

Theorie der Sensorfusion

Udo Frese Lutz Schröder

Stand: 18. Juni 2008

geT_EXed von:

Christoph Hertzberg, Janosch Machowinski, René Wagner und
Pierre Willenbrock

Inhaltsverzeichnis

1	Einführung	1
1.1	Thema der Veranstaltung	1
1.1.1	Beispiele aus der Anwendung	1
1.1.2	Lernziele	3
1.2	Einführungsbeispiel	3
1.2.1	Aussagen	3
2	Wahrscheinlichkeitstheorie (in \mathbb{R})	4
2.1	Grundidee der W-Theorie	4
2.1.1	Intuition	4
2.1.2	Axiome	4
2.1.3	Folgerungen	4
2.1.4	Beispiel	4
2.1.5	Bedingte W'keit	4
2.1.6	Beispiel	5
2.2	Stetige Verteilungen auf \mathbb{R}	5
2.2.1	Wahrscheinlichkeits-Dichte	5
2.2.2	Beispiele	5
2.3	Zufallsvariablen	5
2.4	Erwartungswert und Varianzen	6
2.4.1	Beispiel	7
2.4.2	Erwartungswert einer Zufallsvariablen $X : \Omega \rightarrow \mathbb{R}$	7
2.4.3	Rechenregeln für Erwartungswerte	8
3	Fusion zweier Messwerte: Optimaler Schätzer	10
3.1	Ziel: Herleitung einer Schätzformel $\hat{x}(z_1, \sigma_1, z_2, \sigma_2)$	10
3.1.1	Annahmen	10
3.1.2	Forderungen	10
3.1.3	Herleitung	10
3.1.4	Überprüfung der Aussagen aus 1.2.1	11
3.1.5	Artefakte in der Schätztheorie	11
3.2	Bayes Ansatz	12
3.2.1	Intuition im Kontext von Messungen	12
3.2.2	Beispiel	13
3.2.3	Bayes-Schätzer	13

4 Fortsetzung Wahrscheinlichkeitstheorie	16
4.1 Wiederholung	16
4.2 Varianz	16
4.3 Standardabweichung(Streuung)	17
4.3.1 Eigenschaften	18
4.4 Kovarianz	18
4.4.1 Eigenschaften	18
4.5 Korelation	19
4.5.1 Nachtrag	20
4.6 Konfidenzintervalle	20
4.7 Die Faltung	21
4.8 Unabhängigkeit	21
4.8.1 Beispiel	21
4.8.2 Problem: Zustandsschätzung	22
4.9 Algorithmus „Bayes-Filter“	23
5 (Extended) Kalman Filter	25
5.1 Kalman Filter	25
5.1.1 Annahmen	25
5.1.2 Algorithmus	25
5.2 Extended Kalman Filter	26
5.2.1 Annahmen	26
5.2.2 Algorithmus	27
5.3 Skala der Nichtlinearität	27
5.4 Beispiel Wasserwanne	27
6 Fortsetzung Wahrscheinlichkeitstheorie	29
6.1 Faltungslemma	29
7 Zentraler Grenzwertsatz	30
7.1 Beweisstrategie (Kersting)	30
7.2 Zentraler Grenzwertsatz A	32
7.3 Recall: Taylorreihe	34
7.4 Zentraler Grenzwert Satz B	34
8 Lineare Algebra	38
8.1 Vektor, Vektorraum	38
8.2 Lineare Abbildung	38
8.3 Matrizen	38
8.3.1 Spezielle Matrizen	39
8.3.2 Matrixmultiplikation	39
8.3.3 Matrizen und lineare Abbildungen	39
8.3.4 Weitere Matrixoperationen	40
8.3.5 Rechenregeln	40

8.4	Skalarprodukt und Norm	40
8.4.1	Geometrische Interpretation	40
8.4.2	Eigenschaften	40
8.5	Symmetrische Matrizen	41
8.5.1	Quadratische Funktion	41
8.5.2	Jacobi Matrix	41
9	Wahrscheinlichkeitsrechnung im \mathbb{R}^n	42
9.1	Zufallsvariable	42
9.2	Erwartungswert	42
9.2.1	Eigenschaften	42
9.3	Kovarianzmatrix	43
9.3.1	Eigenschaften	43
9.4	Mehrdimensionale Gaußverteilung	45
10	Mehrdimensionaler (Extended) Kalmanfilter	48
10.1	Lineare Quadratische Ausgleichsrechnung	48
10.1.1	Annahmen	48
10.2	Nichtlineare Ausgleichsrechnung	49
10.2.1	Annahmen	49
10.2.2	Algorithmus	50
10.3	Kalmanfilter	50
10.3.1	Annahmen	52
10.3.2	Algorithmus	52
10.4	Extended-Kalman-Filter	52
10.4.1	Annahmen	52
10.4.2	Algorithmus	52
10.5	Beweis Kalmanfilter (10.3)	53
10.6	Variante des EKF mit Rauschen in U	54
10.6.1	Annahmen	54
10.7	Unscented Kalman Filter	55
10.7.1	Idee	55
10.7.2	Algorithmus	57
11	Transformationen in 3D und Homogene Koordinaten	58
11.1	Projektive Geometrie	58
11.2	Schnitt zweier Geraden	59
11.3	Euklidischen Koordinatentransformationen(EKT) bzw. Starrkörperbewegungen	60
11.4	Koordinatentransformationen, Lage von Körpern im Raum	60
11.5	Rotationen um eine Achse	61
11.6	UKF für beliebige Mannigfaltigkeiten	66
11.6.1	Schätzung einer Orientierung	66

Inhaltsverzeichnis

11.6.2 Darstellung von Mannigfaltigkeiten im UKF	66
11.6.3 Wahrscheinlichkeitsrechnung auf Mannigfaltigkeiten	67
11.6.4 Annahmen	68
11.6.5 Algorithmus	70
12 Starrkörperbewegungen	71
12.1 (Einfacher) Zusammenhang	72
Literaturverzeichnis	a

1 Einführung

1.1 Thema der Veranstaltung

Vorlesung
2006-10-26

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

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

1.1.1 Beispiele aus der Anwendung

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

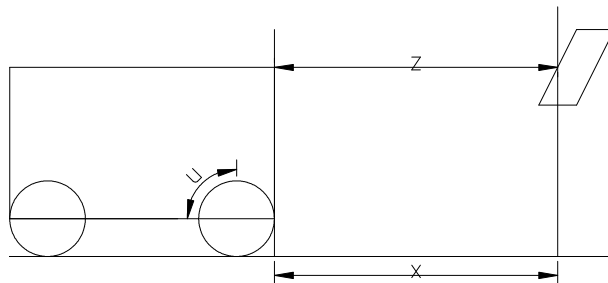


Abbildung 1.1: Beispiel: Schienenfahrzeug

- z (Radar) ist absolut aber ungenau
 - u (Radumdrehung) ist relativ aber 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
 - 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
 - z absolut, relativ ungenau, unvollständig, unzuverlässig

1 Einführung

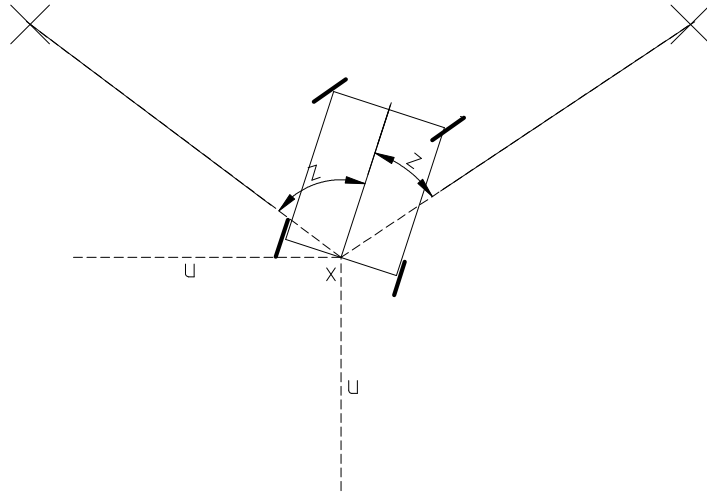


Abbildung 1.2: Beispiel: Fahrzeug in der Ebene

- 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.

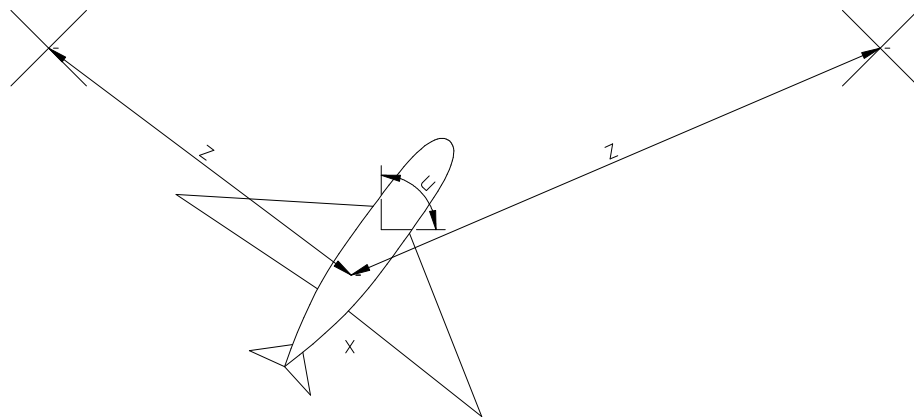


Abbildung 1.3: Beispiel: Flugzeug

1.1.2 Lernziele

- Modellierung fehlerbehafteter Größen
- (Extended, Unscented) Kalman Filter
- Darstellung von Orientierungen im Raum
- Verknüpfung theoretischer Resultate mit intuitiver Vorstellung („Was sagt mir das?“)

1.2 Einführungsbeispiel

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

1.2.1 Aussagen

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

- $\hat{x}(z_1, \sigma_1, z_2, \sigma_2) = \hat{x}(z_2, \sigma_2, z_1, \sigma_1)$
- $\sigma(\hat{x}) \leq \sigma_1, \sigma(\hat{x}) \leq \sigma_2$
- $\sigma_1 \rightarrow \infty \Rightarrow \sigma(\hat{x}) \rightarrow \sigma_2, \hat{x} \rightarrow z_2$
- $\sigma_1 \rightarrow 0 \Rightarrow \sigma(\hat{x}) \rightarrow \sigma_1, \hat{x} \rightarrow z_1$ außer $\sigma_2 \rightarrow 0$
- $\sigma_1 = \sigma_2 \Rightarrow \hat{x} = \frac{z_1 + z_2}{2}$
- je größer σ_1 im Vergleich zu σ_2 , desto näher \hat{x} an z_2
- $\min(z_1, z_2) \leq \hat{x} \leq \max(z_1, z_2)$

2 Wahrscheinlichkeitstheorie (in \mathbb{R})

Vorlesung
2006-10-30

2.1 Grundidee der W-Theorie

Menge Ω von Ergebnissen

Ereignis = Menge $A \subset \Omega$ (i. A. nicht alle Teilmengen möglich)

Wahrscheinlichkeit $P A \in [0, 1]$

2.1.1 Intuition

Wiederholung eines Experiments mit Ergebnis in Ω . Nach „vielen“ Wiederholungen ist $P A$ der Anteil der Fälle mit Ergebnis in A .

2.1.2 Axiome

- $P \emptyset = 0$
- $P \Omega = 1$
- $P(\bigcup_{i \in I} A_i) = \sum_{i \in I} P A_i$, falls I abzählbar und $i \neq j \Rightarrow A_i \cap A_j = \emptyset$

2.1.3 Folgerungen

- $P(A \cup B) = P A + P B - P(A \cap B)$
- $P(\Omega - A) = 1 - P A$

2.1.4 Beispiel

Gezinkter Würfel: $\Omega = \{1, \dots, 6\}$, $P\{1\} = \dots = P\{5\} = \frac{1}{8}$, $P\{6\} = \frac{3}{8}$ Dann ist: $P\{1, 6\} = \frac{4}{8} = \frac{1}{2}$.

2.1.5 Bedingte W'keit

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

$$P(A|B) = \frac{P(A \cap B)}{P B} \quad (2.1)$$

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

2.1.6 Beispiel

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

2.2 Stetige Verteilungen auf \mathbb{R}

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

Wie P angeben? 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.

2.2.1 Wahrscheinlichkeits-Dichte

Definition. $p : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$, integrierbar (z. B. stückweise stetig), mit $\int_{-\infty}^{\infty} p(x)dx = 1$ ist eine Wahrscheinlichkeitsdichte (-funktion).

Damit $P[a, b] = \int_a^b p(x)dx$, insbesondere: $P\{a\} = \int_a^a p(x)dx = 0$.
Allgemeiner: $P A = \int_{x \in A} p(x)dx$ (auch für $A \subset \mathbb{R}^n$)

2.2.2 Beispiele

Stetige Gleichverteilung auf $[a, b]$ $p(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & \text{sonst} \end{cases}$ (2.2)

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!

Die Normalverteilung (Gauß-Verteilung)

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

$$\mathcal{N}(\mu, \sigma^2)(x) = p(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2.3)$$

2.3 Zufallsvariablen

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

$$P(X \in A) = P\{\omega | X(\omega) \in A\} \quad (2.4)$$

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} .

2 Wahrscheinlichkeitstheorie (in \mathbb{R})

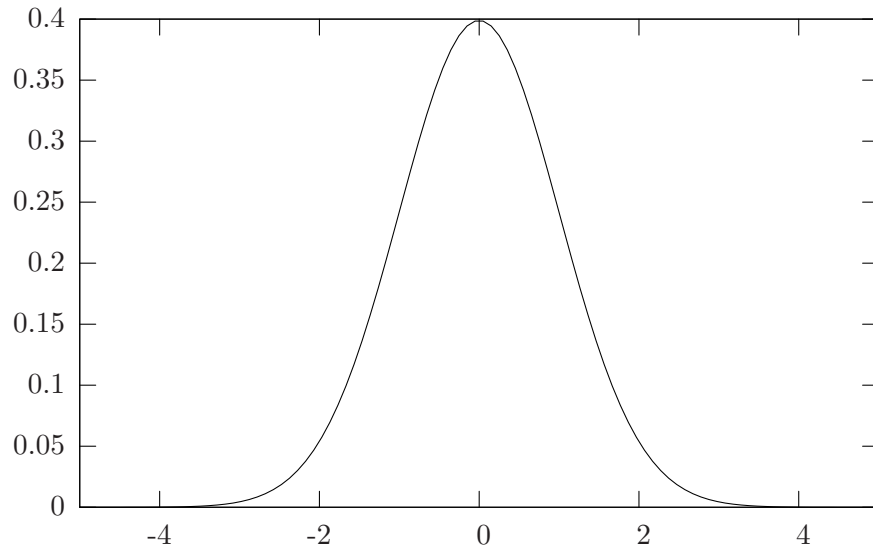


Abbildung 2.1: Gaußverteilung mit $\sigma = 1$ (Standardnormalverteilung)

ZV X, Y heißen unabhängig, wenn

$$\forall x, y \in \mathbb{R} : p(X = x \wedge Y = y) = p(X = x) p(Y = y). \quad (2.5)$$

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

Beweis.

$$P(X \in A \wedge Y \in B) = \int_{x \in A} \int_{y \in B} p(X = x \wedge Y = y) dy dx \quad (2.6)$$

$$= \int_{x \in A} \int_{y \in B} p(X = x) \cdot p(Y = y) dy dx \quad (2.7)$$

$$= \int_{x \in A} p(X = x) \cdot \int_{y \in B} p(Y = y) dy dx \quad (2.8)$$

$$= \int_{x \in A} p(X = x) dx \cdot \int_{y \in B} p(Y = y) dy \quad (2.9)$$

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

2.4 Erwartungswert und Varianzen

Der Erwartungswert $E(X)$ einer ZV X ist def. als: $E X = \int_{-\infty}^{\infty} x \cdot p(X = x) dx$ [Allgemeiner: $\int_{\Omega} x \cdot p(X = x) dx$, typisch $\Omega = \mathbb{R}^n$.]

2.4.1 Beispiel

X gleichverteilt auf $[a, b]$

$$E X = \int_{\mathbb{R}} x \cdot p(X = x) dx = \int_a^b x \cdot \frac{1}{b-a} dx \quad (2.10)$$

$$= \frac{1}{b-a} \int_a^b x dx = \frac{b+a}{2} \quad (2.11)$$

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

$$E X = \mu$$

Beweis. o. E (siehe Übungszettel Aufgabe 2). $\mu = 0$, dann

$$E X = \int_{\mathbb{R}} x \cdot p(X = x) dx \quad (2.12)$$

$$= \int_{-\infty}^0 x \cdot p(X = x) dx + \int_0^{\infty} x \cdot p(X = x) dx \quad (2.13)$$

$$= \int_0^{\infty} -x \cdot p(X = x) dx + \int_0^{\infty} x \cdot p(X = x) dx \quad (2.14)$$

$$= 0 \quad \square$$

2.4.2 Erwartungswert einer Zufallsvariablen $X : \Omega \rightarrow \mathbb{R}$

$$E(X) = \int_{x \in X(\Omega)} x \cdot p(X = x) dx \quad (2.15)$$

*Vorlesung
2006-11-01*

Satz 1 (Maßtheorie). Für ZV X und Funktion f gilt:

$$\int_{x \in X(\Omega)} f(x) \cdot p(X = x) dx = \int_{\omega \in \Omega} f(X(\omega)) \cdot p(\omega) d\omega \quad (2.16)$$

Das heißt, man kann die Berechnung des Erwartungswertes $E(f(X))$ sowohl durch Integration im Wahrscheinlichkeitsraum, also dem Raum der Argumente von X , als auch im Bildraum von X durchführen.

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

2 Wahrscheinlichkeitstheorie (in \mathbb{R})

Berechnung im Bildraum von X .

$$E(X) = \sum_x x \cdot P(X = x) \quad (2.17)$$

$$= 0 \cdot P(X = 0) + 1 \cdot P(X = 1) \quad (2.18)$$

$$= 0 \cdot (1/6 + 1/6 + 1/6) + 1 \cdot (1/6 + 1/6 + 1/6) \quad (2.19)$$

Berechnung im Wahrscheinlichkeitsraum.

$$E(X) = \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega) \quad (2.20)$$

$$= 1 \cdot 1/6 + 0 \cdot 1/6 + 1 \cdot 1/6 + 0 \cdot 1/6 + 1 \cdot 1/6 + 0 \cdot 1/6 \quad (2.21)$$

Im diskreten ist der Unterschied sozusagen eine Anwendung des Distributivgesetzes. Ergebnisse mit gleichem Wert der Zufallsvariablen X werden in (2.17) zuerst zusammengefasst, d. h. ihre Wahrscheinlichkeiten addiert und dann mit dem Funktionswert multipliziert und zur Gesamtsumme addiert. In (2.20) wird jedes Ergebnis einzeln addiert.

2.4.3 Rechenregeln für Erwartungswerte

$$E(X + Y) = \int_{\omega \in \Omega} (X + Y)(\omega) \cdot p(\omega) d\omega = \int_{\omega \in \Omega} ((X(\omega) + Y(\omega)) \cdot p(\omega) d\omega \quad (2.22)$$

$$= \left(\int_{\omega \in \Omega} X(\omega) \cdot p(\omega) d\omega \right) + \left(\int_{\omega \in \Omega} Y(\omega) \cdot p(\omega) d\omega \right) = E(X) + E(Y) \quad (2.23)$$

$$E(\lambda X) = \int_{\omega \in \Omega} (\lambda X)(\omega) \cdot p(\omega) d\omega = \int_{\omega \in \Omega} \lambda X(\omega) \cdot p(\omega) d\omega \quad (2.24)$$

$$= \lambda \int_{\omega \in \Omega} X(\omega) \cdot p(\omega) d\omega = \lambda E(X) \quad (2.25)$$

$$E(\lambda) = \int_{\omega \in \Omega} \lambda \cdot p(\omega) d\omega = \lambda \int_{\omega \in \Omega} p(\omega) d\omega = \lambda \quad (2.26)$$

X, Y **unabhängig**

$$E(XY) = \int_{\omega \in \Omega} (XY)(\omega) \cdot p(\omega) d\omega \stackrel{f: (x,y) \mapsto x \cdot y}{Z:=(X,Y)} \int_{x,y \in \mathbb{R}} x \cdot y \cdot p(X = x \wedge Y = y) dx dy \quad (2.27)$$

$$= \int_{x \in \mathbb{R}} \int_{y \in \mathbb{R}} x \cdot y \cdot p(X = x) \cdot p(Y = y) dy dx \quad (2.28)$$

$$= \int_{x \in \mathbb{R}} x \cdot p(X = x) \cdot \left(\int_{y \in \mathbb{R}} y \cdot p(Y = y) dy \right) dx \quad (2.29)$$

$$= \left(\int_{x \in \mathbb{R}} x \cdot p(X = x) dx \right) \cdot \left(\int_{y \in \mathbb{R}} y \cdot p(Y = y) dy \right) = E(X) \cdot E(Y) \quad (2.30)$$

Gegenbeispiel falls X, Y abhängig

M eine Münze mit $-1/+1$

$$E(M) = 0 \tag{2.31}$$

$$E(MM) = E(M^2) = 1 \neq E(M)^2 \tag{2.32}$$

3 Fusion zweier Messwerte: Optimaler Schätzer

3.1 Ziel: Herleitung einer Schätzformel $\hat{x}(z_1, \sigma_1, z_2, \sigma_2)$

für eine Größe x gegeben zwei Messungen z_1, z_2 mit Genauigkeit σ_1, σ_2

3.1.1 Annahmen

$$\mathbb{E}(Z_1 - x) = 0, \quad \mathbb{E}(Z_2 - x) = 0 \quad (3.1)$$

$$\mathbb{E}((Z_1 - x)^2) = \sigma_1^2 \quad \mathbb{E}((Z_2 - x)^2) = \sigma_2^2 \quad (3.2)$$

Verschiedene Messfehler $Z_1 - x, Z_2 - x$ sind unabhängig.

3.1.2 Forderungen

$$\hat{x} = \alpha z_1 + \beta z_2 \quad (3.3)$$

$$\mathbb{E}(\hat{X} - x) = 0, \quad \mathbb{E}((\hat{X} - x)^2) \rightarrow \min \quad (3.4)$$

3.1.3 Herleitung

$$0 = \mathbb{E}(\hat{X} - x) = \mathbb{E}(\alpha Z_1 + \beta Z_2 - x) = \alpha \mathbb{E}(Z_1) + \beta \mathbb{E}(Z_2) - x \quad (3.5)$$

$$= \alpha x + \beta x - x = (\alpha + \beta - 1)x \quad (3.6)$$

$$\Rightarrow \beta = 1 - \alpha \vee x = 0 \quad (3.7)$$

$$\Rightarrow \beta = 1 - \alpha, \text{ da für alle } x \text{ gültig} \quad (3.8)$$

$$\Rightarrow \hat{x} = \alpha z_1 + (1 - \alpha)z_2 \quad (3.9)$$

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

$$\mathbb{E}((\hat{X} - x)^2) = \mathbb{E}((\alpha Z_1 + (1 - \alpha)Z_2 - x)^2) \quad (3.10)$$

$$= \mathbb{E}((\alpha(Z_1 - x) + (1 - \alpha)(Z_2 - x))^2) \quad (3.11)$$

$$= \alpha^2 \mathbb{E}((Z_1 - x)^2) + 2\alpha(1 - \alpha) \mathbb{E}(Z_1 - x) \mathbb{E}(Z_2 - x) + (1 - \alpha)^2 \mathbb{E}((Z_2 - x)^2) \quad (3.12)$$

$$= \alpha^2 \sigma_1^2 + (1 - \alpha)^2 \sigma_2^2 \quad (3.13)$$

3.1 Ziel: Herleitung einer Schätzformel $\hat{x}(z_1, \sigma_1, z_2, \sigma_2)$

Minimum bestimmen:

$$0 = \frac{d}{d\alpha} [\alpha^2 \sigma_1^2 + (1 - \alpha)^2 \sigma_2^2] = 2\alpha \sigma_1^2 + 2(1 - \alpha) \cdot (-1) \sigma_2^2 \quad (3.14)$$

$$= 2\alpha \sigma_1^2 - 2(1 - \alpha) \sigma_2^2 = 2\alpha(\sigma_1^2 + \sigma_2^2) - 2\sigma_2^2 \quad (3.15)$$

$$\Rightarrow \alpha = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \quad (3.16)$$

$$\Rightarrow \hat{x} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} z_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} z_2 \quad (3.17)$$

$$\sigma_{\hat{x}}^2 = E((\hat{x} - x)^2) = \alpha^2 \sigma_1^2 + (1 - \alpha)^2 \sigma_2^2, \quad \alpha = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \quad (3.18)$$

$$= \frac{\sigma_2^4 \sigma_1^2}{(\sigma_1^2 + \sigma_2^2)^2} + \frac{\sigma_1^4 \sigma_2^2}{(\sigma_1^2 + \sigma_2^2)^2} = \frac{\sigma_1^2 \sigma_2^2 (\sigma_1^2 + \sigma_2^2)}{(\sigma_1^2 + \sigma_2^2)^2} \quad (3.19)$$

$$= \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} = \frac{1}{1/\sigma_1^2 + 1/\sigma_2^2} \quad (3.20)$$

$$\Rightarrow \frac{1}{\sigma_{\hat{x}}^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \quad (3.21)$$

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

3.1.4 Überprüfung der Aussagen aus 1.2.1

- \hat{x} zwischen z_1 und z_2 (ok)
- $\sigma_1 \rightarrow \infty \Rightarrow \hat{x} \rightarrow z_2$ (ok)
- $\sigma_1 \rightarrow 0 \Rightarrow \hat{x} \rightarrow z_1$ (ok)
- $\hat{x}(z_1, \sigma_1, z_2, \sigma_2) = \hat{x}(z_2, \sigma_2, z_1, \sigma_1)$ (ok)
- $\hat{x}(z_1, \sigma, z_2, \sigma) = \frac{z_1 + z_2}{2}$ (ok)
- $\sigma_{\hat{x}}^2 \leq \sigma_1^2, \sigma_{\hat{x}}^2 \leq \sigma_2^2$ (ok)

Vorlesung
2006-11-08

Satz 2. Für zwei erwartungstreue Messungen z_1 , und z_2 mit mittleren quadratischen Fehlern σ_1^2 bzw. σ_2^2 ist der optimale erwartungstreue lineare Schätzer \hat{x} mit mittlerem quadratischen Fehler $\sigma_{\hat{x}}$.

$$\hat{x} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} z_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} z_2, \quad \sigma_{\hat{x}}^2 = \frac{1}{1/\sigma_1^2 + 1/\sigma_2^2}. \quad (3.22)$$

Beweis siehe letzte Vorlesung.

3.1.5 Artefakte in der Schätztheorie

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?

3 Fusion zweier Messwerte: Optimaler Schätzer

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.

3.2 Bayes Ansatz

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. $p(Z_1 = z_1|X = x) = \frac{1}{\sqrt{2\pi\sigma_1}} \cdot e^{-\frac{(z_1-x)^2}{2\sigma_1^2}}$

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 unbekannte) Schätzer sein.

Satz 3 (Bayes). Für Zufallsvariablen X, Y gilt:

$$p(X = x|Y = y) = \frac{p(Y = y|X = x) p(X = x)}{p(Y = y)} \quad (3.23)$$

Beweis.

$$p(X = x|Y = y) \cdot p(Y = y) = p(X = x \wedge Y = y) \quad (3.24)$$

$$= p(Y = y|X = x) \cdot p(X = x) \quad (3.25)$$

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

3.2.1 Intuition im Kontext von Messungen

X ist der uns unbekannte 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 (3.23) 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 das unmögliche ausgeschlossen ist, muss das was übrig bleibt, so unwahrscheinlich es sein mag die Wahrheit sein.“

3.2.2 Beispiel

X = Augensumme zweier Würfel (weiß, rot)

Y = weißer Würfel

$$p(Y = 4|X = 10) = \frac{p(X=10|Y=4)p(Y=4)}{p(X=10)}$$

$$\frac{1}{3} = \frac{\frac{1}{6} \cdot \frac{1}{6}}{\frac{1}{36}}$$

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

3.2.3 Bayes-Schätzer

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 ignorieren äquivalent zu konstante Summanden im Exponenten zu ignorieren.

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

Satz 4. Sind z_1, z_2 zwei Messungen einer a-priori uniform verteilten Größe X mit $\mathcal{N}(0, \sigma_1^2)$ und $\mathcal{N}(0, \sigma_2^2)$ verteiltem Fehler, so ist die Größe a-posteriori ($X|Z_1 = z_1 \wedge Z_2 = z_2$) nach $\mathcal{N}(\hat{x}, \sigma_{\hat{x}}^2)$ verteilt, mit $\hat{x}, \sigma_{\hat{x}}$ wie in **Satz 2**.

$$\hat{x} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} z_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} z_2, \quad \sigma_{\hat{x}}^2 = \frac{1}{1/\sigma_1^2 + 1/\sigma_2^2}. \quad (3.26)$$

Beweis.

$$p(X = x|Z_1 = z_1 \wedge Z_2 = z_2) \stackrel{\text{Bayes}}{=} \frac{p(Z_1 = z_1 \wedge Z_2 = z_2|X = x) p(X = x)}{p(Z_1 = z_1 \wedge Z_2 = z_2)} \quad (3.27)$$

$$\propto p(Z_1 = z_1 \wedge Z_2 = z_2|X = x) p(X = x) \quad (3.28)$$

$$\stackrel{\text{unabh.}}{=} p(Z_1 = z_1|X = x) p(Z_2 = z_2|X = x) p(X = x) \quad (3.29)$$

$$\stackrel{\text{Ann. } p(X=x) \propto 1}{\propto} p(Z_1 = z_1|X = x) p(Z_2 = z_2|X = x) \quad (3.30)$$

$$= \mathcal{N}(0, \sigma_1^2)(z_1 - x) \cdot \mathcal{N}(0, \sigma_2^2)(z_2 - x) \quad (3.31)$$

$$\propto \exp\left(-\frac{(z_1 - x)^2}{2\sigma_1^2}\right) \cdot \exp\left(-\frac{(z_2 - x)^2}{2\sigma_2^2}\right) \quad (3.32)$$

3 Fusion zweier Messwerte: Optimaler Schätzer

$$= \exp\left(-\frac{(z_1 - x)^2}{2\sigma_1^2} - \frac{(z_2 - x)^2}{2\sigma_2^2}\right) \quad (3.33)$$

$$= \exp\left(-\frac{\sigma_2^2(z_1 - x)^2 + \sigma_1^2(z_2 - x)^2}{2\sigma_1^2\sigma_2^2}\right) \quad (3.34)$$

$$= \exp\left(-\frac{\sigma_2^2 z_1^2 - 2\sigma_2^2 z_1 x + \sigma_2^2 x^2 + \sigma_1^2 z_2^2 - 2\sigma_1^2 z_2 x + \sigma_1^2 x^2}{2\sigma_1^2\sigma_2^2}\right) \quad (3.35)$$

$$\propto \exp\left(-\frac{(\sigma_2^2 + \sigma_1^2)x^2 - 2(\sigma_2^2 z_1 + \sigma_1^2 z_2)x}{2\sigma_1^2\sigma_2^2}\right) \quad (3.36)$$

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 $(\sigma_2^2 + \sigma_1^2)$.

$$= \exp\left(-\frac{x^2 - 2\frac{\sigma_2^2 z_1 + \sigma_1^2 z_2}{\sigma_1^2 + \sigma_2^2}x}{2\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}}\right) \quad (3.37)$$

$$\propto \exp\left(-\frac{(x - \frac{\sigma_2^2 z_1 + \sigma_1^2 z_2}{\sigma_2^2 + \sigma_1^2})^2}{2\frac{\sigma_1^2\sigma_2^2}{\sigma_2^2 + \sigma_1^2}}\right) \quad (3.38)$$

$$= \exp\left(-\frac{(x - \hat{x})^2}{2\sigma_{\hat{x}}^2}\right) \quad (3.39)$$

$$\text{mit } \hat{x} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}z_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}z_2, \quad \sigma_{\hat{x}}^2 = \frac{1}{1/\sigma_1^2 + 1/\sigma_2^2} \quad \square$$

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.

Wahre Größe: x

$$E(Z_i) = x$$

$Z_i - x$ sind unabhängig

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

Kasten 3.2.1: Exkurs: Annahmen über Messwerte z_i

$$\hat{x} = \sigma_{\hat{x}}^2 \sum_i^n \frac{z_i}{\sigma_i^2} \qquad \sigma_{\hat{x}}^2 = \frac{1}{\sum_i^n \frac{1}{\sigma_i^2}}$$

Beweis.

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

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

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

Kasten 3.2.2: Exkurs: Der verallgemeinerte optimale erwartungstreue Schätzer

4 Fortsetzung Wahrscheinlichkeitstheorie

Vorlesung
2006-11-15

4.1 Wiederholung

$$E(X) = \int_{\omega \in \Omega} X(\omega) p(\omega) d\omega = \int_{-\infty}^{\infty} x p(X = x) dx \quad (4.1)$$

4.2 Varianz

$$V(X) = E((X - E(X))^2) \quad (4.2)$$

Die Varianz beschreibt die mittlere quadratische Schwankungsbreite von X .

Proposition 4.1. $V(X) = E(X^2) - E(X)^2$

Beweis.

$$V(X) = E((X - E(X))^2) = E(X^2 - 2X E(X) + E(X)^2) \quad (4.3)$$

$$= E(X^2) - E(2X E(X)) + E(E(X)^2) = E(X^2) - 2E(X) E(X) + E(X)^2 \quad (4.4)$$

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

Eigenschaften:

1. $V(\lambda X) = \lambda^2 V(X)$
2. X, Y unabhängig $\Rightarrow V(X + Y) = V(X) + V(Y)$
3. X_1, \dots, X_n unabhängig, $V(X_i) = \sigma^2 \forall i$
 - a) $\Rightarrow V(\sum_{i=1}^n X_i) = n\sigma^2$
 - b) $\Rightarrow V(\frac{1}{n} \sum_{i=1}^n X_i) = \frac{\sigma^2}{n}$

Beweis von 1.

$$V(\lambda X) = E((\lambda X)^2) - E(\lambda X)^2 = E(\lambda^2 X^2) - (\lambda E(X))^2 \quad (4.5)$$

$$= \lambda^2 E(X^2) - \lambda^2 E(X)^2 = \lambda^2 (E(X^2) - E(X)^2) \quad (4.6)$$

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

Beweis von 2.

$$V(X + Y) = E((X + Y)^2) - E(X + Y)^2 \quad (4.7)$$

$$= E(X^2 + 2XY + Y^2) - (E(X) + E(Y))^2 \quad (4.8)$$

$$= E(X^2) + 2E(XY) + E(Y^2) - (E(X)^2 + 2E(X)E(Y) + E(Y)^2) \quad (4.9)$$

$$= E(X^2) + 2E(X)E(Y) + E(Y^2) - E(X)^2 - 2E(X)E(Y) - E(Y)^2 \quad (4.10)$$

$$= E(X^2) + E(Y^2) - E(X)^2 - E(Y)^2 \quad (4.11)$$

$$= E(X^2) - E(X)^2 + E(Y^2) - E(Y)^2 \quad (4.12)$$

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

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

Bei einem Unabhängigkeitsbeweis durch Induktion wäre zu zeigen, dass die resultierende Summe wieder unabhängig ist. Dazu muss man allgemeine Unabhängigkeit der X_i , also $p(X_i = x_i \forall i) = \prod_i p(X_i = x_i)$ voraussetzen. Der folgende Beweis ist eine direkte Rechnung wie oben und benötigt nur paarweise Unabhängigkeit, also $p(X_i = x_i \wedge X_j = x_j) = p(X_i = x_i)p(X_j = x_j) \forall i \neq j$.

$$V\left(\sum_{i=1}^n X_i\right) = E\left(\left(\sum_{i=1}^n X_i\right)^2\right) - E\left(\sum_{i=1}^n X_i\right)^2 \quad (4.13)$$

$$= E\left(\sum_{i=1}^n (X_i^2) + 2 \sum_{i < j} (X_i X_j)\right) - \left(\sum_{i=1}^n E(X_i)\right)^2 \quad (4.14)$$

$$= \sum_{i=1}^n (E(X_i^2)) + 2 \sum_{i < j} (E(X_i) E(X_j)) - \sum_{i=1}^n E(X_i)^2 - 2 \sum_{i < j} (E(X_i) E(X_j)) \quad (4.15)$$

$$= \sum_{i=1}^n (E(X_i^2) - E(X_i)^2) = \sum_{i=1}^n (V(X_i)) = \sum_{i=1}^n (\sigma^2) \quad (4.16)$$

$$= n\sigma^2 \quad (4.17)$$

Es stellt sich heraus, dass nur paarweise Unabhängigkeit nötig ist. \square

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

$$\begin{aligned} V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) &= \frac{1}{n^2} V\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} n\sigma^2 \\ &= \frac{\sigma^2}{n} \quad \square \end{aligned} \quad (4.18)$$

4.3 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.

4.3.1 Eigenschaften

1. $\sigma(\lambda X) = \lambda \sigma(X)$
2. X, Y unabhängig $\Rightarrow \sigma(X + Y) = \sqrt{\sigma(X)^2 + \sigma(Y)^2}$
3. X_i unabhängig, $\sigma(X_i) = \sigma$
 - a) $\Rightarrow \sigma(\sum X_i) = \sqrt{n} \sigma$
 - b) $\Rightarrow \sigma(\frac{1}{n} \sum X_i) = \frac{\sigma}{\sqrt{n}}$

folgen aus den entsprechenden Eigenschaften der Varianz.

4.4 Kovarianz

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

$$\text{Cov}(X, Y) = E((X - E(X))(Y - E(Y))) \quad (4.19)$$

Proposition 4.2. $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$

Beweis. analog Varianz □

4.4.1 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. $\text{Cov}(1, X) = 0$
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. analog zu 4) □

Beweis von 4.

$$\text{Cov}(X + Y, Z) = E((X + Y)Z) - E(X + Y)E(Z) \quad (4.20)$$

$$= E(XZ + YZ) - (E(X) + E(Y))E(Z) \quad (4.21)$$

$$= E(XZ) + E(YZ) - E(X)E(Z) - E(Y)E(Z) \quad (4.22)$$

$$= E(XZ) - E(X)E(Z) + E(YZ) - E(Y)E(Z) \quad (4.23)$$

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

Beweis von 5.

$$\begin{aligned}\text{Cov}(1, Z) &= \text{E}(1Z) - \text{E}(1)\text{E}(Z) = \text{E}(1Z) - 1\text{E}(Z) \\ &= 0\end{aligned}\quad (4.24) \quad \square$$

Beweis von 6.

$$V(X + Y) = \text{E}((X + Y)^2) - (\text{E}(X) + \text{E}(Y))^2 \quad (4.25)$$

$$= \text{E}(X^2 + 2XY + Y^2) - \text{E}(X)^2 - 2\text{E}(X)\text{E}(Y) - \text{E}(Y)^2 \quad (4.26)$$

$$= \text{E}(X^2) + 2\text{E}(XY) + \text{E}(Y^2) - \text{E}(X)^2 - 2\text{E}(X)\text{E}(Y) - \text{E}(Y)^2 \quad (4.27)$$

$$= \text{E}(X^2) - \text{E}(X)^2 + 2\text{E}(XY) - 2\text{E}(X)\text{E}(Y) + \text{E}(Y^2) - \text{E}(Y)^2 \quad (4.28)$$

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

Beweis von 7.

$$X, Y \text{ unabhängig} \Rightarrow V(X) + V(Y) = V(X + Y) = V(X) + 2\text{Cov}(X, Y) + V(Y) \quad (4.29)$$

$$V(X) + V(Y) = V(X) + 2\text{Cov}(X, Y) + V(Y) \Leftrightarrow 0 = \text{Cov}(X, Y) \quad \square$$

4.5 Korelation

Vorlesung
06-11-20

„Normierte Kovarianz“. Falls $\sigma(X), \sigma(Y) > 0$:

$$\text{cor}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)} \quad (4.30)$$

Proposition 4.3.

$$-1 \leq \text{cor}(X, Y) \leq 1 \quad (4.31)$$

[*Zusatz*: $\text{cor}(X, Y) = \pm 1$, gdw. ex. a, b, c mit $a^2 + b^2 > 0$ und $\text{P}(aX + bY + c = 0) = 1$, „ $aX + bY + c = 0$ fast überall“]

Lemma 4.4 (Cauchy-Schwarz-Ungleichung).

$$\text{E}(XY)^2 \leq \text{E}(X^2)\text{E}(Y^2) \quad (4.32)$$

Zusatz: „ $=$ “ gdw. a, b ex. mit $a^2 + b^2 > 0$ und $aX + bY = 0$ f. ü.

Beweis. Mit *Zusatz*:

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

„**nur dann**“: F. a. a, b mit $a^2 + b^2 > 0$ gelte $\text{P}(aX + bY = 0) < 1$. Dann:

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

Wähle:
$$a = -\frac{1}{\sqrt{E(X^2)}}, \quad b = -\frac{1}{\sqrt{E(Y^2)}} \quad (4.33)$$

dann:
$$0 < 1 - 2\frac{E(XY)}{\sqrt{E(X^2)}\sqrt{E(Y^2)}} + 1 \quad (4.34)$$

also:
$$E(XY) < \sqrt{E(X^2)}\sqrt{E(Y^2)} \quad (4.35)$$

analog:
$$E(XY) > -\sqrt{E(X^2)}\sqrt{E(Y^2)} \quad \square$$

Beweis von Prop. 4.3.

$$(\text{Cov}(X, Y))^2 = (E((X - EX)(Y - EY)))^2 \quad (4.36)$$

$$\stackrel{\text{C.S.}}{\leq} (E((X - EX)^2) \cdot E((Y - EY)^2)) \quad (4.37)$$

$$= V(X) V(Y) \quad (4.38)$$

und „=" gdw. a, b ex. mit $a^2 + b^2 > 0$:

$$a(X - EX) + b(Y - EY) = 0 \text{ f. ü.} \quad \square$$

4.5.1 Nachtrag

$$E((x_0 - X)^2) = V X + (x_0 - EX)^2 \quad (4.39)$$

Beweis.

$$E((x_0 - X)^2) = x_0^2 - 2x_0 EX + E(X^2) \quad (4.40)$$

$$= E(X^2) - E(X)^2 + x_0 - 2x_0 EX + E(X)^2 \quad (4.41)$$

$$= V X + (x_0 - EX)^2 \quad \square$$

Vergleiche Übungszettel Aufgabe 3.

4.6 Konfidenzintervalle

Ein Intervall I heißt α -Konfidenzintervall einer ZV X , falls $P(X \in I) \geq 1 - \alpha$.

Satz 5 (Chebyshev-Ungleichung). Für $b > 0$ und ZV X gilt:

$$P(|X - a| \geq b) \leq \frac{E((X - a)^2)}{b^2} \quad (4.42)$$

Korollar 4.5. $[E(X) - k \cdot \sigma(X), E(X) + k \cdot \sigma(X)]$ ist α -Konfidenz-Intervall für $\alpha = \frac{1}{k^2}$.

Beweis. $a = E(X), b = k\sigma(X)$

$$P(|X - EX| \geq k\sigma(X)) \leq \frac{\sigma(X)^2}{k^2\sigma(X)^2} = \frac{1}{k^2} \quad (4.43) \quad \square$$

Bemerkung 4.6. Wenn X normalverteilt, dann sogar:

$$1 - \alpha \approx 68.3\%, 95.4\%, 99.7\% \text{ für } k = 1, 2, 3$$

Beweis Chebyshev.

$$E((X - a)^2) = \int_{\mathbb{R}} (x - a)^2 p(X = x) dx \geq \int_{|x-a| \geq b} (x - a)^2 p(X = x) dx \quad (4.44)$$

$$\begin{aligned} &\geq \int_{|x-a| \geq b} b^2 p(X = x) dx = b^2 \int_{|x-a| \geq b} p(X = x) dx \quad (4.45) \\ &= b^2 P(|X - a| \geq b) \quad \square \end{aligned}$$

4.7 Die Faltung

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

$$(f * g)(z) = \int_{\mathbb{R}} f(x)g(z - x) dx \quad (4.46)$$

Satz 6. Seien X, Y unabh. ZV; dann hat $X + Y$ die W-Dichte $p(X = \cdot) * p(Y = \cdot)$

4.8 Unabhängigkeit

X und Y unabhängig, gdw $p(X = x|Y = y) = p(X = x)$

*Vorlesung
2006-11-22*

$$p(X = x|Y = y) = \frac{p(X = x \wedge Y = y)}{p(Y = y)} = \frac{p(X = x)p(Y = y)}{p(Y = y)} = p(X = x) \quad (4.47)$$

X und Y sind unabhängig gegeben $Z = z$, gdw

$$p(X = x \wedge Y = y|Z = z) = p(X = x|Z = z) p(Y = y|Z = z) \quad (4.48)$$

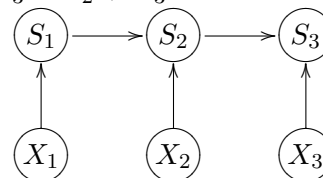
$$\Leftrightarrow p(X = x|Y = y \wedge Z = z) = p(X = x|Z = z) \quad (4.49)$$

4.8.1 Beispiel

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

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

Folgendes Diagramm veranschaulicht die Abhängigkeiten:



Fragen

- a) sind S_1 und S_3 unabhängig?
- b) sind S_1 und S_3 unabhängig gegeben S_2 ?

zu a) S_1 und S_3 abhängig, X_1 abhängig $X_1 + X_2 + X_3$

$$\underbrace{p(S_3 = 17|X_1 = 1)}_{=0} \neq \underbrace{p(S_3 = 17)}_{>0}$$

zu b) $p(S_3 = x|S_1 = a \wedge S_2 = b) = p(X_3 = c - b) = p(S_3 = c|S_2 = b)$

4.8.2 Problem: Zustandsschätzung

Ein System hat einen Zustand X_t der sich in Zeitschritten t ändert. Beobachtet werden Messungen u_t bezüglich des Übergangs X_{t-1} auf X_t und Z_t bezüglich X_t .

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

Beispiel

Siehe **Abbildung 1.1** (Beispiel Schienenfahrzeug).

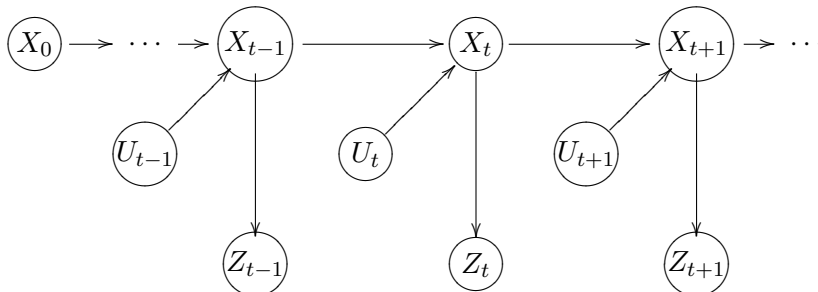
Annahmen

1. X_t ist unabhängig von allen anderen, gegeben X_{t-1}, u_t :

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

2. Die Messung Z_t ist unabhängig von allen anderen, gegebenen X_t :

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



4.9 Algorithmus „Bayes-Filter“

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

Variable: Bel Wahrscheinlichkeitsverteilung Zustand (4.50)

Initialisierung: $Bel(x_0) := p(x_0) \forall x_0$ (4.51)

Dynamik(u_t): $Bel(x_t) := \int p(x_t | u_t, x_{t-1}) Bel(x_{t-1}) dx_{t-1}$ (4.52)

Messung(z_t): $Bel(x_t) := p(z_t | x_t) Bel(x_t) \forall x_t$ (4.53)

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

Satz 7. Bei abwechselndem Aufruf von Dynamik(u_t) und Messung(z_t) für $t' = 1..t$ ist $Bel(x_t) = p(x_t | u_{1..t}, z_{1..t})$.

Vorlesung
2006-11-29

Ursprüngliche Idee war, 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.

Der folgende Satz ist ein Pendant zur Fusion zweier Messwerte, wobei aber die Rolle des ersten Messwerts von der a-priori Verteilung $p(X)$ gespielt wird. Wir nehmen also an, wir hätten schon bestimmte Information über X (z. B. auch in Form früherer Messungen unter denen dann der ganze Wahrscheinlichkeitsraum konditioniert ist). Dieser Satz sagt aus, wie sich die Verteilung ändert, wenn eine Messung Z hinzukommt. Wir führen alles zurück auf die 2 Messungsformel (Satz 4) und formen die Terme in die Predictor-Corrector Form um, die der Kalman Filter verwendet. Außerdem erweitern wir die Formel von direkten Messungen des Zustandes auf die Messung einer linearen Funktion des Zustandes.

Satz 8. Sei X eine $\mathcal{N}(\mu, \sigma^2)$ verteilte Zufallsvariable und $Z = cX + h + \delta, c, h \in \mathbb{R}, \delta$ unabhängig $\mathcal{N}(0, \sigma_\delta^2)$ verteilt. Dann ist für $z \in \mathbb{R}$

$$(X|Z = z) \sim \mathcal{N}(\mu + k(z - (c\mu + h)), \sigma^2 - kc\sigma^2), \quad k = \frac{\sigma^2 c}{\sigma^2 c^2 + \sigma_\delta^2} \quad (4.54)$$

Beweis. Setze $Z_1 = \mu, z_1 = \mu, \sigma_1^2 = \sigma^2, Z_2 = \frac{Z-h}{c}, z_2 = \frac{z-h}{c}, \sigma_2^2 = \frac{\sigma_\delta^2}{c^2}$. Dann folgt nach Satz 4 (2 Messungsformel):

$$p(X|Z_2 = z_2) \quad (4.55)$$

$$= p(X|Z_1 = z_1 \wedge Z_2 = z_2) \quad (4.56)$$

$$= \mathcal{N}\left(\frac{\sigma_2^2 z_1 + \sigma_1^2 z_2}{\sigma_1^2 + \sigma_2^2}, \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right) \quad (4.57)$$

$$= \mathcal{N}\left(\frac{\frac{\sigma_\delta^2}{c^2} \cdot \mu + \sigma^2(z-h)/c}{\sigma^2 + \frac{\sigma_\delta^2}{c^2}}, \frac{\sigma^2 \frac{\sigma_\delta^2}{c^2}}{\sigma^2 + \frac{\sigma_\delta^2}{c^2}}\right) \quad (4.58)$$

$$= \mathcal{N} \left(\frac{\sigma_\delta^2 \mu + \sigma^2 c(z - h)}{\sigma^2 c^2 + \sigma_\delta^2}, \frac{\sigma^2 \sigma_\delta^2}{\sigma^2 c^2 + \sigma_\delta^2} \right) \quad (4.59)$$

$$= \mathcal{N} \left(\mu + \frac{-\mu(\sigma^2 c^2 + \sigma_\delta^2) + \sigma_\delta^2 \mu + \sigma^2 c(z - h)}{\sigma^2 c^2 + \sigma_\delta^2}, \sigma^2 - \frac{\sigma^2(\sigma^2 c^2 + \sigma_\delta^2) - \sigma^2 \sigma_\delta^2}{\sigma^2 c^2 + \sigma_\delta^2} \right) \quad (4.60)$$

$$= \mathcal{N} \left(\mu + \frac{\sigma^2 c(-c\mu + z - h)}{\sigma^2 c^2 + \sigma_\delta^2}, \sigma^2 - \frac{\sigma^4 c^2}{\sigma^2 c^2 + \sigma_\delta^2} \right) \quad (4.61)$$

$$= \mathcal{N} (\mu + k(z - (c\mu + h)), \sigma^2 - k\sigma^2 c); \quad k = \frac{\sigma^2 c}{\sigma^2 c^2 + \sigma_\delta^2} \quad \square$$

Der Faktor k heißt *Kalman-Faktor* oder engl. *Kalman Gain*, die ganze Formel heißt *predictor-corrector*. Der Term $c\mu + h$ ist der Wert, den man a-priori erwartet zu messen. $z - (c\mu + h)$ ist also die Differenz zwischen dem was passierte und dem was man erwartet. Der Kalman Faktor bestimmt, wie stark diese Diskrepanz zu einer Änderung der Schätzung führt.

Man kann diesen Vorgang als einen Regelkreis für den Schätzwert sehen, der so nachgeführt wird, dass der Unterschied zwischen erwarteter Messung und wirklicher Messung gering wird.

5 (Extended) Kalman Filter

5.1 Kalman Filter

5.1.1 Annahmen

1. X_0 ist $\mathcal{N}(\mu_0, \sigma_0^2)$ verteilt
2. $X_t = aX_{t-1} + b_tU_t + g_t + \varepsilon_t; a_t, b_t, g_t \in \mathbb{R}, \varepsilon_t \sim \mathcal{N}(0, \sigma_{\varepsilon t}^2)$
3. $Z_t = c_tX_t + h_t + \delta_t; c, h \in \mathbb{R}, \delta_t \sim \mathcal{N}(0, \sigma_{\delta t}^2)$
4. 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 gerade gemessen, aber natürlich erst zur Laufzeit.

Vergleiche [2][Tabelle 3.1].

5.1.2 Algorithmus

Variablen: μ_t, σ_t^2 (5.1)

Initialisierung: $\mu_0 = E(X_0), \sigma_0^2 = V(X_0)$ (5.2)

Dynamik(U_t): $\bar{\mu}_t = a_t\mu_{t-1} + b_tu_t + g_t$ (5.3)

$$\bar{\sigma}_t^2 = a_t^2\sigma_{t-1}^2 + \sigma_{\varepsilon t}^2$$
 (5.4)

Messung(z_t): $k_t = \frac{\bar{\sigma}_t^2 c_t}{\bar{\sigma}_t^2 c_t^2 + \sigma_{\delta t}^2}$ (5.5)

$$\mu_t = \bar{\mu}_t + k(z_t - (c_t\bar{\mu}_t + h_t))$$
 (5.6)

$$\sigma_t^2 = \bar{\sigma}_t^2 - k_t c_t \bar{\sigma}_t^2$$
 (5.7)

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

5 (Extended) Kalman Filter

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.

Satz 9. Nach Ausführung der Kalman Filter Formeln ist

$$X_t | u_{1..t}, z_{1..t-1} \sim \mathcal{N}(\bar{\mu}_t, \bar{\sigma}_t^2), \quad \text{falls } t > 0 \quad (5.8)$$

$$X_t | u_{1..t}, z_{1..t} \sim \mathcal{N}(\mu_t, \sigma_t^2) \quad (5.9)$$

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 **Satz 8** 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.

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

$$X_{t-1} | u_{1..t-1}, z_{1..t-1} \sim \mathcal{N}(\mu_{t-1}, \sigma_{t-1}^2) \quad (5.10)$$

$$aX_{t-1} | u_{1..t-1}, z_{1..t-1} \sim \mathcal{N}(a\mu_{t-1}, a^2\sigma_{t-1}^2) \quad (5.11)$$

$$b_t u_t + g_t | u_{1..t-1}, z_{1..t-1} \text{ ist konstant} \quad (5.12)$$

$$\varepsilon_t \sim \mathcal{N}(0, \sigma_{\varepsilon t}^2) \quad (5.13)$$

$$X_t | u_{1..t}, z_{1..t-1} \sim \mathcal{N}(a\mu_{t-1} + b_t u_t + g_t, a^2\sigma_{t-1}^2 + \sigma_{\varepsilon t}^2) = \mathcal{N}(\bar{\mu}_t, \bar{\sigma}_t^2) \quad (5.14)$$

Wende **Satz 8** an auf $X | u_{1..t}, z_{1..t-1}$ mit $Z = Z_t$ an.

$$X_t | u_{1..t}, z_{1..t} = X_t | u_{1..t}, z_{1..t-1}, z_t \quad (5.15)$$

$$\sim \mathcal{N}(\bar{\mu}_t + k_t(z_t - (c_t \bar{\mu}_t + h_t)), \bar{\sigma}_t^2 - k_t c_t \bar{\sigma}_t^2) = \mathcal{N}(\mu_t, \sigma_t^2) \quad (5.16)$$

$$k_t = \frac{\bar{\sigma}_t^2 c_t}{\bar{\sigma}_t^2 c_t^2 + \sigma_{\delta t}^2} \quad \square$$

5.2 Extended Kalman Filter

5.2.1 Annahmen

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

5.2.2 Algorithmus

Der Extended Kalman Filter [2][Tabelle 3.3] wendet die Gleichungen des Kalmanfilter an, indem er in jedem Schritt die nichtlinearen Funktionen g und h durch Linearisierungen, also Tangenten am aktuellen Erwartungswert ersetzt.

$$g(x, u) \approx g(\mu_{t-1}, u) + (x - \mu_{t-1})g'(\mu_{t-1}) \quad (5.17)$$

$$h(x) \approx h(\bar{\mu}_t) + (x - \bar{\mu}_t)h'(\bar{\mu}_t), \quad (5.18)$$

Initialisierung: μ_0, σ_0^2 nach 1.

Dynamik: (u_t)

$$\bar{\mu}_t = g_t(\mu_{t-1}, u_t) \quad (5.19)$$

$$\bar{\sigma}_t^2 = a^2 \cdot \sigma_{t-1}^2 + \sigma_{\varepsilon t}^2, \quad a = g'_t(\mu_{t-1}, u_t) \quad (5.20)$$

Messung (z_t)

$$k_t = \frac{\bar{\sigma}_t^2 c_t}{\bar{\sigma}_t^2 c_t^2 + \sigma_{\delta t}^2}, \quad c_t = h'(\bar{\mu}_t) \quad (5.21)$$

$$\mu_t = \bar{\mu}_t + k_t(z_t - h(\bar{\mu}_t)) \quad (5.22)$$

$$\sigma_t^2 = \bar{\sigma}_t^2 - k_t c_t \bar{\sigma}_t^2 \quad (5.23)$$

5.3 Skala der Nichtlinearität

Da der EKF die nichtlineare Funktion durch eine Tangente ersetzt, funktioniert er nur so lange wie die Funktion in der Größenordnung der Zustandsunsicherheit hinreichend linear ist ([2, 3.3-3.5, auf Webseite verfügbar]). Deshalb ist es wichtig eine Intuition dafür zu gewinnen, welche Funktionen auf welcher Größenordnung näherungsweise linear sind.

$\sin x, \cos x$ ungefähr bis $\pi/2$, eher weniger

e^x ungefähr bis 1 (auch eher weniger)

$\sqrt{1-x^2}$ ungefähr bis 1/2

$|\left(\frac{x}{y}\right)| = \sqrt{x^2 + y^2}$ ungefähr bis 1/2

5.4 Beispiel Wasserwanne

Siehe (Abbildung fehlt noch) Ein Abfluss mit Durchflussmesser.

Wie viel Wasser ist in der Wanne?

Abflussmenge: Gesamtumdrehung

Füllhöhe: Wie schnell

5 (Extended) Kalman Filter

x Füllhöhe

u Wassermenge

z Abflussdurchsatz

g Wieviel u den Spiegel fallen lässt

h Füllhöhe \rightarrow Abflussdurchsatz

σ_ε^2 $\frac{\text{Eimer}}{\text{Wannenfläche}} \cdot \text{Häufigkeit}$

σ_δ^2 Messungenauigkeit des Durchflussmessers

6 Fortsetzung Wahrscheinlichkeitstheorie

Vorlesung
2006-12-06

6.1 Faltungslemma

Seien X, Y unabhängige Zufallsvariablen, dann hat $X + Y$ folgende Dichte:

$$p(X + Y = \cdot) = p(X = \cdot) * p(Y = \cdot) \quad (6.1)$$

Allgemein gilt:

$$(f * g)(z) = \int_{x=-\infty}^{\infty} f(x)g(z - x) \quad (6.2)$$

Beweis.

$$P(X + Y \in [a, b]) \quad (6.3)$$

$$\stackrel{\text{s.o.}}{=} \iint_{x+y \in [a, b]} p(X = x \wedge Y = y) dy dx \quad (6.4)$$

$$\stackrel{X, Y \text{ unabh.}}{=} \iint_{x+y \in [a, b]} p(X = x) p(Y = y) dy dx \quad (6.5)$$

$$\stackrel{\substack{z=x+y \\ y=z-x}}{=} \int_{z=a}^b \int_{x=-\infty}^{\infty} p(X = x) p(Y = z - x) dx dz \quad (6.6)$$

□

7 Zentraler Grenzwertsatz

„Eine lange Folge mathematischer Experimente ist annähernd normalverteilt.“

Zur Formalisierung dieser umgangssprachlichen Aussage machen wir folgende Überlegung: Für eine reellwertige Zufallsvariablen X sei die dazugehörige Verteilungsfunktion F definiert als

$$F(x) = P(X \leq x) = \int_{-\infty}^x p(X = x') dx'. \quad (7.1)$$

Insbesondere sei Φ die Verteilungsfunktion der Gaußverteilung φ . Die Verteilungsfunktion bestimmt die Verteilung von X , da

$$P(X \in [a, b]) = F(b) - F(a). \quad (7.2)$$

Für F, G beschränkt setze ¹

$$d(F, G) = \sup_x |F(x) - G(x)| \quad (7.3)$$

Wir formalisieren jetzt „annähernd normalverteilt“ in dem Sinne, dass die $d(\cdot, \Phi)$ Distanz zur Verteilungsfunktion der Normalverteilung gegen 0 geht.

Satz 10 (ZGS). Sind $X_i, i \geq 1$, identisch verteilte unabhängige ZV mit Varianz $0 < \sigma^2 < \infty$ und $E(X_i) = \mu$, so gilt für $S_n := \sum_{i=1}^n X_i, S_n^* = \frac{S_n - n\mu}{\sigma\sqrt{n}}$:

$$d(S_n^*, \Phi) \xrightarrow{n \rightarrow \infty} 0 \quad (7.4)$$

7.1 Beweisstrategie (Kersting)

Reduktion auf einen eingeschränkten Fall

- Erwartungswert 0, Varianz 1 wie Std. Gaußverteilung
- Nur endlich viele verschiedene Werte.
- Gemeinsame obere Schranke aller Werte aller Zufallsvariablen.

¹ $\sup A$ ist die kleinste obere Schranke von $A \subseteq \mathbb{R}$. Dabei ist x obere Schranke von A , falls $\forall y \in A. y \leq x$

Dieser Satz ist eine spezialisierte Version des zentralen Grenzwertsatzes. Die Einschränkungen $E(X_i) = 0, V(X_i) = 1$ sind eher technische Vereinfachungen. Zentral ist die Einschränkung auf endlich viele Werte mit gemeinsamer Schranke. Der allgemeine Satz wird dann durch einen Grenzwertübergang aus diesem Spezialfall hergeleitet, indem der Wertebereich der betrachteten Zufallsvariablen beschränkt und diskretisiert wird, so dass ZGS B anwendbar wird. Dann wird ein Grenzwertübergang durchgeführt in dem die Diskretisierung immer feiner und die Schranken immer laxer werden.

Satz 11 (ZGS B). *Seien X_i ($i \geq 1$) unabhängige ZV mit je endlich vielen Werten, $E(X_i) = 0, V(X_i) = 1, |X_i| \leq B$ für alle $i, S_n := \sum_{i=1}^n X_i, S_n^* = \frac{S_n}{\sqrt{n}}$.*

Dann gilt

$$d(S_n^*, \Phi) \xrightarrow{n \rightarrow \infty} 0 \tag{7.5}$$

Sowohl die Reduktion von ZGS A auf ZGS B als auch der Beweis von ZGS B basieren auf einer Vielzahl von Abschätzungen, wie stark sich der $d(\cdot, \Phi)$ Abstand zur Gaußverteilung erhöhen kann, wenn bestimmte Veränderungen an einer Verteilung vorgenommen werden. Ausgehend von einer von einer Gaußverteilung werden die Abschätzungen mit mehreren Zwischenschritten zusammengeführt, so dass sich am Ende die betrachtete Verteilung von S_n^* ergibt. Alle diese Schranken gehen für $n \rightarrow \infty$ hinreichend schnell gegen 0 womit sich dann der ZGS A ergibt.

Reduktion auf und Beweis von ZGS B benötigen ein Hilfslemma, das es erlaubt aus einer Schranke für $E(Y^2)$ abzuschätzen wie sehr sich der $d(\cdot, \Phi)$ Abstand zur Gaußverteilung erhöht, wenn man Y zu einer Zufallsvariablen addiert. Damit werden $E(\cdot^2)$ Abschätzungen für die Beweisführung nutzbar.

Lemma 7.1. *Sei $E(Y^2) \leq \eta$. Dann*

$$d(X + Y, \Phi) \leq d(X, \Phi) + 2\eta^{\frac{1}{3}} \tag{7.6}$$

Beweis (Lemma 7.1). $\Phi'(x) = \varphi(x)$ wird ² maximal bei $x = 0$, dort $\Phi'(0) = \varphi(0) = \frac{1}{\sqrt{2\pi}} \leq 1$

Satz 12 (MWS). *Sei f stetig differenzierbar auf $[x, x + \delta]$. Dann gilt*

$$f(x + \delta) = f(x) + \delta f'(x + \vartheta\delta) \text{ für ein } \vartheta \in [0, 1] \tag{7.7}$$

Also gilt nach MWS für $\delta > 0$

$$\Phi(x - \delta) + \delta \geq \Phi(x) \geq \Phi(x + \delta) - \delta \tag{7.8}$$

Wenn $X + Y > x$ ist, dann ist für alle δ : $X > x + \delta$ oder $Y > -\delta$, denn sonst wäre $X + Y \leq x + \delta - \delta = x$. Wir haben also für die Wahrscheinlichkeiten

$$P(X + Y \leq x) \leq P(X \leq x + \delta) + P(Y < -\delta) \tag{7.9}$$

² $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \Phi(x) = \int_{-\infty}^x \varphi(z) dz$

7 Zentraler Grenzwertsatz

Analog, wenn $X > x - \delta$ ist, dann muss $X + Y \leq x$ oder $Y > \delta$ sein, weil sonst $X = (X + Y) - Y > x - \delta$ wäre. Es folgt für die Wahrscheinlichkeiten

$$P(X \leq x - \delta) \leq P(X + Y \leq x) + P(Y > \delta) \quad (7.10)$$

Nun wird die linke Hälfte von (7.8) zu (7.9) sowie die rechte Hälfte von (7.8) zu (7.10) addiert und umsortiert. Daraus ergibt sich eine untere und obere Schranke für $P(X + Y \leq x) - \Phi(x)$ also den gesuchten Abstand zwischen der Verteilungsfunktion von $X + Y$ und der Gaußverteilung (an der Stelle x).

$$\begin{aligned} P(X \leq x - \delta) - \Phi(x - \delta) - \delta - P(Y > \delta) \\ \leq P(X + Y \leq x) - \Phi(x) \end{aligned} \quad (7.11)$$

$$\leq P(X \leq x + \delta) - \Phi(x + \delta) + \delta + P(Y < -\delta) \quad (7.12)$$

Man führt $P(Y > \delta)$ und $P(Y < -\delta)$ als $\leq P(|Y| > \delta)$, sowie $|P(X \leq x + \delta) - \Phi(x + \delta)|$ und $|P(X \leq x - \delta) - \Phi(x - \delta)|$ als $\leq d(X, \Phi)$ zusammen.

$$|(P(X + Y \leq x) - \Phi(x)) - d(X, \Phi)| \leq \delta + P(|Y| > \delta) \quad (7.13)$$

$$d(X + Y, \Phi) \leq d(X, \Phi) + \delta + P(|Y| > \delta) \quad (7.14)$$

Setze $\delta := \eta^{\frac{1}{3}}$. Dann $\eta \geq E(Y^2) \geq \delta^2 P(|Y| > \delta) \Rightarrow \eta^{\frac{1}{3}} \geq P(|Y| > \delta)$. Mit (7.14) folgt Behauptung. \square

7.2 Zentraler Grenzwertsatz A

Beweis (ZGS aus ZGS B). Diskretisiere und beschränke die X_i als $Y_{m,i} := f_m(X_i)$ mit Diskretisierungsfunktion f_m

$$f_m(x) = \begin{cases} \frac{k}{m}, & \text{wenn } x \in [\frac{k}{m}, \frac{k+1}{m}), |k| \leq m \\ 0, & \text{sonst} \end{cases} \quad (7.15)$$

Dabei gibt m die Feinheit der Diskretisierung an. Mit wachsendem m wird sie immer feiner und außerdem die Beschränkung immer laxer, so dass sich die diskretisierten Variablen immer mehr den Originalen annähern.

$$\mu_m := E(Y_{m,i}), \quad \sigma_m = \sigma(Y_{m,i}) \quad (7.16)$$

Formal geht mit ein paar Überlegungen der mittlere Quadratische Abstand zwischen X_i und $Y_{m,i}$ gegen 0 und damit konvergieren Erwartungswert und Varianz der Diskretisierung gegen die der Originale (für $m \rightarrow \infty$).

$$E((Y_{m,i} - X_i)^2) \xrightarrow{m \rightarrow \infty} 0 \quad (7.17)$$

$$\Rightarrow \mu_m \xrightarrow{m \rightarrow \infty} \mu \quad (7.18)$$

$$\Rightarrow \sigma_m \xrightarrow{m \rightarrow \infty} \sigma \quad (7.19)$$

Setze $X_{m,i} := \frac{Y_{m,i} - \mu_m}{\sigma_m}$ die auf Erwartungswert 0 und Varianz 1 normierten diskretisierten Zufallsvariablen. Dann erfüllen die $X_{m,i}$ ($i \geq 1$) die Voraussetzungen von ZGS B.

Für $S_{m,n}^* = \frac{1}{\sqrt{n}} \sum_{i=1}^m X_{m,i}$ gilt also $d(S_{m,n}^*, \Phi) \xrightarrow{n \rightarrow \infty} 0$. In Worten: In jeder Diskretisierung konvergiert die normierte Summe der Zufallsvariablen gegen eine Gaußverteilung. Im Folgenden argumentieren wir, dass bei immer feinerer Diskretisierung der Unterschied zur nicht diskretisierten Verteilung immer kleiner wird, so dass auch die Folge nicht diskretisierter Verteilungen gegen die Gaußverteilung geht.

Wir haben schon in (7.17) gesehen, dass der Unterschied im quadratischen Mittel gegen 0 konvergiert, sowie in (7.18), (7.19), dass Erwartungswert und Varianz entsprechend konvergieren. Diese Aussage wandeln wir jetzt über Lemma 7.1 in eine $d(-, \Phi)$ Schranke um.

Der Unterschied zwischen S_n^* und $S_{m,n}^*$ lässt sich in zwei Teile spalten, bei denen der erste vom Unterschied in der Varianz $\frac{\sigma_m}{\sigma}$ und der zweite vom Unterschied im Mittelwert $\mu - \mu_m$ abhängt.

$$S_n^* - S_{m,n}^* = \underbrace{\left(\frac{\sigma_m}{\sigma} - 1\right) S_{m,n}^*}_{A_{m,n}} + \underbrace{\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n [(X_i - \mu) - (Y_{m,i} - \mu_m)]}_{B_{m,n}} \quad (7.20)$$

Der erste geht im quadratischen Mittel gegen 0.

$$\mathbb{E}(A_{m,n}^2) = \left(\frac{\sigma_m}{\sigma} - 1\right)^2 \underbrace{\mathbb{E}(S_{m,n}^{*2})}_{=1} \xrightarrow[\text{glm. } n]{m \rightarrow \infty} 0 \quad (7.21)$$

Dabei ist die Konvergenz gleichmäßig bzgl. n weil der Term unabhängig von n ist. Diese Tatsache brauchen wir später, wenn m und n gleichzeitig gegen ∞ gehen. Für den zweiten Term gilt das gleiche. Da alle X_i gleich verteilt sind, wird die Summe zu einem n ., das sich mit dem $\frac{1}{\sqrt{n}}$ im Vorfaktor wegekürzt. Damit gilt für $i = 1$, z. B.:

$$\mathbb{E}(B_{m,n}^2) = \frac{1}{\sigma^2} \mathbb{E}((X_i - Y_{m,i})^2) \xrightarrow[\text{glm. } n]{m \rightarrow \infty} 0 \quad (7.22)$$

Um die quadratische Differenz $\mathbb{E}((S_n^* - S_{m,n}^*)^2) = \mathbb{E}((A_{m,n} + B_{m,n})^2)$ abzuschätzen, fehlt auf den ersten Blick eine Schranke für den gemischten Summanden $2A_{m,n}B_{m,n}$. Wir bemerken allerdings, dass generell

$$(a + b)^2 \leq (a + b)^2 + (a - b)^2 = a^2 + 2ab + b^2 + a^2 - 2ab + b^2 = 2a^2 + 2b^2 \quad (7.23)$$

ist. Daraus folgt, dass der Unterschied im quadratischen Mittel gegen 0 geht, denn

$$\mathbb{E}((S_n^* - S_{m,n}^*)^2) = \mathbb{E}((A_{m,n} + B_{m,n})^2) \leq 2\mathbb{E}(A_{m,n}^2) + 2\mathbb{E}(B_{m,n}^2) \xrightarrow[\text{glm. } n]{m \rightarrow \infty} 0 \quad (7.24)$$

Jetzt wandeln wir die Schranke im quadratischen Mittel via Lemma 7.1 in eine $d(-, \Phi)$ Schranke um:

$$d(S_n^*, \Phi) = d(S_{m,n}^* + (S_n^* - S_{m,n}^*), \Phi) \leq d(S_{m,n}^*, \Phi) + 2\mathbb{E}((S_n^* - S_{m,n}^*)^2)^{1/3} \quad (7.25)$$

7 Zentraler Grenzwertsatz

Wenn m und n gleichzeitig gegen ∞ gehen, geht der erste Summand gegen 0 wegen ZGS A und der zweite gegen 0 wegen (7.24). Dazu wählt man ein m , das den rechten Term hinreichend klein macht und danach zu diesem m ein n , das den linken hinreichend klein macht. Es folgt das Endergebnis

$$d(S_n^*, \Phi) \xrightarrow{m, n \rightarrow \infty} 0 \quad (7.26)$$

□

7.3 Recall: Taylorreihe

Vorlesung
2006-12-13

Für f hinreichend oft stetig differenzierbar

$$\begin{aligned} f(x+z) &= f(x) + z \cdot f'(x) + \frac{1}{2!} z^2 \cdot f''(x) + \dots + \frac{1}{n!} z^n \cdot f^{(n)}(x) \\ &+ \frac{1}{(n+1)!} z^{n+1} \cdot f^{(n+1)}(x + \vartheta z) \quad \vartheta \in [0, 1] \end{aligned} \quad (7.27)$$

Die Verteilungsfunktion einer ZV X ergibt sich als $x \mapsto P(X \leq x)$ durch Integration der Dichtefunktion $p(X = x)$ von $-\infty$ bis x . Speziell für die Standard-Gaußverteilung $\mathcal{N}(0, 1)$ ist

$$\Phi(x) = \int_{-\infty}^x \varphi(x) dx \quad (7.28)$$

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} \quad (7.29)$$

definiert. Beim Ableiten von φ entstehen nach Ketten- und Produktregel Polynome mal $e^{-\frac{x^2}{2}}$. Da der $e^{-\frac{x^2}{2}}$ Ausdruck schneller gegen 0 strebt, als jeder Polynomfaktor davor ergibt sich ohne formalen Beweis

$$\Phi^{(n)} \text{ beschränkt für alle } n. \quad (7.30)$$

7.4 Zentraler Grenzwert Satz B

Seien $X_i, i \geq 1$ unabhängige ZV mit je endlich vielen Werten, $E(X_i) = 0, V(X_i) = 1$

$$|X_i| \leq B \text{ für alle } i \quad (7.31)$$

Dann gilt

$$d(S_n^*, \Phi) \xrightarrow{n \rightarrow \infty} 0 \quad (7.32)$$

Beweis. Sei $Y_0 \mathcal{N}(0, 1)$ verteilt, unabhängig von den X_i . Sei $N \in \mathbb{N}$. Für $n \geq N$ setze

$$Z_n = \sqrt{\frac{N}{n}} Y_0 + \sqrt{\frac{1}{n}} (S_n - S_N) \quad (7.33)$$

Die Folge der Z_n ist so konstruiert, dass sie für $N = n$ mit $Z_N = Y_0$ also einer Gaußverteilung startet und für $n \rightarrow \infty$ gegen $\sqrt{\frac{1}{n}}S_n = S_n^*$ konvergiert. Außerdem ergibt sich jedes Element Z_{n+1} der Folge aus dem Vorgänger Z_n durch Skalieren und Addieren des entsprechenden X_{n+1} .

$$Z_{n+1} = \sqrt{\frac{n}{n+1}}Z_n + \sqrt{\frac{1}{n+1}}X_{n+1} \quad (7.34)$$

$$Z_n, X_{n+1} \text{ unabhängig} \quad (7.35)$$

Sei F_n die Verteilungsfunktion für Z_n , also $F_n(x) = P(Z_n \leq x)$. Wir wollen abschätzen, wie viel „ungaußiger“, im Sinne von $d(F_{n+1}, \Phi)$, die Z_{n+1} mit jedem Schritt werden können. Dazu leiten wir als erstes eine Formel für die Verteilungsfunktion F_{n+1} in Abhängigkeit von F_n her.

$$F_{n+1}(x) = \sum_{i=1}^k P(X_{n+1} = x_i \wedge Z_{n+1} \leq x) \quad (7.36)$$

$$= \sum_{i=1}^k P(X_{n+1} = x_i \wedge \sqrt{\frac{n}{n+1}}Z_n \leq x - \sqrt{\frac{1}{n+1}}x_i) \quad (7.37)$$

$$\stackrel{\text{unabh.}}{=} \sum_{i=1}^k p_i \cdot P\left(\sqrt{\frac{n}{n+1}}Z_n \leq x - \sqrt{\frac{1}{n+1}}x_i\right) \quad (7.38)$$

$$= \sum_{i=1}^k p_i \cdot F_n\left(\underbrace{\sqrt{\frac{n+1}{n}}x}_{\alpha} - \underbrace{\sqrt{\frac{1}{n}}x_i}_{\beta}\right) \quad (7.39)$$

$$= E(F_n(\alpha x - \beta X_{n+1})) \quad (7.40)$$

Nun verwenden wir diese Formel um nachzurechnen, dass der Abstand von Φ sich beim Übergang von F_n auf F_{n+1} höchstens proportional zu $n^{-3/2}$ erhöht.

$$|F_{n+1}(x) - \Phi(x)| \quad (7.41)$$

$$= |E(F_n(\alpha x - \beta X_{n+1})) - \Phi(\alpha x - \beta X_{n+1}) + \Phi(\alpha x - \beta X_{n+1}) - \Phi(x)| \quad (7.42)$$

$$\leq |E(F_n(\alpha x - \beta X_{n+1})) - \Phi(\alpha x - \beta X_{n+1})| + |E(\Phi(\alpha x - \beta X_{n+1}) - \Phi(x))| \quad (7.43)$$

$$\leq d(F_n, \Phi) + \sup_x |E(\Phi(\alpha x - \beta X_{n+1}) - \Phi(x))| \quad (7.44)$$

also

$$d(F_{n+1}, \Phi) \leq d(F_n, \Phi) + \sup_x |E(\Phi(\alpha x - \beta X_{n+1}) - \Phi(x))| \quad (7.45)$$

Wir möchten jetzt den letzten Summand abschätzen. Dazu entwickeln wir alle auftretenden Terme nach Taylor. Wir wollen darauf hinaus, dass der Unterschied höchstens proportional zu $n^{-3/2}$, also $O(n^{-3/2})$ ist. Daher fassen wir alle Terme kleinerer Ordnung sobald sie auftreten mit dem $O(\dots)$ Symbol zusammen.

7 Zentraler Grenzwertsatz

Wir starten mit der Taylorentwicklung von Φ an der Stelle αx .

$$\Phi(\alpha x - \beta X_{n+1}) = \Phi(\alpha x) - \beta X_{n+1} \Phi'(\alpha x) + \frac{(\beta X_{n+1})^2}{2!} \Phi''(\alpha x) - \underbrace{\frac{(\beta X_{n+1})^3}{6} \Phi'''(\dots)}_{O(n^{-\frac{3}{2}})} \quad (7.46)$$

Analog entwickeln wir α nach Taylor an der Stelle 1 (Formel aus der Formelsammlung entnommen).

$$\alpha = \sqrt{\frac{n+1}{n}} = 1 + \frac{1}{2n} + O(n^{-2}) \quad (7.47)$$

Und setzen ein in die Taylorentwicklung von $\Phi(\alpha x)$ an der Stelle $\Phi(x)$.

$$\Phi(\alpha x) = \Phi(x) + \left(1 + \frac{1}{2n} + O(n^{-2})\right)x\Phi'(x) + \left(1 + \frac{1}{2n} + O(n^{-2})\right)^2 \frac{x^2}{2} \Phi''(\dots) \quad (7.48)$$

Damit

$$\Phi(\alpha x) = \Phi(x) + \frac{x}{2n} \Phi'(x) + O(n^{-2}) \quad (7.49)$$

analog

$$\Phi''(\alpha x) = \Phi''(x) + \frac{x}{2n} \Phi'''(x) + O(n^{-2}) \quad (7.50)$$

Nun

$$E(\Phi(\alpha x - \beta X_{n+1})) = \sum_{i=1}^k \Phi(\alpha x - \beta x_i) p_i \quad (7.51)$$

und somit wegen

$$E(X_{n+1}) = \sum_{i=1}^k x_i \cdot p_i = 0, V(X_{n+1}) = \sum_{i=1}^k x_i^2 \cdot p_i = 1 \quad (7.52)$$

$$E(\Phi(\alpha x - \beta X_{n+1})) = \Phi(\alpha x) + \frac{1}{2n} \Phi''(\alpha x) + O(n^{-\frac{3}{2}}) \quad (7.53)$$

$$= \Phi(x) + \frac{x}{2n} \Phi'(x) + \frac{1}{2n} \Phi''(x) + O(n^{-\frac{3}{2}}) \quad (7.54)$$

Bis zu diesem Punkt ist die Rechnung sehr allgemein und nutzt noch gar nicht aus, dass Φ zu einer Gaußverteilung gehört. Dieser entscheidende Schritt, also, dass die Verteilung gegen eine Gaußverteilung konvergiert und nicht gegen irgendeine andere, folgt nun. Bei einer Gaußverteilung, und nur dort, heben sich nämlich die Φ' und Φ'' Terme auf.

$$\Phi'(x) = \varphi(x) \propto e^{-\frac{x^2}{2}} \quad (7.55)$$

$$\Phi''(x) = \varphi'(x) \propto -xe^{-\frac{x^2}{2}} \quad (7.56)$$

$$\frac{x}{2n} \Phi'(x) = \frac{1}{2n} \Phi''(x) \quad (7.57)$$

$$E(\Phi(\alpha x - \beta X_{n+1})) = \Phi(x) + O(n^{-\frac{3}{2}}) \quad (7.58)$$

Also gibt es eine Konstante k , so dass

$$|d(F_{n+1}, \Phi) - d(F_n, \Phi)| \leq k \cdot n^{-\frac{3}{2}}. \quad (7.59)$$

Wir haben eine Abschätzung für einen Schritt, also wie viel die Abweichung von Φ von F_n nach F_{n+1} wachsen kann. Daraus gewinnen wir nun eine Gesamtabschätzung durch summieren. $Z_N = Y_0$ ist eine Standard-Gaußverteilung, also $F_N = \Phi$ und $d(F_N, \Phi) = 0$. Somit für $n \geq N$

$$d(Z_n, \Phi) = d(f_n, \Phi) \leq k \sum_{i=N}^n i^{-\frac{3}{2}} \leq k \sum_{i=N}^{\infty} i^{-\frac{3}{2}}. \quad (7.60)$$

Der letzte Term ist eine konvergente Reihe mit sich änderndem Startindex N . Das heißt wenn $N \rightarrow \infty$ geht, geht die Summe gegen 0. Dies folgt aus folgender Rechnung:

$$\sum_{i=N}^{\infty} i^{-\frac{3}{2}} = \sum_{i=0}^{\infty} i^{-\frac{3}{2}} - \sum_{i=0}^{N-1} i^{-\frac{3}{2}} \xrightarrow{N \rightarrow \infty} \sum_{i=0}^{\infty} i^{-\frac{3}{2}} - \sum_{i=0}^{\infty} i^{-\frac{3}{2}} = 0 \quad (7.61)$$

Wir wissen nun: Wenn N gegen ∞ geht, wird der Abstand zwischen $Z_n, n \geq N$ und Φ beliebig klein im Sinne von $d(\dots)$. Außerdem wird der Abstand zwischen Z_n und S_n^* beliebig klein im Sinne der mittleren quadratischen Differenz. Wir setzen diese beiden Ergebnisse formal mit Hilfe von Lemma 7.1 zusammen:

Zu $\varepsilon > 0$ existiert also N mit

$$d(Z_n, \Phi) < \frac{\varepsilon}{2} \text{ für alle } n \geq N \quad (7.62)$$

Ferner

$$\mathbb{E}((S_n^* - Z_n)^2) = \mathbb{E} \left(\left(\sqrt{\frac{N}{n}} Y_0 - \sqrt{\frac{1}{n}} S_N \right)^2 \right) \xrightarrow{n \rightarrow \infty} 0 \quad (7.63)$$

Daher für n hinreichend groß

$$\mathbb{E}((S_n^* - Z_n)^2) \leq \frac{1}{2} \left(\frac{\varepsilon}{4} \right)^3 \quad (7.64)$$

also nach Lemma oben

$$d(S_n^*, \Phi) \leq d(Z_n, \Phi) + \frac{\varepsilon}{4} < d(Z_n, \Phi) + \frac{\varepsilon}{2} < \varepsilon \quad (7.65)$$

Damit ist bewiesen, dass $d(S_n^*, \Phi)$ für $n \rightarrow \infty$ gegen 0 konvergiert, also unter den eingeschränkten Voraussetzungen dieses Satzes, S_n^* gegen eine Gaußverteilung. \square

8 Lineare Algebra

8.1 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:

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} + \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} v_1 + w_1 \\ \vdots \\ v_n + w_n \end{pmatrix} \quad \lambda \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} \lambda v_1 \\ \vdots \\ \lambda v_n \end{pmatrix} \quad (8.1)$$

Spezielle Vektoren

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

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

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

8.2 Lineare Abbildung

Eine Abb. $f : \mathbb{R}^m \rightarrow \mathbb{R}^n$ ist *linear*, falls $f(v + w) = f(v) + f(w)$ und $f(\lambda v) = \lambda f(v)$ für alle v, w, λ (Es folgt $f(0) = 0$, denn: $f(0) = f(0 \cdot 0) = 0 \cdot f(0) = 0$)

Eine Abb. g der Form $g(v) = f(v) + a$ mit f linear und $a \in \mathbb{R}^n$ heißt *affin*.

8.3 Matrizen

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

$$A = \begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \vdots \\ A_{m1} & \cdots & A_{mn} \end{pmatrix} \quad (8.2)$$

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.

8.3.1 Spezielle Matrizen

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

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

8.3.2 Matrixmultiplikation

Für $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times p}$ ist $AB \in \mathbb{R}^{m \times p}$ definiert durch $(AB)_{ij} = \sum_{k=1}^n A_{ik}B_{kj}$

8.3.3 Matrizen und lineare Abbildungen

Eine $m \times n$ -Matrix A induziert lineare Abbildung

$$f_A : \mathbb{R}^n \rightarrow \mathbb{R}^m, v \mapsto Av \quad (8.3)$$

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

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

Diese Zuordnungen sind gegenseitig invers.

Beweis.

$$(A^f \cdot v)_i = \sum_{j=1}^n f(e_j)_i v_j = \left(\sum_{j=1}^n v_j f(e_j) \right)_i \stackrel{f \text{ lin.}}{=} f \left(\underbrace{\sum_{j=1}^n v_j e_j}_v \right)_i = f(v)_i \quad (8.4)$$

$$\Rightarrow f_{A^f} = f \quad (8.5)$$

$$f_A(e_j) = A \cdot e_j = A_{\bullet j} \quad (8.6)$$

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

d. h. $m \times n$ -Matrixen „sind“ lineare Abbildungen $\mathbb{R}^n \rightarrow \mathbb{R}^m$.

Dabei entspricht Matrix-Multiplikation gerade der Funktionskomposition:

$$f_{AB} = f_A \circ f_B : \mathbb{R}^p \xrightarrow{f_B} \mathbb{R}^n \xrightarrow{f_A} \mathbb{R}^m \quad (8.7)$$

Beweis.

$$\begin{aligned} f_A(f_B(v))_i &= \sum_{k=1}^p f_A(v)_k = \sum_{k=1}^p A_{ik} \cdot \sum_{j=1}^p B_{kj} v_j = \sum_{j=1}^p \left(\sum_{k=1}^p A_{ik} B_{kj} \right) v_j \\ &= \sum_{j=1}^p (AB)_{ij} v_j = f_{AB}(v)_i \quad \square \end{aligned} \quad (8.8)$$

Folgerung: Matrix-Multiplikation ist assoziativ: $(AB)C = A(BC)$
 Ferner $f_1 = \text{id}$ (bzw. $A^{\text{id}} = I$), also $IA = A = AI$

8.3.4 Weitere Matrixoperationen

Transponierte

$$(A^\top)_{ij} = A_{ji}, \quad A \in \mathbb{R}^{m \times n} \Rightarrow A^\top \in \mathbb{R}^{n \times m} \quad (8.9)$$

Inverse

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

Ein Inverses existiert nicht immer, ist aber bei Existenz eindeutig. Schreibe dann $B = A^{-1}$; es gilt auch $A^{-1}A = I$.

8.3.5 Rechenregeln

$$(A + B)C = AC + BC, \quad C(A + B) = CA + CB \quad (8.10)$$

$$(AB)^\top = B^\top A^\top, \quad \text{_}^\top \text{ ist linear} \quad (8.11)$$

$$(AB)^{-1} = B^{-1}A^{-1}, \quad (\lambda A)^{-1} = \frac{1}{\lambda} \cdot A^{-1}, \quad (8.12)$$

$$(A^{-1})^{-1} = A, \quad (A^{\top\top}) = A, \quad (8.13)$$

$$(A^\top)^{-1} = (A^{-1})^\top \quad (8.14)$$

$$\text{im Allgemeinen } AB \neq BA \quad (8.15)$$

8.4 Skalarprodukt und Norm

Für $v, w \in \mathbb{R}^n$ setze

$$\langle v, w \rangle = v^\top w = \sum_{i=1}^n v_i w_i \quad \text{Skalarprodukt} \quad (8.16)$$

$$\|v\| = \sqrt{\langle v, v \rangle} = \sqrt{\sum_{i=1}^n v_i^2} \quad \text{Norm} \quad (8.17)$$

8.4.1 Geometrische Interpretation

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

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

8.4.2 Eigenschaften

$\langle _, _ \rangle$ ist bilinear (d. h. linear in beiden Komponenten) und symmetrisch, d. h. $\langle v, w \rangle = \langle w, v \rangle$

8.5 Symmetrische Matrizen

$A \in \mathbb{R}^{n \times n}$ heißt symmetrisch, falls $A^\top = A$. Mit A, B sind $A + B, \lambda A, A^{-1}$ symmetrisch. Für A symmetrisch ist $B^\top AB$ symmetrisch: $(B^\top AB)^\top = B^\top A^\top (B^\top)^\top = B^\top AB$ Insbes. BB^\top symmetrisch.

8.5.1 Quadratische Funktion

$A \in \mathbb{R}^{n \times n}$ symmetrisch induziert quadratische Funktion

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, v \mapsto v^\top Av = \sum_{i,j} v_i A_{ij} v_j = \sum_i A_{ii} v_i^2 + 2 \sum_{i < j} A_{ij} v_i v_j \quad (8.18)$$

Funktion 2ter Ordnung:

$$g : v \mapsto v^\top Av + b^\top v + \gamma \quad A \in \mathbb{R}^{n \times n}, b \in \mathbb{R}^n, \gamma \in \mathbb{R} \quad (8.19)$$

8.5.2 Jacobi Matrix

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

$$\frac{\partial f(v)}{\partial v} = \left(\frac{\partial f^i(v)}{\partial v_j} \right)_{ij} \quad (8.20)$$

Speziell für f quadratische Funktion bzw g Funktion 2. Ordnung gilt:

$$\left. \frac{\partial f(v)}{\partial v} \right|_v = 2v^\top A \quad (8.21)$$

$$\left. \frac{\partial g(v)}{\partial v} \right|_v = 2v^\top A + b^\top \quad (8.22)$$

Beweis.

$$\left(\frac{\partial f(v)}{\partial v} \right)_k = 2A_{kk}v_k + 2 \sum_{j \neq k} A_{ij}v_j = 2 \sum_j v_j A_{jk} = (2v^\top A)_k \quad \square$$

9 Wahrscheinlichkeitsrechnung im \mathbb{R}^n

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

9.1 Zufallsvariable

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

9.2 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.

$$E(X) = \int_{\omega \in \Omega} X(\omega) p(\omega) d\omega \quad (9.1)$$

$$= \int_{x \in \mathbb{R}^n} x p(X = x) dx \quad (9.2)$$

$$E(U) = \int_{\omega \in \Omega} U(\omega) p(\omega) d\omega \quad (9.3)$$

$$= \int_{u \in \mathbb{R}^{n \times m}} u p(U = u) du \quad (9.4)$$

9.2.1 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.

$$E(X)_i = E(X_i) \qquad E(U)_{i,j} = E(U_{i,j}) \quad (9.5)$$

Beweis.

$$E(U)_{i,j} = \left(\int_{\omega \in \Omega} U(\omega) p(\omega) d\omega \right)_{i,j} \quad (9.6)$$

$$= \int_{\omega \in \Omega} U(\omega)_{i,j} p(\omega) d\omega \quad (9.7)$$

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

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

9.3 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:

$$\text{Cov}(X) = E((X - E(X))(X - E(X))^\top) \quad (9.8)$$

$$\text{Cov}(X, Y) = E((X - E(X))(Y - E(Y))^\top) \quad (9.9)$$

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 zu vergleichen ist Übungsaufgabe.

9.3.1 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, obwohl sie als Vektorausdruck definiert ist, aus einer Matrixanordnung der Kovarianzen der Komponenten von X und Y .

$$\text{Cov}(X, Y)_{i,j} = \text{Cov}(X_i, Y_j) \quad (9.10)$$

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

$$(uv^\top)_{ij} = \sum_{k=1}^1 u_{i,k}v_{j,k} = u_i v_j \quad (9.11)$$

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

$$\text{Cov}(X, Y)_{i,j} = E(((X - E(X))(Y - E(Y))^\top)_{i,j}) \quad (9.12)$$

$$= E((X_i - E(X_i))(Y_j - E(Y_j))) \quad (9.13)$$

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

Weitere Eigenschaften sind (Beweis: Übungsaufgabe):

1. $\text{Cov}(X, Y) = \mathbb{E}(XY^\top) - \mathbb{E}(X) \mathbb{E}(Y)^\top$ für $X : \Omega \rightarrow \mathbb{R}^n, Y : \Omega \rightarrow \mathbb{R}^m$
2. $\text{Cov}(X) = \mathbb{E}(XX^\top) - \mathbb{E}(X) \mathbb{E}(X)^\top$ 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 $\text{Cov}(X)$ Aussagen ergibt sich durch Einsetzen in die entsprechenden $\text{Cov}(X, Y)$ Aussagen. 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 Kalman Filter etc., enthält die Kovarianzmatrix also alle Information, die man benötigt.

Lemma 9.1. *Die von $\text{Cov}(X)$ induzierte quadratische Form $r \mapsto r^\top \text{Cov}(X)r$ ordnet jeder Richtung r die Varianz von X projiziert in Richtung r zu.*

Beweis.

$$\underbrace{V(r^\top X)}_{V(\sum_k r_k X_k)} = \text{Cov}(r^\top X) \tag{9.14}$$

$$\begin{aligned} &= r^\top \text{Cov}(X) r^{\top\top} \\ &= r^\top \text{Cov}(X) r \end{aligned} \tag{9.15} \quad \square$$

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.

Definition. Eine symmetrische Matrix $A \in \mathbb{R}^{n \times m}$, $A^\top = A$ heißt positiv (semi-)definit, wenn $r^\top A r \geq 0, \forall r$ bzw. $r^\top A r > 0, \forall r \neq 0$.

Für die induzierte quadratische Funktion bedeutet dies, eine Matrix ist positiv (semi) definit, genau dann, wenn die dazugehörige quadratische Form > 0 (≥ 0) ist, für alle Argumente außer dem 0 Vektor.

Korollar 9.2. $\text{Cov}(X)$ ist symmetrisch positiv semidefinit.

Zusammenfassend kann man die Kovarianzmatrix aus 3 verschiedenen Blickwinkeln sehen. Die Kovarianzmatrix ist

- eine als $E(\dots)$ eines äußeren Produktes definierte Größe
- eine systematische Zusammenstellung der Kovarianzen der Komponenten
- eine Funktion die für beliebige Linearkombinationen $\rightarrow \mathbb{R}$ die Varianz zuordnet

Der erste Punkt ermöglicht, mit der Definition durch Anwenden von Matrix- und $E(\dots)$ -Regeln zu rechnen und garantiert, dass der gewählte Begriff unabhängig von Koordinatensystemen ist.

Der zweite Punkt zeigt, dass 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.

Zum Schluss 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, als es ja unendlich viele Richtungen gibt und man für die Varianz in allen Richtungen erst mal unendlich viele Zahlen bräuchte. Die dritte Aussage besagt, dass die n^2 vielen Zahlen der Kovarianzmatrix schon genug sind.

9.4 Mehrdimensionale Gaußverteilung

Wir wollen jetzt die Gaußverteilung auf mehrere Dimensionen generalisieren. 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 Erwartungswerte und beliebiger Kovarianz.

Definition. $X : \Omega \rightarrow \mathbb{R}^n$ ist Standard-Gaußverteilt, falls X_i unabhängig und Standard-Gaußverteilt sind.

Lemma 9.3. Die Dichte der \mathbb{R}^n Standard-Gaußverteilung ist

$$p(X = x) = \frac{1}{\sqrt{2\pi}^n} \exp\left(-\frac{1}{2}x^\top I x\right) = \frac{1}{\sqrt{2\pi}^n} \exp\left(-\frac{1}{2} \sum_i x_i^2\right) \quad (9.16)$$

9 Wahrscheinlichkeitsrechnung im \mathbb{R}^n

Beweis.

$$p(X = x) \stackrel{\text{unabhängig}}{=} \prod_{i=1}^n p(X_i = x_i) \quad (9.17)$$

$$= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}X_i^2\right) \quad (9.18)$$

$$= \frac{1}{\sqrt{2\pi}^n} \exp\left(-\frac{1}{2} \sum_i X_i^2\right) \quad (9.19)$$

$$= \frac{1}{\sqrt{2\pi}^n} \exp\left(-\frac{1}{2}x^\top I x\right) \quad \square$$

Lemma 9.4. *Die Standard-Gaußverteilung X hat $E(X) = 0$ und $\text{Cov}(X) = I$.*

Beweis.

$$E(X)_i = E(X_i) = 0 \quad (9.20)$$

$$\text{Cov}(X)_{i,j} = \text{Cov}(X_i, X_j) = \begin{cases} 0, & i \neq j, \text{ weil unabhängig} \\ 1, & i = j, \text{ weil Std.Gauß} \end{cases} \quad (9.21)$$

$$\Rightarrow \text{Cov}(X) = I \quad \square$$

Das folgende Lemma ergibt sich aus der Transformation von Integralen und soll hier ohne Beweis benutzt werden. 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 9.5. *Seien $X : \Omega \rightarrow \mathbb{R}^n$, $A \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$ und $Y = AX + b$ dann*

$$p(Y = y) = \frac{p(X = A^{-1}(y - b))}{|A|} \quad (9.22)$$

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 9.6. *(Cholesky): $C \in \mathbb{R}^{n \times n}$, symmetrisch positiv definit, dann gibt es $L \in \mathbb{R}^{n \times n}$, so dass $LL^\top = C$.*

Mit diesen Hilfsmitteln zeigen wir, dass es zu jedem Erwartungswert und jeder positiven definiten Kovarianzmatrix eine Gaußglocke gibt und leiten deren Formel her.

Satz 13. *Zu jedem $\mu \in \mathbb{R}^n$ und symmetrisch positiv definiten $C \in \mathbb{R}^{n \times n}$ ist Y ein affines Bild der Standard-Gaußverteilung mit Erwartungswert μ und Kovarianz C und Dichte*

$$p(Y = y) = \frac{1}{\sqrt{2\pi}^n \sqrt{|C|}} \exp\left(-\frac{1}{2}(y - \mu)^\top C^{-1}(y - \mu)\right) \quad (9.23)$$

Beweis. Setze $LL^\top = C$ und $Y = LX + \mu$ für X Standard-Gaußverteilt. Dann:

$$E(Y) = E(LX + \mu) = LE(X) + \mu \quad (9.24)$$

$$= \mu \quad (9.25)$$

$$\text{Cov}(Y) = \text{Cov}(LX + \mu) = \text{Cov}(LX) + \text{Cov}(\mu) \quad (9.26)$$

$$= \text{Cov}(LX) = L \text{Cov}(X)L^\top = LL^\top \quad (9.27)$$

$$= C \quad (9.28)$$

Damit stimmen Erwartungswert und Varianz. Nachzurechnen ist noch, ob die Verteilung die oben angegebene Form hat:

$$p(Y = y) = \frac{p(X = x)}{|L|} \quad \text{wobei: } x = L^{-1}(y - \mu) \quad (9.29)$$

$$= \frac{1}{\sqrt{2\pi}^n \sqrt{|L| \cdot |L^\top|}} \exp\left(-\frac{1}{2}x^\top I x\right) \quad (9.30)$$

$$= \frac{1}{\sqrt{2\pi}^n \sqrt{|C|}} \exp\left(-\frac{1}{2}(L^{-1}(y - \mu))^\top I (L^{-1}(y - \mu))\right) \quad (9.31)$$

$$= \frac{1}{\sqrt{2\pi}^n \sqrt{|C|}} \exp\left(-\frac{1}{2}(y - \mu)^\top L^{-\top} L^{-1}(y - \mu)\right) \quad (9.32)$$

$$= \frac{1}{\sqrt{2\pi}^n \sqrt{|C|}} \exp\left(-\frac{1}{2}(y - \mu)^\top (LL^\top)^{-1}(y - \mu)\right) \quad (9.33)$$

$$= \frac{1}{\sqrt{2\pi}^n \sqrt{|C|}} \exp\left(-\frac{1}{2}(y - \mu)^\top C^{-1}(y - \mu)\right) \quad \square$$

10 Mehrdimensionaler (Extended) Kalmanfilter

Vorlesung
2007-01-15

10.1 Lineare Quadratische Ausgleichsrechnung

Die lineare quadratische Ausgleichsrechnung ist das n -dimensionale Pendant zur Zwei-Messungs Formel ([Satz 4](#)) 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.

Diese Formel wird sehr häufig genutzt. Sie ist die Grundlage, z. B. für Ausgleichsgeraden oder -kurven und Kalibrierung von Messgeräten. Außerdem wurde sie historisch 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 10.1. *Ist $X : \Omega \rightarrow \mathbb{R}^n$ verteilt als Exponential einer quadratischen Funktion $p(X = x) \propto \exp(-\frac{1}{2}(x^\top Ax + x^\top b + \gamma))$ 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*

$$E(X) = \frac{-A^{-1}b}{2} \qquad \text{Cov}(X) = A^{-1} \qquad (10.1)$$

Beweis.

$$\begin{aligned} & \exp\left(-\frac{1}{2}\left(\frac{x + A^{-1}b}{2}\right)^\top A \left(\frac{x + A^{-1}b}{2}\right)\right) \\ & \propto \exp\left(-\frac{1}{2}(x^\top Ax + b^\top A^{-1}Ax/2 + x^\top A^{-1}Ab/2)\right) \\ & = \exp\left(-\frac{1}{2}(x^\top Ax + x^\top b)\right) \quad \square \end{aligned} \qquad (10.2)$$

10.1.1 Annahmen

1. Zustand $X : \Omega \rightarrow \mathbb{R}^n$ uniform verteilt, gesucht
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 beobachtet

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.

Lemma 10.2. *Unter den obigen Annahmen ist $X|Z = z$ Gaußsch mit*

$$E(X|Z = z) = (A^\top \Sigma_\varepsilon^{-1} A)^{-1} A^\top \Sigma_\varepsilon^{-1} (z - b) \quad (10.3)$$

$$\text{Cov}(X|Z = z) = (A^\top \Sigma_\varepsilon^{-1} A)^{-1} \quad (10.4)$$

Wäre A invertierbar, würde sich (10.3) 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 (10.3) nicht in die Klammer hineinziehen.

Beweis.

$$p(X = x|Z = z) = \frac{p(Z = z|X = x) p(X = x)}{p(Z = z)} \propto p(Z = z|X = x) \quad (10.5)$$

$$= p(\varepsilon = z - (Ax + b)) \quad (10.6)$$

$$\propto \exp\left(-\frac{1}{2}(z - (Ax + b))^\top \Sigma_\varepsilon^{-1} (z - (Ax + b))\right) \quad (10.7)$$

$$\propto \exp\left(-\frac{1}{2}\left(x^\top \underbrace{A^\top \Sigma_\varepsilon^{-1} A}_{A \text{ Lemma 10.1}} x + x^\top \underbrace{A^\top \Sigma_\varepsilon^{-1} (b - z)}_{b \text{ Lemma 10.1}} \cdot (2)\right)\right) \quad (10.8)$$

$$\Rightarrow X|Z = z \text{ Gaußsch} \quad (10.9)$$

$$E(X|Z = z) = (A^\top \Sigma_\varepsilon^{-1} A)^{-1} A^\top \Sigma_\varepsilon^{-1} (z - b) \quad (10.10)$$

$$\text{Cov}(X|Z = z) = (A^\top \Sigma_\varepsilon^{-1} A)^{-1} \quad \square$$

10.2 Nichtlineare Ausgleichsrechnung

Die nichtlineare Ausgleichsrechnung erweitert die Annahmen von oben auf einen nichtlinearen Zusammenhang zwischen Zustand und Messung.

10.2.1 Annahmen

1. X uniform verteilt $X : \Omega \rightarrow \mathbb{R}^n$, gesucht
2. $Z = f(X) + \varepsilon$, $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $\varepsilon : \Omega \rightarrow \mathbb{R}^m \sim \mathcal{N}(0, \Sigma_\varepsilon)$, gemessen

Idee: f linearisieren:

$$f(x) \approx f(x_0) + \left. \frac{\partial f(x)}{\partial x} \right|_{x=x_0} \cdot (x - x_0) \quad (10.11)$$

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. Wiederholt man dies bis zur Konvergenz, so ist die Lösung zwar nicht der exakte nichtlineare Erwartungswert aber zumindest ein lokales Maximum der exakten nichtlinearen Wahrscheinlichkeitsdichte.

10.2.2 Algorithmus

Startwert μ

Wiederhole bis μ konvergiert:

$$A := \left. \frac{\partial f(x)}{\partial x} \right|_{x=\mu} \quad (10.12)$$

$$b := f(\mu) - A\mu \quad (10.13)$$

$$\mu := (A^\top \Sigma_\varepsilon^{-1} A)^{-1} A^\top \Sigma_\varepsilon^{-1} (z - b) \quad (10.14)$$

Dann ist $\Sigma = (A^\top \Sigma_\varepsilon^{-1} A)^{-1}$.

Es gibt noch verbesserte Varianten des Algorithmus, z. B. den Levenbergh-Marquardt Algorithmus [8, §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.

10.3 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. Zunächst benötigen wir den Begriff einer Blockmatrix.

Bemerkung 10.3. *Setzt man Matrizen Blockweise zusammen, gelten analoge Rechenregeln für die Blöcke wie bei einzelnen Zahlen. Insbesondere:*

$$\begin{pmatrix} x \\ z \end{pmatrix}^\top \begin{pmatrix} P & Q \\ R & S \end{pmatrix} \begin{pmatrix} x \\ z \end{pmatrix} = \begin{pmatrix} x \\ z \end{pmatrix}^\top \begin{pmatrix} Px & Qz \\ Rx & Sz \end{pmatrix} = x^\top Px + x^\top Qz + z^\top Rx + z^\top Sz \quad (10.15)$$

ist $\begin{pmatrix} P & Q \\ R & S \end{pmatrix}$ symmetrisch: $P = P^\top, R = Q^\top, S = S^\top$, dann

$$= x^\top Px + 2x^\top Qz + z^\top Sz \quad (10.16)$$

Lemma 10.4 (Block-Matrix-Inversions-Lemma). *Es ist:*

$$\begin{pmatrix} P & Q \\ R & S \end{pmatrix}^{-1} = \begin{pmatrix} \tilde{P} & \tilde{Q} \\ \tilde{R} & \tilde{S} \end{pmatrix} \quad (10.17)$$

Mit:

$$\tilde{P} = (P - QS^{-1}R)^{-1} \quad \text{Schur-Komplement} \quad (10.18)$$

$$\tilde{Q} = -\tilde{P}(QS^{-1}) \quad (10.19)$$

$$\tilde{R} = (S^{-1}R)\tilde{P} \quad (10.20)$$

$$\tilde{S} = S^{-1} + (S^{-1}R)\tilde{P}(QS^{-1}) \quad (10.21)$$

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

Das folgende Lemma leitet die Formel her, die der Kalman Filter zur Integration einer Messung benutzt. Sie ermöglicht den Erwartungswert von X unter einer Bedingung $Z = z$ aus Kovarianzen zwischen X und Z zu berechnen.

Lemma 10.5 (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*

$$E(X|Z = z) = E X + \text{Cov}(X, Z) \text{Cov}(Z)^{-1}(z - E Z) \quad (10.22)$$

$$\text{Cov}(X|Z = z) = \text{Cov} X - \text{Cov}(X, Z) \text{Cov}(Z)^{-1} \text{Cov}(X, Z)^\top \quad (10.23)$$

Beweis. Reduktion auf $E X = E Z = 0$. Zuerst den Spezialfall

$$p(X = x|Z = z) = \frac{p(X = x \wedge Z = z)}{p(Z = z)} \quad (10.24)$$

$$\propto p(X = x \wedge Z = z) = p\left(\begin{pmatrix} X \\ Z \end{pmatrix} = \begin{pmatrix} x \\ z \end{pmatrix}\right) \quad (10.25)$$

$$\propto \exp\left(-\frac{1}{2} \begin{pmatrix} x \\ z \end{pmatrix}^\top \begin{pmatrix} \text{Cov}(X) & \text{Cov}(X, Z) \\ \text{Cov}(X, Z)^\top & \text{Cov}(Z) \end{pmatrix}^{-1} \begin{pmatrix} x \\ z \end{pmatrix}\right) \quad (10.26)$$

Nach BM-Inversionslemma:

$$= \exp\left(-\frac{1}{2} \begin{pmatrix} x \\ z \end{pmatrix}^\top \begin{pmatrix} \tilde{P} & \tilde{Q} \\ \tilde{R} & \tilde{S} \end{pmatrix} \begin{pmatrix} x \\ z \end{pmatrix}\right) \quad (10.27)$$

$$\propto \exp\left(-\frac{1}{2} x^\top \tilde{P} x + x^\top (2\tilde{Q}z)\right) \quad (10.28)$$

also nach Lemma 10.1 Gaußsch

$$E(X|Z = z) = \tilde{P}^{-1}\tilde{Q}z = \tilde{P}^{-1}\tilde{P}QS^{-1}z \quad (10.29)$$

$$= QS^{-1}z = \text{Cov}(X, Z) \text{Cov}(Z)^{-1}z \quad (10.30)$$

$$\text{Cov}(X|Z = z) = \tilde{P}^{-1} = P - QS^{-1}R \quad (10.31)$$

$$= \text{Cov}(X) - \text{Cov}(X, Z) \text{Cov}(Z)^{-1} \text{Cov}(X, Z)^\top \quad (10.32)$$

Allgemeinfall: Setze $X' = X - E X, Z' = Z - E Z, z' = z - E Z$

$$E(X|Z = z) = E(X'|Z' = z') + E X = E X + \text{Cov}(X, Z) \text{Cov}(Z)^{-1}(z - E Z) \quad (10.33)$$

$$\text{Cov}(X|Z = z) = \text{Cov}(X'|Z' = z') = \text{Cov}(X) - \text{Cov}(X, Z) \text{Cov}(Z)^{-1} \text{Cov}(X, Z)^\top \quad \square$$

10.3.1 Annahmen

- $X_t : \Omega \rightarrow \mathbb{R}^n$ Zustand zur Zeit t
 - $U_t : \Omega \rightarrow \mathbb{R}^p$ Zustandsübergangsmessung 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 ε_t, δ_t unabhängig

10.3.2 Algorithmus

Variablen: $\mu_t \in \mathbb{R}^n, \Sigma_t \in \mathbb{R}^{n \times n}$ (10.34)

Initialisierung: $\mu_0 = E(X_0), \Sigma_0 = \text{Cov}(X_0)$ (10.35)

Dynamik($U_t = u_t$): $\bar{\mu}_t = A_t \mu_{t-1} + B_t u_t + g_t$ (10.36)

$$\bar{\Sigma}_t = A_t \Sigma_{t-1} A_t^\top + \Sigma_{\varepsilon t}$$
 (10.37)

Messung($Z_t = z_t$): $K_t = \bar{\Sigma}_t C_t^\top (C_t \bar{\Sigma}_t C_t^\top + \Sigma_{\delta t})^{-1}$ (10.38)

$$\mu_t = \bar{\mu}_t + K_t (z_t - (C_t \bar{\mu}_t + h_t))$$
 (10.39)

$$\Sigma_t = \bar{\Sigma}_t - K_t C_t \bar{\Sigma}_t$$
 (10.40)

10.4 Extended-Kalman-Filter

Der Extended Kalman Filter ersetzt wie im eindimensionalen die linearen Funktionen für Zustandsübergang und Messung durch allgemeine 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.

10.4.1 Annahmen

1. X_0 ist ungefähr $\mathcal{N}(\mu_0, \Sigma_0)$
2. $X_t = g_t(X_{t-1}, U_t) + \varepsilon_t$; $g_t : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}^n$, $\varepsilon_t \sim \mathcal{N}(0, \Sigma_{\varepsilon t})$
3. $Z_t = h_t(X_t) + \delta_t$; $h_t : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $\delta_t \sim \mathcal{N}(0, \Sigma_{\delta t})$
4. Alle ε_t, δ_t sind unabhängig.

10.4.2 Algorithmus

Dynamik($U_t = u_t$): $\bar{\mu}_t = g(\mu_{t-1}, u_t)$ (10.41)

$$\bar{\Sigma}_t = A_t \Sigma_{t-1} A_t^\top + \Sigma_{\varepsilon t}, \quad A = \left. \frac{\partial g(x, u)}{\partial x} \right|_{x=\mu_{t-1}}$$
 (10.42)

$$\text{Messung}(Z_t = z_t): \quad C_t = \left. \frac{\partial h(x)}{\partial x} \right|_{x=\bar{\mu}_t} \quad (10.43)$$

$$K_t = \bar{\Sigma}_t C_t^\top (C_t \bar{\Sigma}_t C_t^\top + \Sigma_{\delta t})^{-1} \quad (10.44)$$

$$\mu_t = \bar{\mu}_t + K_t(z_t - h(\bar{\mu}_t)) \quad (10.45)$$

$$\Sigma_t = \bar{\Sigma}_t - K_t C_t \bar{\Sigma}_t \quad (10.46)$$

10.5 Beweis Kalmanfilter (10.3)

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 14. *Nach Ausführung der KF Formel ist*

$$X_t | u_{1..t}, z_{1..t-1} \sim \mathcal{N}(\bar{\mu}_t, \bar{\Sigma}_t) \quad \forall t > 0 \quad (10.47)$$

$$X_t | u_{1..t}, z_{1..t} \sim \mathcal{N}(\mu_{t-1}, \Sigma_t) \quad \forall t \geq 0 \quad (10.48)$$

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.

$$X_{t-1} | u_{1..t-1}, z_{1..t-1} \sim \mathcal{N}(\mu_{t-1}, \Sigma_{t-1}) \quad (10.49)$$

$$A_t X_{t-1} | u_{1..t-1}, z_{1..t-1} \sim \mathcal{N}(A_t \mu_{t-1}, A_t \Sigma_{t-1} A_t^\top) \quad (10.50)$$

$$\varepsilon \sim \mathcal{N}(0, \Sigma_{\varepsilon t}) \quad (10.51)$$

$$X_t = A_t X_{t-1} + B_t U_t + g_t + \varepsilon_t | u_{1..t}, z_{1..t-1} \quad (10.52)$$

$$\sim \mathcal{N}(A_t \mu_{t-1} + B_t u_t + g_t, A_t \Sigma_{t-1} A_t^\top + \Sigma_{\varepsilon t}) \quad (10.53)$$

$$\sim \mathcal{N}(\bar{\mu}_t, \bar{\Sigma}_t) \quad (10.54)$$

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

$$X_t | *, z_t = X_t | u_{1..t}, z_{1..t} \text{ Gaußsch} \quad (10.55)$$

10 Mehrdimensionaler (Extended) Kalmanfilter

Um die Terme für Erwartungswert und Kovarianz zusammzusetzen, berechnen wir zuerst die im Lemma auftretenden Kovarianzausdrücke aus den im Algorithmus auftretenden Variablen.

$$\text{Cov}(X_t, Z_t) = \text{Cov}(X_t, C_t X_t + h_t + \delta_t | *) \quad (10.56)$$

$$= \text{Cov}(X_t, C_t X_t | *) \quad (10.57)$$

$$= \text{Cov}(X_t, X_t | *) C_t^\top \quad (10.58)$$

$$= \bar{\Sigma}_t C_t^\top \quad (10.59)$$

$$\text{Cov}(Z_t | *) = \text{Cov}(C_t X_t + h_t + \delta_t | *) \quad (10.60)$$

$$= \text{Cov}(C_t X_t | *) + \text{Cov}(\delta_t | *) \quad (10.61)$$

$$= C_t \text{Cov}(X_t | *) C_t^\top + \Sigma_{\delta t} \quad (10.62)$$

$$= C_t \bar{\Sigma}_t C_t^\top + \Sigma_{\delta t} \quad (10.63)$$

$$\text{Cov}(X, Z | *) \text{Cov}(Z | *)^{-1} = K_t \quad (10.64)$$

$$\text{E}(Z_t | *) = \text{E}(C_t X_t + h_t + \delta_t | *) \quad (10.65)$$

$$= C_t \bar{\mu}_t + h_t \quad (10.66)$$

Nun setzen wir nach dem Konditionierungslemma zusammen:

$$\text{E}(X_t | u_{1..t}, z_{1..t}) = \text{E}(X_t | *, z_t) \quad (10.67)$$

$$= \bar{\mu}_t + \text{Cov}(X_t, Z_t | *) \text{Cov}(Z_t | *)^{-1} (z_t - \text{E}(Z_t | *)) \quad (10.68)$$

$$= K (z_t - (C_t \bar{\mu}_t + h_t)) \quad (10.69)$$

$$\text{Cov}(X_t | u_{1..t}, Z_{1..t}) = \text{Cov}(X_t | *, z_t) \quad (10.70)$$

$$= \text{Cov}(X_t | *) - \text{Cov}(X_t, Z_t | *) \text{Cov}(Z_t | *)^{-1} \text{Cov}(X, Z | *)^\top \quad (10.71)$$

$$= \bar{\Sigma}_t - K C_t \bar{\Sigma}_t = \Sigma_t \quad \square$$

10.6 Variante des EKF mit Rauschen in U

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 . 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 unterscheidet zwischen der Zustandsübergangsmessung U und dem wahren Wert V der in Zustandsübergang eingeht.

10.6.1 Annahmen

- $X_t : \Omega \rightarrow \mathbb{R}^n$ Zustand zur Zeit t
- $U_t : \Omega \rightarrow \mathbb{R}^p$ Zustandsübergangsmessung t

- $V_t : \Omega \rightarrow \mathbb{R}^p$ Zustandsübergang wahrer Wert t
 - $Z_t : \Omega \rightarrow \mathbb{R}^m$ Messung t
1. $X_0 \sim \mathcal{N}(\mu_0, \Sigma_0)$
 2. $X_t = g_t(X_{t-1}, V_t) + \varepsilon_t; \varepsilon_t \sim \mathcal{N}(0, \Sigma_{\varepsilon t}) \quad U_t = V_t + \gamma_t; \gamma_t \sim \mathcal{N}(0, \Sigma_{\gamma t})$
 3. $Z_t = h(X_t) + \delta_t; \delta_t \sim \mathcal{N}(0, \Sigma_{\delta t})$
 4. Alle $\varepsilon_t, \delta_t, \gamma_t$ unabhängig

Im Algorithmus ändert sich nur der Dynamikschritt. Es kommt ein Term hinzu der die Unsicherheit addiert, die sich aus $\Sigma_{\gamma t}$ durch g ergibt.

$$\text{Dynamik}(U_t = u_t): \quad \bar{\Sigma}_t = A_t \Sigma_{t-1} A_t^\top + B_t \Sigma_{\gamma t} B_t^\top + \Sigma_{\varepsilon t} \quad (10.72)$$

$$B_t = \left. \frac{\partial g(x, u)}{\partial u} \right|_{u=u_t, x=\mu_{t-1}} \quad (10.73)$$

10.7 Unscented Kalman Filter

Vorlesung
2007-01-22

10.7.1 Idee

EKF Formeln propagieren $E(\dots)$ und $\text{Cov}(\dots)$ durch nichtlineare Funktionen mit Hilfe von $f(\dots)$ und $\frac{\partial f}{\partial x}$. Stattdessen $f(\dots)$ an mehreren „repräsentativen“ Stellen $f(\dots)$ auswerten. Annahmen wie EKF.

Definition (Sigma-Punkte-Propagation). Sei $\mu \in \mathbb{R}^n, \Sigma \in \mathbb{R}^{n \times n}$ symmetrisch positiv definit $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$. Sei ferner LL^\top die Cholesky-Zerlegung von Σ . Dann sind definiert:

$$\mu_f = \frac{1}{2n+1} \left[f(\mu) + \sum_{i=1}^n (f(\mu + L_{\bullet i}) + f(\mu - L_{\bullet i})) \right] \quad (10.74)$$

$$\begin{aligned} \Sigma_f = & \frac{1}{2} \left[(f(\mu) - \mu_f)(f(\mu) - \mu_f)^\top \right. \\ & + \sum_{i=1}^n (f(\mu + L_{\bullet i}) - \mu_f)(f(\mu + L_{\bullet i}) - \mu_f)^\top \\ & \left. + \sum_{i=1}^n (f(\mu - L_{\bullet i}) - \mu_f)(f(\mu - L_{\bullet i}) - \mu_f)^\top \right] \end{aligned} \quad (10.75)$$

$$\Sigma_{fX} = \frac{1}{2} \left[\sum_{i=1}^n (f(\mu + L_{\bullet i}) - \mu_f)(L_{\bullet i})^\top + \sum_{i=1}^n (f(\mu - L_{\bullet i}) - \mu_f)(-L_{\bullet i})^\top \right] \quad (10.76)$$

Idee: Hat $X : \Omega \rightarrow \mathbb{R}^n$ $E(X) = \mu, \text{Cov}(X) = \Sigma$, so hat $E(f(X)) \approx \mu_f, \text{Cov}(f(X)) \approx \Sigma_f, \text{Cov}(f(X), X) \approx \Sigma_{fX}$.

Die Argumente an denen f ausgewertet wird, also μ und $\mu \pm L_{\bullet i}$ heißen Sigma-Punkte. Die Cholesky-Zerlegung spielt in diesem Fall dieselbe Rolle, wie im eindimensionalen Fall

die Standard Abweichung. Dort sind die Sigma-Punkte dann $\mu - \sigma, \mu, \mu + \sigma$. Durch diese Konstruktion streuen die Sigma Punkte „repräsentativ“ im Bereich der durch μ, Σ beschriebenen Unsicherheit. 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.

Lemma 10.6. *Idee gilt exakt, wenn f linear, also $f(X) = AX + b$*

Beweis.

$$\mu_f = \frac{1}{2n+1} \left[A\mu + b + \sum_{i=1}^n (A(\mu + L_{\bullet i}) + b) + \sum_{i=1}^n (A(\mu - L_{\bullet i}) + b) \right] \quad (10.77)$$

$$= A\mu + b = E(f(X)) \quad (10.78)$$

$$\begin{aligned} \Sigma_{ff} &= \frac{1}{2} \left[((A\mu + b) - (A\mu + b))((A\mu + b) - (A\mu + b))^\top \right. \\ &\quad + \sum_{i=1}^n (A(\mu + L_{\bullet i}) + b) - (A\mu + b) ((A(\mu + L_{\bullet i}) + b) - (A\mu + b))^\top \\ &\quad \left. + \sum_{i=1}^n (A(\mu - L_{\bullet i}) + b) - (A\mu + b) ((A(\mu - L_{\bullet i}) + b) - (A\mu + b))^\top \right] \quad (10.79) \end{aligned}$$

$$= \sum_{i=1}^n (AL_{\bullet i})(AL_{\bullet i})^\top = A \left(\sum_{i=1}^n L_{\bullet i} L_{\bullet i}^\top \right) A^\top \quad (10.80)$$

$$= ALL^\top A^\top = A\Sigma A^\top = \text{Cov}(AX) \quad (10.81)$$

$$\text{N.R.: } (LL^\top)_{ij} = \sum_{k=1}^n L_{ik} L_{jk} = \left(\sum_{k=1}^n L_{\bullet k} L_{\bullet k}^\top \right)_{ij}$$

$$\begin{aligned} \Sigma_{fX} &= \frac{1}{2} \left[\sum_{i=1}^n (A(\mu + L_{\bullet i}) + b - (A\mu + b)) L_{\bullet i}^\top \right. \\ &\quad \left. + \sum_{i=1}^n (A(\mu - L_{\bullet i}) + b - (A\mu + b)) (-L_{\bullet i})^\top \right] \quad (10.82) \\ &= \sum_{i=1}^n AL_{\bullet i} L_{\bullet i}^\top = ALL^\top = A\Sigma = \text{Cov}(f(X), X) \quad \square \end{aligned}$$

Es stellt sich die Frage nach der Rolle des ersten Terms, also des Sigmapunktes μ , wo sich sein Beitrag bei linearen Funktionen wie oben gesehen herauskürzt. Im Endeffekt ist dieser Term der Unterschied zwischen Funktionswert des Mittelwertes und Mittelwert der Funktionswerte ins Quadrat erhoben. Im linearen Fall ist dies 0 und es wird tendenziell je größer je nichtlinearer die Funktion wird. Dadurch steigt Σ_f ohne, dass sich Σ_{fX} ändert. In der Kalman Filter Formel (10.85) ist dies äquivalent zu der Annahme eines höheren Rauschens, d. h. der Filter schätzt vorsichtiger. In gewissem Maße merkt der

Filter also selbsttätig, wie nichtlinear die Funktion ist und arbeitet diese Unsicherheit wie Rauschen in die Schätzung ein. Das ist eine sehr wünschenswerte Eigenschaft im Gegensatz zum EKF der fehlerhaft rechnet aber trotzdem der Meinung ist, er wüsste das Ergebnis genau.

10.7.2 Algorithmus

Dynamik($U_t = u_t$): Wende σ -Punkte-Propagation auf $g, \mu_{t-1}, \Sigma_{t-1}$ an

$$\bar{\mu}_t = \mu_g \quad (10.83)$$

$$\bar{\Sigma}_t = \Sigma_g + \Sigma_{\varepsilon t}, \quad (10.84)$$

Messung($Z_t = z_t$): Wende σ -Punkte-Propagation auf $h, \bar{\mu}_t, \bar{\Sigma}_t$ an

$$K = \Sigma_{hX}^\top (\Sigma_{hh} + \Sigma_{\delta t})^{-1} \quad (10.85)$$

$$\mu_t = \bar{\mu}_t + K(z_t - \mu_h) \quad (10.86)$$

$$\Sigma_t = \bar{\Sigma}_t - K \Sigma_{hX} \quad (10.87)$$

11 Transformationen in 3D und Homogene Koordinaten

Vorlesung
2007-01-24

Referenz [9].

Definition. Der projektive Raum ist $\mathbb{P}^n = (\mathbb{R}^{n+1} \setminus \{0\}) / (\mathbb{R} \setminus \{0\})$

Das heißt ein Raum von $(n + 1)$ -D Vektoren bei denen skalare Vielfache als dasselbe betrachtet werden.

$$[x] = \{\lambda \cdot x \mid \lambda \in \mathbb{R} \setminus \{0\}\}, \mathbb{P}^n = \{[x] \mid x \in \mathbb{R}^{n+1}\} \quad (11.1)$$

Betrachte $\mathbb{R}^n \subset \mathbb{P}^n$ vermöge $x \mapsto \begin{bmatrix} x \\ 1 \end{bmatrix}$.

Umgekehrt $\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{pmatrix} x_1/x_{n+1} \\ \vdots \\ x_n/x_{n+1} \end{pmatrix} x_{n+1} \neq 0$, da

$$\begin{pmatrix} x_1/x_{n+1} \\ \vdots \\ x_n/x_{n+1} \end{pmatrix} = \begin{bmatrix} x_1/x_{n+1} \\ \vdots \\ x_n/x_{n+1} \\ 1 \end{bmatrix} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \\ x_{n+1} \end{bmatrix}$$

Beispiel:

$$\begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} = \begin{bmatrix} 2 \\ 3 \\ 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \\ 8 \\ 2 \end{bmatrix} = \begin{bmatrix} -4 \\ -6 \\ -8 \\ -2 \end{bmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ 2 \\ 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 6 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ -2 \\ -3 \\ 0 \end{bmatrix} \neq \begin{pmatrix} x \\ y \\ z \end{pmatrix} \forall x, y, z \in \mathbb{R}$$

Das heißt $\mathbb{P}^n = \mathbb{R}^n \cup (\mathbb{P}^{n-1} \times \{0\})$. In Worten: Der projektive Raum der Dimension n \mathbb{P}^n enthält die gewöhnlichen „endlichen“ Vektoren der Dimension n , die man normalerweise als Punkte auffasst. Außerdem enthält er alle Richtungen des n -dimensionalen Raumes ins Unendliche, anschaulich den Horizont des \mathbb{R}^n .

11.1 Projektive Geometrie

Durch Formulierung in projektiver Geometrie lassen sich viele geometrische Aussagen mit weniger Sonderfällen hinschreiben. Besonders nützlich ist dies in der Photogrammetrie oder geometrischen Bildverarbeitung, wo man von Bildpositionen auf Raumpositionen schließen will [9].

Geraden und Punkte in 2D \mathbb{P}^2 . In der gewöhnlichen (nicht projektiven) ebenen Geometrie wird eine Gerade durch eine Gleichung der Form

$$ax + by + c = 0 \quad (11.2)$$

beschrieben. Dabei sind x, y die Koordinaten des Punktes und a, b, c die Parameter der Gerade. Verschiedene Parameter führen im Allgemeinen zu verschiedenen Geraden, aber ein lineares Skalieren der Parameter ändert die Gerade nicht. Daher liegt es nahe, den projektiven Vektor

$$g = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \in \mathbb{P}^2 \quad (11.3)$$

als die Gerade aufzufassen. Der Punkt kann mittels der üblichen Einbettung projektiv gesehen werden.

$$p = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \in \mathbb{P}^2 \quad (11.4)$$

Dann ist $g^\top p = 0$ die (11.2) entsprechende Gleichung in \mathbb{P}^2 .

11.2 Schnitt zweier Geraden

$p \in \mathbb{P}^2$ ist Schnittpunkt von $g_1 \in \mathbb{P}$ und $g_2 \in \mathbb{P}$, wenn $p^\top g_1 = p^\top g_2 = 0$ oder äquivalent $p = g_1 \times g_2$. Beweis: $g_1 \times g_2$ ist senkrecht auf g_1 und g_2 als gewöhnlicher \mathbb{R}^3 Vektor gesehen.

Bemerkenswert an dieser Aussage ist, dass alle Paare verschiedener Geraden einen eindeutigen Schnittpunkt haben. Auch parallele. Betrachten wir, was passiert in zwei Beispielen:

$$\begin{matrix} x=1 \\ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \end{matrix} \times \begin{matrix} y=1 \\ \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \end{matrix} = \begin{matrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \end{matrix}, \quad \begin{matrix} x=-1 \\ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \end{matrix} \times \begin{matrix} x=1 \\ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \end{matrix} = \begin{matrix} \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix} \end{matrix} \quad (11.5)$$

Im ersten Fall gibt es einen normalen endlichen Schnittpunkt $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$. Im zweiten Fall sind die Geraden parallel, es gibt keinen endlichen Schnittpunkt. Statt dessen einen unendlichen, nämlich die Richtung $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ die parallel zu beiden Geraden ist.

Diese Fähigkeit, parallele Geraden ohne Sonderfall zu behandeln vereinfacht in vielen Fällen erheblich die Programmierung. Dies ist ein großer Vorteil einer projektiven Formulierung.

Definition. Eine Kollinearität ist eine Abbildung $f : \mathbb{P}^n \rightarrow \mathbb{P}^m$ die Geraden erhält, $n, m \geq 2$.

Satz 15 (Fundamentalsatz der projektiven Geometrie). Jede Kollinearität $f : \mathbb{P}^n \rightarrow \mathbb{P}^{m+1}$ ist eine lineare Abbildung in den unterliegenden Vektorräumen. $f([x]) = [Ax]$

Dieser Satz ist ähnlich bemerkenswert wie die entsprechende Charakterisierung der linearen Abbildungen durch Matrizen. Die Menge der Abbildungen zwischen zwei Vektorräumen an sich ist nämlich etwas unendlich dimensionales, d. h. man braucht unendlich viele Zahlen um eine Abbildung eindeutig darzustellen. Damit lassen sich Abbildungen an sich nicht ohne Tricks, Approximationen oder Einschränkungen im Rechner

darstellen. Dieser Satz sagt nun, dass für projektive Abbildungen dies Problem nicht besteht. Jede projektive Abbildung im n -dimensionalen Raum lässt sich als $(n+1) \times (n+1)$ Matrix repräsentieren und damit im Rechner abspeichern.

Für das praktische Rechnen ist dies extrem nützlich und in der Tat unterliegt ein Datentyp für projektive \mathbb{P}^3 Abbildung als 4×4 Matrix den meisten 3D Computergrafik Paketen.

11.3 Euklidischen Koordinatentransformationen(EKT) bzw. Starrkörperbewegungen

Seine Stärken entfaltet die projektive Sichtweise vor allem im Bereich Computergrafik oder Bildverarbeitung. Die Zentralperspektive, also der Übergang Raum \mapsto Bild ist nämlich auch eine projektive Abbildung. Im Rahmen dieser Veranstaltung ist unser Ziel aber nur die Lage von Objekten im Raum zu beschreiben. Das führt auf eine wesentlich kleinere Klasse von projektiven Abbildung.

Definition. Eine EKT in 3D ist $f : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ mit $|f(x) - f(y)| = |x - y|$

Satz 16. Jede EKT f ist im projektiven Raum \mathbb{P}^3 eine lineare Abbildung der Form

$$f \left(\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} \right) = \left[\left(\begin{array}{c|c} Q & t \\ \hline 0 & 1 \end{array} \right) \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} \right], Q \in \mathbb{R}^{3 \times 3}, Q^T Q = I, t \in \mathbb{R}^3 \quad (11.6)$$

An der unteren Zeile 0, 0, 0, 1 sieht man, dass eine EKF die 4. Komponente des Vektors einfach durchreicht. Damit treten grundsätzlich nur 1 für Punkte oder 0 für Richtungen auf. Praktisch bedeutet dass, für eine Richtung mit $w = 0$ wird einfach $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ mit Q multipliziert. Für Richtungen ist die Abbildung linear in \mathbb{R}^3 . Für Punkte wird zusätzlich t addiert. Damit ist die Abbildung affin. D. h. bei EKT treten „projektive Effekte“ eigentlich gar nicht auf. Der Wert der Sichtweise als projektive Abbildung liegt darin, dass die ganze Abbildung durch eine Matrix dargestellt wird statt durch zwei Terme $Q \begin{pmatrix} x \\ y \\ z \end{pmatrix} + t$.

Dem entsprechend enthält $Q \in \mathbb{R}^{3 \times 3}$ als zugehörige Abbildung der Richtungen die Information über die Orientierung, während t die Information über die Position enthält.

11.4 Koordinatentransformationen, Lage von Körpern im Raum

Lage und Bewegung von Körpern werden über 3D-Koordinatensysteme dargestellt. Meist ein unbewegtes (weltfestes) Welt-System und mehrere sich mitbewegende (körperfeste) Körpersysteme. Die Abbildung, die einem Vektor in A-Koordinatensystem den geometrisch selben Vektor in B-Koordinaten zuordnet heißt T_B^A oder A2B und wird als 4×4 Matrix dargestellt.

Erinnern wir uns: Eine Matrix beschreibt eine lineare Abbildung und die Spalten einer Matrix sind die Bilder der Einheitsvektoren dieser Abbildung. Will man also die

Transformation `Styropor2Tisch` aufstellen, sind deren Spalten die X,Y,Z Achse bzw. der Ursprung des `Styropor` Koordinatensystems in `Tisch` Koordinaten. Wenn man sich dies überlegt, kann man die Matrix von Hand hinschreiben oder zumindest eine vom Programm berechnete Matrix überprüfen.

$$\text{Styropor2Tisch} = \begin{pmatrix} 0 & -1/\sqrt{2} & 1/\sqrt{2} & 1m \\ 0 & -1/\sqrt{2} & -1/\sqrt{2} & 1m \\ 1 & 0/\sqrt{2} & 0/\sqrt{2} & 0m \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (11.7)$$

Nach gleichem Schema ermittelt man: `Tisch2World`, `Udo2World`

Umrechnen von V_{Styropor} in V_{Udo} :

$$V_{\text{Udo}} = \underbrace{\text{Udo2World}^{-1}}_{\text{World2Udo}} \cdot \underbrace{\text{Tisch2World} \cdot \text{Styropor2Tisch} \cdot V_{\text{Styropor}}}_{V_{\text{World}}} \cdot \underbrace{V_{\text{Tisch}}}_{V_{\text{Tisch}}} \quad (11.8)$$

Beim Arbeiten mit Koordinatentransformation kann man sich sehr leicht vertun. Fehler entstehen oft durch die Reihenfolge beim Multiplizieren und die Frage Inverse oder nicht. Es scheint oft verlockend so lange an dem Term zu ändern, bis das Ergebnis richtig aussieht. Leider produzieren viele halbrichtige Terme Resultate, die nur bei aufmerksamer Beobachtung als falsch erkennbar sind. Außerdem ist dann vielleicht das Endergebnis richtig aber interne Größen sind falsch, d. h. abgespeicherte Transformation sind nicht das was dokumentiert wurde.

Es ist deshalb enorm wichtig, strikt auf die Richtigkeit aller Teile, Variablen, etc. zu achten. Dabei hilft enorm, alle Punkte mit einem Suffix zu versehen, auf welches Koordinatensystem sie sich beziehen und alle Transformationen **A2B** zu nennen, wenn sie einen Punkt in **A** Koordinaten auf den geometrisch selben Punkt in **B** Koordinaten abbilden. Beherrzt man dies, so ist an den Ausdrücken unmittelbar ablesbar ob sie konsistent sind, indem man die Koordinatensysteme von rechts nach links durchläuft (siehe Beispiel oben).

11.5 Rotationen um eine Achse

Die besondere Form der EKF ist die Rotation um eine vorgegebene Achse im Raum. Wir definieren diese Rotation axiomatisch über.

Definition. $\text{Rot}(v, \alpha, p)$ für $v \in \mathbb{R}^3$, $|v| = 1$, $\alpha \in \mathbb{R}$, $x \in \mathbb{R}^3$ (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

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 einen Satz aus der Theorie der Differentialgleichungen (DGL).

Bemerkung 11.1. *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 11.2. *Rot(v, \alpha, p) ist linear in p, d. h. es lässt sich durch eine Matrix Rot(v, \alpha) darstellen.*

$$\text{Rot}(v, \alpha, p) = \underbrace{[\text{Rot}(v, \alpha, e_1), \text{Rot}(v, \alpha, e_2), \text{Rot}(v, \alpha, e_3)]}_{\text{Rot}(v, \alpha)} \cdot p \quad (11.9)$$

Beweis. $\text{Rot}'(v, \alpha, p)$ wie in (11.9) und prüfe die beiden Rotationsaxiome.

$$1.: \quad \text{Rot}'(v, 0, p) = \text{Rot}(v, 0)p = [\text{Rot}(v, 0, e_1), \text{Rot}(v, 0, e_2), \text{Rot}(v, 0, e_3)]p \quad (11.10)$$

$$= [e_1, e_2, e_3]p = Ip = p \quad (11.11)$$

$$2.: \quad \frac{\partial \text{Rot}'(v, \alpha, p)}{\partial \alpha} = \frac{\partial (\text{Rot}(v, \alpha) \cdot p)}{\partial \alpha} = \sum_{i=1}^3 \frac{\partial \text{Rot}'(v, \alpha, e_i)}{\partial \alpha} \cdot p_i \quad (11.12)$$

$$= \sum_{i=1}^3 (v \times \text{Rot}(v, \alpha, e_i))p_i = v \times \sum_{i=1}^3 (\text{Rot}(v, \alpha)e_i p_i) \quad (11.13)$$

$$= v \times \text{Rot}(v, \alpha) \sum_{i=1}^3 (e_i p_i) = v \times \text{Rot}(v, \alpha) \cdot p \quad (11.14)$$

$$= v \times \text{Rot}(v, \alpha, p) \quad (11.15)$$

Damit folgt $\text{Rot}' = \text{Rot}$ und damit (11.9) aus dem Eindeutigkeitsatz für DGL. \square

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 heißt skalierte Achse und stellt eine Rotation durch einen einzelnen Vektor dar.

Definition (Skalierte Axis).

$$\text{Rot}(v) = \text{Rot}(v/|v|, |v|) \tag{11.16}$$

Drehung um v mit Winkel $|v|$

Das erste Problem stellt dar, dass man nicht mehr zwischen negiertem Winkel und negierter Achse unterscheiden kann. 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. Das folgende Lemma beweist, dass in der Tat $\text{Rot}(v)$ als Funktion von v gesehen sogar differenzierbar um 0 ist und gibt einen Ausdruck für die Ableitung an.

Lemma 11.3.

$$\text{Rot}(v) = \begin{pmatrix} 1 & -v_3 & v_2 \\ v_3 & 1 & -v_1 \\ -v_2 & v_1 & 1 \end{pmatrix} + O(|v|^2) \tag{11.17}$$

Beweis. Taylorentwicklung von $\text{Rot}(v, \alpha)$ nach α um $\alpha = 0$

$$\text{Rot}(v, \alpha)p \stackrel{\text{Taylor}}{=} \text{Rot}(v, 0)p + \alpha \left. \frac{\partial(\text{Rot}(v, \alpha)p)}{\partial \alpha} \right|_{\alpha=0} + O(\alpha^2) \tag{11.18}$$

$$\stackrel{\text{Def.}}{=} p + \alpha v \times p + O(\alpha^2) \tag{11.19}$$

$$\text{Rot}(v)p = p + v \times p + O(|v|^2) \tag{11.20}$$

$$= \begin{pmatrix} 1 & -v_3 & v_2 \\ v_3 & 1 & -v_1 \\ -v_2 & v_1 & 1 \end{pmatrix} p + O(|v|^2) \tag{11.21}$$

Dabei ergibt sich die letzte Gleichung aus der Tatsache, dass $p + v \times p$ linear in p ist. Die Spalten der Matrix erhält man durch Auswerten von $p + v \times p$ auf e_1, e_2, e_3 . \square

Lemma 11.4 (Rodriguez-Formel).

$$\text{Rot} \left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}, \alpha \right) = \begin{pmatrix} (1-c)x^2 + c & (1-c)xy - sz & (1-c)xz + sy \\ (1-c)xy + sz & (1-c)y^2 + c & (1-c)yz - sx \\ (1-c)xz - sy & (1-c)yz + sx & (1-c)z^2 + c \end{pmatrix}, \tag{11.22}$$

$$\text{mit } c = \cos \alpha, \quad s = \sin \alpha \tag{11.23}$$

Beweis. Übungszettel Aufgabe 15. \square

11 Transformationen in 3D und Homogene Koordinaten

Mit der Rodriguez-Formel können wir jede Rotation um eine Achse als Matrix schreiben. Im Folgenden wollen wir die Umkehrung nachweisen, nämlich, dass sich jede Koordinatentransformation in 3D als Rotation um eine Achse darstellen lässt. Hierzu definieren wir noch einmal formal die Menge (Gruppe) der Koordinatentransformationen von freien Vektoren (die Richtungen darstellen) $SO(3)$ und von Ortsvektoren $SE(3)$.

Definition.

$$SO(n) = \{Q \in \mathbb{R}^{n \times n} \mid Q^T Q = I, |Q| = 1\} \quad (11.24)$$

$$SE(n) = \left\{ \begin{pmatrix} Q & t \\ 0 & 1 \end{pmatrix} \mid Q \in SO(n), t \in \mathbb{R}^n \right\} \quad (11.25)$$

Die $Q^T Q = I$ Bedingung bedeutet, dass die Spalten $Q_{\bullet j}$ Länge 1 haben und aufeinander senkrecht stehen, wie man es für Koordinatentransformationen erwartet. Dadurch bleiben Längen und Winkel erhalten. Die zusätzliche Bedingung $|Q| = 1$ erzwingt, dass die Koordinatentransformationen die Händigkeit erhalten. Ansonsten könnten auch $|Q| = -1$ Transformationen wie Spiegelungen auftreten. Wir sehen also, eine Koordinatentransformation für freie 3D Vektoren lässt sich durch 9 Zahlen darstellen, die aber 6 Zwangsbedingungen unterliegen (Länge 1, Länge 2, Länge 3, 1 und 2 senkrecht, 2 und 3 senkrecht, 1 und 3 senkrecht).

Hat man $SO(n)$ durchschaut, ist $SE(n)$ die Menge (Gruppe) der Euklidischen Koordinatentransformationen einfach zu begreifen. Sie besteht im Wesentlichen aus einer $SO(3)$ Transformation für die freien Vektoren mit 0 als 4. Koordinate plus einer Spalte t die für Ortsvektoren mit 1 als 4. Koordinate eine Translation addiert.

Wir möchten im Folgenden die $SO(3)$ Transformationen charakterisieren und überprüfen deshalb zuerst, wie Rotation mit Koordinatentransformation vertauscht. Da Rotation ja ein geometrischer Vorgang ist, sollte man erwarten, dass es egal ist, ob man einen Vektor in Koordinatensystem A rotiert und das Ergebnis nach B transformiert, oder ob man Vektor und Achse nach B transformiert und dort rotiert. Dieses beweist das folgende Lemma.

Lemma 11.5. *Sei $Q \in SO(3)$, $v \in \mathbb{R}^3$, dann ist*

$$\text{Rot}(v) = Q^T \text{Rot}(Qv)Q \quad (11.26)$$

Im oben diskutierten Sinne der Koordinatentransformation heißt das:

$$\text{Rot}(v_A) = A_{2B}^T \text{Rot}(A_{2B} \cdot v_A)A_{2B}. \quad (11.27)$$

Beweis. Der Beweis erfolgt wieder durch nachprüfen der beiden Drehaxiome für die rechte Seite von (11.26). Setze $\text{Rot}'(v, \alpha, p) = Q^T \text{Rot}(Qv, \alpha)Qp$

$$1.: \quad \text{Rot}'(v, 0, p) = Q^T \text{Rot}(Qv, 0)Qp = Q^T I Q p = Q^T Q p = p \quad (11.28)$$

$$2.: \quad \frac{\partial \text{Rot}'(v, \alpha, x)}{\partial \alpha} = \frac{\partial(Q^\top \text{Rot}(Qv, \alpha) \cdot Qp)}{\partial \alpha} \quad (11.29)$$

$$= Q^\top \frac{\partial(\text{Rot}(Qv, \alpha) \cdot Qp)}{\partial \alpha} \quad (11.30)$$

$$= Q^\top ((Qv) \times (\text{Rot}(Qv, \alpha) \cdot Qp)) \quad (11.31)$$

$$\stackrel{Q(a \times b) = (Qa) \times (Qb)}{=} Q^\top Q(v \times (Q^\top \text{Rot}(Qv, \alpha) Qp)) \quad (11.32)$$

$$= v \times \text{Rot}'(v, \alpha)p \quad (11.33)$$

Wie in Lemma 11.2 folgt aus dem Eindeutigkeitsatz für DGL, dass $\text{Rot}' = \text{Rot}$ und damit (11.26). \square

Das folgende Lemma klassifiziert $SO(2)$ also die 2D Koordinatentransformationen mit dem Ergebnis, dass es nur Drehungen gibt.

Lemma 11.6. $Q \in SO(2)$, dann $Q = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}$ für ein α .

Den eigentlichen Kern der Klassifikation von $SO(3)$ importieren wir aus der linearen Algebra.

Satz 17 (Normalform für Orthonormale Matrizen). Sei $A \in SO(n)$, dann gibt es $Q \in SO(n)$, so dass $Q^\top A Q$ nur aus 1×1 und 2×2 Diagonalklöcken besteht.

Dieser Satz ist eine starke Einschränkung für orthonormale Matrizen. Die Matrix Q hat die Rolle einer Koordinatentransformation. Er sagt damit, dass man für eine orthonormale Matrix stets Koordinatensysteme wählen kann, so dass die Matrix unabhängig auf einzelnen Koordinaten oder 2-er Blöcken von Koordinaten wirkt.

Damit klassifizieren wir jetzt die Koordinatentransformation in 3D mit dem Ergebnis, dass jede eine Drehung um eine bestimmte Achse ist. ACHTUNG: Diese Aussage gilt nicht in 4D.

Satz 18. Jede Matrix $A \in SO(3)$ ist $A = \text{Rot}(v)$, $v \in \mathbb{R}^3$.

Beweis. Nach Satz 17 gibt es $Q \in SO(3)$, so dass

$$Q^\top A Q \stackrel{\text{Satz 17}}{=} \begin{pmatrix} 1 & 0 & 0 \\ 0 & * & * \\ 0 & * & * \end{pmatrix} \stackrel{\text{Lemma 11.6}}{=} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & -\sin \alpha \\ 0 & \sin \alpha & \cos \alpha \end{pmatrix} = \text{Rot}\left(\alpha \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\right) \quad (11.34)$$

$$\Rightarrow A = Q(Q^\top A Q)Q^\top = Q \text{Rot}\left(\alpha \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\right)Q^\top = \text{Rot}\left(Q \begin{pmatrix} \alpha \\ 0 \\ 0 \end{pmatrix}\right) \quad (11.35)$$

$$= \text{Rot}(\alpha Q \bullet_1) \quad \square$$

Damit können wir die Umkehrfunktion der Rotation, also sozusagen die Arcus-Rotation definieren. Wir haben damit gezeigt, dass sich jede $SO(3)$ Koordinatentransformation durch einen 3D Vektor der skalierten Achse darstellen kann. $SO(3)$ hat also, wie die Zahl der 6 Nebenbedingungen schon vermuten ließ, genau 3 Freiheitsgrade.

Definition. $\text{aRot}(Q) = v \Leftrightarrow \text{Rot}(v) = Q \wedge |v| \leq \pi$

11.6 UKF für beliebige Mannigfaltigkeiten

11.6.1 Schätzung einer Orientierung

Wir wollen uns dem Problem widmen mit Hilfe eines UKF die Orientierung eines Körpers zu schätzen. Das heißt, der zu schätzende Zustand X_t beinhaltet eine Orientierung. Es stellt sich die Frage, wie diese dargestellt werden kann.

Eine naheliegende Möglichkeit wäre aus der 3×3 Matrix $\in SO(3)$ einen 9er Vektor zu machen und einfach 9 Zahlen im Zustand zu haben. Die Zahlen unterliegen dann aber nichtlinearen Nebenbedingungen (Länge 1, senkrecht) von denen der UKF nichts weiß. D.h. wenn der UKF den Mittelwert nach einer Messung anpasst sind die Zahlen in der Regel keine Koordinatentransformation mehr. Das ist nicht überzeugend.

Eine zweite naheliegende Möglichkeit wäre [Satz 18](#) zu nutzen und die 3×3 Matrix durch 3 Parameter $v \in \mathbb{R}^3$ 3 Parameter als $\text{Rot}(v)$ darzustellen. Es gäbe keine Nebenbedingungen, da jedes v eine gültige Transformationen $\text{Rot}(v)$ darstellt. Leider gibt es doch ein Problem, dass sich tiefgründig aus der topologischen Struktur von $SO(3)$ ergibt. $SO(3)$ hat in jedem Punkt drei Freiheitsgrade, weil man immer eine kleine Drehung $\text{Rot}(\delta v)$ draufmultiplizieren kann. Damit der Filter funktioniert, muss man dies durch kleine Änderung der Parameter v realisieren können. D.h. $\text{Rot}(v)\text{Rot}(\delta v) = \text{Rot}(v')$ sein, wobei $v - v'$ gegen 0 geht, wenn δv gegen 0 geht. Also in Worten: Kleine Änderungen in der Orientierung müssen durch kleine Änderungen in den Parametern darstellbar sein. Dies ist nicht der Fall. Alle Vektoren v mit $|v| = 2\pi$ stellen 360° Drehungen um verschiedene Achsen, also alle die Einheitsmatrix dar. Damit haben für $|v| = 2\pi$ zwar Änderungen der Länge eine Wirkung aber Änderungen, die die Länge gleich lassen nicht. An dieser Stelle hat die Parametrisierung nur noch 1 und nicht 3 Freiheitsgrade.

In der Tat gibt es keine Parametrisierung von $SO(3)$ mit 3 Parametern ohne Singularitäten (wird noch bewiesen).

11.6.2 Darstellung von Mannigfaltigkeiten im UKF

Die Idee ist, solche Zustandsräume 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 die Operationen, die kleine Änderungen im Zustand als Vektor darstellen und kleine als Vektor dargestellte Änderungen auf einen Zustand draufrechnen. Alle generischen Formeln des UKF (Sigma-Punkt Propagation, Kalman Gain, Zustandsaktualisierung) stützen sich auf dieser vektoriellen Darstellung der Änderung relativ zum Mittelwert ab.

Formal ist die Zustandsmenge eine allgemeine Menge S mit den Operationen $\oplus : S \times \mathbb{R}^n \rightarrow S$, $\ominus : S \times S \rightarrow \mathbb{R}^n$, $X : \Omega \rightarrow S$, so dass $d \mapsto s \oplus d$ an jedem Zustand s eine Bijektion

einer Umgebung von s mit einer Umgebung von 0 ist. Nach dem Einbettungssatz von Whitney ist so eine Mannigfaltigkeit S stets isomorph zu einer Teilmenge des \mathbb{R}^p . D. h. die Zustände sind wie Vektoren durch endlich viele Zahlen beschreibbar für die allerdings zusätzliche Randbedingungen gelten. In unserem konkreten Beispiel von $S = SO(3)$ ist dies von vornherein der Fall. Eine 3×3 Matrix lässt sich natürlich auch als ein Vektor von 9 Zahlen betrachten. Die Matrix muss allerdings auch noch orthonormal sein und das definiert 6 Randbedingungen für die 9 Zahlen.

Zusammenfassend: Der Zustandsraum S ist durch reelle Zahlen mit Nebenbedingungen beschreibbar aber ansonsten für den UKF eine Blackbox. Zugriff erfolgt nur durch die Dynamik- und Messmodelle und \oplus, \ominus . Die Operation \oplus wendet auf einen Zustand $s \in S$ eine kleine parametrisierte Änderung d an. Umgekehrt ergibt $s_2 \ominus s_1$ die parametrisierte Änderung, die s_1 in s_2 überführt. Es ist also

$$s_1 \oplus (s_2 \ominus s_1) = s_2 \quad (11.36)$$

Bei Vektorräumen ist einfach $\oplus = +, \ominus = -$ 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 \oplus, \ominus definiert werden.

Eine einfache Parametrisierung besteht darin, die relative Orientierung als skalierte Achse darzustellen.

Definition. *Parametrisierung von $S = SO(n), n = 3$ durch Relativrotationen in lokalen Koordinaten.*

$$s \oplus d = s \cdot \text{Rot}(d) \quad (11.37)$$

$$s_2 \ominus s_1 = \text{aRot}(s_1^{-1} s_2) \quad (11.38)$$

Bei dieser Definition wird die Achse in Körperkoordinaten dargestellt. Es ist daher schwer sich die Zahlen z. B. in Kovarianzmatrizen vorzustellen. Einfacher für den Menschen verständlich wird es, wenn man die Achse der relativen Orientierung in Weltkoordinaten darstellt. Dies führt zu folgender Definition.

Definition. *Parametrisierung von $S = SO(n), n = 3$ durch Relativrotationen in Weltkoordinaten.*

$$s \oplus d = s \cdot \text{Rot}(s^{-1} d) \quad (11.39)$$

$$s_2 \ominus s_1 = s_1 \cdot \text{aRot}(s_1^{-1} s_2) \quad (11.40)$$

11.6.3 Wahrscheinlichkeitsrechnung auf Mannigfaltigkeiten

Wir generalisieren jetzt unsere probabilistischen Standardbegriffe – Erwartungswert, Kovarianz und Gaußverteilung – mit Hilfe von \oplus und \ominus auf allgemeine Mannigfaltigkeiten. Es muss bemerkt werden, dass diese Begriffe leider nicht mehr die mathematische Strenge besitzen, wie ihre Vektorraum Pendanten. 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.

$$\mathbb{E}(X) : \mathbb{E}(X \ominus \mathbb{E}(X)) = 0 \quad (11.41)$$

$$\text{Cov}(X, Y) = \mathbb{E}((X \ominus \mathbb{E}(X))(Y \ominus \mathbb{E}(Y))^\top) \quad (11.42)$$

$$\mathcal{N}(\mu, \Sigma) = \mu \oplus \mathcal{N}(0, \Sigma) \quad (11.43)$$

Die Definitionen der S -Kovarianz (11.42) ist einfach. Sie ist eh definiert über die Differenz zum Mittelwert ins äußere Produkt erhoben. Diese Differenz wird einfach über \ominus statt $-$ errechnet. Schwieriger ist die Definition des S -Erwartungswertes, weil S i. A. keine Möglichkeit bieten, Elemente „aufzusummieren“. Statt dessen können wir aber die \ominus Differenz zu einem festen Referenzelement nehmen. Diese ist eine normaler Vektor und darüber können wir wie üblich einen Erwartungswert bilden. Das Ergebnis sollte die \ominus Differenz des S -Erwartungswertes zum Referenzelement sein. Insbesondere sollte das Ergebnis 0 sein, wenn man als Referenzelement den Erwartungswert selbst nimmt. Diese Überlegung führt zur Definition des S -Erwartungswertes durch die implizite Gleichung (11.41).

Da mit (11.41) $\mathbb{E}(X)$ nicht explizit definiert ist, stellt sich die Frage, wie man ihn berechnet. Das folgende Lemma liefert einen Hinweis.

Lemma 11.7 (Ausrechnen des Erwartungswerts). *Rechnet man iterativ:*

$$\mu_{i+1} = \mu_i \oplus \mathbb{E}(X \ominus \mu) \quad (11.44)$$

So gilt, falls μ_i konvergiert:

$$\lim \mu_i = \mathbb{E}(X) \quad (11.45)$$

Mit diesen Begriffen definieren wir jetzt die Annahmen die wir über unser zu schätzendes System in S machen. Die Abhängigkeit von der konkreten Zustandsmangfaltigkeit versteckt sich in den Dynamik- und Messfunktionen g und h , die ja eh problemabhängig sind. Ansonsten wird nur die Ungenauigkeit durch \oplus Addition von Rauschen modelliert.

11.6.4 Annahmen

1. $X_0 \sim \mathcal{N}(\mu_0, \Sigma_0)$, $\mu_0 \in S$, $\Sigma_0 \in \mathbb{R}^{n \times n}$
2. $X_{t+1} = g(X_t, U_t) \oplus \varepsilon_t$, $\varepsilon_t \sim \mathcal{N}(0, \Sigma_{\varepsilon t})$
3. $Z_t = h(X_t) + \delta_t$, $\delta_t \sim \mathcal{N}(0, \Sigma_{\delta t})$
4. ε_t, δ_t unabhängig.

Definition (Mannigfaltigkeits-Sigma-Propagation). *Sei $\mu \in S$, $\Sigma \in \mathbb{R}^{n \times n}$, $f : S \rightarrow T$ eine Abbildung zwischen zwei Mannigfaltigkeiten, die mit entsprechenden Operationen \oplus_S , \ominus_S , \oplus_T , \ominus_T ausgestattet sind. Sei ferner $LL^\top = \Sigma$ die Cholesky-Zerlegung von Σ .*

$$\mu_{f0} = f(\mu) \quad (11.46)$$

$$\mu_{f,i+1} = \mu_{fi} \oplus_T \frac{1}{2n+1} \left[(f(\mu) \ominus_T \mu_{fi}) + \right. \quad (11.47)$$

$$\left. \sum_{j=1}^n (f(\mu \oplus_S L_{\bullet j}) \ominus_T \mu_{fi}) + \right. \quad (11.48)$$

$$\left. \sum_{j=1}^n (f(\mu \oplus_S -L_{\bullet j}) \ominus_T \mu_{fi}) \right]$$

$$\mu_f = \lim_{i \rightarrow \infty} \mu_{fi} \quad (11.49)$$

$$\Sigma_f = \frac{1}{2} \left[(f(\mu) \ominus_T \mu_f)(f(\mu) \ominus_T \mu_f)^\top + \right. \quad (11.50)$$

$$\left. \sum_{j=1}^n (f(\mu \oplus_S L_{\bullet j}) \ominus_T \mu_f)(f(\mu \oplus_S L_{\bullet j}) \ominus_T \mu_f)^\top + \right.$$

$$\left. \sum_{j=1}^n (f(\mu \oplus_S -L_{\bullet j}) \ominus_T \mu_f)(f(\mu \oplus_S -L_{\bullet j}) \ominus_T \mu_f)^\top \right]$$

$$\Sigma_{fx} = \frac{1}{2} \left[\sum_{j=1}^n (f(\mu \oplus_S L_{\bullet j}) \ominus_T \mu_f)(L_{\bullet j})^\top + \sum_{j=1}^n (f(\mu \oplus_S -L_{\bullet j}) \ominus_T \mu_f)(-L_{\bullet j})^\top \right] \quad (11.51)$$

Im Vergleich zu der Vektorraum-Sigma-Propagation gibt es drei wesentliche Unterschiede.

- Die Sigma-Punkte, also $\mu, \mu \oplus_S L_{\bullet j}, \mu \oplus_S -L_{\bullet j}$ werden durch aufaddieren eines „Sigmas“ mit \oplus statt $+$ gewonnen.
- Für die Kovarianz werden die Differenzen zwischen dem Funktionswert der Sigma-Punkte und dem Mittelwert der Funktionswertswerte benötigt. Diese werden mit \ominus an Stelle von $-$ berechnet.
- Die Formel für μ_f ist deutlich anders. Da man in T nicht summieren kann, werden alle Funktionswerte von Sigmapunkten mit \ominus_S relativ zu einem Referenzwert parametrisiert, über diesen wird der Mittelwert gebildet und das Ergebnis wieder mit \oplus_S auf den Referenzwert draufgeschlagen. Das Ergebnis hängt (leicht) vom Referenzwert ab, deshalb wird es bis zur Konvergenz iteriert.

Vom UKF wird die Sigma-Punkt Propagation einmal im Dynamikschritt mit g und einmal im Messschritt mit f benutzt. Im Dynamikschritt ist daher $S = T$. Im Messschritt ist T der Messraum, ein Vektorraum, so dass $\oplus_T = +$ und $\ominus_T = -$.

Ansonsten ändert sich der UKF nur darin, dass die Mannigfaltigkeits-Sigma-Punkt-Propagation benutzt wird und bei der Aktualisierung von μ_t im Messschritt die Änderung mit \oplus auf den Zustand addiert wird.

11.6.5 Algorithmus

Dynamik($U_t = u_t$):

Wende Mannigfaltigkeits- σ -Punkte-Propagation auf $g, \mu_{t-1}, \Sigma_{t-1}$ an, $T = S$

$$\bar{\mu}_t = \mu_g \quad (11.52)$$

$$\bar{\Sigma}_t = \Sigma_g + \Sigma_{\varepsilon t}, \quad (11.53)$$

Messung($Z_t = z_t$):

Wende Mannigfaltigkeits- σ -Punkte-Propagation auf $h, \bar{\mu}_t, \bar{\Sigma}_t$ an, $\oplus_T = +, \ominus_T = -$

$$K = \Sigma_{hX}^\top (\Sigma_{hh} + \Sigma_{\delta t})^{-1} \quad (11.54)$$

$$\mu_t = \bar{\mu}_t \oplus K(z_t - \mu_h) \quad (11.55)$$

$$\Sigma_t = \bar{\Sigma}_t - K \Sigma_{hX} \quad (11.56)$$

12 Starrkörperbewegungen

Vorlesung
2007-02-07

Fakt 1. Jede längenerhaltende Transformation f des \mathbb{R}^3 lässt sich projektiv (d. h. in \mathbb{P}^3) durch eine Matrix der Form $\begin{pmatrix} Q & t \\ 0 & 1 \end{pmatrix}$ darstellen mit $t \in \mathbb{R}^3$, $Q \in O(3)$.

Zusatz: $Q \in SO(3)$, falls f orientierungstreu

Beweisskizze.

1. f ist auch winkeltreu (f bewahrt $\langle -, - \rangle$)

[Denn: $\|a + b\|^2 - \|a - b\|^2 = 4 \langle a, b \rangle$, siehe Wikipedia]

2. f ist geradentreu

[Denn nach Cauchy-Schwarz: $\langle a, b \rangle^2 \leq \|a\|^2 \|b\|^2$, „ $=$ “ gdw. a, b linear abhängig.

a, b, c kollinear $\Leftrightarrow \|a - b\| + \|c - b\| = \|a - c\|$]

3. Nach Hauptsatz der projektiven Geometrie existiert $A \in \mathbb{R}^{4 \times 4}$, so dass:

$f(v) = A \begin{pmatrix} v \\ 1 \end{pmatrix}$ in Projektion

Sei $A = \begin{pmatrix} Q & t \\ a & b \end{pmatrix}$

rechte Spalte: Bild von $\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$, also „endlich“ also $b \neq 0$, also auf 1 normierbar.

Annahme: $a \neq 0$ Dann existiert ein endlicher Vektor $\begin{pmatrix} v \\ 1 \end{pmatrix}$ mit $A \begin{pmatrix} v \\ 1 \end{pmatrix} = \begin{pmatrix} \vdots \\ 0 \end{pmatrix}$ unendlich.

Widerspruch.

Also ist $f(v) = A \begin{pmatrix} v \\ 1 \end{pmatrix} = Qv + t$, also $Q \in O(3)$ □

Fakt 2. Es existiert keine singularitätenfreie „Parametrisierung“ $f : \mathbb{R}^3 \rightarrow SO(3)$

Was heißt das?

singularitätenfrei = stetig ($SO(3) \hookrightarrow \mathbb{R}^{3 \times 3}$, mit Topologie aus $\mathbb{R}^{3 \times 3}$:

$$f : _ \rightarrow SO(3) \text{ stetig} \Leftrightarrow f : _ \rightarrow SO(3) \hookrightarrow \mathbb{R}^{3 \times 3} \text{ stetig}$$

f soll „lokal“ \mathbb{R}^3 in $SO(3)$ deformieren.

Formal: suchen Überlagerung

das heißt: zu „kleiner“ Umgebung U von $A \in SO(3)$ besteht das Urbild $f^{-1}U$ aus disjunkten „Blättern“, die wie U aussehen

Beispiel: \mathbb{R} überlagert $SO(2)$ (siehe [Abbildung 12.1](#))

Behauptung. Es existiert keine Überlagerung $\mathbb{R}^3 \rightarrow SO(3)$

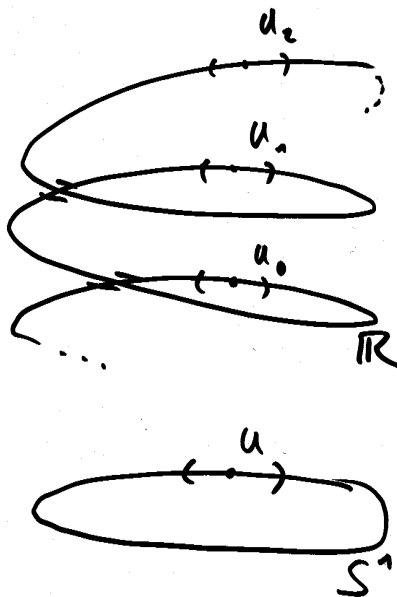


Abbildung 12.1: Die reellen Zahlen \mathbb{R} bilden eine Überlagerung von S^1 . Die entsprechende stetige Abbildung ergibt sich im Bild durch „herunterprojektion“ der Spirale \mathbb{R} auf den darunter liegenden Kreis S^1 . Jede Umdrehung des Kreises entspricht einer Schraubenwindung in der Spirale. Da Urbild der offenen Umgebung eines Punktes $u \in S^1$ sind damit unendlich viele u_i in \mathbb{R} die jeweils um eine Schraubenwindung versetzt sind. Da \mathbb{R} einfach zusammenhängend ist, ist dies eine universelle Überlagerung.

12.1 (Einfacher) Zusammenhang

Definition (Pfad). Sei X ein (topologischer) Raum, $x, y \in X$. Ein Pfad σ von x nach y ist eine stetige Abbildung $\sigma : [0, 1] \rightarrow X$ mit $\sigma(0) = x$, $\sigma(1) = y$.

Definition (homotop). σ_1, σ_2 homotop $\Leftrightarrow \exists h : [0, 1] \times [0, 1] \rightarrow X$, $h(0, -) = \sigma_1$, $h(1, -) = \sigma_2$, $h(-, 0) = x$, $h(-, 1) = y$

Definition. Ein Nullpfad $x_0 \rightarrow x_0$ ist eine Abbildung $\sigma_0(t) = x_0$

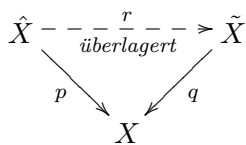
Eine Schleife bei $x_0 =$ ist ein Pfad $\sigma_{x_0} \rightarrow x_0$, σ ist zusammenziehbar, gdw σ, σ_0 homotop.

Definition. Ein Raum X ist zusammenhängend $\Leftrightarrow \forall x, y \in X \exists$ Pfad $x \rightarrow y$.

X ist einfach zusammenhängend(1-zshd) \Leftrightarrow Jede Schleife in X ist zusammenziehbar.

Definition. Eine Überlagerung $p : \tilde{X} \rightarrow X$ heißt universell, wenn \tilde{X} 1-zshd (Abb. 12.1).

Satz 19. Für X hinreichend zshd. überlagert eine universelle Überlagerung $\hat{X} \xrightarrow{p} X$ alle Überlagerungen q von X .



Lemma 12.1. Pfade liften eindeutig vom Basisraum in den überlagernden Raum bei Festlegung des Anfangspunkts.

Beweis. vgl [Abbildung 12.1](#) □

Rotation um φ in \mathbb{R}^2 entspricht komplexer Multiplikation mit $e^{i\varphi}$.

Darstellung von Rotationen im \mathbb{R}^3 durch Einheitsquaternionen $a + bi + ck + dk$, $\|(a, b, c, d)\| = 1 \hat{=} S^3$ (siehe Wikipedia). $S^3 \rightarrow SO(3)$ Überlagerung. S^3 ist 1-zshd.

Beweis von Fakt 2. (Widerspruchsbeweis)

Angenommen es ex. Überlagerung $\mathbb{R}^3 \rightarrow SO(3)$, dann ex. Überlagerung $S^3 \rightarrow \mathbb{R}^3$.

S^3 ist kompakt (beschränkt und abgeschlossen). Stetige Bilder kompakter Mengen sind kompakt. Damit wäre \mathbb{R}^3 kompakt. Widerspruch. □

Literaturverzeichnis

- [1] U. Frese and L. Schröder. Theorie der Sensorfusion: Merkzettel Mathematische Grundlagen. <http://www.informatik.uni-bremen.de/~ufrese/teaching/tds06/tdsmerkblatt.pdf>, 2006.
- [2] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics*. MIT Press, 2005.
- [3] Hafner. *Wahrscheinlichkeitsrechnung und Statistik*. Springer Verlag, 1989.
- [4] D. Kehlmann. *Die Vermessung der Welt*. Rowohlt, Reinbek, 2005.
- [5] S. Thrun, W. Burgard, and D. Fox. Probabilistic Robotics – Web Site. <http://www.probablistic-robotics.org>, 2005.
- [6] S. Thrun, W. Burgard, and D. Fox. Probabilistic Robotics – Kalman Filter Slides. <http://robots.stanford.edu/probablistic-robotics/ppt/kalman.ppt>.
- [7] U. Krengel. *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Vieweg Verlagsgesellschaft, 2005.
- [8] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes, Second Edition*. Cambridge University Press, Cambridge, 1992.
- [9] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004.