu_k\|^2 \quad (16.36)$$

Das heisst, solange sich die Iteration verändert wird sie auch besser und am Optimum ist $E(X \boxminus \mu) = 0$. Zudem ist die Iteration bei Vektorräumen in einem Schritt fertig, so dass sie schnell konvergieren sollte, wenn man mit einem geeigneten Startwert startet.

16.4.4. Gaußverteilung

Für die Gaußverteilung wird einfach eine reelle Gaussverteilung $\mathcal{N}(0, \Sigma)$ mit Hilfe von \boxplus auf den \mathcal{S} -Erwartungswert μ addiert. Das passt dazu, dass der \mathcal{S} -Erwartungswert $\in \mathcal{S}$ ist, aber die \mathcal{S} Kovarianz eine $n \times n$ Matrix.

Die Definition für die Gaussglocke auf \boxplus -Systemen hat ein Problem: Man wünscht sich, dass die so definierte Kovarianz der so definierten Gaussglocke wieder Σ ist. Das stimmt aber nur für den Teil von $\mathcal{N}(0, \Sigma)$ der in V liegt, für den also \boxplus eindeutig ist. In diesem Bereich liefert das \boxminus in der Kovarianz wieder den ursprünglichen Wert aus der

Q 44

16. \boxplus -Mannigfaltigkeiten

vektoriellen Gaussglocke. Ausserhalb von V (z.B. für einen Winkel 190°) wird von \boxplus ein anderer kleinerer Wert zurückgeliefert wird (im Beispiel -170°). Das wird dazu führen, dass z.B. die probabilistische Aussagen über quadratische Ausgleichsrechnung nur unter der Vorbedingung gilt, dass alle Gaußschen Rauschwerte in V liegen.

In den praktischen Anwendungen haben wir aber eh Unsicherheiten, die viel kleiner als 180° sind, weil sonst die Linearisierung des UKF nicht mehr genau genug ist. Für entsprechend kleine Ungenauigkeiten, d.h. Kovarianzen ist der Teil der Gaussglocke außerhalb V verschwindend klein und kann vernachlässigt werden, so dass die oben definierte Gaussglocke fast genau eine Kovarianz von Σ hat.

16.4.5. Mahalonobis-Distanz auf \boxplus -Mannigfaltigkeiten

Analog zur Gaußverteilung lässt sich die Mahalonobis-Distanz für \boxplus -Mannigfaltigkeiten definieren:

Definition 25. Mahalonobis-Distanz für \boxplus -Mannigfaltigkeiten Sei $X : \Omega \mapsto \mathcal{S}$ eine Zufallsvariable mit $E(X) = \mu$ und $\text{Cov}(X) = \Sigma$ und $(\mathcal{S}, \boxplus, \boxminus, V)$ ist eine \boxplus -Mannigfaltigkeit, dann ist die Mahalonobis-Distanz von X zu einer Realisierung x bestimmt durch:

$$\Delta(x, X) = \sqrt{(x \boxminus \mu)^T \Sigma^{-1} (x \boxminus \mu)}$$

.

16.4.6. Konsistenz für \boxplus -Mannigfaltigkeiten

Mit der Mahalonobis-Distanz und dem \boxminus -Operator lassen sich naheliegende Konsistenzkriterien für \boxplus -Mannigfaltigkeiten bilden. Ähnlich wie bei beim EKF, kommen durch die Linearisierung Fehler hinzu, wodurch die Idealwerte kaum zu erreichen sind. Zusätzlich, entspricht die Kovarianz der Gaußverteilung auf \boxplus -Mannigfaltigkeiten nicht Σ . Dies sorgt für weitere Abweichungen.

Konsistenzkriterien für \boxplus -Mannigfaltigkeiten

1. Erwartungstreue Schätzung des wahren Wertes $E(X \boxminus \hat{x}) = 0$
2. Übereinstimmung von geschätzter und wahrer Varianz $E(\Delta(X, \hat{X})^2) \approx n$
3. Innovations $(z - E(Z|\bar{\mu}_t))$ ebenfalls erwartungstreu und übereinstimmend
4. Innovations sind unabhängig und Gaußverteilt

16.5. Ausgleichsrechnung in \boxplus -Mannigfaltigkeiten

Algorithmus 16.1 zeigt die Adaption des Gauß-Newton Algorithmus aus Sektion 12.2.3 auf \boxplus -Mannigfaltigkeiten.

Klassischer Gauß-Newton
Gauß-Newton auf \boxplus -Mannigfaltigkeiten

$$f : \mathbb{R}^n \rightarrow \mathbb{R}^m \qquad f : \mathcal{S} \rightarrow \mathcal{M} \qquad (16.37)$$

$$f(X) \sim \mathcal{N}(z, \Sigma) \Leftrightarrow f(X) - z \sim \mathcal{N}(0, \Sigma) \qquad f(X) \sim z \boxplus_M \mathcal{N}(0, \Sigma) \stackrel{*}{\Leftrightarrow} f(X) \boxminus_M z \sim \mathcal{N}(0, \Sigma) \qquad (16.38)$$

Iterationen von Startwert x_0 bis x_i konvergiert:

$$J_{\bullet k} := \frac{f(x_i + \varepsilon e_k) - f(x_i - \varepsilon e_k)}{2\varepsilon} \qquad J_{\bullet k} := \frac{(f(x_i \boxplus_S \varepsilon e_k) \boxminus_M z) - (f(x_i \boxplus_S -\varepsilon e_k) \boxminus_M z)}{2\varepsilon} \qquad (16.39)$$

$$x_{i+1} := x_i - (J^\top \Sigma^{-1} J)^{-1} J^\top \Sigma^{-1} (f(x_i) - z) \qquad x_{i+1} := x_i \boxplus_S - (J^\top \Sigma^{-1} J)^{-1} J^\top \Sigma^{-1} (f(x_i) \boxminus_M z) \qquad (16.40)$$

Tabelle 16.1.: Mit kleinen Änderungen lässt sich der klassische Gauß-Newton Algorithmus für quadratische Ausgleichsrechnung (links) auf \boxplus -Mannigfaltigkeiten (rechts) adaptieren. Veränderung an den Parametern werden mit \boxplus appliziert, sowohl für die numerische Jakobi als auch für das eigentliche Update. Messungen werden im Messraum mit \boxminus verglichen. Anders als im Vektorraum fällt kürzt sich das $\boxminus z$ nicht raus. Die mit * markierte Äquivalenz ist eine Näherung. Ein derartiges Vorgehen ist z.B. in der Bibliothek Ceres [2] als `LocalParametrization` implementiert.

16.6. Sigmapunktpropagation

16.6.1. Generierung

In einem Vektorraum werden Sigmapunkte generiert, indem man \sqrt{n} -fache der Spalten der Choleskyzerlegung der Kovarianzmatrix $LL^T = \Sigma \Leftrightarrow L = \sqrt{\Sigma}$ zum Erwartungswert addiert und von ihm abzieht. Dies wird symbolisch notiert als

$$\mathcal{X} = \left(\mu \quad \mu \quad \mu + \sqrt{(n+1)\Sigma_i} \Big|_{i=1}^n \quad \mu - \sqrt{(n+1)\Sigma_i} \Big|_{i=1}^n \right) \quad (16.41)$$

$$= \left(\mu \quad \mu \quad \mu + \sqrt{(n+1)\Sigma} \quad \mu - \sqrt{(n+1)\Sigma} \right) \quad (16.42)$$

$$= \begin{pmatrix} 2 \times \\ \mu \quad \mu \pm \sqrt{(n+1)\Sigma} \end{pmatrix} \quad (16.43)$$

Die einzelnen Sigmapunkte sind dann $\mathcal{X}^{[i]}$, $i = 1 \dots 2n$. Das Generieren der Sigmapunkte geschieht analog für \boxplus -Mannigfaltigkeiten mit dem \boxplus -Operator:

$$\mathcal{X} = \left(\mu \quad \mu \quad \mu \boxplus \sqrt{n\Sigma_i} \Big|_{i=1}^n \quad \mu \boxplus (-\sqrt{n\Sigma_i}) \Big|_{i=1}^n \right) \quad (16.44)$$

$$= \left(\mu \quad \mu \quad \mu \boxplus \sqrt{n\Sigma} \quad \mu \boxplus (-\sqrt{n\Sigma}) \right) \quad (16.45)$$

$$= \begin{pmatrix} 2 \times \\ \mu \quad \mu \boxplus (\pm\sqrt{(n+1)\Sigma}) \end{pmatrix} \quad (16.46)$$

16.6.2. Propagation

Diese Sigma-Punkte \mathcal{X} werden dann durch die entsprechende Modellfunktion f propagiert (Dynamikfunktion g oder Messfunktion h). Als Modellfunktion ist f ja spezifisch für das Problem und weiß mit den Mannigfaltigkeiten umzugehen. Dies notieren wir als

$$\bar{\mathcal{X}} = f(\mathcal{X}) \quad \Leftrightarrow \quad \bar{\mathcal{X}}^{[i]} = f(\mathcal{X}^{[i]}) \quad (16.47)$$

16.6.3. Erwartungswert und Kovarianzberechnung

Aus den transformierten Sigmapunkten $\bar{\mathcal{X}}$ werden dann der Erwartungswert $E(f(X))$ die Kovarianz $\text{Cov}(f(X))$ und die Kovarianz $\text{Cov}(f(X), X)$ näherungsweise zurückgewonnen. Im Vektorfall geschieht dies mit den Formeln.

$$\bar{\mu}_f = \frac{1}{|\bar{\mathcal{X}}|} \sum_{i=1}^{|\bar{\mathcal{X}}|} \bar{\mathcal{X}}^{[i]} \quad (16.48)$$

$$\bar{\Sigma}^f = \frac{1}{|\bar{\mathcal{X}}|} \sum_{i=1}^{|\bar{\mathcal{X}}|} (\bar{\mathcal{X}}^{[i]} - \bar{\mu}_f)(\bar{\mathcal{X}}^{[i]} - \bar{\mu}_f)^T \quad (16.49)$$

$$\bar{\Sigma}^{xf} = \frac{1}{|\bar{\mathcal{X}}|} \sum_{i=1}^{|\bar{\mathcal{X}}|} (\mathcal{X}^{[i]} - \mu)(\bar{\mathcal{X}}^{[i]} - \bar{\mu}_f)^T \quad (16.50)$$

⊞-Manifold-MeanOfSigmaPoints

Input:

$$\mathcal{Y}^{[i]}, \text{ where } \mathcal{Y}^{[0]} = f(\mu) \quad (16.54)$$

Determine mean μ' :

$$\mu'_0 = \mathcal{Y}^{[0]} \quad (16.55)$$

$$\mu'_{k+1} = \mu'_k \boxplus \frac{1}{|\mathcal{Y}|} \sum_{i=1}^{|\mathcal{Y}|} \mathcal{Y}^{[i]} \boxminus \mu'_k \quad (16.56)$$

$$\mu' = \lim_{k \rightarrow \infty} \mu'_k \quad (16.57)$$

Tabelle 16.2.: MeanOfSigmaPoints berechnet den Mittelwert einer Menge von \boxplus -Mannigfaltigkeits Sigmapunkten \mathcal{Y} . In der Praxis wird der Grenzwert in (16.57) als Schleife implementiert, die abgebrochen wird, wenn die Norm der letzten Änderung unter einen Schwellwert fällt. Als Startwert der Iteration benutzt man den Funktionswert des Erwartungswertes, also den ersten Sigmapunkt. Die Anzahl Sigmapunkte muss nicht der Dimension des Raums entsprechen.

Für den Erwartungswert muss analog zu Kapitel 16.4.3 iteriert werden. Dies ist in Tabelle 16.2 dargestellt. Die Berechnung der Kovarianz erfolgt analog, so wie in (16.26), mit \boxminus an Stelle von $-$.

$$\bar{\mu}_f = \text{MeanOfSigmaPoints}(\bar{\mathcal{X}}) \quad (16.51)$$

$$\bar{\Sigma}^f = \frac{1}{|\bar{\mathcal{X}}|} \sum_{i=1}^{|\bar{\mathcal{X}}|} (\bar{\mathcal{X}}^{[i]} \boxminus \bar{\mu}_f)(\bar{\mathcal{X}}^{[i]} \boxminus \bar{\mu}_f)^T \quad (16.52)$$

$$\bar{\Sigma}^{xf} = \frac{1}{|\bar{\mathcal{X}}|} \sum_{i=1}^{|\bar{\mathcal{X}}|} (\mathcal{X}^{[i]} \boxminus \mu)(\bar{\mathcal{X}}^{[i]} \boxminus \bar{\mu}_f)^T \quad (16.53)$$

16.6.4. Beispiel (am Modell gezeigt)

Man betrachte ein Beispiel mit einer Kamera (C) die sich in einer bestimmten Orientierung $Q_{W \leftarrow C}$ relativ zum Weltkoordinatensystem (W) (vgl. Modell). Die Position halten

16. \boxplus -Mannigfaltigkeiten

wir fest und betrachten sie hier nicht.

$$E(Q_{W \leftarrow C}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & +1 \\ 0 & -1 & 0 \end{pmatrix} = Q \quad (16.58)$$

$$\text{Cov}(Q_{W \leftarrow C}) = \begin{pmatrix} (5^\circ)^2 & 0 & 0 \\ 0 & (5^\circ)^2 & 0 \\ 0 & 0 & (5^\circ)^2 \end{pmatrix} \quad (16.59)$$

$f(Q)$ sei die Funktion $SO(3) \rightarrow \mathbb{R}^2$ die jeder Orientierung die Bildposition eines festen 3D Punktes im Kamerabild zuordnet (dargestellt mit dem Laserpointer). Im UKF kann das später z.B. das Messmodell sein.

Wir wollen jetzt mit Sigmapunktpropagation $E(f(Q_{W \leftarrow C}))$ und $\text{Cov}(f(Q_{W \leftarrow C}))$ näherungsweise ausrechnen.

- Choleskyzerlegung der Kovarianz

$$L = \begin{pmatrix} 5^\circ & 0 & 0 \\ 0 & 5^\circ & 0 \\ 0 & 0 & 5^\circ \end{pmatrix} \quad (16.60)$$

- Sigmapunkte (am Modell gezeigt): $\sqrt{n+1} = \sqrt{4} = 2$

$$\mathcal{X} = \begin{pmatrix} Q & Q \\ Q \boxplus \begin{pmatrix} 10^\circ \\ 0 \\ 0 \end{pmatrix} & Q \boxplus \begin{pmatrix} 0 \\ 10^\circ \\ 0 \end{pmatrix} & Q \boxplus \begin{pmatrix} 0 \\ 0 \\ 10^\circ \end{pmatrix} \\ Q \boxplus \begin{pmatrix} -10^\circ \\ 0 \\ 0 \end{pmatrix} & Q \boxplus \begin{pmatrix} 0 \\ -10^\circ \\ 0 \end{pmatrix} & Q \boxplus \begin{pmatrix} 0 \\ 0 \\ -10^\circ \end{pmatrix} \end{pmatrix} \quad (16.61)$$

$$= \begin{pmatrix} Q & Q \\ Q \text{Rot} \begin{pmatrix} 10^\circ \\ 0 \\ 0 \end{pmatrix} & Q \text{Rot} \begin{pmatrix} 0 \\ 10^\circ \\ 0 \end{pmatrix} & Q \text{Rot} \begin{pmatrix} 0 \\ 0 \\ 10^\circ \end{pmatrix} \\ Q \text{Rot} \begin{pmatrix} -10^\circ \\ 0 \\ 0 \end{pmatrix} & Q \text{Rot} \begin{pmatrix} 0 \\ -10^\circ \\ 0 \end{pmatrix} & Q \text{Rot} \begin{pmatrix} 0 \\ 0 \\ -10^\circ \end{pmatrix} \end{pmatrix} \quad (16.62)$$

- Propagation durch f (am Modell mit Laserpointer, Punkte in der Bildebene mit Stift markiert) Zahlen sind hier beispielhaft für einen Punkt in Z Richtung und eine 200 * 200-Pixelkamera wo $1^\circ \approx 4$ Pixeln entspricht und im Bild x nach rechts und y nach oben zeigt:

$$\bar{\mathcal{X}} = \left(\begin{pmatrix} 100px \\ 100px \end{pmatrix} \begin{pmatrix} 100px \\ 100px \end{pmatrix} \begin{pmatrix} 100px \\ 60px \end{pmatrix} \begin{pmatrix} 140px \\ 100px \end{pmatrix} \begin{pmatrix} 100px \\ 100px \end{pmatrix} \begin{pmatrix} 100px \\ 140px \end{pmatrix} \begin{pmatrix} 60px \\ 100px \end{pmatrix} \begin{pmatrix} 100px \\ 100px \end{pmatrix} \right) \quad (16.63)$$

- Erwartungswert und Kovarianz in \mathbb{R}^2 (am Modell in der Bildebene abgelesen):

$$E(f(Q_{W \leftarrow C})) \approx \begin{pmatrix} 100px \\ 100px \end{pmatrix} \quad (16.64)$$

$$\text{Cov}(f(Q_{W \leftarrow C})) \approx \begin{pmatrix} 400px^2 & 0 \\ 0 & 400px^2 \end{pmatrix} \quad (16.65)$$

$$\text{Cov}(Q_{W \leftarrow C}, f(Q_{W \leftarrow C})) \approx \begin{pmatrix} 0 & -100^\circ px \\ 100^\circ px & 0 \\ 0 & 0 \end{pmatrix} \quad (16.66)$$

16.7. UKF auf \boxplus -Mannigfaltigkeiten

Mit diesen Werkzeugen ausgestattet kann der UKF Filter für \boxplus -Mannigfaltigkeiten generalisiert werden. Im wesentlichen verwendet er die Sigmapunktpropagation für \boxplus -Mannigfaltigkeiten wie in Kapitel 16.6, \boxminus um den Unterschied zwischen Sigmapunkt und Mittelwert auszurechnen und \boxplus um beim Messschritt die Änderung auf den Zustand anzuwenden. Letzteres ist eine Näherung, die funktioniert, wenn \boxplus in beiden Argumenten stetig ist, wie in unserem Fall. Im Allgemeinfall muss eine weitere Sigmapunktpropagation angewandt werden [19, Table 2].

A.26

A.27

16. \boxplus -Mannigfaltigkeiten

Klassischer UKF

UKF auf \boxplus -Mannigfaltigkeiten

Annahmen:

$$\mu_{t-1} \in \mathbb{R}^n, \Sigma_{t-1} \in \mathbb{R}^{n \times n}, u_t \in \mathbb{R}^p, z_t \in \mathbb{R}^m \quad \mu_{t-1} \in \mathcal{S}, \Sigma_{t-1} \in \mathbb{R}^{n \times n}, u_t \in \mathbb{R}^p, z_t \in \mathcal{M} \quad (16.67)$$

$$\begin{aligned} g: \mathbb{R}^n \times \mathbb{R}^p &\rightarrow \mathbb{R}^n, & g: \mathcal{S} \times \mathbb{R}^p &\rightarrow \mathcal{S}, & (16.68) \\ X_t = g(X_{t-1}, u_t) &+ \mathcal{N}(0, \Sigma_{ct}) & X_t = g(X_{t-1}, u_t) &\boxplus_S \mathcal{N}(0, \Sigma_{ct}) & (16.69) \end{aligned}$$

$$h: \mathbb{R}^n \rightarrow \mathbb{R}^m, z_t = h(X_t) + \mathcal{N}(0, \Sigma_{\delta t}) \quad h: \mathcal{S} \rightarrow \mathcal{M}, z_t = h(X_t) \boxplus_M \mathcal{N}(0, \Sigma_{\delta t}) \quad (16.70)$$

Dynamikschritt (Prediction) Step:

$$\mathcal{X}_{t-1} = \left(\begin{array}{c} 2 \times \\ \mu_{t-1} \end{array} \quad \mu_{t-1} \pm \sqrt{(n+1)\Sigma_{t-1}} \right) \quad \mathcal{X}_{t-1} = \left(\begin{array}{c} 2 \times \\ \mu \end{array} \quad \mu_{t-1} \boxplus_M (\pm \sqrt{(n+1)\Sigma_{t-1}}) \right) \quad (16.71)$$

$$\bar{\mathcal{X}}_t^* = g(\mathcal{X}_{t-1}, u_t) \quad \bar{\mathcal{X}}_t^* = g(\mathcal{X}_{t-1}, u_t) \quad (16.72)$$

$$\bar{\mu}_t = \frac{1}{|\bar{\mathcal{X}}_t^*|} \sum_{i=1}^{|\bar{\mathcal{X}}_t^*|} \bar{\mathcal{X}}_t^{*[i]} \quad \bar{\mu}_t = \text{MeanOfSigmaPoints}(\bar{\mathcal{X}}_t^*) \quad (16.73)$$

$$\bar{\Sigma}_t = \frac{1}{|\bar{\mathcal{X}}_t^*|} \sum_{i=1}^{|\bar{\mathcal{X}}_t^*|} (\bar{\mathcal{X}}_t^{*[i]} - \bar{\mu}_t)(\bar{\mathcal{X}}_t^{*[i]} - \bar{\mu}_t)^\top + \Sigma_{ct} \quad \bar{\Sigma}_t = \frac{1}{|\bar{\mathcal{X}}_t^*|} \sum_{i=1}^{|\bar{\mathcal{X}}_t^*|} (\bar{\mathcal{X}}_t^{*[i]} \boxminus \bar{\mu}_t)(\bar{\mathcal{X}}_t^{*[i]} \boxminus \bar{\mu}_t)^\top + \Sigma_{ct} \quad (16.74)$$

Messschritt (Correction) Step:

$$\bar{\mathcal{X}}_t = \left(\begin{array}{c} 2 \times \\ \bar{\mu}_t \end{array} \quad \bar{\mu}_t \pm \sqrt{(n+1)\bar{\Sigma}_t} \right) \quad \bar{\mathcal{X}}_t = \left(\begin{array}{c} 2 \times \\ \bar{\mu}_t \end{array} \quad \bar{\mu}_t \boxplus (\pm \sqrt{(n+1)\bar{\Sigma}_t}) \right) \quad (16.75)$$

$$\bar{\mathcal{Z}}_t = h(\bar{\mathcal{X}}_t) \quad \bar{\mathcal{Z}}_t = h(\bar{\mathcal{X}}_t) \quad (16.76)$$

$$\hat{z}_t = \frac{1}{|\bar{\mathcal{Z}}_t|} \sum_{i=1}^{|\bar{\mathcal{Z}}_t|} \bar{\mathcal{Z}}_t^{[i]} \quad \hat{z}_t = \text{MeanOfSigmaPoints}(\bar{\mathcal{Z}}) \quad (16.77)$$

$$\Sigma_t^z = \frac{1}{|\bar{\mathcal{Z}}_t|} \sum_{i=1}^{|\bar{\mathcal{Z}}_t|} (\bar{\mathcal{Z}}_t^{[i]} - \hat{z}_t)(\bar{\mathcal{Z}}_t^{[i]} - \hat{z}_t)^\top + \Sigma_{\delta t} \quad \Sigma_t^z = \frac{1}{|\bar{\mathcal{Z}}_t|} \sum_{i=1}^{|\bar{\mathcal{Z}}_t|} (\bar{\mathcal{Z}}_t^{[i]} \boxminus_M \hat{z}_t)(\bar{\mathcal{Z}}_t^{[i]} \boxminus_M \hat{z}_t)^\top + \Sigma_{\delta t} \quad (16.78)$$

$$\bar{\Sigma}_t^{x,z} = \frac{1}{|\bar{\mathcal{Z}}_t|} \sum_{i=1}^{|\bar{\mathcal{Z}}_t|} (\bar{\mathcal{X}}_t^{[i]} - \bar{\mu}_t)(\bar{\mathcal{Z}}_t^{[i]} - \hat{z}_t)^\top \quad \bar{\Sigma}_t^{x,z} = \frac{1}{|\bar{\mathcal{Z}}_t|} \sum_{i=1}^{|\bar{\mathcal{Z}}_t|} (\bar{\mathcal{X}}_t^{[i]} \boxminus_S \bar{\mu}_t)(\bar{\mathcal{Z}}_t^{[i]} \boxminus_M \hat{z}_t)^\top \quad (16.79)$$

$$K_t = \bar{\Sigma}_t^{x,z} (\Sigma_t^z)^{-1} \quad K_t = \bar{\Sigma}_t^{x,z} (\Sigma_t^z)^{-1} \quad (16.80)$$

$$\mu_t = \bar{\mu}_t + K_t(z_t - \hat{z}_t) \quad \mu_t = \bar{\mu}_t \boxplus_S K_t(z_t \boxminus_M \hat{z}_t) \quad (16.81)$$

$$\Sigma_t = \bar{\Sigma}_t - K_t(\Sigma_t^{x,z})^\top \quad \Sigma_t = \bar{\Sigma}_t - K_t(\Sigma_t^{x,z})^\top \quad (16.82)$$

Tabelle 16.3.: Klassischer UKF und UKF auf \boxplus -Mannigfaltigkeiten. Vergleiche Text.

17. Beobachtbarkeit von Zuständen

Leitfrage: *Kann ich messen was ich messen will?*

Wenn man ein System auslegt, muss man sich immer die Frage stellen, ob es leisten kann was es leisten soll. Im Falle der Sensorfusion, soll der Algorithmus den gesuchten Zustand möglichst genau schätzen können. Häufig, wird das Ergebnis der Zustandsschätzung weiterverwendet. Z.B. um einen Heißluftballon möglichst sanft zu landen oder ein Auto autonom zu fahren. In diesem Fall definiert die Anwendung, wie genau die Schätzung sein muss.

Wir haben schon beim 1D Kohlenkarren auf der Schiene festgestellt, dass Messungen nicht überall gleich genau sind, auch wenn die Sensoren immer das gleiche Messrauschen aufweisen. Das liegt beim 1D Beispiel daran, dass ein Winkelfehler von 1° rechts auf der Schiene einen größeren Positionsfehler verursacht als links. Der EKF konnte diesen Zusammenhang über die Ableitungen des Messmodells selbstständig berücksichtigen. Am Ende ist es so: Je flacher die Kurve (also je kleiner die Ableitung) desto größer wird ein Fehler in der Messung sich im Zustand zeigen. Was passiert aber, wenn die Messung ganz flach wird, also die Ableitung 0 wird? Theoretisch wird dadurch auch ein kleiner Messfehler zu einem unendlichen Fehler im Zustand. Das zeigt uns schonmal eine erste Situation, in denen die Schätzung ungenau wird. Aber es gibt noch weitere.

In den meisten Anwendungen wird der gesuchte Zustand nicht direkt gemessen. Erinnern wir uns an den Kohlekarren in 2D: Statt die 2D-Position zu messen wird der Winkel zu den Funkbaken gemessen. Der EKF/UKF erleichtert uns daraus die Position auszurechnen, da er auf dem Messmodell $z = h(x)$ arbeitet. Dieses ist in den meisten Fällen leichter aufzustellen, als das inverse Modell $x = h^{-1}(z)$. In der Simulation des 2D Kohlenkarrens kann man feststellen, dass die Schätzung der Pose schlecht wird, wenn der Kohlenkarren nur eine der Funkbaken sieht. Sieht er mehrere oder sogar alle nacheinander (man dreht sich auf der Stelle), erhält man aber eine sehr genaue Schätzung der Pose. Diese Beobachtung impliziert, dass die Messung zu einer Funkbarke nicht ausreicht, um den Zustand korrekt zu rekonstruieren. Hier stellt sich die Frage, wie viele Funkbaken eigentlich nötig sind, um die Pose rekonstruieren zu können. Also wie viele Funkbaken muss man aufstellen, damit ein solcher Kohlenkarren sich immer gut Orten kann.

Q 46

17.1. Beobachtbarkeit

In diesem Kapitel wollen wir diese beiden Phänomene betrachten: Also was bedeutet es, dass die Ableitung der Messung 0 wird und wie findet man raus wie viele Funkbaken nötig sind um die Pose zu schätzen.

Dazu benutzen wir den Formalismus der Beobachtbarkeit. Wenn man prüft ob ein System Beobachtbar ist, sucht man die Antwort auf die Frage:

Kann man den Zustand X mit den gegebenen Messungen U_t und Z_t eindeutig bestimmen ?

Wir sagen also, dass ein System beobachtbar ist, wenn es nur einen Zustand X gibt, der die Messungen U_t und Z_t erklärt. „Eindeutig“ ist dabei das Schlüsselwort. Formal definieren wir Beobachtbarkeit als:

Definition 26. Beobachtbarkeit Ein Input-Output System ist beobachtbar an einem Punkt $x(0) \in \mathcal{M}$ wenn es mindestens einen Satz Inputs u_t gibt, sodass $x(0)$ aus den Outputs $z(t)$ und den Inputs u_t in dem Zeitintervall \mathcal{I} eindeutig bestimmt werden kann. Zusätzlich ist das System auf einer gegebenen Menge $\mathcal{U} \subseteq \mathcal{M}$ beobachtbar, wenn es an jedem $x \in \mathcal{U}$ beobachtbar ist.

Nach: [29]

In der Definition ist von einem Zeitintervall die Rede. Für Beobachtbarkeit ist es nicht notwendig, dass der Zustand aus einer einzelnen Messung eindeutig bestimmt werden kann. Es geht vielmehr darum, dass es zu einer Reihe von Messungen (zeitlich nacheinander) nur eine plausible Reihe von Zuständen gibt.

Schauen wir uns ein Beispiel an: Wir betrachten wieder den Kohlenkarren in der 2D-Ebene. Zur Vereinfachung nehmen wir statt einer Winkelmessung eine Abstandsmessung zur Funkbake (z.B. als Time-of-Flight Messung wie im GPS).

Bei unserer Betrachtung ob das System beobachtbar ist, machen wir 2 Änderungen zu unseren bisherigen Modellierungen. Zum einen betrachten wir statt die Modelle zeitkontinuierlich statt zeitdiskret. Wir nehmen also an, wir erhalten Messungen zu jeder Zeit t anstatt sie lediglich in Intervallen zu erhalten. Zudem nehmen wir an, dass wir Messungen ohne Sensorrauschen erhalten. Diese beiden Annahmen sind notwendig, damit überhaupt Eindeutigkeit erreicht werden kann. Würde man mit Sensorrauschen argumentieren, wäre jeder Zustand bei jeder Messung möglich, da die Gaußverteilung jedem Wert eine Wahrscheinlichkeit über 0 zuordnet. In einem diskretisierten System werden die Messungen U_t über Zeitintervalle gemittelt, wodurch beliebige Trajektorien zwischen 2 Messpunkten möglich wären, die alle nicht unterscheidbar sind.

Überprüft man die Beobachtbarkeit des Zustandes mit diesen beiden Vereinfachungen führt das dazu, dass der Zustand eines beobachtbaren Systems dennoch nicht perfekt geschätzt werden kann. Zu der grundsätzlichen Eignung der Messungen um den Zustand zu bestimmen mischen sich noch die Messfehler und die Diskretisierungsfehler.

Mit diesen Vereinfachungen benutzen wir das Zeitkontinuierliche Modell des Kohlen-

$$\text{karrens in 2D: } X_t = \begin{pmatrix} p_x \\ p_y \\ \theta \end{pmatrix}, U_t = \begin{pmatrix} v \\ \omega \end{pmatrix}, Z_t = \sqrt{p_x^2 + p_y^2}, \dot{X}_t = \begin{pmatrix} v \cos(\theta) \\ v \sin(\theta) \\ \omega \end{pmatrix}$$

Überprüfen wir nun, ob der Zustand aus den Messungen eindeutig rekonstruierbar ist. Grundsätzlich ist eine Abstandsmessung unabhängig von der Richtung in der der Kohlenkarren zur Funkbake steht. Daher reicht eine einzelne Messung nicht um die Position eindeutig zu bestimmen. Jeder Punkt der den selben Abstand zur Bake hat (ein Kreis um die Bake) wäre eine mögliche Position (siehe Abbildung 17.1).

Wie ist es aber wenn wir eine Zeitreihe von Abstandsmessungen nehmen? Auch in diesem Fall gibt es für jede weitere weitere Abstandsmessung eine Vielzahl an möglichen

Sorry, this version does not include copyrighted Material.

Abbildung 17.1.: Beobachtbarkeit des 2D Kohlenkarrens mit einer Abstandsmessung.
Copyright [29]

Positionen die den gleichen Abstand zur Bake haben. Es gibt also für eine Reihe von Messungen unendlich viele mögliche Reihen von Zuständen. Diese sind die richtige Trajektorie verdreht um die Funkbake.

Es ist allerdings möglich, dass die Übergangsmessung U_t dazu führen, dass der Zustand beobachtbar wird. Die Idee dabei ist folgende: Es mag zwar mehrere Reihen an Zuständen geben, die mit der Reihe an Messungen übereinstimmen, aber die Übergänge innerhalb der Zustandsreihe müssen auch mit den Übergangsmessungen übereinstimmen. Wenn z.B. eine Fahrstrecke von 1 m gemessen wird, kann man alle Reihen von Zuständen ausschließen, wo die Positionen mehr als 1 m auseinander liegen müssten um mit den Messungen übereinzustimmen. Gleiches gilt auch für Verdrehungen, Geschwindigkeiten etc.

In unserem Beispiel ist dies allerdings nicht der Fall. Wenn man nicht nur die Positionen um die Funkbake dreht, sondern auch noch die Startorientierung, passen Relationen der Positionen und Orientierungen weiterhin zu den Übergangsmessungen. Daher ist der Zustand nicht beobachtbar.

Nicht beobachtbar heißt aber nicht, dass gar keine Information über den Zustand vorliegt. Der Zustand ist lediglich unvollständig schätzbar. In dem Beispiel haben wir herausgefunden, dass wir das ganze System um die Funkbake herum drehen können, ohne dass sich die Messungen ändern. Diese Drehung des Systems können wir als Transformation des Systems auffassen und ein neues System aufstellen:

$$\tilde{X}_t = \begin{pmatrix} \cos(\tau)p_x - \sin(\tau)p_y \\ \sin(\tau)p_x + \cos(\tau)p_y \\ \theta + \tau \end{pmatrix} \Rightarrow Z_t|X_t = Z_t|\tilde{X}, U_t|X_t = U_t|\tilde{X}$$

Das System \tilde{X}_t ist aus dem Ursprungssystem entstanden, durch eine Rotation um den Winkel τ . Es erzeugt aber dieselben Messungen und Übergangsmessungen wie das System x_t . Eine solche Transformation nennen wir die unbeobachtbare Transformation.

Definition 27. Ist der Zustand eines Input-Output-System $x_t \in \mathcal{M}$ durch gegebene Inputs u_t und Outputs $z_t = h(x_t)$ mit $\frac{\partial x_t}{\partial t} = g(x_t, u_t)$ eines Intervalls \mathcal{I} nicht beobachtbar, dann nennen wir eine Transformation $A : \mathcal{M} \mapsto \mathcal{M}$ die unbeobachtbare Transformation wenn gilt:

$$h(x_t) = h(A(x_t)), \quad \frac{\partial A(x_t)}{\partial t} = g(A(x_t), u_t)$$

Die Anzahl der Parameter der Transformation A entspricht dabei der Anzahl der unbeobachtbaren Freiheitsgrade des Zustandes x_t .

Die gezeigte Transformation hängt nur von einem Parameter τ ab, der den Winkel der Rotation beschreibt. Die Anzahl der frei wählbaren Parameter in einer solchen Transformation kommt die besondere Rolle zu, auszusagen, wie viele Freiheitsgrade unbeobachtbar sind. D.h. in unserem Beispiel, bleibt 1 Freiheitsgrad unbeobachtbar. Dieser

17. Beobachtbarkeit von Zuständen

Sorry, this version does not include copyrighted Material.

Abbildung 17.2.: Beobachtbarkeit des 2D Kohlenkarrens mit einer Winkelmessung in Körperkoordinaten. Copyright [29]

Freiheitsgrad muss nicht immer zu einem Zustand gehören. Er kann auch ausdrücken, das wir z.B. die Differenz von x und y kennen, aber nicht die absoluten Werte. Der Freiheitsgrad gibt einen Hinweis darauf, welche Messung man dem System hinzufügen muss, damit es beobachtbar wird: In diesem Fall eine Messung die einem den absoluten Winkel zum Ursprung erlaubt.

Weitere Beispiele (Kohlekarren mit anderer Messung):

1. Messung $Z_t = \beta$ (Winkel zu einer Funkbake in Körperkoordinaten)

Prinzipiell heißt eine Winkelmessung, dass man beliebig weit weg von der Bake sein kann und die gleiche Messung erhält. Daher wird dieser Fall zuerst betrachtet. Man Stelle sich zwei Startpositionen A und B vor, die auf dem gleichen Winkel zur Bake liegen (und daher auch die gleiche Orientierung haben da der Winkel in Körperkoordinaten gemessen wird). A ist dabei die wahre Position und B eine falsche (siehe Abbildung 17.2). Wenn man von beiden Positionen A und B aus dieselbe Strecke U weiterfährt, dann sind die Punkte an denen man ankommt auf unterschiedlichen Winkeln zur Funkbake. Da über die Dynamikmessungen die Fahrstrecke gemessen wird, kann man damit B ausschließen. Daher ist der Abstand nicht unbeobachtbar. Das liegt daran, dass das Fahrzeug wenn man weiter weg ist von der Bake einen weiteren Weg fahren muss, um dieselbe Winkeländerung zu erhalten und einen kürzeren wenn es näher an der Bake ist. Demnach enthält die Winkelmessung zusammen mit den Übergangsmessungen auch die Abstandsmessungen.

Nun ist das Problem, dass die Messung in Körperkoordinaten ist. Dadurch kann man erneut das System um die Funkbake rotieren, sodass sich die Position ändert und der gemessene relative Winkel gleich bleibt. Man erhält dieselbe unbeobachtbare Transformation wie bei der Abstandsmessung.

2. Messung $Z_t = (\beta, \rho)$ (Winkelmessung in Körperkoordinaten und Abstandsmessung)

Spannenderweise wird der Zustand auch nicht beobachtbar, wenn man Abstands- und Winkelmessung kombiniert. Da beide Messungen die Drehung um die Bake nicht beobachten können, kann man das System wieder um die Funkbake drehen, ohne dass sich die Messungen ändern würden. Es bringt daher nichts, Messungen zu kombinieren, die die gleiche unbeobachtbare Transformation haben.

3. Messung $Z_t = (\rho \cos(\beta), \rho \sin(\beta))$ Messung der Position der Funkbake in Kartesischen Koordinaten

Diese Messung ist prinzipiell dieselbe wie die vorherige. Es ändert sich lediglich die Darstellung der Zahlen, bzw. das Koordinatensystem. Statt in Polarkoordinaten wird die Messung in kartesischen Koordinaten dargestellt. Daher kann man auch

Sorry, this version does not include copyrighted Material.

Abbildung 17.3.: Beobachtbarkeit des 2D Kohlenkarrens mit einer Winkelmessung in Weltkoordinaten. Copyright [29]

hier wieder eine Rotation um die Bake ausführen ohne die Messung zu ändern. Eine Änderung des Koordinatensystems ändert also ebenfalls nichts an der Beobachtbarkeit, da man die Messung in einem Koordinatensystem vollständig aus der Messung im anderen Koordinatensystem berechnen kann und somit keine neuen Informationen im System vorliegen.

4. Messung $Z_t = \varphi = \text{atan2}(p_y, p_x)$ (Winkelmessung in Weltkoordinaten)
Dieses System ist das einzige von den Beispielen, das beobachtbar ist. Der Unterschied zur Winkelmessung in Körperkoordinaten ist, dass ein Kohlekarren, der den selben Winkel misst damit automatisch auch auf demselben Winkel steht, da er dies nicht durch seine Orientierung ausgleichen kann. Dafür ist die Orientierung durch eine einzelne Messung nicht festgelegt. Nimmt man wieder 2 Startpunkte A und B (wobei A der echte Zustand ist), und eine Fahrstreckenmessung U könnte der Startzustand B zur selben 2. Winkelmessung gelangen indem die Startorientierung entsprechend angepasst wird (siehe Abbildung 17.3). Bei einer 3. Messung mit einer bekannten Fahrstrecke ist dies aber nicht mehr der Fall. Daher ist auch hier der Abstand zur Funkbake aus Messung und Übergangsmessung herleitbar. Da der Abstand der Bake und der Winkel bekannt sind, kann die Position eindeutig berechnet werden.

17.2. Schwache Beobachtbarkeit

Die geometrischen Argumente des letzten Kapitels sind anschaulich, dafür aber schwer zu finden. Zudem gibt es keine Garantie, dass man durch geometrische Argumente die unbeobachtbare Transformation auch findet. D.h. nur weil man keine Transformation gefunden hat, gibt es auch keine. Das macht diese Art der Prüfung unsicher. Ebenso schwierig ist es Beobachtbarkeit zu zeigen, wenn man dafür mehrere Messungen berücksichtigen muss, wie im letzten Beispiel. Deshalb wird ein Prüfkriterium benötigt, mit dem man eine klare Ja/Nein Aussage treffen kann ob ein System beobachtbar ist.

Zunächst benötigen wir eine Möglichkeit die Messgleichungen Systematisch darzustellen und dabei zu berücksichtigen, dass wir die Messungen eines ganzen Zeitintervalls benutzen können. Dazu ziehen wir das Taylor-Theorem ran:

$$z(t) = \sum_{k=0}^{\infty} \frac{d^k z(t)}{dt^k} \Big|_{t=0} \frac{t^k}{k!}$$

Dieses sagt aus, dass wir eine Funktion korrekt darstellen können wenn wir alle (unendlich viele) Ableitungen der Funktion zum Zeitpunkt $t = 0$ kennen. Daher können wir unsere Messungen aus einem Zeitintervall auch als Vektor der unendlich vielen Ableitungen

17. Beobachtbarkeit von Zuständen

gen darstellen. Klingt erst einmal komplizierter erleichtert aber den Beobachtbarkeits-test ungemein. Die zeitlichen Ableitungen berechnen wir über den Umweg, dass wir das Messmodell nach den Zuständen ableiten und mit der zeitlichen Ableitung der Zustände multiplizieren (Kettenregel). Die zeitliche Ableitung der Zustände entspricht dabei dem kontinuierlichen Dynamikmodell:

$$Z(t) = \begin{pmatrix} z(t) \\ z^{(1)}(t) \\ \vdots \\ z^{(\infty)}(t) \end{pmatrix} = \begin{pmatrix} h(X(t)) \\ \frac{dh(X(t))}{dX(t)} g(x, u) \\ \frac{d^2z}{dx^2} \cdot g(x, u) + \frac{dz}{dx} \cdot \left(\frac{dg(x, u)}{dx} \frac{dx}{dt} + \frac{dg(x, u)}{du} \frac{du}{dt} \right) \\ \vdots \end{pmatrix}$$

Dieser Vektor ist jetzt eine Funktion von dem Zustand und den Übergangsmessungen, da diese im Dynamikmodell vorkommen. Dadurch werden in dieser Darstellung auch automatisch die Informationen die durch die Übergangsmessungen hinzukommen berücksichtigt. Da es sich um eine Funktion vom Zustand handelt, können wir den Zustand erhalten, indem wir die Inverse der Funktion anwenden:

$$Z(t) = H(X(t)) \Rightarrow X(t) = H^{-1}(H(X(t)))$$

Demnach ist der Zustand beobachtbar, wenn die Inverse existiert. Und damit führen wir die Beobachtbarkeit auf ein klassisches Problem der Mathematik zurück: Das Inverse Function Theorem.

Definition 28. Inverse Function Theorem Sei $f : D \mapsto M^d$ eine freie polynomielle Abbildung. Die folgenden Aussagen sind gleichwertig:

1. $Df(X)$ ist eine nicht-singuläre Abbildung $X \in D$.
2. f ist injektiv.
3. f^{-1} existiert und ist eine freie Abbildung.

Nach: [33]

Der Kern des Theorems ist, dass die inverse einer Funktion **lokal** existiert, wenn die Jacobi-Matrix vollen Rang hat. Das Wort lokal macht hier auf eine interessante Eigenschaft aufmerksam: Viele Funktionen sind periodisch. D.h. sie haben wie zum Beispiel der Sinus ein immer wieder kehrendes Muster. Innerhalb einer Periodendauer, kann man von der Höhe auf den entsprechenden Winkel schließen. Aber es gibt unendlich viele Winkel, die dieselbe Höhenmessung erzeugen. Demnach kann man eine Inverse nur unter der Voraussetzung finden, dass man die Periode kennt, in der sich der Winkel befindet. Oder anders gesagt: Man braucht eine ausreichend genaue Schätzung des Winkels um die Inverse finden zu können. Dies führt ebenfalls zu einer schwächeren Form der Beobachtbarkeit:

Definition 29. Schwache Beobachtbarkeit Ein Input-Output System ist schwach beobachtbar an einem Punkt $x(0) \in \mathcal{M}$ wenn es eine offene Menge B von x_0 gibt, sodass es mit dem Wissen $x(0) \in B$ mindestens einen Satz Inputs u_t gibt, sodass $x(0)$ aus den

Outputs $z(t)$ und den Inputs u_t in dem Zeitintervall \mathcal{I} eindeutig bestimmt werden kann. Zusätzlich ist das System auf einer gegebenen Menge $\mathcal{U} \subseteq \mathcal{M}$ schwach beobachtbar, wenn es an jedem $x \in \mathcal{U}$ schwach beobachtbar ist.

Nach: [29]

Warum macht eine solche Definition für unseren Kontext sinn? Wenn der Kohlenkarren Winkelmessungen zu 2 Funkbaken erhält, dann lassen sich über 2 aufeinander folgende Zeitschritte alle Optionen auf der selben Seite der Baken ausschließen (Siehe Abbildung 17.4). Aber es gibt noch die gegenüberliegende Seite, die man durch eine Punktspiegelung erhält. Sie ergibt die gleichen Messungen. In diesem Fall würde aber eine sehr grobe Information wie „Der Kohlenkarren befindet sich oberhalb der Baken“ ausreichen um den Zustand eindeutig zu rekonstruieren. Im Sensorfusionskontext haben wir diese grobe Schätzung, da wir mit der Initialisierung der Pose die ungefähre Position und Orientierung erhalten. Durch das iterative Vorgehen der KF-Varianten gibt es zu jeder Zeit eine grobe Schätzung die ausreichen sollte, um zu unterscheiden ob der Karren ober- oder unterhalb der Baken ist. Das ist genau, was die Schwache Beobachtbarkeit erlaubt. Es ist also eine Definition der Beobachtbarkeit die hervorragend zu den Sensorfusionsproblemen mit passt.

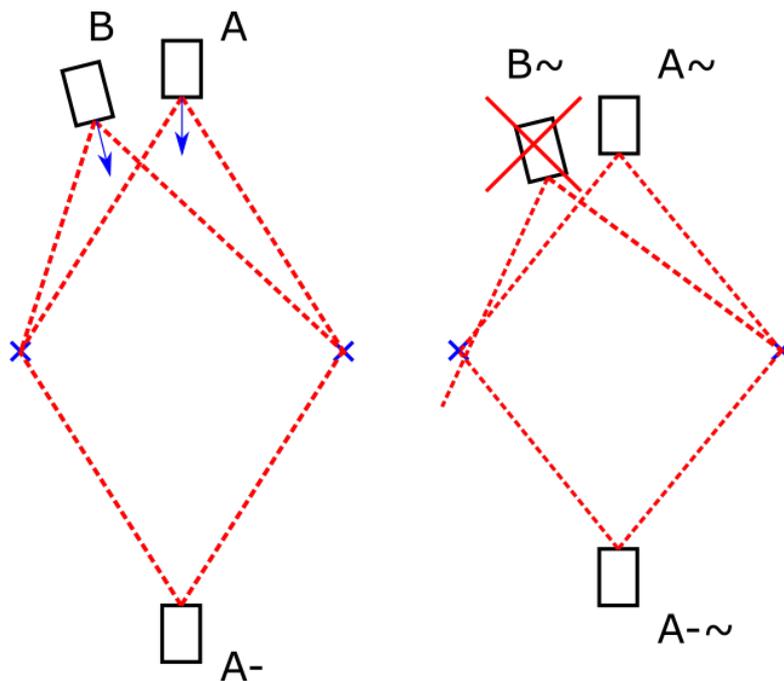


Abbildung 17.4.: Uneindeutigkeit der Pose bei Winkelmessung zu 2 Funkbaken (Blaue Kreuze). Posen auf der selben Seite von den Baken (B) lassen sich durch die Übergangsmessung und 2 aufeinander folgende Messungen ausschließen. Die Punktspiegelung (A-) passt allerdings zu den Messungen.

17. Beobachtbarkeit von Zuständen

Das Inverse Function Theorem liefert direkt eine Methode, um auf Schwache Beobachtbarkeit zu prüfen:

Korollar 17.1. *Test: Schwache Beobachtbarkeit* Der Zustand x ist schwach beobachtbar an $x(0)$, wenn die Jacobi Matrix $\left. \frac{\partial Z(x)}{\partial x} \right|_{x=x(0)}$ vollen Rang hat.

Q 49

Nach: [1]

Um diesen Test praktisch durchzuführen, kann man verschiedene Rangtests aus der linearen Algebra benutzen. Eine Matrix hat vollen Rang wenn:

1. $\det(M) \neq 0$
2. Alle n Spalten von M sind linear unabhängig
3. n Zeilen von M sind linear unabhängig
4. Der Kern(Nullspace) von M ist leer ($v \in \text{Kern}(M) \Rightarrow M * v = \vec{0}$)

Auch wenn Test 4. am kompliziertesten klingt ist er eigentlich der anschaulichste, um zu zeigen was schwach beobachtbar bedeutet: Die Jacobi-Matrix hat eine ähnliche Eigenschaft wie die Kovarianzmatrix. Mit ihr lassen sich Richtungsableitungen bestimmen. Die Jacobi-Matrix J multipliziert mit einem Richtungsvektor v , $v^T v = 1$ ergibt die Änderung der Funktion Z , wenn man sich in Richtung v bewegt. Der sogenannte Kern der Jacobi-Matrix enthält nun alle Richtungsvektoren, bei denen die Operation $J * v = 0$ ist, d.h. alle Richtungen in die sich der Zustand ändern kann, ohne dass sich die Funktion Z ändert. Sollte es solche Richtungen geben, kann sich der Zustand also ändern, ohne dass dies in den Messungen sichtbar ist. Dementsprechend wäre die Änderung im Zustand nicht beobachtbar und somit auch nicht der Zustand. In der praktischen Durchführung ist es allerdings schwierig den Kern der Jacobi-Matrix zu finden, besonders wenn man zeigen möchte, dass er leer ist (das System also beobachtbar).

Q 50

17.3. Beobachtbarkeitsrangkriterium (Observability rank criteria)

Das bisherige Rangkriterium hat 2 wesentliche Probleme: Zum einen, werden die Ableitungen sehr schnell sehr lang, da für die zeitlichen Ableitungen immer wieder mit dem Dynamikmodell multipliziert werden muss und die Übergangsmessungen immer weiter abgeleitet werden. Die Kettenregel führt dabei zu schlecht handhabbaren langen Ausdrücken. Zum anderen eignet sich der Test schlecht, um zu beweisen das ein System unbeobachtbar ist. Theoretisch müsste man unendlich viele Ableitungen bilden, denn bei nichtlinearen Systemen ist es möglich, dass sich die Beobachtbarkeit erst durch spätere Ableitungen ergibt. Unendlich viele Ableitungen zu bilden ist jedoch unpraktikabel.

Abhilfe für diese beiden Probleme schafft das Observability rank Criteria von [29]. Daher gehen wir dieses in diesem Kapitel durch.

17.3.1. Neue Systemgleichung

Für diesen Test benutzen wir eine neue Grundform der Systemgleichungen:

$$\dot{x} = f_0(x) + \sum_{i=1}^p f_i(x)u_i \quad (17.1)$$

$$z = (h_1(x), \dots, h_m(x))^T \quad (17.2)$$

Diese sagt aus, dass jede Komponente der Übergangsmessung linear in die (kontinuierliche) Ableitung eingeht. Dies ist eine Einschränkung des ansonsten sehr generischen Dynamikmodells, ist aber für die meisten Sensorfusionsprobleme umsetzbar. Auch wenn die Komponenten selbst linear eingehen, können die Funktionen f_i (inklusive f_0) nicht-linear sein. f_0 haben wir in bisherigen Darstellung als den Drift kennengelernt, also die Änderung des Zustandes die unabhängig von den Übergangsmessungen stattfindet (z.B. Beschleunigung durch Gravitation)

Beispiel:

$$\dot{x} = \begin{pmatrix} v \cos(\theta) \\ v \sin(\theta) \\ \omega \end{pmatrix} = \underbrace{\begin{pmatrix} \cos(\theta) \\ \sin(\theta) \\ 0 \end{pmatrix}}_{f_1(x)} v + \underbrace{\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}}_{f_2(x)} \omega$$

Wichtig: Diese Gleichungsform wird ausschließlich für den Beobachtbarkeitstest benötigt. Für den bisherigen Teil der Vorlesung (Modellierung für E/U-KF) benutzt man die allgemeine Darstellung.

17.3.2. Lie-Ableitung

Die Lie-Ableitung ist eine andere Form der Ableitung als die Zeitableitung. Sie ist definiert als die Ableitung eines Vektorfeldes in Richtung eines anderen Vektorfeldes. Sie gibt damit an, wie sich eine Funktion (die auch einen Vektor oder eine Matrix als Ergebnis haben kann) sich ändert, wenn sich die Eingabevariablen in eine definierte Richtung ändern.

Für unsere Zwecke ist die Lie-Ableitung die Richtungsableitung:

$$\mathcal{L}_f^1 h = \nabla h \cdot f \quad (17.3)$$

Dabei bedeutet diese Schreibweise die Lie-Ableitung der Funktion h in Richtung f erster Ordnung. Berechnet wird sie wie die Richtungsableitung durch das Skalarprodukt aus dem Gradienten von h (∇ - der sog. Nabla-Operator) und dem Richtungsvektor f . f kann dabei eine Funktion sein. Es gibt auch höhere Ordnung der Lie-Ableitungen (equivalent zur 2. oder 3. Zeitableitung) mit der man fassen kann, wie sich die Funktion h ändert wenn man sich entlang eines Pfades f bewegt:

$$\mathcal{L}_f^2 h = \mathcal{L}_f \mathcal{L}_f^1 h = \nabla \mathcal{L}_f^1 h \cdot f \quad (17.4)$$

17. Beobachtbarkeit von Zuständen

Dabei gibt es die praktische Regel, dass man wenn man die Lie-Ableitung über den Gradienten einer Funktion bildet, man stattdessen den Gradienten der Lie-Ableitung der Funktion berechnen kann. Dies macht die Berechnung des neuen Rangkriteriums deutlich leichter. Auch die 0. Ordnung der Lie-Ableitung existiert und ist bestimmt durch:

$$\mathcal{L}_f^0 h = h \quad (17.5)$$

Q 51

Die 0. Ordnung ist also einfach die Originalfunktion h .

17.3.3. Rangkriterium mit Lie-Ableitungen

Mithilfe der Lie-Ableitungen können wir eine genauere Aussage über die Änderung der Messungen treffen. Durch das Dynamikmodell sind die möglichen Änderungen des Zustandes je nach Zustand beschränkt. Der Kohlenkarren kann z.B. nur geradeausfahren und kann sich daher nur in Richtung seiner Orientierung weiterbewegen. Wendet man nun die Lie-Ableitung auf die Messung in Richtung dieser möglichen Bewegung an, erfährt man, wie sich die Messung überhaupt an diesem Zustand ändern kann. Diese Information steckt auch in den Zeitableitungen, ist aber aufgrund der vielen Ableitungen der Übergangsmessungen und der Anwendungen der Kettenregel „Schwierig zu sehen“. Bildet man also alle Lie-Ableitungen $\mathcal{L}_{f_i}^1 h_j$ erhält man eine genaue Aussage, wie sich die Funktion h_j ändern kann.

Um längere Bewegungen zu berücksichtigen, kann man auch hier Lie-Ableitungen höherer Ordnung bilden wie z.B. $L_{f_k f_i}^2 h_j = \mathcal{L}_{f_k}^1 \mathcal{L}_{f_i}^1 h_j$. Dabei leitet man nacheinander in die verschiedenen Richtungen ab. Nimmt man nun alle Lie-Ableitungen bis zur unendlichen Ordnung zusammen, erhält man wieder denselben Informationsgehalt wie durch die unendlich vielen Zeitableitungen. Man kann also den Beobachtbarkeitstest auf den Lie-Ableitungen statt auf den Zeitableitungen durchführen. Der Vorteil dabei: Die dauernde Multiplikation mit den zeitlichen Ableitungen der Inputs entfällt. Die Inputs tauchen generell nicht mehr in den Lie-Ableitungen auf. Sie werden nur noch implizit berücksichtigt, wodurch die Gleichungen deutlich simpler werden.

Einen Nachteil hat diese Methode jedoch: Die Lie-Ableitung in eine Richtung f_i hat nur eine Bedeutung, wenn die zugehörige Übergangsmessung nicht 0 ist. Das ist so zu verstehen: Die Lie-Ableitung in eine Richtung f_i stellt den Informationsgewinn durch die Übergangsmessung u_i dar. Wenn allerdings keine Bewegung in die Richtung f_i stattfindet, da $u_i = 0$ kann auch keine Information aus dieser Bewegung gezogen werden.

17.3.4. Observable Codistribution

Mit der Anwendung der Lie-Ableitung lässt sich das erste Problem lösen und die Gleichungen deutlich vereinfachen. Jedoch werden immer noch unendliche viele Ableitungen benötigt. Dieses Problem lässt sich mithilfe einer Beobachtung umgehen. Die unendlich vielen Ableitungen sorgen für unendlich viele Zeilen in der Jacobi-Matrix. Von diesen unendlich vielen Zeilen sind aber nur Zeilen interessant, die linear unabhängig von den anderen sind. Das liegt daran, dass die Jacobi-Matrix den Informationsgehalt über

17.3. Beobachtbarkeitsrangkriterium (Observability rank criteria)

das System der Lie-Ableitungen angeben. Wenn Gradienten linear abhängig voneinander sind, enthalten sie dieselben Informationen. Daher können diese Zeilen ignoriert werden. Es ist sogar so: Wenn beim Aufstellen der Lie-Ableitung der Gradient linear abhängig von den bisherigen Gradienten ist, dann hat diese Lie-Ableitung keinen weiteren Informationsgehalt- auch nicht durch weitere Ableitungen. Daher kann man das Ableiten dieses Pfades stoppen. Daraus ergibt sich ein praktischer Algorithmus um die linear unabhängigen Zeilen (die sog. Observable Codistribution) zu berechnen: Set $\Omega_0 = \text{span } \nabla h$

Set $k = 0$

do

Set $k=k+1$

Set $\Omega_k = \Omega_{k-1} \oplus \mathcal{L}_{f_0}^1 \Omega_{k-1} \oplus \bigoplus_{i=1}^p \mathcal{L}_{f_i}^1 \Omega_{k-1}$

while $\dim(\Omega_k) > \dim(\Omega_{k-1})$

Set $\Omega = \Omega_k$

Wenn $\dim(\Omega) = n$ dann ist das System schwach beobachtbar, ansonsten sind $n - \dim(\Omega)$ Freiheitsgrade unbeobachtbar

Nach [29]

Der Algorithmus arbeitet iterativ. Zunächst werden die Gradienten der Messfunktion h in die Codistribution aufgenommen. Der Ausdruck $\text{span}(\dots)$ bedeutet dabei, dass bereits hier auf lineare Abhängigkeit geprüft wird. Dies ist nur interessant, wenn man mehrere Messungen hat (z.B. Abstand zu 2 Funkbaken). Danach wird iteriert. In jeder Iteration werden alle Zeilen der Codistribution mithilfe der Lie-Ableitung in alle Richtungen f_i ($i \in [0, p]$) abgeleitet. Die so entstehenden Gradienten werden geprüft ob sie linear abhängig von den bisherigen Zeilen der Codistribution sind und hinzugefügt falls sie unabhängig sind (dies gibt vereinfacht gesagt der Operator \oplus an). Die Iteration wird gestoppt, wenn in einer Iteration keine weiteren Zeilen der Codistribution hinzugefügt wurden also wenn $\Omega_{k+1} = \Omega_k$. Dies ist spätestens der Fall, wenn sie n (Anzahl der Zustände) Zeilen enthält. Daher benötigt dieser Algorithmus maximal n Iterationen.

Die Codistribution besteht am Ende lediglich aus linear unabhängigen Zeilen, wodurch sich der Rang ablesen lässt: Er entspricht der Anzahl der Zeilen. Das System ist daher Beobachtbar wenn n -Zeilen vorhanden sind. Für jede Zeile die bis zur Beobachtbarkeit fehlt, hat das System einen unbeobachtbaren Freiheitsgrad, d.h. man kann eine unbeobachtbare Transformation mit $n - \dim(\Omega)$ freien Parametern finden.

Beispiel:

Fahrzeug in der Ebene mit einer Abstandsmessung:

Die Berechnung ist leichter in Polarkoordinaten:

$$\rho = \sqrt{p_x^2 + p_y^2}, \quad \varphi = \text{atan2}(p_y, p_x), \quad x = \begin{pmatrix} \rho \\ \varphi \\ \theta \end{pmatrix}, \quad h = \rho, \quad (17.6)$$

17. Beobachtbarkeit von Zuständen

$$\dot{x} = \begin{pmatrix} v \cos(\theta - \varphi) \\ \frac{v \sin(\theta - \varphi)}{\rho} \\ \omega \end{pmatrix}, f_0 = 0, f_1 = \begin{pmatrix} \cos(\theta - \varphi) \\ \frac{\sin(\theta - \varphi)}{\rho} \\ 0 \end{pmatrix}, f_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad (17.7)$$

$$\Omega_0 = \nabla h = (1 \ 0 \ 0) \quad (17.8)$$

$$\mathcal{L}_{f_1}^1 \Omega_0 = \mathcal{L}_{f_1}^1 \nabla h = \nabla \mathcal{L}_{f_1}^1 h = \nabla (\nabla h * f_1) = \nabla (\cos(\theta - \varphi)) \quad (17.9)$$

$$= (0 \ \sin(\theta - \varphi) \ -\sin(\theta - \varphi)) \quad (17.10)$$

$\mathcal{L}_{f_1}^1 \Omega_0$ ist linear unabhängig von Ω_0

$$\mathcal{L}_{f_2}^1 \Omega_0 = \mathcal{L}_{f_2}^1 \nabla h = \nabla \mathcal{L}_{f_2}^1 h = \nabla (\nabla h * f_2) = \nabla (0) \quad (17.11)$$

$\mathcal{L}_{f_2}^1 \Omega_0$ ist 0 und damit in jedem Fall linear abhängig von Ω_0 :

$$\Omega_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sin(\theta - \varphi) & -\sin(\theta - \varphi) \end{pmatrix} \quad (17.12)$$

Da wir Zeile 1 schon abgeleitet haben, führen wir den Algorithmus nur für neue Zeilen aus.

$$\mathcal{L}_{f_1}^1 (0 \ \sin(\theta - \varphi) \ -\sin(\theta - \varphi)) = \mathcal{L}_{f_1}^1 \nabla (\cos(\theta - \varphi)) = \nabla \mathcal{L}_{f_1}^1 \cos(\theta - \varphi) \quad (17.13)$$

$$= \nabla (\nabla \cos(\theta - \varphi) * f_1) = \nabla \left(\frac{\sin(\theta - \varphi)^2}{\rho} \right) \quad (17.14)$$

$$= \left(-\frac{\sin^2(\theta - \varphi)}{\rho^2} \quad -2\frac{\sin(\theta - \varphi)\cos(\theta - \varphi)}{\rho} \quad 2\frac{\sin(\theta - \varphi)\cos(\theta - \varphi)}{\rho} \right) \quad (17.15)$$

Dieser Gradient ist linear abhängig von Ω_1 . Er ergibt sich indem man Zeile 1 mit $-\frac{\sin^2(\theta - \varphi)}{\rho^2}$ multipliziert und Zeile 2 mal $-2\frac{\cos(\theta - \varphi)}{\rho}$ dazuaddiert.

$$\mathcal{L}_{f_2}^1 (0 \ \sin(\theta - \varphi) \ -\sin(\theta - \varphi)) = \nabla ((0 \ \sin(\theta - \varphi) \ -\sin(\theta - \varphi)) * f_2) \quad (17.16)$$

$$= \nabla (-\sin(\theta - \varphi)) = (0 \ \cos(\theta - \varphi) \ -\cos(\theta - \varphi)) \quad (17.17)$$

Auch dieser Gradient ist linear abhängig von Zeile 2 durch die Multiplikation mit $\frac{\cos(\theta - \varphi)}{\sin(\theta - \varphi)}$. Daher ändert sich Omega nicht und der Algorithmus ist konvergiert mit 2 Zeilen in Ω

$$\Omega = \Omega_2 = \Omega_1 \quad (17.18)$$

Das System ist also unbeobachtbar und hat eine unbeobachtbare Transformation mit einem freien Parameter.

17.3. Beobachtbarkeitsrangkriterium (Observability rank criteria)

Dieser Algorithmus lässt sich vergleichsweise leicht in symbolischen Mathetoolboxen wie Matlab, Octave oder sympy (die beiden letztgenannten sind frei erhältlich) implementieren und damit automatisieren. Das erlaubt es einem, automatisiert die Beobachtbarkeit eines Systems zu prüfen. Hat man Systeme mit mehreren Outputs (z.B. 2 Abstandsmessung) startet der Algorithmus mit einem größeren Ω_0 , wobei lineare Unabhängigkeit geprüft werden muss. Für 2 Winkelmessungen wie in unserem 2D Beispiel ergibt der Algorithmus diese Codistribution:

$$\Omega_1 = \begin{pmatrix} \frac{-y}{x^2+y^2} & \frac{x}{x^2+y^2} & -1 \\ \frac{-y}{(x-1)^2+y^2} & \frac{x-1}{(x-1)^2+y^2} & -1 \\ \frac{(y^2-x^2)\sin(\theta)+2xy\cos(\theta)}{(x^2+y^2)^2} & \frac{(y^2-x^2)\sin(\theta)-2xy\cos(\theta)}{(x^2+y^2)^2} & \frac{x\cos(\theta)+y\sin(\theta)}{x^2+y^2} \end{pmatrix}$$

Da sie 3 Zeilen hat ist der Zustand im Leitbeispiel schwach beobachtbar, solange man 2 Funkbaken sehen kann (es reicht auch nacheinander). Zu beachten ist hierbei, dass diese Matrix nur gültig ist, solange beide Übergangsmessungen ungleich 0 sind. Daher müsste man den Algorithmus nochmal durchführen, wenn man z.B. testen möchte ob eine Geradeausfahrt ($\omega = 0$) zur Beobachtbarkeit ausreicht. Dabei würde man nicht mehr in Richtung f_2 ableiten, da diese Richtung zu ω gehört.

18. Review der Sensorfusionsliteratur

Dieses Kapitel ist als Ergänzung zur Vorlesung gedacht. Es besteht kein Anspruch, dass hier alles verstanden wird und es ist auch nicht Prüfungsrelevant. Die Übersicht über die Literatur der Sensorfusion soll den Horizont über die Methoden dieser Vorlesung hinaus erweitern. Dabei besteht kein Anspruch auf Vollständigkeit. Die in der Vorlesung gezeigten Algorithmen sind zwar ein guter Start für viele Sensorfusionsprobleme, aber die Literatur hat noch deutlich mehr zu bieten. Die meisten hier angerissenen Algorithmen benötigen das Handwerkszeug aus dieser Vorlesung: Zustände, Mess- und Dynamikmodelle und natürlich Messrauschen. Die Bayessche Sicht, das Messmodell gegeben den Zustand $p(Z|X)$ zu verwenden ist weit verbreitet in der Community.

Zurzeit beinhaltet dieses Kapitel nur die Modellbasierte Sensorfusion. Aber auch in der Sensorfusion kommt Deep Learning allmählich an. Die Quellen an den Kapiteln sind als weiterer Einstieg in das Thema gedacht.

Für eine eigene, weiterführende Literaturrecherche lohnt es sich bei Surveys[4, 7, 9, 45, 17, 39] anzufangen, da diese viele relevante Arbeiten zitieren und einem die passenden Suchbegriffe liefern. Wir sind offen für weitere Vorschläge von Quellen der Sensorfusionsliteratur oder Algorithmen.

18.1. Abwandlungen des Kalman Filters

Hier findet ihr einige Abwandlungen des Kalman Filters bzw. EKF die versuchen bessere Schätzungen zu liefern. Viele dieser Abwandlungen sind noch nicht kompatibel mit \boxplus -Zuständen.

18.1.1. Second Order EKF (SO-EKF)

[15] Wir haben den EKF hergeleitet, indem wir mithilfe des Taylor-Polynoms nicht-lineare Funktionen linearisiert haben. Damit konnten wir ungefähre Rechenregeln für nichtlineare Modelle herleiten. Man kann auch mehr Teile des Taylor-Polynoms als die ersten beiden dazunehmen. Nimmt man den quadratischen Anteil dazu, erhält man den Second-Order EKF (SO-EKF). Dazu benötigt man die sogenannte Hesse Matrix (Hessian) G_{xx} von g bzw. H_{xx} von h . Die Gleichungen für den Dynamikschritt sind:

$$\bar{\mu}_k = g(\mu_{k-1}, u_{k-1}) + \frac{1}{2} \sum_i e_i \operatorname{tr} \left\{ G_{xx}^{(i)}(\mu_{k-1}, u_{k-1}) \Sigma_{k-1} \right\} \quad (18.1)$$

$$\begin{aligned} \bar{\Sigma}_k &= g_x(\mu_{k-1}, u_{k-1}) \Sigma_{k-1} G_x^T(\mu_{k-1}, u_{k-1}) \\ &+ \frac{1}{2} \sum_{i,i'} e_i e_{i'}^T \operatorname{tr} \left\{ G_{xx}^{(i)}(\mu_{k-1}, u_{k-1}) \Sigma_{k-1} G_{xx}^{(i')}(\mu_{k-1}, u_{k-1}) \Sigma_{k-1} \right\} + Q_{k-1} \end{aligned} \quad (18.2)$$

Und für den Update Schritt:

$$v_k = Z_k - h(\mu_k^-) - \frac{1}{2} \sum_i e_i \text{tr} \{ H_{xx}^{(i)}(\bar{\mu}_k) \bar{\Sigma}_k \} \quad (18.3)$$

$$S_k = H_x(\bar{\mu}_k) \bar{\Sigma}_k H_x^T(\bar{\mu}_k) + \frac{1}{2} \sum_{i,i'} e_i e_{i'}^T \text{tr} \{ H_{xx}^{(i)}(\bar{\mu}_k) \bar{\Sigma}_k H_{xx}^{(i')}(\bar{\mu}_k) \bar{\Sigma}_k \} + R_k \quad (18.4)$$

$$K_k = \bar{\Sigma}_k H_X^T(\bar{\mu}_k) S_k^{-1} \quad (18.5)$$

$$\mu_k = \bar{\mu}_k + K_k v_k \quad (18.6)$$

$$P_k = \bar{\Sigma}_k - K_k S_k K_k^T \quad (18.7)$$

Der SO-EKF ist durch die bessere Approximierung in vielen Fällen genauer als der standard EKF (First order EKF). Die Berechnung der Hesse Matrix ist aber eine zusätzliche Hürde bei der Verwendung.

18.1.2. Iterated Extended Kalman Filter

[4] Ein grundlegendes Problem des EKF ist die Linearisierung. Die Linearisierung ist nur in einem kleinen Bereich um den wahren Zustand herum gültig. Da wir aber den wahren Zustand nicht kennen, linearisieren wir um unseren aktuellen Erwartungswert. Daher wird die Linearisierung schlechter, je schlechter der EKF schätzt. Im Endeffekt führt der falsche Linearisierungspunkt zu einer falschen Berechnung des Kalman Gains.

Die Idee beim Iterated EKF ist nun, dass wir den Zustand nach der Messung genauer kennen als vorher. Daher können wir nach der Messung besser linearisieren und den Kalman Gain besser berechnen. Um das zu nutzen, linearisiert man nach dem Update um die neue Schätzung herum und führt das Update von vorne aus. D.h. man führt es nicht vom neuen sondern vom alten Erwartungswert aus, da man sonst die Messung mehrfach verwenden und so stärker als gewollt gewichten würde. Dies kann man iterativ durchführen, bis sich nichts mehr ändert (oder aber die Rechenzeit aufgebraucht ist). Man kann die Iteration auch auf das Ergebnis der letzten Iteration zurückführen und erhält so die Gleichungen des Iterated EKF:

$$\begin{aligned} h^i &= h(\bar{\mu}^i) \\ H^i &= \left. \frac{\partial h_i(x)}{\partial x} \right|_{x=\mu} \\ \mu^{i+1} &= \mu^i + \Sigma^i (H^i)^T \Sigma_\delta^{-1} (z - h^i) - \Sigma^i \bar{\Sigma}^{-1} (\mu^i - \bar{\mu}) \\ \Sigma^i &= \bar{\Sigma} - \bar{\Sigma} (H^i)^T (H^i \bar{\Sigma} (H^i)^T + \Sigma_\delta)^{-1} H^i \bar{\Sigma} \end{aligned}$$

Mit dem Iterated EKF erhält man deutlich bessere Ergebnisse als mit dem EKF im Austausch für mehr Rechenzeit. Im wesentlichen werden die Fehler durch den falschen Linearisierungspunkt reduziert. Beim EKF wurde dann nämlich mit dem Schätzwert $\bar{\mu}_t = E(x_t | z_{1..t-1}, u_{1..t})$ vor der Messung linearisiert, beim IEKF mit dem Schätzwert $\mu_t = E(x_t | z_{1..t}, u_{1..t})$ nach der Messung. Das ist allerdings immer noch schlechter als bei quadratischer Ausgleichsrechnung, die *alle* Messungen von Anfang bis Ende mit

dem Schätzwert $E(x_t|z_{1..T}, u_{1..T})$ nach *allen* Messungen linearisiert. Den größten Vorteil bringt das in der Situation, wo die Unsicherheit durch eine Messung viel kleiner wird. Dies passiert oft beim Start des Filters.

18.1.3. \boxplus -EKF

[32] Anstelle des UKFs lässt sich auch der EKF auf \boxplus -Mannigfaltigkeiten verwenden. Die Ableitungen durch den \boxplus -Operator sind allerdings oft kompliziert, so dass man am besten eine automatische Differenzierungsbibliothek nutzt. Dafür ist der \boxplus -EKF schneller, da er das Modell seltener auswertet. Man tauscht also die einfache Benutzung und gute Approximierung des UKF gegen die Geschwindigkeit des EKFs aus.

Die Ableitungen für den \boxplus -EKF sollte man mit einem Tool wie Octave, Matlab, Mathematica etc. berechnen, oder man benutzt direkt automatische Differenzierung. Dies erleichtert die Verwendung des \boxplus -EKF ungemein. Eine C++ Implementierung mit automatischer Differenzierung findet ihr unter <https://github.com/TomLKoller/ADEKF>.

Prediction:

$$\bar{\mu}_t = g(\mu_{t-1}, u_t, \vec{0}) \quad (18.8)$$

$$A_t = \left. \frac{\partial g(\mu_{t-1} \boxplus \delta, u_k, \vec{0}) \boxminus \bar{\mu}_t}{\partial \delta} \right|_{\delta=\vec{0}} \quad (18.9)$$

$$B_t = \left. \frac{\partial g(\mu_{t-1}, u_k, \epsilon) \boxminus \bar{\mu}_t}{\partial \epsilon} \right|_{\epsilon=\vec{0}} \quad (18.10)$$

$$\bar{\Sigma}_t = A_t \Sigma_t A_t^T + B_t Q_t B_t^T \quad (18.11)$$

Update:

$$C_t = \left. \frac{\partial h(\bar{\mu}_t \boxplus \delta) \boxminus h(\bar{\mu}_t)}{\partial \delta} \right|_{\delta=\vec{0}} \quad (18.12)$$

$$S_t = H_t \bar{\Sigma}_t H_t^T + R_t \quad (18.13)$$

$$K_t = \bar{\Sigma}_t H_t^T S_t^{-1} \quad (18.14)$$

$$r_t = z_k \boxminus h(\bar{\mu}_t) \quad (18.15)$$

$$\mu_t = \bar{\mu}_t \boxplus K_t r_t \quad (18.16)$$

$$J_t = \left. \frac{\partial (\bar{\mu}_t \boxplus (r_t + \delta) \boxminus \mu_t)}{\partial \delta} \right|_{\delta=\vec{0}} \quad (18.17)$$

$$\Sigma_t = J_t (\bar{\Sigma}_t - K_t C_t \bar{\Sigma}_t) J_t^T \quad (18.18)$$

18.1.4. Multiplicative EKF

[28] Neben unserer \boxplus -Methode gibt es noch andere Wege mit Rotationen im Kalman Filter umzugehen. Eine Methode ist der Multiplicative-EKF (MEKF). Dieser ist konzeptionell sehr ähnlich zu \boxplus , aber speziell für Rotationen entwickelt. Die Idee des MEKF

ist, dass man die aktuelle Orientierung immer in Bezug zu einer Referenz darstellt (im Normallfall das Weltkoordinatensystem):

$$q = q_{ref} \cdot Rot(\delta) \quad (18.19)$$

Statt jetzt die ganze Orientierung im Zustand zu haben, wird nur der Vektor δ gespeichert und mit obiger Formel, wird die Orientierung in den Modellfunktionen bei Bedarf berechnet. Wie wir allerdings gelernt haben, führt diese Minimalparametrisierung mit nur 3 Komponenten zu einer Singularität, wenn der $\|\delta\| \approx 2\pi$. Um dies zu vermeiden, wird öfter mal das δ in die Referenz geschoben:

$$q_{ref} \leftarrow q_{ref} \cdot Rot(\delta) \quad \delta \leftarrow 0 \quad (18.20)$$

Dies Vorgehen ist (faast) identisch zu \boxplus , wo in μ das Mannigfaltigkeitselement entsprechend Qq_{ref} drin steckt und (18.19) vom \boxplus gemacht wird. Die Darstellung in (18.19) ist etwas flexibler, weil sie erlaubt, einen Mean in δ zu haben, während im \boxplus -E/UKF der Mean immer 0 ist. Dadurch muss der \boxplus -E/UKF jedesmal den Mean von δ nach q_{ref} schieben, näherungsweise als Update $\$mu_t = \bar{\mu}_t \boxplus \dots$ oder mit einer zusätzlichen Sigmapunktpropagation.

18.1.5. Error State Kalman Filter

[27, 49, 54] Die Filter die wir kennen gelernt haben, haben immer einen Schätzfehler, dessen Verteilung sie mit der Kovarianz angeben. Diesen Schätzfehler \tilde{x} kann man berechnen, indem man von der Schätzung \hat{x} den wahren Zustand x abzieht:

$$\tilde{x} = \hat{x} - x$$

Die eigentliche Aussage der Kovarianz ist also, dass der Schätzfehler Gaußverteilt ist:

$$\tilde{x} \sim \mathcal{N}(0, \Sigma)$$

Die Idee beim Error State Kalman Filter (ESKF) ist es, nur den Schätzfehler zu schätzen. Der Filter wird also nur benutzt, um die Messfehler im Dynamikmodell zu schätzen. Der eigentliche Zustand wird außerhalb des Filters mitgeschätzt.

Im Prinzip benutzt der Filter 2 Dynamikmodelle. Ein Dynamikmodell, dass die Änderung des Zustandes gegeben die Übergangsmessung darstellt, und ein Dynamikmodell, dass nur aussagt wie die Fehler sich entwickeln. Also wie sich die Kovarianz weiter entwickelt. Der eigentliche Zustand des Systems wird außerhalb des KF durch das erste Dynamikmodell aufintegriert. Im KF wird nur das Fehlerdynamikmodell benutzt, dass nur die Kovarianz verändert.

Beim update Schritt des KF wird geschätzt, wie falsch der Zustand ist, der außerhalb des KF aufintegriert wird. Dieser Schätzfehler wird dann auf den Zustand angewandt, um ihn zu korrigieren und der KF setzt seine Fehlerschätzung wieder auf 0 zurück (behält aber die Kovarianz bei).

Dieses Vorgehen hat mehrere Vorteile:

1. Der Zustand muss nicht durch Addition des Fehlers korrigiert werden. Auch andere Funktionen sind möglich. Bei Quaternionen z.B. ist der Fehler wieder eine Euler-Winkel-Achse die mit Rot auf den Zustand korrigiert wird. Daher erlaubt diese Variante die Nutzung von Quaternionen.
2. Fehlerdynamiken sind häufig linear, auch wenn es die Modelle nicht sind. So steigt der Schätzfehler der Orientierung linear zum Messrauschen eines Gyrometers. Dadurch lassen sich die Schätzfehler in einigen Problem optimal (ohne Approximierungen wie im UKF oder EKF) schätzen, auch wenn dies beim Zustand nicht möglich ist.
3. Fehlerdynamiken sind meist langsamer und einfacher zu berechnen, als Zustandsdynamiken. So entwickelt sich der Orientierungsfehler mit einem Gyrometer gleichmäßig mit der Zeit, während der Zustand von der Reihenfolge der Rotationen abhängt. Dies ist von Vorteil, wenn man z.B. ein Gyrometer mit 1000Hz und eine Zustandsmessung mit 1Hz hat. Statt die Fehlerdynamik 1000mal pro Sekunde zu berechnen, fasst man sie zusammen und berechnet sie einmal, wenn eine neue Zustandsmessung verfügbar ist. Man verliert trotzdem keine Genauigkeit beim Zustand, da man das Zustandsdynamikmodell trotzdem mit 1000Hz ausführt. Der Trick ist also Kovarianz und Zustand getrennt voneinander zu betrachten.

Der Nachteil des ESKF ist, dass man neben dem Zustandsdynamikmodell auch das Fehlerdynamikmodell braucht. Diese sind aber für viele Trackinganwendungen in der Literatur verfügbar.

Der ESKF ist konzeptionell ähnlich zur \boxplus -Methode. Er betrachtet die Fehler oder „kleinen Änderungen“ aber explizit, während sie bei der \boxplus -Methode vor dem Anwender versteckt werden. Dadurch hat er die genannten Vorteile. Der ESKF erfreut sich großer Beliebtheit bei Luftfahrtanwendungen, da dort IMUs mit hohen Messfrequenzen und Radarmessungen mit niedrigen Messfrequenzen eingesetzt werden.

18.1.6. Information Filter (IF)

[31, 34] Die Filter die wir bislang kennen gelernt haben speichern die Verteilung als Erwartungswert und Kovarianz. Es gibt aber noch eine andere Variante die Verteilung zu speichern, die man den Information space bestehend aus y_t, Y_t nennt. Dieser verhält sich folgendermaßen zur bisherigen Darstellung:

$$y_t = \Sigma_t^{-1} \mu_t \quad (18.21)$$

$$Y_t = \Sigma_t^{-1} \quad (18.22)$$

Speichert man die Verteilung im Information space ändert sich der Messschritt (lineares Messmodell $Z_t = C_t X_t + h_t + \delta_t$) zu:

$$y_t = \bar{y}_t + C_t^\top \Sigma_{\delta t}^{-1} (z_t - h_t) \quad (18.23)$$

$$Y_t = \bar{Y}_t + C_t^\top \Sigma_{\delta t}^{-1} C_t \quad (18.24)$$

Sorry, this version does not include copyrighted Material.

Abbildung 18.1.: Verteilte Sensorknoten. Copyright IEEE 2017 [36]

Diesen Filter nennt man den Information Filter (IF). Man nennt das ganze den Information space, da der Messschritt so geformt ist, dass man die Informationen, die durch die Messung dazu kommen einfach aufaddieren kann. Im Gegensatz zur Kovarianz steigt die Informationsmatrix Y_t mit jeder Messung. Sie steigt also mit mehr Informationen über den Zustand. Der große Vorteil dieser Darstellung ist, dass der linke Summand in beiden Formeln unabhängig von den bisher erhaltenen Informationen ist. Dies stellt die dazugewonnene Information dar.

Der große Vorteil zeigt sich bei Sensorsystemen, bei denen die Sensoren über viele Standorte verteilt sind. Eine systematische Darstellung eines solchen Systems ist in Abbildung 18.1 gezeigt. Es wäre eine zu große Rechenlast, alle Messungen zu einem Knoten zu schicken und dort zu fusionieren. Um aber in einem klassischen KF die Messungen verteilt zu verarbeiten, müsste man nach jeder Messung die neue Kovarianz an alle Knoten verteilen, was genauso utopisch ist. Der IF erlaubt es, dass jeder Knoten nur die neue Information durch die Messung (die zweiten Summanden) weiterschickt. Die Zwischenknoten addieren die Informationen auf und senden die Gesamtinformation weiter. So werden die Messungen optimal und effizient verteilt ausgewertet.

18.1.7. Smoothing

[4, 43] Der Grundsatz der KF-Varianten ist es nicht nur, die Messungen mehrerer Sensoren miteinander zu fusionieren, sondern auch die Messungen der Vergangenheit zu berücksichtigen. Über das Dynamikmodell wird die Information aus vergangenen Messungen weitergetragen, sodass sie mit der aktuellen Messung fusioniert werden kann. Beim Smoothing (Glätten) nutzt man nicht nur die Messungen aus der Vergangenheit, sondern auch aus der Zukunft. Da diese zum Messpunkt nicht verfügbar sind, arbeitet man retrospektiv.

Grundsätzlich verbessert Smoothing Schätzungen eines Filters (z.B. EKF) von vergangenen Zuständen durch die Fusion mit allen neueren Messungen. Frei nach dem Prinzip „Hinterher ist man immer schlauer“. Probabilistisch ausgedrückt berechnen die bekannten Filter die Verteilung $p(X_t|Z_{1:t})$ also den Zustand zum Zeitpunkt t gegeben alle Messungen bis zum Zeitpunkt t . Smoother berechnen hingegen die Verteilung $p(X_t|Z_{1:T})$, also den Zustand zum Zeitpunkt t gegeben alle Messungen bis zum Ende $T \geq t$.

Smoother lassen sich auf verschiedene Arten kategorisieren, aber alle Smoother verbessern die Schätzung immens, da auf mehr Informationen zurückgegriffen werden kann. Besonders gut sind smoother, wenn Zustandsmessungen nur selten vorhanden sind. Wir stellen hier 2 Kategorisierungen vor:

Einteilung nach Ziel-/Fragestellung [4]:

18. Review der Sensorfusionsliteratur

1. Fixed point-smoother: Verbessert die Schätzung eines Zustandes zu einem festen (fixed) Zeitpunkt (point). Dies kann man benutzen, wenn der Zustand zu einem Zeitpunkt relevant ist, z.B. um rauszufinden, welches Kind unserem Roboter Doggy den Ball zugeworfen hat.
2. Fixed lag-smoother: Verbessert die Schätzung eines Zustandes der eine feste (fixed) Zeit in der Vergangenheit (lag) liegt, z.B. eine Sekunde zurück. Diese Variante ist sinnvoll, wenn man die Zustandsschätzung auch verzögert gebrauchen kann. Auch die wenigen Messungen von z.B. einer Sekunde aus der Zukunft, können die Schätzung des Zustandes schon deutlich verbessern. Insbesondere, wenn man nur niedrigfrequente Messungen erhält. Solange keine konkrete Regelung von der Schätzung abhängt, ist dieses Verfahren sehr gut anwendbar. Ein hypothetisches Beispiel wäre die Visualisierung von der eigenen Position auf einem GPS Gerät. Durch den Glättungseffekt des Smoothings, werden Sprünge durch Ausreißermessungen reduziert und die Trajektorie sieht „richtig“ aus.
3. Fixed intervall-smoother: Statt nur einen Zustand zu einem Zeitpunkt zu smoothen, kann man auch ein ganzes Intervall an Zuständen smoothen. Diese Variante benutzt man, wenn man nach einem Experiment möglichst genaue Schätzungen der Zustände eines Zeitintervalls (oder auch des ganzen Experiments) benötigt. Hierbei zielt man also auf maximale Ausnutzung aller Messungen ab, um die bestmögliche Schätzung aller Zustände zu erhalten.

Einteilung nach Funktionsprinzip: [43]

1. Two-Filter Smoother: Ein Filter läuft wie gehabt vorwärts durch die Zeit von 1 bis T, während ein zweiter Filter die Zustände rückwärts durch die Zeit also von T bis 1 schätzt. Das Ergebnis des Smoothings erhält man, indem man die Schätzungen beider Filter zu jedem Zeitpunkt t über die Formel

$$p(X_k|Z_{1:T}) \propto p(X_k|Z_{1:t})p(Z_{t+1:T}|X_k)$$

fusioniert. Diese Variante ist sehr einfach, da die Schätzungen der beiden Filter zu jedem Zeitpunkt k unabhängig voneinander sind, da ein Filter alle Messungen zeitlich vor/gleich t und der zweite Filter alle Messungen nach t verwendet. Problematisch hierbei ist, dass für den rückwärtslaufenden Filter ein zeitinverses Dynamikmodell benötigt wird, also ein Modell, mit dem die Bewegung rückwärts durch die Zeit beschrieben wird. Dies ist insbesondere unter Berücksichtigung der Fehler nicht immer möglich. Zudem benötigt man, damit die Unabhängigkeit tatsächlich gilt, nicht nur die Verteilung des Startzustandes X_0 sondern auch die Verteilung des Endzustandes X_T . Häufig wird X_T durch den Vorwärtsfilter berechnet, allerdings verletzt dies streng genommen die Unabhängigkeit, da das X_T aus dem ersten Filter bereits die Informationen aller Messungen enthält.

2. Forward Filter-Backward Smoother: Bei diesem Prinzip wird auf ein inverses Dynamikmodell verzichtet. Ein Filter schätzt in einem Vorwärtsdurchlauf die Zustände

zu allen Zeitpunkten t (allerdings nur mit den Messungen bis t). Anschließend iteriert ein Smoother rückwärts und verbessert die Schätzung zum Zeitpunkt t mit der Schätzung zum Zeitpunkt $t + 1$. Dabei ist zu beachten, dass diese beiden Schätzungen voneinander abhängig sind und daher die Kovarianz in der Fusion berücksichtigt werden muss. Probabilistisch rechnet man die Verteilung:

$$p(x_t | y_{1:T}) = p(x_t | y_{1:t}) \int \left[\frac{p(x_{t+1} | x_t) p(x_{t+1} | y_{1:T})}{p(x_{t+1} | y_{1:t})} \right] dx_{t+1}$$

aus. In [40] wurden die Rauch-Tung-Striebel (RTS) Gleichungen vorgestellt, mit denen man einen Smoother für den KF erhält. Die Gleichungen des RTS-Smoother sind (Start bei $k = N - 1$):

$$\begin{aligned} \bar{\mu}_{k+1} &= f(\mu_k, u_k) \\ \Sigma_{k+1} &= F_x(\mu_k, u_k) \Sigma_k F_x^T(\mu_k, u_k) + Q_k \\ C_k &= \Sigma_k F_x^T(\mu_k, u_k) [\bar{\Sigma}_{k+1}]^{-1} \\ \mu_k^s &= \mu_k + C_k [\mu_{k+1}^s - \bar{\mu}_{k+1}] \\ \Sigma_k^s &= \Sigma_k + C_k [\Sigma_{k+1}^s - \bar{\Sigma}_{k+1}] C_k^T \end{aligned}$$

Die RTS Gleichungen kommen ohne inverses Dynamikmodell und ohne den Endzustand aus und werden daher häufig angewandt. Im linearen Fall, sind sie ebenso optimal wie der KF.

18.2. Andere Filtertypen

In diesem Kapitel befassen wir uns mit Algorithmen, die grundsätzlich anders funktionieren als der Kalman Filter. Dennoch werden bekannte Konzepte aus der Vorlesung verwendet.

18.2.1. Batch Least Squares (Quadratische Ausgleichsrechnung)

[13] Batch Least Squares, meistens Bundle Adjustment (BA) genannt, ist im Prinzip die Quadratische Ausgleichsrechnung wie wir sie kennen gelernt haben. Allerdings fallen hierunter auch Algorithmen, die besser funktionieren als der vorgestellte Gauß-Newton Algorithmus (z.B. Levenberg-Marquardt). Insgesamt geht es also darum, einen Zustand gegeben eine Menge an Messungen zu schätzen.

Die Ergebnisse von BA Verfahren sind im für nichtlineare Modelle meist deutlich besser als die von Filtern, da sie durch iterative Ansätze besser mit den Nichtlinearitäten umgehen können. Für BA Probleme gibt es einige nutzbare Programme, die die Algorithmen implementieren und lediglich die Zustands- und Messmodelle benötigen (ceres solver, Matlab, scipy, GTSAM).

Prinzipiell lassen sich nicht nur statische Probleme wie Kalibrierungen oder die Vermessung mit BA Algorithmen lösen, sondern auch die klassische Anwendung vom KF lässt sich damit lösen. Der Trick dabei ist, dass man die Systemzustände der kompletten

Zeitreihe (also ein Systemzustand je Zeitpunkt) in den Zustand des BA aufnimmt [22]. Die Zustandsmessungen zu den verschiedenen Zeitpunkten basieren dann immer auf dem Systemzustand der zu diesem Zeitpunkt gehört. Die Verknüpfung über die Zeit kommt wieder vom Dynamikmodell. Dieses wird als eine Messung umformuliert. Das zugehörige Messmodell für je zwei aufeinander folgende Zustände ist dann:

$$f(X) = g(x_t) \boxplus x_{t+1}$$

Das heißt, man führt ein Messmodell ein, das aussagt, dass die Vorhersage basierend auf x_t mit x_{t+1} übereinstimmen muss. Als Messkovarianz nimmt man das Dynamikrauschen.

Ein solcher BA Algorithmus erbringt deutlich bessere Ergebnisse, als ein KF, da er sämtliche Messungen auf einmal einbezieht. Ein Smoothing wird daher automatisch durchgeführt. Ist das System linear, ist diese Lösung aber äquivalent zum KF mit dem RTS-Smoother, da diese optimal sind. BA für ganze Zeitreihen benötigt sehr viel Rechenzeit und ist daher nur für Auswertungen nach einem Experiment geeignet. Auch \boxplus -Mannigfaltigkeiten können im Zustand verwendet werden, wenn das Framework es zulässt. Der ceres-solver löst dies mit sogenannten lokalen Parametrisierungen [3].

18.2.2. Moving Horizon Estimator (MHE)

[18, 38, 45] Der Moving Horizon Estimator (MHE) ist eine Variante die sehr gute Leistung des BA Echtzeitfähig zu machen. Statt die ganze Zeitreihe seit Beginn zu optimieren, werden nur die letzten N Zeitschritte optimiert. N wird dabei so gewählt, dass die Optimierung auf dem System in Echtzeit läuft. Die Schätzung der Vergangenheit (also vor den letzten N Messungen) wird als eine Gaußverteilung zusammengefasst und als Startverteilung ausgewählt.

Durch dieses Verfahren erreicht man einen sehr guten Kompromiss aus Laufzeit und Qualität der Schätzung, wobei dennoch sehr hohe Rechenkapazitäten nötig sind. Der verbreitet sich zurzeit in der Community, aufgrund seiner guten Schätzungen trotz nicht-linearer Modelle.

18.2.3. Grid/Histogram Filter

[44, 35] Der Grid- oder auch Histogrammfilter (Literaturname: Bayesian Occupancy Filter (BOF)) kommt ohne die Gaussannahme der Verteilung aus. Er setzt daher direkt auf dem Bayes-Filter auf. Da Erwartungswert und Kovarianz nur für Gaußverteilungen sinnvoll sind, approximiert der Gridfilter die Verteilung Gitteransatz. Der Zustandsraum wird wie bei einem Histogramm in viele Bins bzw. ein Gitter unterteilt. Durch die Gitterdarstellung, können wir Wahrscheinlichkeiten (anstelle von Wahrscheinlichkeitsdichten) für einzelne Zellen berechnen. Jede Zelle des Gitters, erhält also einen Wert, der aussagt wie wahrscheinlich sich der Zustand in der Zelle befindet. Mit diesem Ansatz kann man jede beliebige Verteilung darstellen. Je feiner das Gitter wird, desto genauer sind die Verteilungen darstellbar.

Bei einer Dynamikmessung werden die Wahrscheinlichkeiten entsprechend des Dynamikmodells verschoben. Beim Messschritt, werden die Wahrscheinlichkeiten der einzelnen Zellen entsprechend des Bayes updates neu berechnet. Prinzipiell führt man also

den Bayes-Filter auf der diskretisierten Darstellung der Wahrscheinlichkeitsverteilung aus. Unter <https://www.youtube.com/watch?v=qSof0p0bPJ8> findet sich eine schöne Visualisierung des Gridfilters.

Problematisch bei diesem Ansatz ist, dass die Anzahl der nötigen Zellen im Gitter mit der Anzahl der Zustände rapide ansteigt, wenn man die Genauigkeit erhalten möchte. Zudem sind Wahrscheinlichkeitsverteilungen meist „dicht“, d.h. nur ein relativ kleiner Teil des möglichen Zustandsraumes beinhaltet den Großteil der Wahrscheinlichkeitsverteilung (z.B: Gaußlocke $1,96\sigma$ enthält 95%). Daher wird viel Rechenleistung auf unwahrscheinliche Zellen verbraucht.

18.2.4. Partikelfilter

[11] Eine sehr gute graphische Erklärung des Partikelfilters findet man von Matlab unter https://www.youtube.com/watch?v=NrzMH_yerBU.

Ähnlich wie der Gridfilter approximiert der Partikelfilter beliebige Verteilungen. Dafür benutzt er allerdings einen Sampling Ansatz. Statt den Zustandsraum in Zellen zu unterteilen, wird die Verteilung mithilfe vieler Partikel, ähnlich den Sigmapunkten des UKF, dargestellt. Diese Partikel werden allerdings zufällig aus der Verteilung gezogen, statt der systematischen Vorgehensweise der unscented transform.

Im Dynamikschritt werden alle Partikel mit dem Dynamikmodell weiterbewegt, allerdings wird das Dynamikrauschen nicht auf 0 gesetzt, sondern zufällig aus der Verteilung des Dynamikrauschens gezogen. Im anschließenden Messschritt wird für jedes Partikel berechnet, wie wahrscheinlich es war, d.h. wie gut es zu der erhaltenen Messung passt. Die so errechnete Wahrscheinlichkeit interpretiert man als Gewicht dieses Zustandes in der Verteilung. Die Gesamtwahrscheinlichkeit muss immer 1 ergeben, was durch Normalisierung erzwungen wird. Aus den gewichteten Partikeln kann man z.B. einen Erwartungswert und eine Kovarianz berechnen, mit denen man leichter handhaben kann.

Theoretisch ist der Partikelfilter optimal, wenn man unendlich viele Partikel verwendet dies ist aber nicht erreichbar. Üblicherweise wählt man eine Anzahl an Partikeln, die Echtzeitfähig ist. Hochdimensionale Zustände erfordern dabei mehr Partikel, um die Verteilung korrekt darzustellen.

Partikelfilter (Pseudocode)

Initialize:

Sample N Particles P^n with Weights W^n from X_0

Dynamic:

Foreach n :Sample noise ε_n Dynamic propagation: $\bar{P}_k^n = g(P_k^n, u_k, \varepsilon^n)$ Weight recalculation: $\bar{W}_k^n = W_k^n \cdot p(\varepsilon^n)$

Update:

Foreach n $W_{k+1}^n = \bar{W}_k^n \cdot p(z_{k+1} | \bar{P}_k^n)$

Normalise Weights

Resample

Auch der Partikelfilter leidet unter dem Problem, dass viele Partikel sehr unwahrscheinlich werden. Um mehr Partikel im Bereich mit einer hohen Wahrscheinlichkeitsdichte zu haben, führt man das sogenannte Resampling aus. Dabei werden neue Partikel aus der alten Partikelmenge gezogen. Je wahrscheinlicher ein Partikel in der Ursprungsverteilung war, desto wahrscheinlicher kommt es in der neuen Partikelmenge vor (auch mehrfach). Dadurch werden unwahrscheinliche Partikel gelöscht. Am häufigsten benutzt man das Systematische Resampling, dass eine Zufallszahl zieht und darauf basierend alle Partikel auswählt:

Systematic Resampling

$$U_1 \sim \mathcal{U}\left[0, \frac{1}{N}\right], U_i = U_1 + \frac{i-1}{N}, i \in [2, N]$$

$$P_{k+1}^i = \left\{ U_j : \sum_{n=1}^{i-1} W_k^n \leq U_j \leq \sum_{n=1}^i W_k^n \right\}$$

Weitere findet man in [11].

18.2.5. Robust Estimation

[48] Auch wenn die Gaußannahme der Messfehler meistens zutrifft, so gibt es doch immer wieder Ausreißermessungen, die nicht zu der Verteilung passen. Da die Gaußglocke jedoch jeder Messung eine Wahrscheinlichkeit größer als 0 gibt, zerstören Ausreißer schnell die Schätzung. Ein prominentes Beispiel ist das Multipath Problem beim GPS, dass deutlich längere Flugzeitmessungen erhält, wenn der direkte Weg zum Satelliten blockiert ist.

Das Feld der Robust Estimation versucht mit solchen Problemen umzugehen, dabei versucht man Ausreißer zu erkennen und anders zu behandeln. Sei es durch Ignorieren der Messung, starkes Erhöhen des Messrauschens oder durch die Neuentwicklung von Filtern durch die Annahme passenderer Verteilungen. Dieses Feld befasst sich also, mit den praktischen Problemen der Sensorfusion, die durch die fehlerhaften/vereinfachten Annahmen entstehen.

18.3. Pseudomessungen

[45, 39] Informationen zu einem System müssen nicht immer nur von Sensoren kommen. In einigen Fällen, gelten für ein System zusätzliche Einschränkungen. Man nehme z.B. unseren Kohlenkarren auf der Schiene in Abbildung 18.2. Statt auf einer graden Schiene wurde er auf einer Kurvenstrecke ähnlich einer Achterbahn platziert und mit einer IMU ausgestattet. Wir könnten das Standarddynamikmodell für IMUs verwenden. Dieses würde allerdings ignorieren, dass sich der Kohlenkarren nur auf der Schiene (mit bekannter Topologie/Form) bewegen kann. Diese Information wollen wir jetzt nutzen, um die Schätzung zu verbessern.

Sorry, this version does not include copyrighted Material.

Abbildung 18.2.: Schienenfahrzeug mit IMU. Copyright IEEE 2019 [24]

Man nehme eine Funktion $f(X)$, die berechnet wie weit der Kohlenkarren von der Bahnstrecke entfernt ist. Es muss gelten, dass dieser Abstand 0 ist, da der Kohlenkarren sich auf der Bahn befindet. Aus dieser Überlegung können wir eine Messung mit Modell konstruieren. Das Messmodell ist $f(X)$ und der Messwert ist immer z_t . Das Messrauschen müssen wir wählen. Dieses stellt dar, wie sicher wir uns sind, dass der Kohlenkarren wirklich auf der Strecke sein muss. Bei einer Schiene wählen wir ein sehr kleines Messrauschen. Bei einem Auto auf einer Straße würden wir ein größeres Messrauschen wählen.

Diese konstruierte Messung können wir jetzt mit dem Messschritt im EKF/UKF verwenden, um den Karren auf die Bahn zu zwingen. Diese Messung bezeichnet man auch als Pseudomessung und im Falle von 0 Kovarianz auch als Perfekte Messung. Die Verwendung einer Pseudomessung erlaubt es einem zusätzliche Informationen über den Zustand einfach in einen KF oder eine Ausgleichsrechnung einzubringen. Diese Informationen können sogar die Beobachtbarkeit des Zustandes verbessern (Analyse mit den vorgestellten Kriterien, als wäre die Pseudomessung eine tatsächliche Messfunktion).

Das große Problem bei diesem Vorgehen ist, dass der Fehler dieser Pseudomessung meistens nicht Gaußverteilt ist und die Pseudomessungen über die Zeit korreliert sind. Wird der Karren von der Schiene gehoben und 1 m daneben abgestellt, ist der Messfehler der Pseudomessung immer 1 m und damit stark korreliert. Um dieses Problem abzuferdern, wählt man meistens eine deutlich höhere Kovarianz als nötig.

Es gibt auch andere Mittel die zusätzlichen Informationen einzubringen (z.B. Wahl einer Zustandsdarstellung die nur Positionen auf der Strecke erlaubt), diese sind aber deutlich komplizierter umzusetzen als eine Pseudomessung.

Insgesamt ermöglichen es Pseudomessungen die Schätzqualität deutlich zu verbessern, allerdings sind die Schätzungen aufgrund der korrelierten Pseudomessungen nicht mehr erwartungstreu. Die Systeme wirken zudem intelligenter, da sie Kontextwissen (Vorwissen) mit berücksichtigen. Anwendungen finden sich im Road Matching [30, 8, 26] und bei Motion Capture Suits [41, 55] auf IMU Basis sowie in den weiteren Unterkapiteln. Wenn nicht weiter angegeben, wird davon ausgegangen, dass das System nur mit einer IMU und keinen weiteren Sensoren ausgestattet ist.

18.3.1. Vehicle Constraints

[10, 42] Fahrzeuge auf Rädern können sich nur in Lenkrichtung fortbewegen (siehe Abbildung 18.3). Messgleichung:

$$\vec{v}_b = \begin{pmatrix} |\vec{v}_w| \\ 0 \\ 0 \end{pmatrix}$$

- Kovarianz: Drift/Sliding in Kurven wird über das Messrauschen modelliert
- Durch das Vorwissen werden IMU biase, die Geschwindigkeit in Körperkoordinaten, und Roll und Pitch schwach beobachtbar.

Sorry, this version does not include copyrighted Material.

Abbildung 18.3.: Fahrzeug mit IMU. Copyright IEEE 2019 [24]

18.3.2. Topologiekarten

[23] Fahrzeuge, die nicht fliegen können, müssen sich auf einem Untergrund fortbewegen. Die Räder müssen also den Boden berühren (außer das Fahrzeug fällt/stürzt z.B. beim Mountain Biking).

- Messgleichung: 3D-Abstand vom Rad oder der IMU Position zur Oberfläche ist 0
- Kovarianz: Bildet ab, wie sicher man ist, dass sich das Rad auf dem Boden befindet, aber auch wie genau die Oberflächentopologie bekannt ist.
- Die Pose (Position und Orientierung) kann schwach beobachtbar werden, wenn die Oberfläche nicht-trivial ist (also z.B. eine Ebene oder Schräge Ebene)

Sorry, this version does not include copyrighted Material.

Abbildung 18.4.: Radrennbahn Sixdays. Copyright IEEE 2019 [23]

18.3.3. Zero-Velocity Update

[12, 6, 47] Bei Gangmustern, werden die Füße immer wieder auf dem Boden abgestellt. Dabei ist die Geschwindigkeit des Fußes mindestens für einen Moment 0 (siehe Abbildung 18.5). Der Zeitpunkt, zu dem dies zutrifft, muss aus den IMU-Daten bestimmt werden (z.B. durch Klassifikation). Messgleichung:

$$\vec{v}_W = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Sorry, this version does not include copyrighted Material.

Abbildung 18.5.: Zero-Velocity-Update. Copyright IEEE 2019 [24]

- Kovarianz: Fehler der Zero-Velocity-Annahme. Erkannter Zeitpunkt ist meist leicht versetzt und daher ist die Geschwindigkeit nicht wirklich 0.
- Durch das Vorwissen werden IMU biase, die Geschwindigkeit in Körperkoordinaten, und Roll und Pitch schwach beobachtbar.

18.3.4. Gebäudekarten

[16, 17] Innerhalb von Gebäuden, können sich auch Fußgänger nur an zugänglichen Stellen befinden und sie können sich nicht durch Wände bewegen (siehe Abbildung 18.6). Dieses Vorwissen ist schwierig im Pseudomessungstil zu lösen, da es sehr stark von einer

Sorry, this version does not include copyrighted Material.

Abbildung 18.6.: Zero-Velocity-Update. Copyright IEEE 2008 [6]

Gaußverteilung abweicht. Im Prinzip, ist es eine Messung, die eine Uniformverteilung über den gesamten begehbaren Bereich bildet. Daher wird es häufig im Partikelfilter gelöst. Partikel die sich durch Wände bewegt haben, oder außerhalb der begehbaren Fläche liegen erhalten eine Wahrscheinlichkeit von 0 und werden damit beim Resampling gelöscht.

Literatur

- [1] Jürgen Adamy. *Nichtlineare Systeme und Regelungen*. Springer, 2018.
- [2] Sameer Agarwal, Keir Mierle u. a. *Ceres Solver*. <http://ceres-solver.org>.
- [3] Sameer Agarwal, Keir Mierle u. a. *Ceres Solver*. URL: <http://ceres-solver.org>.
- [4] Yaakov Bar-Shalom, X. Rong Li und Thiagalingam Kirubarajan. *Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software*. en. Google-Books-ID: xz9nQ4wdXG4C. John Wiley & Sons, Apr. 2004. ISBN: 978-0-471-46521-8.
- [5] Timothy D Barfoot. *State Estimation for Robotics*. Cambridge University Press, 2017.
- [6] S. Beauregard, Widyawan und M. Klepal. “Indoor PDR Performance Enhancement using Minimal Map Information and Particle Filters”. In: *2008 IEEE/ION Position, Location and Navigation Symposium*. ISSN: 2153-358X. Mai 2008, S. 141–147. DOI: [10.1109/PLANS.2008.4570050](https://doi.org/10.1109/PLANS.2008.4570050).
- [7] Julia Breßler u. a. “GNSS positioning in non-line-of-sight context—A survey”. In: *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*. ISSN: 2153-0017. Nov. 2016, S. 1147–1154. DOI: [10.1109/ITSC.2016.7795701](https://doi.org/10.1109/ITSC.2016.7795701).
- [8] Li Cuiping u. a. “Tracking ground targets with road constraint using multiple hypotheses tracking”. In: *2010 2nd International Conference on Signal Processing Systems*. Bd. 2. Juli 2010, S. V2–265–V2–269. DOI: [10.1109/ICSPS.2010.5555498](https://doi.org/10.1109/ICSPS.2010.5555498).
- [9] Pavel Davidson und Robert Piché. “A Survey of Selected Indoor Positioning Methods for Smartphones”. en. In: 19.2 (2017), S. 24.
- [10] G. Dissanayake u. a. “The aiding of a low-cost strapdown inertial measurement unit using vehicle model constraints for land vehicle applications”. In: *IEEE Transactions on Robotics and Automation* 17.5 (Okt. 2001), S. 731–747. ISSN: 2374-958X. DOI: [10.1109/70.964672](https://doi.org/10.1109/70.964672).
- [11] Arnaud Doucet und Adam Johansen. “A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later”. In: *Handbook of Nonlinear Filtering* 12 (Jan. 2009).
- [12] E. Foxlin. “Pedestrian tracking with shoe-mounted inertial sensors”. In: *IEEE Computer Graphics and Applications* 25.6 (Nov. 2005), S. 38–46. ISSN: 1558-1756. DOI: [10.1109/MCG.2005.140](https://doi.org/10.1109/MCG.2005.140).
- [13] Giorgio Grisetti u. a. “Least Squares Optimization: from Theory to Practice”. In: *arXiv:2002.11051 [cs]* (Feb. 2020). arXiv: 2002.11051. URL: <http://arxiv.org/abs/2002.11051> (besucht am 29. 04. 2020).

- [14] Hafner. *Wahrscheinlichkeitsrechnung und Statistik*. Springer Verlag, 1989.
- [15] Haitao Zhang, Jian Rong und Xiaochun Zhong. “The performance comparison and algorithm analysis of first order EKF, second order EKF and smoother for GPS/DR navigation”. In: *2008 11th IEEE International Conference on Communication Technology*. Nov. 2008, S. 432–437. DOI: [10.1109/ICCT.2008.4716286](https://doi.org/10.1109/ICCT.2008.4716286).
- [16] Michael Hardegger, Daniel Roggen und Gerhard Tröster. “3D ActionSLAM: wearable person tracking in multi-floor environments”. en. In: *Personal and Ubiquitous Computing* 19.1 (Jan. 2015), S. 123–141. ISSN: 1617-4909, 1617-4917. DOI: [10.1007/s00779-014-0815-y](https://doi.org/10.1007/s00779-014-0815-y). URL: <http://link.springer.com/10.1007/s00779-014-0815-y> (besucht am 14.10.2020).
- [17] Robert Harle. “A Survey of Indoor Inertial Positioning Systems for Pedestrians”. In: *IEEE Communications Surveys Tutorials* 15.3 (2013). Conference Name: IEEE Communications Surveys Tutorials, S. 1281–1293. ISSN: 1553-877X. DOI: [10.1109/SURV.2012.121912.00075](https://doi.org/10.1109/SURV.2012.121912.00075).
- [18] Eric L Haseltine und James B Rawlings. “Critical evaluation of extended Kalman filtering and moving-horizon estimation”. In: *Industrial & engineering chemistry research* 44.8 (2005), S. 2451–2460.
- [19] C. Hertzberg u. a. “Integrating Generic Sensor Fusion Algorithms with Sound State Representations through Encapsulation of Manifolds”. In: *Information Fusion* (2012). <http://dx.doi.org/10.1016/j.inffus.2011.08.003>.
- [20] Google Inc. *Tensorflow Probability*. online, Stand 2.12.2022. URL: [%5Curl %7Bhttps://www.tensorflow.org/probability/%7D](https://www.tensorflow.org/probability/).
- [21] D. Kehlmann. *Die Vermessung der Welt*. Rowohlt, Reinbek, 2005.
- [22] Tom L. Koller und Udo Frese. “State Observability through Prior Knowledge: Analysis of the Height Map Prior for Track Cycling”. en. In: *Sensors* 20.9 (Jan. 2020). Number: 9 Publisher: Multidisciplinary Digital Publishing Institute, S. 2438. DOI: [10.3390/s20092438](https://doi.org/10.3390/s20092438). URL: <https://www.mdpi.com/1424-8220/20/9/2438> (besucht am 01.07.2020).
- [23] Tom L. Koller und Udo Frese. “State Observability through Prior Knowledge: Tracking Track Cyclers with Inertial Sensors”. In: *2019 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. ISSN: 2162-7347. Sep. 2019. DOI: [10.1109/IPIN.2019.8911757](https://doi.org/10.1109/IPIN.2019.8911757).
- [24] Tom L. Koller, Tim Laue und Udo Frese. “State Observability through Prior Knowledge A Conceptual Paradigm in Inertial Sensing”. In: *Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*. 2019.
- [25] U. Krenzel. *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Vieweg Verlagsgesellschaft, 2005.

Literatur

- [26] Jaime López-Araquistain u. a. “A new approach to map-assisted Bayesian tracking filtering”. en. In: *Information Fusion* 45 (Jan. 2019), S. 79–95. ISSN: 1566-2535. DOI: [10.1016/j.inffus.2018.01.002](https://doi.org/10.1016/j.inffus.2018.01.002). URL: <https://www.sciencedirect.com/science/article/pii/S1566253517300933> (besucht am 09.02.2021).
- [27] Venkatesh Madyastha u. a. “Extended Kalman Filter vs. Error State Kalman Filter for Aircraft Attitude Estimation”. In: *AIAA Guidance, Navigation, and Control Conference*. American Institute of Aeronautics und Astronautics. DOI: [10.2514/6.2011-6615](https://doi.org/10.2514/6.2011-6615). URL: <https://arc.aiaa.org/doi/abs/10.2514/6.2011-6615> (besucht am 04.03.2020).
- [28] F. Landis Markley. “Attitude Error Representations for Kalman Filtering”. In: *Journal of Guidance, Control, and Dynamics* 26.2 (2003). Publisher: American Institute of Aeronautics and Astronautics. eprint: <https://doi.org/10.2514/2.5048>, S. 311–317. DOI: [10.2514/2.5048](https://doi.org/10.2514/2.5048). URL: <https://doi.org/10.2514/2.5048> (besucht am 01.07.2020).
- [29] Agostino Martinelli. *Observability: A New Theory Based on the Group of Invariance*. en. Google-Books-ID: Nx_5DwAAQBAJ. SIAM, Aug. 2020. ISBN: 978-1-61197-625-0.
- [30] M. Mertens und M. Ulmke. “Precision GMTI tracking using road constraints with visibility information and a refined sensor model”. In: *2008 IEEE Radar Conference*. ISSN: 2375-5318. Mai 2008, S. 1–6. DOI: [10.1109/RADAR.2008.4720781](https://doi.org/10.1109/RADAR.2008.4720781).
- [31] Arthur G. O. Mutambara. *Decentralized Estimation and Control for Multisensor Systems*. en. Google-Books-ID: Z1YfUGkG8poC. CRC Press, Jan. 1998. ISBN: 978-0-8493-1865-8.
- [32] David Nakath, Joachim Clemens und Kerstin Schill. “Multi-Sensor Fusion and Active Perception for Autonomous Deep Space Navigation”. In: *2018 21st International Conference on Information Fusion (FUSION)*. Juli 2018, S. 2596–2605. DOI: [10.23919/ICIF.2018.8455788](https://doi.org/10.23919/ICIF.2018.8455788).
- [33] J. E. Pascoe. “The inverse function theorem and the Jacobian conjecture for free analysis”. en. In: *Mathematische Zeitschrift* 278.3-4 (Dez. 2014), S. 987–994. ISSN: 0025-5874, 1432-1823. DOI: [10.1007/s00209-014-1342-2](https://doi.org/10.1007/s00209-014-1342-2). URL: <http://link.springer.com/10.1007/s00209-014-1342-2> (besucht am 21.12.2020).
- [34] F. Pfaff u. a. “Information form distributed Kalman filtering (IDKF) with explicit inputs”. In: *2017 20th International Conference on Information Fusion (Fusion)*. Juli 2017, S. 1–8. DOI: [10.23919/ICIF.2017.8009724](https://doi.org/10.23919/ICIF.2017.8009724).
- [35] Florian Pfaff, Kailai Li und Uwe D Hanebeck. “Estimating Correlated Angles Using the Hypertoroidal Grid Filter”. en. In: (Sep. 2020), S. 7.
- [36] Florian Pfaff u. a. “Information Form Distributed Kalman Filtering (IDKF) with Explicit Inputs”. en. In: (), S. 8.
- [37] W.H. Press u. a. *Numerical Recipes, Second Edition*. Cambridge University Press, Cambridge, 1992.

- [38] Christopher V. Rao und James B. Rawlings. “Constrained process monitoring: Moving-horizon approach”. en. In: *AIChE Journal* 48.1 (Jan. 2002), S. 97–109. ISSN: 00011541, 15475905. DOI: [10.1002/aic.690480111](https://doi.org/10.1002/aic.690480111). URL: <http://doi.wiley.com/10.1002/aic.690480111> (besucht am 19.03.2020).
- [39] Ghulam Rasool. “Constrained State Estimation-A Review”. In: *arXiv preprint arXiv:1807.03463* (2018).
- [40] H. E. Rauch, F. Tung und C. T. Striebel. “Maximum likelihood estimates of linear dynamic systems”. en. In: *AIAA Journal* 3.8 (Aug. 1965), S. 1445–1450. ISSN: 0001-1452, 1533-385X. DOI: [10.2514/3.3166](https://doi.org/10.2514/3.3166). URL: <https://arc.aiaa.org/doi/10.2514/3.3166> (besucht am 08.02.2021).
- [41] D. Roetenberg, H. Luinge und P. Slycke. “Xsens MVN: Full 6DOF Human Motion Tracking Using Miniature Inertial Sensors”. In: 2009.
- [42] Yaacov Rothman, Itzik Klein und Sagi Filin. “Analytical Observability Analysis of INS with Vehicle Constraints”. In: *Navigation: Journal of The Institute of Navigation* 61.3 (2014). Publisher: Wiley Online Library, S. 227–236.
- [43] Simo Särkkä. “Unscented Rauch–Tung–Striebel Smoother”. In: *IEEE Transactions on Automatic Control* 53.3 (Apr. 2008), S. 845–849. ISSN: 2334-3303. DOI: [10.1109/TAC.2008.919531](https://doi.org/10.1109/TAC.2008.919531).
- [44] Marcelo Saval-Calvo u. a. “A Review of the Bayesian Occupancy Filter”. en. In: *Sensors* 17.2 (Feb. 2017). Number: 2 Publisher: Multidisciplinary Digital Publishing Institute, S. 344. DOI: [10.3390/s17020344](https://doi.org/10.3390/s17020344). URL: <https://www.mdpi.com/1424-8220/17/2/344> (besucht am 09.02.2021).
- [45] D. Simon. “Kalman filtering with state constraints: a survey of linear and nonlinear algorithms”. en. In: *IET Control Theory & Applications* 4.8 (Aug. 2010), S. 1303–1318. ISSN: 1751-8652. DOI: [10.1049/iet-cta.2009.0032](https://doi.org/10.1049/iet-cta.2009.0032). URL: <https://digital-library.theiet.org/content/journals/10.1049/iet-cta.2009.0032> (besucht am 19.02.2020).
- [46] Dan Simon. *Optimal State Estimation: Kalman, H_∞ , and Nonlinear Approaches*. en. Hoboken, NJ, USA: John Wiley & Sons, Inc., Mai 2006. ISBN: 978-0-470-04534-3 978-0-471-70858-2. DOI: [10.1002/0470045345](https://doi.org/10.1002/0470045345). URL: <http://doi.wiley.com/10.1002/0470045345> (besucht am 15.07.2020).
- [47] Isaac Skog u. a. “Zero-Velocity Detection—An Algorithm Evaluation”. en. In: *IEEE Transactions on Biomedical Engineering* 57.11 (Nov. 2010), S. 2657–2666. ISSN: 0018-9294. DOI: [10.1109/TBME.2010.2060723](https://doi.org/10.1109/TBME.2010.2060723). URL: <http://ieeexplore.ieee.org/document/5523938/> (besucht am 15.10.2020).
- [48] H. E. Söken und C. Hajiyev. “REKF and RUKF for pico satellite attitude estimation in the presence of measurement faults”. In: *Journal of Systems Engineering and Electronics* 25.2 (Apr. 2014). Conference Name: Journal of Systems Engineering and Electronics, S. 288–297. ISSN: 1004-4132. DOI: [10.1109/JSEE.2014.00033](https://doi.org/10.1109/JSEE.2014.00033).

Literatur

- [49] Joan Solà. “Quaternion kinematics for the error-state Kalman filter”. In: *arXiv:1711.02508 [cs]* (Nov. 2017). arXiv: 1711.02508. URL: <http://arxiv.org/abs/1711.02508> (besucht am 11.12.2020).
- [50] S. Thrun, W. Burgard und D. Fox. *Probabilistic Robotics*. MIT Press, 2005. ISBN: 0262201623.
- [51] S. Thrun, W. Burgard und D. Fox. *Probabilistic Robotics – Kalman Filter Slides*. <http://robots.stanford.edu/probabilistic-robotics/ppt/kalman.ppt>.
- [52] S. Thrun, W. Burgard und D. Fox. *Probabilistic Robotics – Web Site*. <http://www.probabilistic-robotics.org>. 2005.
- [53] J. Väisälä. “A Proof of the Mazur-Ulam Theorem”. In: *The American Mathematical Society Monthly* 110.7 (2003).
- [54] Rachel V. Vitali, Ryan S. McGinnis und Noel C. Perkins. “Robust Error-State Kalman Filter for Estimating IMU Orientation”. In: *IEEE Sensors Journal* (2020). Conference Name: IEEE Sensors Journal, S. 1–1. ISSN: 1558-1748. DOI: [10.1109/JSEN.2020.3026895](https://doi.org/10.1109/JSEN.2020.3026895).
- [55] Felix Wenk. “Inertial Motion Capturing Rigid Body Pose and Posture Estimation with Inertial Sensors”. PhD Thesis. University of Bremen, 2017.
- [56] Wikipedia. *Landau-Symbole* — *Wikipedia, die freie Enzyklopädie*. [Online; Stand 23. November 2022]. 2021. URL: [%5Curl%7Bhttps://de.wikipedia.org/w/index.php?title=Landau-Symbole&oldid=217959469%7D](https://de.wikipedia.org/w/index.php?title=Landau-Symbole&oldid=217959469%7D).
- [57] Wikipedia. *Normalverteilung* — *Wikipedia, die freie Enzyklopädie*. [Online; Stand 24. November 2022]. 2022. URL: [%5Curl%7Bhttps://de.wikipedia.org/w/index.php?title=Normalverteilung&oldid=227828907%7D](https://de.wikipedia.org/w/index.php?title=Normalverteilung&oldid=227828907%7D).

A. Quizfragen und Arbeitsblätter

Frage Q 1. Jede Zeile zeigt eine Wahrscheinlichkeitsfunktion P eines Münzwurfs. Welche erfüllt alle Axiome?

	$P(\emptyset)$	$P\{\text{Kopf}\}$	$P\{\text{Zahl}\}$	$P\{\text{Kopf,Zahl}\}$
A	0	0.6	0.6	1.2
B	0.2	0.6	0.6	1.0
C	0	0.4	0.6	1.0
D	0	0.6	0.6	1.0

Frage Q 2. Ist als Verallgemeinerung von Folgerung 1 $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(C \cap A) + P(A \cap B \cap C)$...

A	$-P(A \cap B) - P(B \cap C) - P(C \cap A)$	B	$-P(A \cap B) - P(B \cap C) - P(C \cap A) - P(A \cap B \cap C)$
C	$-P(A \cap B) - P(B \cap C) - P(C \cap A) + P(A \cap B \cap C)$	D	$-P(A \cap B \cap C)$

Frage Q 3. Die Wahrscheinlichkeit, einen Viererpassch gewürfelt zu haben, wenn die Augensumme 8 ist, ist...

A	$\frac{1}{5}$	B	$\frac{1}{3}$
C	$\frac{8}{36}$	D	$\frac{1}{36}$

Frage Q 4. Welche Einheit hat die Wahrscheinlichkeitsdichte, wenn das Ergebnis $\in \Omega$ eine 1D Position im Raum ist?

A	m	B	1/m
C	m^2	D	1

A. Quizfragen und Arbeitsblätter

Frage Q 5. Seien X und Y zwei Würfel. Welche Paare von Zufallsvariablen sind unabhängig?

A	X, Y	B	$X + Y, X - Y$
C	$X \bmod 2, X \bmod 3$	D	$X, X + Y$

Frage Q 6. Seien X und Y zwei Würfel, der $E(XY)$ Erwartungswert des Produktes ist

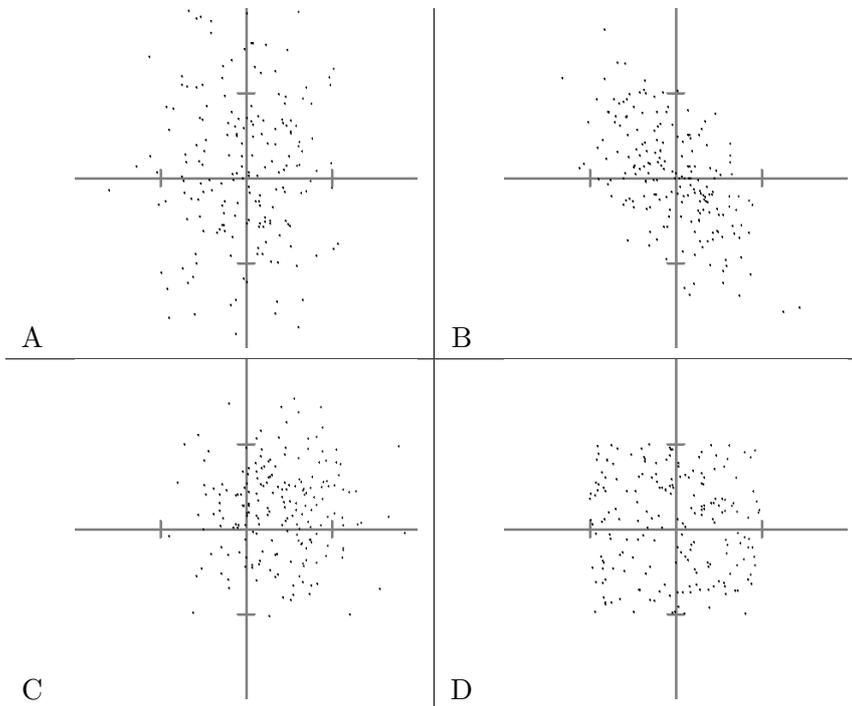
A	36	B	37/2
C	0	D	49/4

A.1. Erwartungswerte

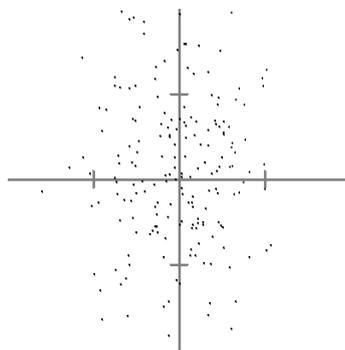
Die Zufallsvariable X hat eine uniforme Verteilung zwischen 0 und 10, $\mathcal{U}(0, 10)$.

- Was ist die Wahrscheinlichkeitsdichte $p(X = x)$ als Formel?
- Berechne den Erwartungswert $E(X)$!
- Berechne den Erwartungswert $E(X^2)$!
- Vergleiche $E(X)^2$ und $E(X^2)$. Siehst Du einen anschaulichen Zusammenhang?

Frage Q 7. Die Grafiken zeigen zufällige Stichproben zweier Messungen Z_1, Z_2 , wobei der wahre Wert in der Mitte liegt. Welche davon erfüllen die Anforderungen aus der Vorlesung?



Frage Q 8. Welche Werte für $E((Z_1 - x)^2) = \sigma_1^2$ und $E((Z_2 - x)^2) = \sigma_2^2$ hat die Messungen deren stichprobenhafte Verteilung unten gezeigt ist? Die Striche geben jeweils ± 1 an.



A	$\sigma_1 = 2, \sigma_2 = 1$	B	$\sigma_1 = \frac{1}{10}, \sigma_2 = \frac{1}{5}$
C	$\sigma_1 = \frac{1}{2}, \sigma_2 = 1$	D	$\sigma_1 = 1, \sigma_2 = 1$

Frage Q 9. Um welchen Faktor kann die Fusion zweier Messwerte (nach der Formel aus der Vorlesung) bestenfalls genauer sein, als die bessere der beiden? ($\max \frac{\min(\sigma_1, \sigma_2)}{\sigma_{\bar{x}}}$)

A	2	B	$\sqrt{2}$
C	$\frac{4}{3}$	D	∞

A.2. Eigenschaften der 2-Messwerte-Formel

$$\hat{x}(z_1, \sigma_1, z_2, \sigma_2) = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} z_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} z_2, \quad \sigma_{\hat{x}}^2(\sigma_1, \sigma_2) = \frac{1}{1/\sigma_1^2 + 1/\sigma_2^2}. \quad (\text{A.1})$$

Prüfe, welche der von uns intuitiv vermuteten Eigenschaften der Fusion zweier Messwerte für die konkrete Formel aus der Vorlesung gelten.

- \hat{x} zwischen z_1 und z_2
- $\sigma_1 \rightarrow \infty \Rightarrow \hat{x} \rightarrow z_2$
- $\sigma_1 \rightarrow 0 \Rightarrow \hat{x} \rightarrow z_1$
- $\hat{x}(z_1, \sigma_1, z_2, \sigma_2) = \hat{x}(z_2, \sigma_2, z_1, \sigma_1)$
- $\hat{x}(z_1, \sigma, z_2, \sigma) = \frac{z_1 + z_2}{2}$
- $\sigma_{\hat{x}}^2 \leq \sigma_1^2, \sigma_{\hat{x}}^2 \leq \sigma_2^2$

A.3. Generalisierung auf n -Messwerte

Für zwei Messwerte lautet die Fusionsformel

$$\hat{x} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} z_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} z_2, \quad \sigma_{\hat{x}}^2 = \frac{1}{1/\sigma_1^2 + 1/\sigma_2^2}.$$

Versuche durch Intuition und Logik eine allgemeine Formel für n -Messwerte zu erraten. Beachte dabei die folgenden Eigenschaften die für die Formel gelten müssen:

- \hat{x} ist eine gewichtete Summe von z_1 bis z_n .
- Für zwei Messwerte muss sie mit obiger Formel übereinstimmen.
- Geht $\sigma_i \rightarrow \infty$ muss dessen Gewicht in der Formel gegen 0 gehen.
- \hat{x} muss zwischen $\min(z_1, \dots, z_n)$ und $\max(z_1, \dots, z_n)$ liegen. Was heißt das für die Summe der Gewichte von z_1 bis z_n ?

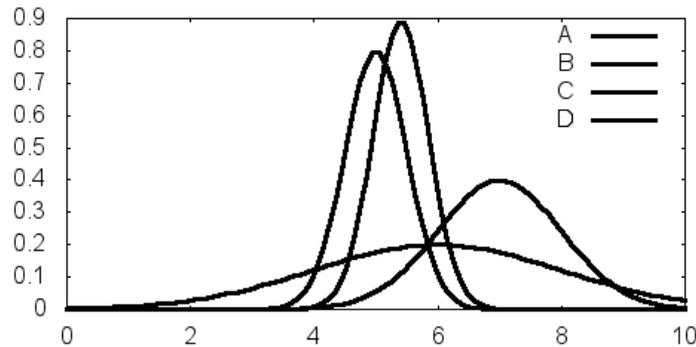
A.3. Generalisierung auf n -Messwerte

Frage Q 10. Es sei X ein roter, Y ein weißer Würfel und $Z = X + Y$ die Augensumme. Für welche Augensumme z ist es am wahrscheinlichsten, eine weiße vier gewürfelt zu haben? ($Y = 4$)

$$P(Y = 4|Z = z) = \text{-----} \quad (\text{A.2})$$

A	$z = 2$	B	$z = 10$
C	$z = 5$	D	$z = 7$

Frage Q 11. Welche der folgenden Gaußlocken A-D ist (bis auf Proportionalität) das Produkt zweier anderen?



Frage Q 12. Nehmen wir an, der Messfehler N_i ist nicht Gauß'sch ($\mathcal{N}(0, \sigma_i^2)$), sondern uniform ($\mathcal{U}(-\sigma_i, +\sigma_i)$) also in einem Intervall verteilt. Sei der Einfachheit halber $z_1 < z_2$, $z_1 - \sigma_1 < z_2 - \sigma_2$ und $z_1 + \sigma_1 < z_2 + \sigma_2$. Was ergibt sich dann für eine a-posteriori Verteilung? (Grafik an Tafel dazu machen)

A	$\mathcal{U}\left(\frac{z_1\sigma_2+z_2\sigma_1}{\sigma_1\sigma_2} \mp \frac{\sigma_1\sigma_2}{\sigma_1+\sigma_2}\right)$	B	$\mathcal{U}(z_2 - \sigma_2, z_1 + \sigma_1)$
C	keine uniforme Verteilung	D	$\mathcal{U}(z_1 - \sigma_1, z_2 + \sigma_2)$

A. Quizfragen und Arbeitsblätter

Frage Q 13. Welche Varianz hat eine ± 1 Münze?

A	1		B	$\frac{1}{2}$
C	0		D	$\frac{1}{3}$

Frage Q 14. Welche Standardabweichung hat ungefähr ein Würfel?

A	≈ 1.5		B	≈ 1
C	≈ 3		D	≈ 6

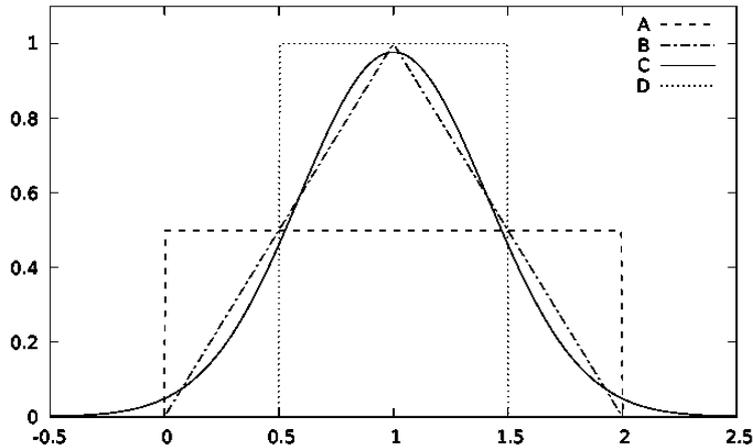
Frage Q 15. Bildet man aus 100 unabhängigen ZV X_i mit Standardabweichung $\sigma(X_i) = 1$ (z.B. Messungen) den Mittelwert, sinkt die Standardabweichung auf $\frac{1}{10}$. Was ist die Standardabweichung, wenn die Größen "100% abhängig sind", weil $X_i = X_1$?

A	$\frac{1}{10}$		B	1
C	$\sqrt{\frac{1}{10}}$		D	10

Frage Q 16. X hat Varianz 1 und Y Varianz 2 und beide sind möglicherweise abhängig, was ist dann $\text{Cov}(X + Y, X - Y)$?

A	-1		B	verschieden
C	0		D	+3

Frage Q 19. Seien $X, Y \sim \mathcal{U}(0, 1)$ unabhängig. Was ist die Verteilung von $X + Y$?



A.4. Umfang korreliert mit Fläche

Die Kantenlängen eines Rechtecks X, Y sind unabhängig $\mathcal{N}(1, 2)$ und $\mathcal{N}(2, 1)$ verteilt. Seien U der Umfang und F die Fläche. Was sind $E U, V U, E F, V F, \text{Cov}(U, F)$? Die Reihenfolge entspricht steigender Schwierigkeit. Ihr benötigt nur die untenstehenden Rechenregeln.

Auszug aus Kapitel 1. des Skripts

Der **Erwartungswert** einer Zufallsvariablen X gibt an, welchen Wert die Zufallsgröße "im Mittel" annimmt. $E(X + Y) = E(X) + E(Y)$, $E(\lambda X) = \lambda E(X)$, $E(1) = 1$, $E(XY) = E(X)E(Y)$, falls X und Y unabhängig.

Die **Varianz** einer Zufallsvariablen ist $V(X) = E((X - E(X))^2) = E(X^2) - E(X)^2$. Sie gibt an, wie stark die Zufallsgröße um den Erwartungswert schwankt. $V(X + Y) = V(X) + V(Y)$, wenn X und Y unabhängig. $V(\lambda X) = \lambda^2 V(X)$.

Die **Kovarianz** zweier Zufallsvariablen ist $\text{Cov}(X, Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y)$. Sie gibt an, wie stark die Zufallsgrößen gekoppelt um den Erwartungswert schwanken. Die Kovarianz einer Zufallsgröße mit sich selbst ist die Varianz der Größe $\text{Cov}(X, X) = V(X)$. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$, $\text{Cov}(\lambda X, Y) = \lambda \text{Cov}(X, Y)$, $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$, $\text{Cov}(1, X) = 0$, $V(X + Y) = V(X) + 2 \text{Cov}(X, Y) + V(Y)$

A. Quizfragen und Arbeitsblätter

Frage Q 20. Welche der folgenden Zufallsvariablen in 7.1.1 sind unabhängig unter der angegebenen Bedingung?

A	X_1, X_3	B	S_1, S_3
C	$S_1, S_3 X_2$	D	$S_1, S_3 S_2$

Frage Q 21. Welche der folgenden Zufallsvariablen sind unabhängig unter der angegebenen Bedingung?

A	X_1, X_3	B	$Z_1, Z_3 X_3$
C	$X_1, X_3 X_2, U_3$	D	$X_1, X_3 Z_{1..3}, U_{1..3}$

A.5. Schnelltests einer Pflegekraft

Im Kampf gegen eine Pandemie sollen Pflegekräfte mit Schnelltests getestet werden. Man geht davon aus, dass sich bereits 3% der Pflegekräfte angesteckt haben. Die Pflegekräfte können sich jederzeit mit dem Virus anstecken. Pro Tag geschieht dies bei etwa 1% der Pflegekräfte. Von den Kranken werden pro Tag 20% wieder gesund (wir nehmen dies als zufällig an, eine diskussionswürdige aber praktische Annahme). Pro Tag wird ein Schnelltest durchgeführt. Dieser ist zwar günstig und schnell, aber dafür sehr ungenau. Bei einem Erkrankten ist der Test lediglich zu 60% positiv. Bei Gesunden ist das Ergebnis in 1% der Fälle positiv. Modelliere die Wahrscheinlichkeitsverteilung einer einzelnen Pflegekraft erkrankt zu sein, für den Einsatz im Bayes-Filter!

- Was sind X_t, Z_t, U_t und $p(x_0)$?
- Was sind $p(x_t | x_{t-1}, u_t)$ und $p(z_t | x_t)$?
- Wie sieht die Wahrscheinlichkeitsverteilung nach 1 Tag für eine Pflegekraft mit positivem Test aus?

Frage Q 22. Welche der folgenden 4 Ereignisse gehört bei der Bayesschen Hasenjagd zu welcher Änderung von $Bel(x_t) = p(X_t = x_t | Z_{t'} = z_{t'} \wedge U_{t'} = u_{t'} \forall t' \leq t)$?

1	Initialisierung	2	Dynamikschritt (D)
3	Messschritt $z_t = \text{false}(-)$	4	Messschritt $z_t = \text{true}(+)$
A	Bel wird links schwächer, wandert dadurch nach rechts	B	Bel hat einen Eintrag mit 100% sonst 0%
C	Bel wandert nach rechts und auseinander	D	Bel wird rechts schwächer, wandert dadurch nach links

Frage Q 23.

$$\begin{pmatrix} x \\ y \end{pmatrix}^T \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = ? \tag{A.3}$$

A $\begin{pmatrix} ax^2 & bxy \\ bxy & cy^2 \end{pmatrix}$	B $\begin{pmatrix} ax+by \\ bx+cy \end{pmatrix}$
C $ax^2 + 2bxy + cy^2$	D $ax^2 + bx + cy^2$

Frage Q 24.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = ? \tag{A.4}$$

A $\frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$	B $\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$
C $\frac{1}{ad-bc} \begin{pmatrix} a & b \\ c & d \end{pmatrix}$	D $\begin{pmatrix} a^{-1} & b^{-1} \\ c^{-1} & d^{-1} \end{pmatrix}$

A.6. Statistik aus der Lehre

Wir betrachten statistische Daten der Veranstaltung "Grundlagen der Medieninformatik 1" vom WS17/18, deren Prüfungsleistung benotete Übungszettel (tlw. einzeln, tlw. Gruppe) und ein benotetes Fachgespräch beinhaltet.

		E(.)	V(.)	Cov(.,.)	cor(.,.)
ÜZ-Note	Z	1,796366279	0,579323492	0,247231377	0,300893459
FG-Note	X	1,897342193	1,332526246		

Vornote / Pkt	0,7	1	1,3	1,7	2	2,3	2,7	3	3,3	3,7	4	5
1	14	22	12	12	6	1	4	1	2	1	0	2
1,3	16	16	15	10	5	3	0	3	3	0	2	2
1,7	5	6	6	2	6	6	3	4	1	2	1	1
2	3	4	7	4	4	2	2	4	1	4	1	4
2,3	2	3	3	3	2	3	2	0	1	1	0	3
2,7	0	2	4	2	2	0	0	1	0	1	1	3
3	1	0	0	2	1	2	0	0	0	0	0	0
3,3	2	0	0	1	2	2	1	0	0	0	3	2
3,7	0	2	0	0	1	2	0	2	1	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0

- Welche Note kann ich laut Konditionierungslemma im FG erwarten, wenn ich eine 1.0 in den Übungszetteln habe?
- Welche bei einer 3.0?
- Welche laut Verteilung? (Tabelle)
- Wie valide sind die Annahmen der Rechnung?
- Zum Vergleich: Was für eine Korrelation hätten wir, wenn man annimmt, die wahre Leistung eines Studis wäre wie eine Würfel verteilt, im FG zeigt sich die Einzelleistung und im ÜZ der Mittelwert von der 3er-Gruppe.

A. Quizfragen und Arbeitsblätter

Frage Q 25. Sei im 1-D Schienenbeispiel X die Position und Z die Zeit, die das Radarecho braucht um zum Radar zurückzukommen. Welche Einheiten haben dann c und k ?

A	$c[\text{m}], k[\text{s}]$	B	$c[\frac{\text{s}}{\text{m}}], k[\frac{\text{m}}{\text{s}}]$
C	$c[\frac{\text{m}}{\text{s}}], k[1]$	D	$c[\frac{\text{m}}{\text{s}}], k[\frac{\text{s}}{\text{m}}]$

A.7. Der landende Ballon

Ein Ballon soll autonom landen. Dazu soll seine Sinkgeschwindigkeit mit Hilfe eines nach unten gerichteten Dopplerradar bestimmt werden. So ein Dopplerradar misst die Verschiebung der Frequenz des Echos und bestimmt daraus die Relativgeschwindigkeit zwischen Radar und detektiertem Objekt, also dem Boden. Der Effekt ist analog zu der Frequenzverschiebung der Sirene eines herannahenden Krankenwagens. Leider sind die Messungen ziemlich verrauscht. Deshalb misst zusätzlich ein Accelerometer am Ballon die vertikale Beschleunigung. Beide Sensoren sollen fusioniert werden um zu einer genaueren Sinkgeschwindigkeit zu gelangen.

Modelliert das Problem mit einem Kalman Filter und gebt insbesondere an:

- Zustand X_t , Messung Z_t , Zustandsübergangsmessung U_t
- Die Parameter a_t , b_t und g_t des Dynamikmodells, das (X_{t-1}, U_t) auf X_t abbildet und das Dynamikrauschen $\sigma_{\epsilon t}$
- Die Parameter c_t und h_t des Messmodells, das X_t auf Z_t abbildet und das Messrauschen $\sigma_{\delta t}$.

Welche Annahmen trifft man, die in der Realität Probleme bereiten könnten?

A.8. Der fortgeschritten landende Ballon

Schnell kommt man drauf, dass für einen autonom landenden Ballon nicht nur die Sinkgeschwindigkeit sondern auch der Abstand zum Boden wichtig ist. Ein normaler nach unten schauender Radar misst genau diese Information, aber leider verrauscht.

Also soll aus Radar und Accelerometer eine möglichst gute Höhenschätzung fusioniert werden. Versucht das Problem mit einem Kalman Filter zu modellieren, auf welches Problem stoßt ihr?

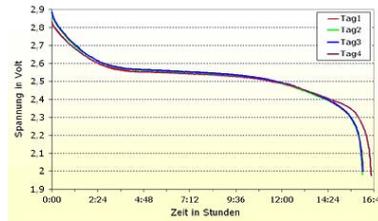
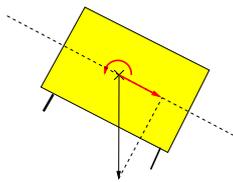


Abbildung A.1.: **Links:** Ein Zug in (übertriebener) Neigung. Ein Beschleunigungsmesser misst den Teil der Gravitation, der in Zugrichtung von rechts nach links verläuft und damit indirekt die Neigung (Pfeil). Ein Gyroskop misst die Drehung um die Vorne-Hinten-Achse (Drehpfeil). **Rechts:** Entladekurve $q(t)$ (Spannung als Funktion der Zeit) eines Akkus in einem elektronischen Gerät bei 100mA Entladestrom.

A.9. Neigungsschätzung im Zug

In einem fahrenden Zug soll die momentane Neigung quer zur Fahrtrichtung geschätzt werden, z.B. als Teil eines Systems um Züge automatisch in die Kurve zu neigen (Neigetechnik, Pendolino). Dazu ist an dem Zug ein Beschleunigungsmesser (Accelerometer) in links-rechts Richtung angebracht. Dieser misst den Teil der Erdbeschleunigung, der in links-rechts-Richtung in Zugkoordinaten zeigt und damit indirekt die Neigung. Leider misst er auch Beschleunigungen in Kurven und das Schlingern des Zuges um das Gleis, den sogenannten Sinuslauf.

Deshalb soll die Schätzung der Neigung durch Sensorfusion verbessert werden. Dazu wird ein Drehratenmesser (Gyroskop) hinzugenommen, der die Drehgeschwindigkeit um die Vor-Zurück-Achse und damit die Änderung der Neigung misst (Abb. A.1, links).

Die beiden Sensoren sollen mit einem Extended Kalman Filter (EKF) fusioniert werden. Modelliert das Problem im Rahmen eines EKF und gibt insbesondere an:

- Zustand X_t , Messung Z_t , Zustandsübergangsmessung U_t
- Dynamikfunktion g_t die (X_{t-1}, U_t) auf X_t abbildet und das Dynamikrauschen $\sigma_{\epsilon t}$
- Messfunktion h_t die X_t auf Z_t abbildet und das Messrauschen $\sigma_{\delta t}$.

A.10. Anzeige der Akkulaufzeit

Ein elektronisches Gerät wird mit einem Akku betrieben. Das Gerät misst den Stromverbrauch und die Batteriespannung. Abb. A.1 (rechts) zeigt einen typische Verlauf der Batteriespannung während des Entladens. Es soll die Restladung (in Coulomb = As) geschätzt werden. Spannung kann an sich sehr genau gemessen werden, aber die Akkuspannung wird auch von der Last und vom Alter des Akkus beeinflusst. Deshalb soll ein EKF die Schätzung verbessern.

Modelliert das Problem mit dem EKF. Gebt die selben Größen wie in Aufgabe 1 an. Was würde sich ändern, wenn man den Innenwiderstand der Batterie berücksichtigt?

A.11. Zweimal nichtlinear konditioniert

Es seien $X \sim \mathcal{N}(2, 2)$, $f: \mathbb{R} \rightarrow \mathbb{R}, x \mapsto x^2$, $Z = f(X)$ und $z = 1$. Wir fragen uns, was die verschiedenen Näherungsverfahren von EKF und UKF für die im Konditionierungslemma vorkommenden Größen $E Z, V Z, \text{Cov}(X, Z), E(X|Z = z)$ und $V(X|Z = z)$ für Ergebnisse liefern. Im Kontext von Sensorfusion entspricht dies dem Messupdate, wobei X der gesuchte Zustand und Z eine Messung ohne Rauschen mit Messfunktion f ist.

Die Aufgabe ist der Anfang von Aufgabe 16 aus Übungszettel 4.

- Berechne Näherungswerte der 5 Größen mit den ableitungsbasierten Rechenregeln in Kap. 9.1 und dem Konditionierungslemma 8.3, so wie ein EKF.

$$E(f(X)) \approx f(EX) \quad (\text{A.5})$$

$$V(f(X)) \approx (f'(EX))^2 V X \quad (\text{A.6})$$

$$\text{Cov}(f(X), X) \approx f'(EX) V(X) \quad (\text{A.7})$$

$$E(X|Z = z) \approx EX + \frac{\text{Cov}(X, Z)}{V Z} (z - EZ) \quad (\text{A.8})$$

$$V(X|Z = z) \approx V X - \frac{\text{Cov}(X, Z)^2}{V Z}. \quad (\text{A.9})$$

- Berechne Näherungswerte der 5 Größen mit der Sigmapunktpropagation in Kap. 9.4 und dem Konditionierungslemma 8.3, so wie ein UKF.

$$\mu = E X, \quad \sigma^2 = V X \quad (\text{A.10})$$

$$\mathcal{X} = (\mu + \sqrt{2}\sigma, \mu, \mu, \mu - \sqrt{2}\sigma) \quad (\text{A.11})$$

$$\bar{\mathcal{X}} = f(\mathcal{X}) \Leftrightarrow \bar{\mathcal{X}}^{[i]} = f(\mathcal{X}^{[i]}) \quad (\text{A.12})$$

$$E(f(X)) \approx \bar{\mu}_f = \frac{1}{4} \sum_{i=1}^4 \bar{\mathcal{X}}^{[i]} \quad (\text{A.13})$$

$$V(f(X)) \approx \bar{\sigma}_f^2 = \frac{1}{4} \sum_{i=1}^4 (\bar{\mathcal{X}}^{[i]} - \bar{\mu}_f)^2 \quad (\text{A.14})$$

$$\text{Cov}(f(X), X) \approx \bar{\sigma}_{fX}^2 = \frac{1}{4} \sum_{i=1}^4 (\bar{\mathcal{X}}^{[i]} - \bar{\mu}_f)(\mathcal{X}^{[i]} - \mu) \quad (\text{A.15})$$

$$E(X|Z = z) \approx EX + \frac{\text{Cov}(X, Z)}{V Z} (z - EZ) \quad (\text{A.16})$$

$$V(X|Z = z) \approx V X - \frac{\text{Cov}(X, Z)^2}{V Z}. \quad (\text{A.17})$$

A.12. Anschauliche Bedeutung von E und Cov in 2D

Frage Q 26. Sei $X \sim \mathcal{N}(0, 1)$ und $Y \sim \mathcal{N}(0, 2)$ und $Z = X + Y + \mathcal{N}(0, 1)$ mit unabhängigen Gaußverteilungen. Was ist $\text{Cov} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$?

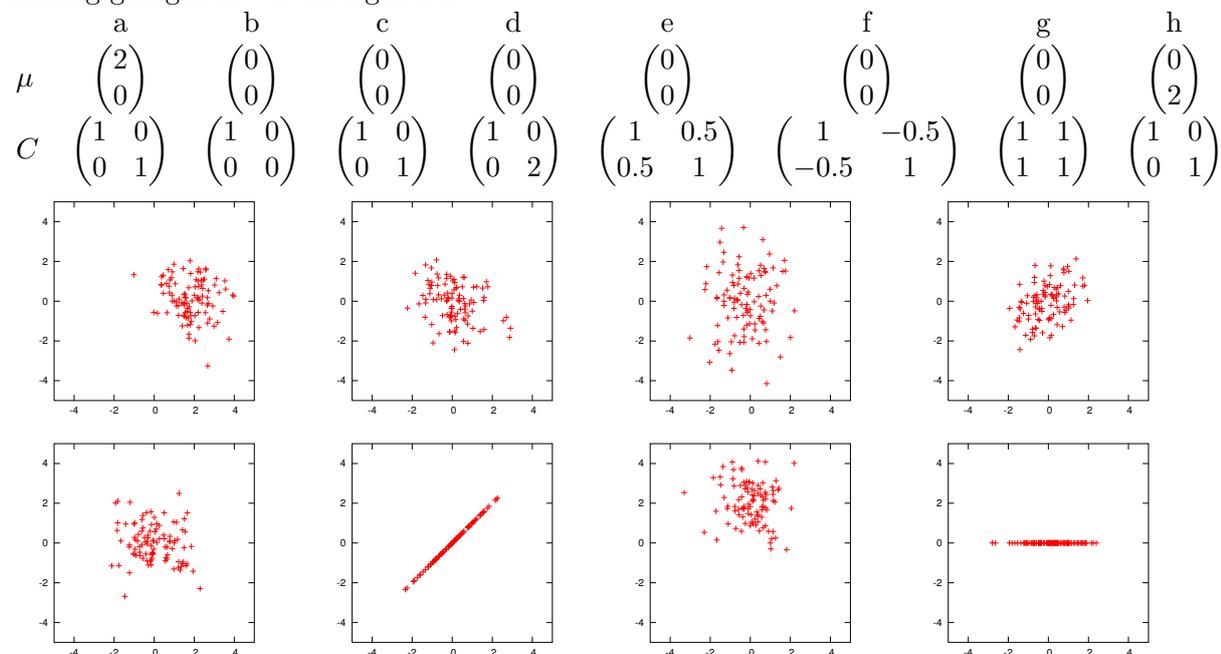
A	$(1, 2, 4)$	B	$\begin{pmatrix} 1, 0, 1 \\ 0, 2, 2 \\ 1, 2, 4 \end{pmatrix}$
C	$\begin{pmatrix} 1, 0, 0 \\ 0, 2, 0 \\ 0, 0, 4 \end{pmatrix}$	D	$\begin{pmatrix} 1, 0, 1 \\ 0, 2, 2 \\ 1, 2, 1 \end{pmatrix}$

Frage Q 27. Eine zweidimensionale Zufallsvariable X hat die Kovarianz $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. Welche Richtung r , $\|r\| = 1$ hat die geringste Varianz $V(r^T X) = r^T \text{Cov}(X)r$?

A	$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$	B	$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$
C	$\begin{pmatrix} 1 \\ 1 \end{pmatrix} / \sqrt{2}$	D	$\begin{pmatrix} 1 \\ -1 \end{pmatrix} / \sqrt{2}$

A.12. Anschauliche Bedeutung von E und Cov in 2D

Ordne die folgenden Paare aus Erwartungswert μ und Kovarianz C den untenstehenden zufällig gezogenen Verteilungen zu.



A.13. Höherdimensionale Kovarianzmatrizen

Charakterisiere sprachlich die 4D-Unsicherheiten folgender Kovarianzmatrizen:

$\begin{pmatrix} 9 & 0 & 0 & 0 \\ 0 & 9 & 0 & 0 \\ 0 & 0 & 9 & 0 \\ 0 & 0 & 0 & 9 \end{pmatrix}$	$\begin{pmatrix} 9 & 9 & 9 & 9 \\ 9 & 9 & 9 & 9 \\ 9 & 9 & 9 & 9 \\ 9 & 9 & 9 & 9 \end{pmatrix}$	$\begin{pmatrix} 9 & 1 & 1 & 1 \\ 1 & 9 & 1 & 1 \\ 1 & 1 & 9 & 1 \\ 1 & 1 & 1 & 9 \end{pmatrix}$	$\begin{pmatrix} 9 & 8 & 0 & 0 \\ 8 & 9 & 0 & 0 \\ 0 & 0 & 9 & 8 \\ 0 & 0 & 8 & 9 \end{pmatrix}$
--	--	--	--

A.14. Mittelwertvektor und Kovarianzmatrix abgeleiteter Größen

Sei $X \mathcal{N}(1, 2)$ verteilt und unabhängig $Y \mathcal{N}(2, 1)$ verteilt. Was ist $E(Z)$, $\text{Cov}(Z)$ für den Vektor $Z = (X, Y, X + Y)^T$? Benutze die Rechenregeln!

Frage Q 28. Was ist die mittlere quadratische Länge $\sqrt{E(\|X\|^2)} = \sqrt{E(X_1^2 + X_2^2 + X_3^2)}$ eines Standard-Gaußverteilten 3D Vektors X ?

$$\begin{array}{cc|cc} \text{A} & 1 & \text{B} & \frac{3}{\pi} \\ \text{C} & \sqrt{3} & \text{D} & 3 \end{array}$$

Frage Q 29. Was ist die Choleskyzerlegung von $\begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}$?

$$\begin{array}{cc|cc} \text{A} & \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} & \text{B} & \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \\ \text{C} & \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} & \text{D} & \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \end{array}$$

Frage Q 30. Wenn A nicht positiv semidefinit ist ($\exists x : x^T A x < 0$). Dann ist $p(X = x) \propto \exp(-\frac{1}{2}(x^T A x + x^T b + \gamma))$ keine Wahrscheinlichkeitsverteilung, weil ...

A $p(X = x) < 0$ wird	B $\int p(X = x) dx = \infty$
C A nicht invertierbar ist	D die Dichte ∞ wird

Frage Q 31. Welche Matrix Σ_ε drückt im Beispiel aus, dass das Rauschen proportional zum Gewicht ist?

A $\Sigma_\varepsilon = \begin{pmatrix} 25 & & & \\ & 100 & & \\ & & 225 & \\ & & & 400 \end{pmatrix}$	B $\Sigma_\varepsilon = \begin{pmatrix} \frac{1}{5} & & & \\ & \frac{1}{10} & & \\ & & \frac{1}{15} & \\ & & & \frac{1}{20} \end{pmatrix}$
C $\Sigma_\varepsilon = \begin{pmatrix} 5 & & & \\ & 10 & & \\ & & 15 & \\ & & & 20 \end{pmatrix}$	D $\Sigma_\varepsilon = \begin{pmatrix} \frac{1}{25} & & & \\ & \frac{1}{100} & & \\ & & \frac{1}{225} & \\ & & & \frac{1}{400} \end{pmatrix}$

A.15. Parabelfit

Ein Messaufbau liefert regelmäßig Messwerte für $t = -3, -2, -1, 0, 1, 2, 3$. Die Messwerte sehen im Graphen parabelförmig aus, deshalb soll eine Parabel an die Messwerte gefittet werden. Formalisiert dies Problem als lineare Ausgleichsrechnung. Was ist X , was Z ? Schreibt A , b und Σ_ε konkret für die obigen 7 Zeitschritte hin.

A.16. Tune in - Least Square Version

Ein Gerät erhält als Signal eine verrauschte Sinusschwingung. Es soll deren Amplitude und Frequenz bestimmt werden. Formuliert das Problem als nichtlineare quadratische Ausgleichsrechnung.

- Was ist X ? Aus welchen Komponenten besteht X als Vektor?
- Was ist Z ? Aus welchen Komponenten besteht Z als Vektor?
- Was ist f ? Aus welchen Komponenten besteht f als eine Funktion die von X nach Z abbildet? Was sind die Formeln für die einzelnen Komponenten?
- Was ist Σ_ϵ ? Aus welchen Komponenten besteht Σ_ϵ ? Wie kommt man zu den Zahlenwerten?

A.17. GPS in a nutshell

Bei der Satellitennavigation GPS senden Satelliten an Orten p_i zu Zeitpunkten t_i ein Signal aus. Beide Größen sind bekannt, weil sie der GPS Satellit mitsendet. Ein GPS-Empfänger empfängt das Signal zu einem leicht späteren Zeitpunkt t'_i und berechnet aus den Laufzeiten $t'_i - t_i$ die Position des Empfängers.

Formuliere das Problem, die Position des Empfängers zu schätzen als nichtlineare quadratische Ausgleichsrechnung. Zuerst unter der Annahme, dass die Uhr im Empfänger perfekt genau ist.

- Was ist X ? Aus welchen Komponenten besteht X als Vektor?
- Was ist Z ? Aus welchen Komponenten besteht Z als Vektor?
- Was ist f ? Aus welchen Komponenten besteht f als eine Funktion die von X nach Z abbildet? Was sind die Formeln für die einzelnen Komponenten?
- Was ist Σ_ϵ ? Aus welchen Komponenten besteht Σ_ϵ ? Wie kommt man zu den Zahlenwerten?

Was ändert sich, wenn man annimmt, dass die Uhr im Empfänger leicht ungenau ist?

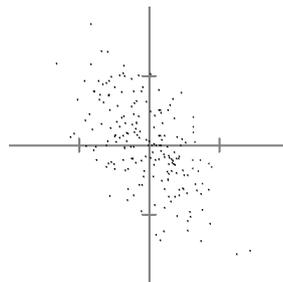
A.18. Der Kompass dreht sich nichtlinear im Kreis

In einem Auto ist ein 2D Magnetsensor eingebaut, der als elektronischer Kompass die Himmelsrichtung anzeigen soll (wo ist Norden?). Technisch misst er das Erdmagnetfeld b_t in der Ebene als 2D Vektor. Leider misst er nicht nur das externe Magnetfeld (b_{ext}) der Erde, sondern auch ein Magnetfeld (b_{int}) das vom Metall des Autos ausgeht und sich mit dem Auto und dem Sensor mitdreht. Dieser Effekt verfälscht die angezeigte Himmelsrichtung und soll rauskalibriert werden. Als Kalibrierungsprozedur fährt das Auto eine 360° Kurve, während derer Messungen des Magnetsensors aufgezeichnet werden.

Diese Aufgabe war schon im Übungszettel, dort stand ein Sensor zur Verfügung, der die Orientierung genau genug misst. Hier soll einer solcher Zusatzsensor nicht benutzt werden, sondern die Kalibrierung nur aus den Daten des Magnetsensors während einer nicht direkt gemessenen Drehung berechnet werden.

- Was sind X und Z sowie ihre Komponenten?
- Was ist Σ_ϵ und wo kommen die Zahlen her?
- Was ist das Messmodell $Z = \dots$? Formuliere es zuerst in der Schreibweise der Domäne, so dass der physikalisch / geometrische Zusammenhang klar wird.
- Wo liegt der Unterschied zu der Situation mit einem Orientierungssensor für die Kalibrierung?
- In der naheliegendsten Formulierung ist ein Freiheitsgrad nicht beobachtbar? Welcher?
- Wie kann man den Freiheitsgrad beobachtbar machen?
- Nehmen wir an, wir haben ein zusätzliches Gyroskop um die Hochachse. Wie liessen sich dessen Messungen in die Fusion integrieren?

Frage Q 32. X und Z haben Erwartungswert 0 und eine gemeinsame Kovarianz von $\begin{pmatrix} 1 & -0.7 \\ -0.7 & 1.49 \end{pmatrix}$ wie in der Grafik dargestellt (Striche bei 2). Was ist $E(X|Z = 2)$?



A	$\approx +0.94$	B	≈ 0
C	≈ -0.94	D	≈ -2

Frage Q 33. In einer Variante des 1-D Kohlenkarre soll die Position des Radars (Z_t) gestützt werden durch einen Beschleunigungsmesser (links/rechts, U_t). Dazu wird die Geschwindigkeit mit in den Zustand genommen ($X_t = \begin{pmatrix} p_t \\ v_t \end{pmatrix}$). Was sind die Matrizen A_t , B_t und C_t ?

A. Quizfragen und Arbeitsblätter

A $A_t = \begin{pmatrix} 1 & \Delta t \\ 0 & 1 \end{pmatrix}, B_t = \begin{pmatrix} \frac{1}{2}\Delta t^2 \\ \Delta t \end{pmatrix}, C_t = (1 \ 0)$	B $A_t = (1), B_t = (\Delta t), C_t = (1)$
C $A_t = \begin{pmatrix} \Delta t & \Delta t^2 \\ \Delta t^2 & \Delta t \end{pmatrix}, B_t = (1, 0), C_t = (0 \ 1)$	D $A_t = \begin{pmatrix} 1 & \Delta t \\ 0 & 1 \end{pmatrix}, B_t = \begin{pmatrix} 0 \\ \Delta t \end{pmatrix}, C_t = (1 \ 0)$

Frage Q 34. Ein Fahrzeug in der Ebene misst den Winkel (Peilung) zu einer bekannten punktförmigen Landmarke. Welches Format haben die Matrizen in der Kalman Gain Formel $\bar{\Sigma}C^T(C\bar{\Sigma}C^T + \Sigma_\varepsilon)^{-1}$?

A $\begin{pmatrix} * & * & * \\ * & * & * \\ * & * & * \end{pmatrix} \begin{pmatrix} * & * & * \end{pmatrix} \left(\begin{pmatrix} * \\ * \\ * \end{pmatrix} \begin{pmatrix} * & * & * \\ * & * & * \\ * & * & * \end{pmatrix} \begin{pmatrix} * & * & * \end{pmatrix} + \begin{pmatrix} * \end{pmatrix} \right)^{-1}$	B $(*) \begin{pmatrix} * & * & * \end{pmatrix} \left(\begin{pmatrix} * \\ * \\ * \end{pmatrix} \begin{pmatrix} * \end{pmatrix} \begin{pmatrix} * & * & * \end{pmatrix} + \begin{pmatrix} * & * & * \\ * & * & * \\ * & * & * \end{pmatrix} \right)^{-1}$
C $\begin{pmatrix} * & * & * \\ * & * & * \\ * & * & * \end{pmatrix} \begin{pmatrix} * \\ * \\ * \end{pmatrix} \left(\begin{pmatrix} * & * & * \end{pmatrix} \begin{pmatrix} * & * & * \\ * & * & * \\ * & * & * \end{pmatrix} \begin{pmatrix} * \\ * \\ * \end{pmatrix} + \begin{pmatrix} * \end{pmatrix} \right)^{-1}$	D $\begin{pmatrix} * \\ * \\ * \end{pmatrix} \begin{pmatrix} * & * & * \end{pmatrix} \left(\begin{pmatrix} * \\ * \\ * \end{pmatrix} \begin{pmatrix} * & * & * \end{pmatrix} \begin{pmatrix} * \\ * \\ * \end{pmatrix} + \begin{pmatrix} * \\ * \\ * \end{pmatrix} \right)^{-1}$

A.19. Der landende Ballon 2

Wir betrachten erneut einen landenden Ballon der mit einem nach unten schauenden Radar und einem Beschleunigungsmesser ausgerüstet ist. Diesmal handelt es sich aber nicht um einen Dopplerradar der die Sinkgeschwindigkeit misst, sondern um einen gewöhnlichen Radar der die Entfernung zum Boden misst.

Modelliert das Problem als Kalmanfilter!

- Was ist X_t ? Was Z_t ? Was U_t ?
- Was ist das Messmodell $Z_t = \dots$? Wie wird es durch C_t und h_t beschrieben?
- Was ist das Dynamikmodell $X_t = \dots$? Wie wird es durch A_t, B_t, g_t beschrieben?
- Woher kommen die Unsicherheiten $\Sigma_{\epsilon t}$ und $\Sigma_{\delta t}$?

Frage Q 35. Ein Fahrzeug fährt einen Meter vorwärts. Was ist die Jakobimatrix der Position (x', y') danach nach der Position (x, y) und Orientierung (θ) davor (zusammengefasst als Vektor $\begin{pmatrix} x \\ y \\ \theta \end{pmatrix}$)?

A	$\begin{pmatrix} 1 & 1 & \cos \theta \end{pmatrix}$	B	$\begin{pmatrix} 1 & 0 & -\sin \theta \\ 0 & 1 & \cos \theta \end{pmatrix}$
C	$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$	D	$\begin{pmatrix} 1 & 0 & \cos \theta \\ 0 & 1 & \sin \theta \end{pmatrix}$

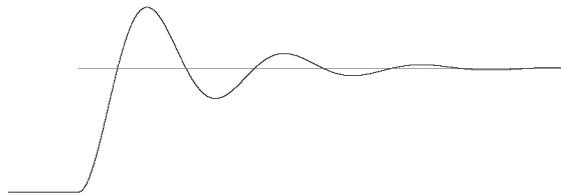
A.20. Die schnelle Waage

Eine Wägezelle, der Sensor einer Waage, misst vereinfacht gesagt, wie stark eine Feder unter dem Gewicht der zu messenden Masse komprimiert wird. Legt man plötzlich ein Gewicht auf die Waage, setzt zuerst eine Schwingung ein die erst abklingen muss, bevor man das Gewicht ablesen kann.

In einer elektronischen Waage soll dieses Problem gelöst werden, indem ein (Extended/Unscented) Kalman Filter die Masse aus dem Verlauf der Messungen des Wägesensors schätzt.

Modelliert dies Problem! Was ist der Zustand X_t , Messung Z_t und Zustandsübergangsmessung U_t ? Was ist das Dynamikmodell g_t und das Messmodell h_t ?

In den Modellen kommen Parameter vor die fest, aber spezifisch für die Waage und erst einmal unbekannt sind. Wie könnte man die bestimmen?



A.21. Tune in EKF Version

Ein Gerät erhält als Signal eine verrauschte Sinusschwingung. Es soll deren Amplitude und Frequenz bestimmt werden. Modelliert dieses Problem! Was ist der Zustand X_t , Messung Z_t und Zustandsübergangsmessung U_t ? Was ist das Dynamikmodell g_t und das Messmodell h_t ?

A.22. Fahrender Zug mit Bias

In einem fahrenden Zug soll die momentane Neigung quer zur Fahrtrichtung geschätzt werden, z.B. als Teil eines Systems um Züge automatisch in die Kurve zu neigen (Neigetechnik, Pendolino). Dazu ist an dem Zug ein Beschleunigungsmesser (Accelerometer) in links-rechts Richtung angebracht. Dieser misst den Teil der Erdbeschleunigung, der in links-rechts-Richtung in Zugkoordinaten zeigt und damit indirekt die Neigung. Leider misst er auch Beschleunigungen in Kurven und das Schlingern des Zuges um das Gleis, den sogenannten Sinuslauf.

Deshalb soll die Schätzung der Neigung durch Sensorfusion verbessert werden. Dazu wird ein Drehratenmesser (Gyroskop) hinzugenommen, der die Drehgeschwindigkeit um die Vor-Zurück-Achse und damit die Änderung der Neigung misst. Allerdings hat das Gyroskop einen bias also einen Mittelwertsfehler (die Messung des Gyroskops ist nicht mehr ω sonder $\omega + b$). Dieser ist annähernd konstant, daher kann er anfangs herauskalibriert werden. Mit der Zeit ändert sich aber der bias, weshalb er mitgeschätzt werden muss.

Die beiden Sensoren sollen mit einem Extended Kalman Filter (EKF) fusioniert werden. Modelliert das Problem im Rahmen eines EKF. Was ist der Zustand X_t , Messung Z_t und Zustandsübergangsmessung U_t ? Was ist das Dynamikmodell g_t und das Messmodell h_t ?

Es gibt zwei Varianten für das Modell, wie sich der Bias verändert:

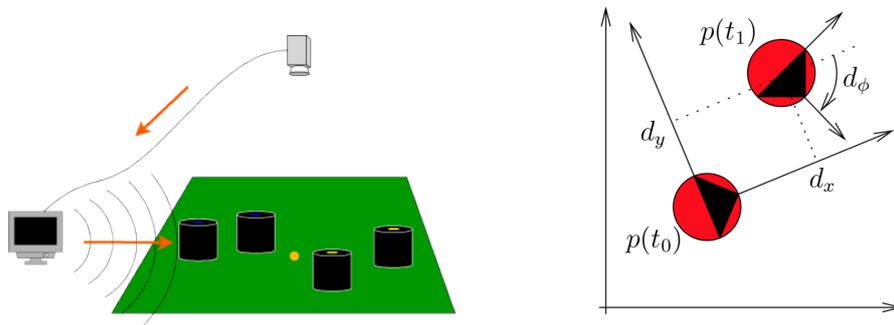
a) *Random Walk*: In einem Zeitraum von Δt akkumuliert der Bias ein Rauschen von $\mathcal{N}(0, \Delta t \sigma_b^2)$. Das bedeutet, dass a-priori der Bias immer größer wird.

b) *Farbiges Rauschen*: Zu jedem Zeitpunkt hat der Bias a-priori eine Standardabweichung von σ_b . Der Bias zu zwei δt entfernten Zeitpunkten hat eine Korrelation von $\exp(-\frac{\Delta t}{\tau})$. Das bedeutet, dass der Bias im vorgegebenen Bereich hin- und her wandert und τ ist die Zeit in der sich die Korrelation auf $1/3 \approx 0.37$ reduziert.

A.23. RoboCup Small Size League I

Beim Roboterfußballwettbewerb “RoboCup” in der Small Size League spielen radgetriebene Roboter Fußball. Die Roboter werden von einem Zentralrechner per Funk gesteuert. Auf Grund spezieller Räder können sich die Roboter nicht nur vorwärts-rückwärts, sondern auch quer bewegen und um beliebige Punkte drehen. Diese Radbewegungen werden gemessen und aus den rohen Messwerten berechnet, wie sich der Roboter relativ zum letzten Zeitschritt bewegt hat. Die Roboter werden von einer Deckenkamera beobachtet, die aber hier nur einen Bildpunkt pro Roboter, keine Orientierung liefert.

Modelliert dieses Problem! Was ist der Zustand X_t , Messung Z_t und Zustandsübergangsmessung U_t ? Was ist das Dynamikmodell g_t und das Messmodell h_t ? Woran beobachtet der Filter anschaulich die Orientierung?



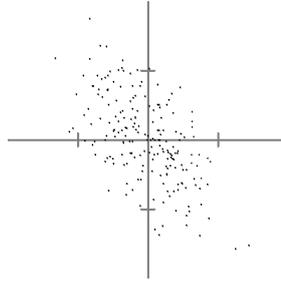
Bildquellen: http://robocupssl.cpe.ku.ac.th/_media/dataflow.png, eigene Arbeit.

A.24. RoboCup Small Size League II

Nehmen wir im obigen Beispiel an, dass die Fahrkommandos sehr ungenau ausgeführt werden, z.B. durch Schlupf zwischen Rädern und Boden. Daher sollen die Raddrehsensoren nicht benutzt werden. Um dies zu kompensieren wird jeder Roboter mit einem 2D Accelerometer (Beschleunigungsmesser) und einem 1D Gyrometer (Drehgeschwindigkeitsmesser) ausgestattet, dessen Daten zum Zentralrechner übertragen werden. Modelliert dies Problem! Was ist der Zustand X_t , Messung Z_t und Zustandsübergangsmessung U_t ? Was ist das Dynamikmodell g_t und das Messmodell h_t ?

A. Quizfragen und Arbeitsblätter

Frage Q 36. Zeichnet ungefähr die Sigma-Punkte in die unten stehende Verteilung ein!



A.25. Sigmas Cousinen vom Grade null bis drei

In dieser Aufgabe wollen wir das zentrale Argument dafür, dass die Sigmapunktpropagation korrekt vom Grad 3 ist am Beispiel nachrechnen.

Sei X standardgaußverteilt in 2 Dimensionen.

1. Schreibe die zu der Verteilung von X passenden Sigmapunkte \mathcal{X} auf.
2. Sei $f_0 : \mathbb{R}^2 \rightarrow \mathbb{R}, \begin{pmatrix} x \\ y \end{pmatrix} \mapsto 1$ ein beispielhaftes Monom vom Grad 0.
Was ist $E(f(X))$? Was $E(f(\mathcal{X}))$?
3. Sei $f_1 : \mathbb{R}^2 \rightarrow \mathbb{R}, \begin{pmatrix} x \\ y \end{pmatrix} \mapsto x$ ein beispielhaftes Monom vom Grad 1.
Was ist $E(f(X))$? Was $E(f(\mathcal{X}))$?
4. Seien $f_{2a} : \mathbb{R}^2 \rightarrow \mathbb{R}, \begin{pmatrix} x \\ y \end{pmatrix} \mapsto x^2$ und $f_{2b} : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto xy$ zwei beispielhafte Monome vom Grad 2. Was ist jeweils $E(f(X))$? Was $E(f(\mathcal{X}))$?
5. Seien $f_{3a} : \mathbb{R}^2 \rightarrow \mathbb{R}, \begin{pmatrix} x \\ y \end{pmatrix} \mapsto x^3$ und $f_{3b} : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto xy^2$ zwei beispielhafte Monome vom Grad 3. Was ist jeweils $E(f(X))$? Was $E(f(\mathcal{X}))$?

Frage Q 37. Welche Funktion ist keine Isometrie in \mathbb{R}^2 ?

A	Drehung um $(2, 1)$ um 23° .	B	$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 2x \\ y \end{pmatrix}$
C	$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} -y \\ x \end{pmatrix}$	D	Spiegelung an $y = 0.3x + 2$.

Frage Q 38. In welcher Norm gibt es mehr als einen Punkt, der sowohl 1 von $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ als auch 1 von $\begin{pmatrix} 2 \\ 0 \end{pmatrix}$ entfernt ist?

A	$\ \begin{pmatrix} x \\ y \end{pmatrix} \ _2 = \sqrt{x^2 + y^2}$	B	$\ \begin{pmatrix} x \\ y \end{pmatrix} \ _1 = x + y $
C	$\ \begin{pmatrix} x \\ y \end{pmatrix} \ _3 = \sqrt[3]{ x ^3 + y ^3}$	D	$\ \begin{pmatrix} x \\ y \end{pmatrix} \ _\infty = \max\{ x , y \}$

Frage Q 39. Seien die Transformationen Aircraft-in-World $T_{W \leftarrow A}$ und Runway-in-World $T_{W \leftarrow R}$ gegeben. Wie berechnet man daraus Aircraft-in-Runway $T_{R \leftarrow A}$?

A	$T_{W \leftarrow A} T_{W \leftarrow R}^{-1}$	B	$T_{W \leftarrow R}^{-1} T_{W \leftarrow A}$
C	$T_{W \leftarrow R} T_{W \leftarrow A}^{-1}$	D	$T_{W \leftarrow A}^{-1} T_{W \leftarrow R}$

Frage Q 40. Warum ist mit der angegebenen Axiomatisierung $\text{Rot}(v, \alpha, v) = v$, v also eine Achse?

Frage Q 41. Betrachten wir die Kugel $\| (x \ y \ z) \| = 2\pi$ mit Radius 2π . Alle Drehungen $\text{Rot}(x, y, z)$ auf dieser Kugel sind "einmal herumgedreht" also I . Betrachten wir jetzt konkret $\text{Rot}(x, y, z)$ an der Stelle $x = 2\pi, y = 0, z = 0$. Was passiert anschaulich, wenn wir y ein klein bisschen erhöhen?

A	fast nix	B	Drehung um y
C	Drehung um x	D	Drehung um z

A. Quizfragen und Arbeitsblätter

Frage Q 42. Was passiert, wenn man eine Orientierung $Q \in SO(3)$ schätzen will und dazu dies als $Q = \text{Rot}(v)$ parametrisiert und v in den Zustand eines EKF packt?

A	es funktioniert	B	Problem sofort
C	Problem im Dynamikschritt bei $ v \approx 2\pi$	D	Problem im Messschritt bei $ v \approx 2\pi$

Frage Q 43. Was passiert, wenn man eine Orientierung $Q \in SO(3)$ schätzen will, also eine Matrix $Q \in \mathbb{R}^{3 \times 3}$ mit $Q^T Q = I$ und dafür einfach in einem EKF die 9 Zahlen aus der Matrix in den Zustand packt?

A	nach Dynamikschritt $Q^T Q \neq I$	B	Kein $(\dots)^{-1}$ im Messschritt
C	Alles funktioniert	D	nach Messschritt $Q^T Q \neq I$

Frage Q 44. X sei eine zufällige Orientierung im Raum: $X : \Omega \rightarrow SO(3)$, die $X \sim \mathcal{N}(I, \sigma^2 I)$ verteilt ist. Was ist $\text{Cov}(X)$?

A	$\sigma^2 I$	B	$\sigma'^2 I, \sigma' < \sigma$
C	$\sigma'^2 I, \sigma' > \sigma$	D	Kein Vielfaches von I

Frage Q 45. Es soll eine 1D-Drehung als 2×2 -Matrix in einer \boxplus -Mannigfaltigkeit modelliert werden. Wie kann man \boxplus und \boxminus wählen? (2 richtige Antworten)

A	$x \boxplus \delta = x \exp(\delta), \quad y \boxminus x = \log(x^{-1}y)$	$\exp(\delta) = \begin{pmatrix} \cos(\delta) & -\sin(\delta) \\ \sin(\delta) & \cos(\delta) \end{pmatrix}$ $\log(x) = \text{atan2}(x_{21}, x_{11})$ $\log(\exp(\delta)) = \delta$
B	$x \boxplus \delta = x \exp(\delta), \quad y \boxminus x = \log(y^{-1}x)$	
C	$x \boxplus \delta = \exp(\delta)x, \quad y \boxminus x = \log(y^{-1}x)$	
D	$x \boxplus \delta = \exp(\delta)x, \quad y \boxminus x = \log(x^{-1}y)$	

$$(x \boxplus 0 = x) \tag{A.18a}$$

$$\forall y \in \mathcal{S} : x \boxplus (y \boxminus x) = y \tag{A.18b}$$

$$\forall \delta \in V : (x \boxplus \delta) \boxminus x = \delta \tag{A.18c}$$

$$\forall \delta_1, \delta_2 \in \mathbb{R}^n : \|(x \boxplus \delta_1) \boxminus (x \boxplus \delta_2)\| \leq \|\delta_1 - \delta_2\|. \tag{A.18d}$$

Sorry, this version does not include copyrighted Material.

Abbildung A.2.: Auf einem Schuh montierter Inertialsensor. (Quelle: Park and Suh, Sensors, 2010.) Das Magnetometer betrachten wir hier nicht.

A.26. Der tapsende Inertialsensor

Eine Person trägt einen Schuh mit darauf montiertem Inertialsensor. Es soll aus den Inertialsensordaten die Bewegung der Person in einem ebenen Gebäude verfolgt werden.

Dynamikmodell eines Inertialsensors

Wir wollen ein Dynamikmodell für einen Inertialsensor aufstellen. Dabei wollen wir erst mal annehmen, dass der Inertialsensor sich völlig frei im Raum bewegt. Ein Inertialsensor enthält ein Gyrometer, welches die momentane Drehachse ω_t in Sensorkoordinaten mit Drehrate (rad/s) $|\omega_t|$ misst.

Außerdem enthält es ein Accelerometer, welches die Beschleunigung a_t , also die zweite Ableitung der Position (in Weltkoordinaten) umgerechnet in Sensorkoordinaten misst. Allerdings misst es aus grundlegenden physikalischen Gründen nicht die Erdbeschleunigung. D.h. ein frei fallender Inertialsensor misst 0 (Schwerelosigkeit), ein liegender Inertialsensor misst die Beschleunigung nach oben mit der der Tisch ihn am fallen hindert.

Was ist Zustandsraum X_t , Zustandsübergangsmessung U_t , Dynamikmodell $g(x, u)$ und Dynamikrauschen für diesen frei fliegenden Inertialsensor?

Zustandsraum als \boxplus -Mannigfaltigkeit

Was sind die \boxplus und \boxminus -Operatoren die den Zustandsraum zu einer \boxplus -Mannigfaltigkeit machen?

Sensorfusion in Anschauung

Überlege, wieso das Ganze überhaupt ein Sensorfusionsproblem ist! Es gibt ja nur einen Sensor! Was könnte eine zweite Quelle von Information sein? Wie können man genauer werden als durch Aufakkumulation was für ein paar Sekunden geht.

Was sind Messung Z_t und Messmodell $h(x)$?

A. Quizfragen und Arbeitsblätter

Sorry, this version does not include copyrighted Material.

Abbildung A.3.: Ein Roboterarm hält eine Kamera in der Hand. Die Pose der Kamera relativ zur "Hand" des Roboters (hier TCP genannt) soll kalibriert werden, zusammen mit den Parametern die die Kamera beschreiben.
Quelle: K. Strobl., Optimal Hand-Eye Calibration, IROS 2006

A.27. Das Auge in die Hand nehmen

Ein Roboter trägt eine Kamera in der "Hand" und soll später damit Objekte manipulieren. Dazu muss er den räumlichen Bezug zwischen dem Kamerabild und der "Hand" also der Pose des Roboters die man steuert, dem sogenannten TCP kennen. Deshalb soll diese Vorab mit einem Schachbrett kalibriert werden, dass undefiniert in den Arbeitsraum gelegt wird.

Der Roboter bewegt sich durch den Arbeitsraum und nimmt das Schachbrett aus unterschiedlichen Perspektiven k auf, während er die Pose $T_{B \leftarrow T_k}$ abspeichert. Aus dem Bild erkennt eine Bildverarbeitungssoftware die Schachbrettecken ij an Bildposition u_{kij}, v_{kij} . Wobei die Kantenlänge eines Schachbrettkästchens d ist und i zur Schachbrett X-Achse und j zur Schachbrett Y-Achse korrespondiert.

Für die Kamera ist ein Modell $c(p^C, \theta)$ gegeben, welches angibt, wo ein 3D-Punkt p^C (relativ zur Kamera) im Bild zu sehen ist. Das Modell verwendet dabei Parameter θ (z.B. die Brennweite). Die Details des Modells interessieren hier nicht, wir betrachten es als Black-Box.

Modelliert das Problem mit quadratischer Ausgleichsrechnung auf \boxplus -Mannigfaltigkeiten.

- Was ist X ? Was ist Z ?
- Was ist f ? Was ist Σ ?
- Was sind \boxplus und \boxminus auf der Mannigfaltigkeit von X ?

A.28. Axiomatik anschaulich erklärt

Kapitel 16.3.1 vom Skript beweist, dass eine \boxplus -Mannigfaltigkeit eine natürlich Metrik $d(x, y)$ hat, nämlich die Norm $\|y \boxminus x\|$. Der Beweis nutzt zentral Axiom 4 ($\|(x \boxplus \delta_1) \boxminus (x \boxplus \delta_2)\| \leq \|\delta_1 - \delta_2\|$).

Erklärt den Beweis anschaulich mit Hilfe einer Skizze analog zu Abbildung 16.2 im Skript. D.h., fertigt eine analoge Skizze an, die die Situation im Beweis darstellt und erklärt dann die Schritte des Beweises an Hand dieser Skizze. Der Beweis besteht aus drei Teilen für die drei Metrikeigenschaften (16.16-18). Die erste Eigenschaft ist unkompliziert. Macht dementsprechend eine Skizze jeweils für die 2. und 3.Eigenschaft.

Frage Q 46. Wie viele Funkbaken benötigt man, um die Position eines Kohlekarrens auf einer Ebene eindeutig zu bestimmen, wenn jede Funkbake den Abstand zum Kohlekarren misst?

A	1	B	2
C	3	D	4

Frage Q 47. Mit welcher Messung ist das System $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, $\dot{x} = 0$ beobachtbar?

A	$z = \begin{pmatrix} x_1 \\ 2x_1 \end{pmatrix}$	B	$z = \begin{pmatrix} x_1 - x_2 \\ x_2 - x_1 \end{pmatrix}$
C	$z = \begin{pmatrix} x_1^2 + x_2^2 \\ \text{atan2}(x_1/x_2) \end{pmatrix}$	D	$z = \begin{pmatrix} x_1/x_2 \\ x_2/x_1 \end{pmatrix}$

Frage Q 48. Was ist die unbeobachtbare Transformation des Systems: $X_t = \begin{pmatrix} p_x \\ p_y \\ \theta \end{pmatrix}$, $U_t =$

$$\begin{pmatrix} v \\ \omega \end{pmatrix}, Z_t = p_x + p_y, \dot{X}_t = \begin{pmatrix} v \cos(\theta) \\ v \sin(\theta) \\ \omega \end{pmatrix}$$

A	$A(x_t) = \begin{pmatrix} p_x + \tau \\ p_y + \tau \\ \theta \end{pmatrix}$	B	$A(x_t) = \begin{pmatrix} p_x + p_y \\ 0 \\ \theta \end{pmatrix}$
C	$A(x_t) = \begin{pmatrix} p_x + \tau \\ p_y - \tau \\ \theta + \tau \end{pmatrix}$	D	$A(x_t) = \begin{pmatrix} p_x + \tau \\ p_y - \tau \\ \theta \end{pmatrix}$

Frage Q 49. Was ist die Jacobi-Matrix $\frac{\partial Z(x)}{\partial x}$ des Systems: $X_t = \begin{pmatrix} p_x \\ p_y \\ \theta \end{pmatrix}$, $U_t = \begin{pmatrix} v \\ \omega \end{pmatrix}$, $Z_t =$

$$p_x + p_y, \dot{X}_t = \begin{pmatrix} v \cos(\theta) \\ v \sin(\theta) \\ \omega \end{pmatrix} \text{ mit 2 Ableitungen?}$$

A	$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & -\sin(\theta)v + v \cos(\theta) \\ 0 & 0 & -v \cos(\theta) - v \sin(\theta) \end{pmatrix}$	B	$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & -\sin(\theta)v + v \cos(\theta) \\ 0 & 0 & -v\omega \cos(\theta) - v\omega \sin(\theta) \end{pmatrix}$
C	$\begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & \sin(\theta)v - v \cos(\theta) \\ 0 & 0 & v\omega \cos(\theta) + v\omega \sin(\theta) \end{pmatrix}$	D	$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$

Frage Q 50. Hat die Matrix $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & v(\cos(\theta) - \sin(\theta)) \\ 0 & 0 & -v\omega(\cos(\theta) + \sin(\theta)) \end{pmatrix}$ vollen Rang?

A. Quizfragen und Arbeitsblätter

A Ja | B Nein

Frage Q 51. Was ist die 1. Lie-Ableitung von $h = \sin(x) + y^2$ in Richtung $f = \begin{pmatrix} x \\ 2y \end{pmatrix}$?

A	$\cos(x)\dot{x} + 2y\dot{y}$	B	$\begin{pmatrix} x \sin(x) \\ 2y^3 \end{pmatrix}$
C	$x \cos(x) + 4y^2$	D	$\begin{pmatrix} \cos(x) & 2y \\ x & 2y \end{pmatrix}$

A.29. Beobachtbarkeit mit einer Winkelmessung

Beim Kohlenkarren in der Ebene erhält der Kohlenkarren Winkelmessungen zu Funkbaken. Überprüft, ob es reicht, eine Funkbake aufzustellen, um den Zustand des Karrens zu schätzen. Nutzt dazu die Systemgleichungen: $x = \begin{pmatrix} \rho \\ \varphi \\ \theta \end{pmatrix}$, $h = \varphi - \theta$, $\dot{x} = \begin{pmatrix} v \cos(\theta - \varphi) \\ \frac{v \sin(\theta - \varphi)}{\rho} \\ \omega \end{pmatrix}$

Und das Observability-Rank-Criteria:

Set $\Omega_0 = \text{span } \nabla h$

Set $k = 0$

do

Set $k=k+1$

Set $\Omega_k = \Omega_{k-1} \oplus \mathcal{L}_{f_0}^1 \Omega_{k-1} \oplus \bigoplus_{i=1}^p \mathcal{L}_{f_i}^1 \Omega_{k-1}$

while $\dim(\Omega_k) > \dim(\Omega_{k-1})$

Set $\Omega = \Omega_k$

Wenn $\dim(\Omega) = n$ dann ist das System schwach beobachtbar, ansonsten sind $n - \dim(\Omega)$ Freiheitsgrade unbeobachtbar

A.30. Beobachtbarkeit mit einer Winkel- und Abstandsmessung

Beim Kohlenkarren in der Ebene erhält der Kohlenkarren Winkelmessungen zu Funkbaken. Überprüft, ob es reicht, eine Funkbake aufzustellen, um den Zustand des Karrens zu schätzen, wenn man auch noch eine Abstandsmessung erhält. Nutzt dazu die Systemgleichungen: $x = \begin{pmatrix} \rho \\ \varphi \\ \theta \end{pmatrix}$, $h_1 = \varphi - \theta$, $h_2 = \rho$, $\dot{x} = \begin{pmatrix} v \cos(\theta - \varphi) \\ \frac{v \sin(\theta - \varphi)}{\rho} \\ \omega \end{pmatrix}$

Und das Observability-Rank-Criteria.

A.30. Beobachtbarkeit mit einer Winkel- und Abstandsmessung

A. Quizfragen und Arbeitsblätter

Antworten zu den Quizfragen:

Frage	Antwort
Q1	C
Q2	C
Q3	A
Q4	B
Q5	A,C
Q6	D
Q7	A,D
Q8	C
Q9	B
Q10	B
Q11	B
Q12	B
Q13	B
Q14	A
Q15	B
Q16	A
Q19	B
Q20	A,D
Q21	B,C
Q22	1:B, 2:C, 3:D, 4:A
Q23	C
Q24	A
Q25	B
Q26	B
Q27	D
Q28	C
Q29	A
Q30	B
Q31	A
Q32	C
Q33	A
Q34	C
Q35	B
Q41	A
Q42	D
Q43	D
Q44	B
Q45	A,C
Q46	C
Q47	C
Q48	D
Q49	A
Q50	B
Q51	C