

Bayes-Filter für dynamische Koordinatenmessungen

Elmar Garcia und Tino Hausotte

Lehrstuhl für Fertigungsmesstechnik, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), 91052 Erlangen, Deutschland, elmar.garcia@fau.de, tino.hausotte@fau.de

Kurzfassung

Das Theorem von Thomas Bayes ist einer der meistgenutzten Ansätze zur Lösung probabilistischer Inferenzprobleme beliebiger, dynamischer Systeme und bildet außerdem die Grundlage der Messunsicherheitsermittlung nach GUM. Für die Schätzung stochastischer Größen wurden in der langjährigen Forschung auf dem Gebiet der bayesschen Estimationsverfahren zahlreiche Algorithmen entwickelt und erfolgreich umgesetzt. Zu den bekanntesten Algorithmen zählen hierbei die Familie der Kalman-Filter (KF), die gitterbasierten Filterverfahren und jüngst zunehmend die Klasse der Partikel-Filter (PF). Über die letzten 15 Jahre, wurden die PF zunehmend Gegenstand von Forschung und ingenieurstechnischen Anwendungen, wie die Verarbeitung dynamischer Koordinatenmessungen, die Rekonstruktion und Prädiktion von Signalen aus verrauschten Messdaten sowie die Ermittlung von Messunsicherheiten.

Der Erfolg der PF liegt in ihrer Fähigkeit nicht lineare und/oder nicht gaußsche Systeme zu behandeln sowie in ihrer programmiertechnisch sehr einfachen Umsetzbarkeit. PF sind Realisierungen stichprobenbasierter sequentieller Monte-Carlo-Methoden, welche eine Menge gewichteter Stichproben zur rekursiven Berechnung von bayesschen a posteriori Wahrscheinlichkeitsdichtefunktion (WDF) verwenden.

In dieser Publikation werden Bayes-Filter zur Genauigkeitsverbesserung und Unsicherheitsermittlung von dynamischen Koordinatenmessungen erarbeitet. Dazu werden zunächst die Grundlagen der bayesschen Estimations- und Filtertheorie präsentiert. Ausgehend vom Bayes-Theorem wird das theoretische Fundament der KF und PF hergeleitet und analysiert. Am Beispiel der dynamischen Koordinatenmesstechnik werden hierbei sowohl allgemeine als auch spezifische Defizite und Anwendungsprobleme diskutiert.

1 Bayes-Filter

Bayes-Filter sind Umsetzungen der bayesschen Formel zur Berechnung der Wahrscheinlichkeit eines unbekanntes Ereignisses oder einer Hypothese H bei Beobachtung oder Evidenz eines weiteren Ereignisses E

$$p(H|E) = \frac{p(E|H) \cdot p(H)}{p(E)} \quad (1)$$

mit $p(H|E)$ – a posteriori WDF,
 $p(H)$ – a priori WDF,
 $p(E|H)$ – Likelihood-Funktion,
 $p(E)$ – Normierungskonstante.

Dieses, die bayessche Statistik begründende, Theorem ist eine fundamentale wahrscheinlichkeitstheoretische Regel für den Rückschluss von der Beobachtung E auf ein unbekanntes Ereignis H [1]. Mit Hilfe der a posteriori WDF können Wahrscheinlichkeitsaussagen über H gemacht werden, welche auf einer tatsächlich beobachteten Realisierung von E basieren. Sie beschreibt, was aus den beobachteten Daten E über die a priori Einschätzung von

H hinzugelernt werden konnte. Diese Möglichkeit ist nur in der bayesschen Statistik gegeben, da die Wahrscheinlichkeitsaussagen der klassischen Statistik stets auf dem Zufallscharakter von E beruhen und sich damit nicht auf eine tatsächlich realisierte Beobachtung, sondern zwangsläufig auf alle denkbaren Realisierungen von E beziehen. Sie erweitert somit den Wahrscheinlichkeitsbegriff von der klassischen, frequentistischen Interpretation als relative Häufigkeit hin zu einem Grad des Dafürhaltens (engl. degree of belief).

Die rechte Seite der Gleichung (1) beschreibt die Transformation des Wissensstandes über H vor der Messung, verkörpert durch die a priori WDF $p(H)$, in den Wissensstand nach der Messung von E , dargestellt durch die a posteriori WDF $p(H|E)$. Der Nenner $p(E)$ ist unabhängig von H und eine Normierungskonstante, so dass sich die a posteriori WDF zu Eins integriert. Die Likelihood-Funktion beschreibt die Wahrscheinlichkeit der Vereinbarkeit der Messung E mit der Hypothese H .

Im Folgenden wird die Anwendung des Bayes-Theorems für die Schätzung einer Messgröße aus abweichungsbehafteten Messungen erörtert. Der hierbei zu Grunde liegende Messprozess wird allgemeingültig als nichtlineares dynamisches System mit zeitlich veränderlichen Zuständen behandelt. Es gilt nun die bayessche a posteriori Dichte nicht mehr nur für eine statische Zufallsvariable mit bekannter Statistik zu berechnen, sondern zeitlich fortlaufend bestmöglich den Zustand eines dynamischen Systems aus verrauschten Messungen durch sequentielle Lösung des Bayes-Theorems zu ermitteln. Die Grundlage bildet hierbei eine stochastische Modellierung des Systems, welche sowohl die mit Unsicherheiten behaftete zeitliche Entwicklung des Systemzustandes als auch die Unsicherheit der Beziehung von Zustand und Messwerten erfasst. Diese fortlaufende Extraktion von Informationen über eine Größe zum Zeitpunkt t unter Einbeziehung aller Messdaten bis einschließlich t wird als Filterung bezeichnet. Die Umsetzungen dieser Datenverarbeitung im Kontext der bayesschen Inferenz zur dynamischen Zustandsschätzung ergibt die Bayes-Filter.

2 Bayessche Inferenz in dynamischen Systemen

Jedes dynamische System kann als Zustandsraummodell beschrieben werden, welches aus zwei Differentialgleichungen zur Beschreibung der Prozessdynamik und der beobachtbaren Messungen besteht [2]. Da die zu erarbeitenden Verfahren auf Digitalrechnern umgesetzt werden sollen, werden im Folgenden ausschließlich zeitdiskrete Zustandsraummodelle betrachtet. Ohne Beschränkung der Allgemeinheit werden der Einfachheit halber keine exogenen Einflussgrößen berücksichtigt¹. Das zeitdiskrete Zustandsraummodell ist gegeben durch

$$\begin{aligned} \mathbf{x}_k &= \mathbf{f}_{k-1}(\mathbf{x}_{k-1}, \mathbf{v}_{k-1}) \\ \mathbf{y}_k &= \mathbf{h}_k(\mathbf{x}_k, \mathbf{w}_k). \end{aligned} \quad (2)$$

Die Entwicklung des Zustandes wird zu diskreten, nicht notwendigerweise äquidistanten Zeitpunkten t_0, t_1, t_2, \dots betrachtet. Der Zustandsvektor $\mathbf{x}_k := \mathbf{x}(t_k)$, $\mathbf{x}_k \in \mathbb{R}^{n_x}$ mit Dimension n_x entwickelt sich zeitlich entsprechend der nichtlinearen Prozessgleichung \mathbf{f}_{k-1} aus dem letzten Zustand \mathbf{x}_{k-1} und dem Prozessrauschen \mathbf{v}_{k-1} . Das Prozessrauschen beschreibt die unvollständige Kenntnis der tatsächlichen Prozessdynamik, die selbst bei gegebenem Anfangszustand nicht als exakte Beschreibung der tatsächlichen Dynamik angenommen werden kann. Die Beobachtungen $\mathbf{y}_k \in \mathbb{R}^{n_y}$ mit Dimension n_y sind durch einen Messprozess generierte Transformationen der Zustände gemäß der nichtlinearen Messgleichung \mathbf{h}_k und dem Messrauschen \mathbf{w}_k . Der durch die Messung bedingte Abbildungsfehler wird durch das Messrauschen dargestellt.

Der Systemzustand \mathbf{x}_k ist zumeist nicht direkt beobachtbar und kann nur über die verrauschten Messungen \mathbf{y}_k rekonstruiert werden. Er wird daher auch als verdeckter oder verborgener Zustand bezeichnet (engl. hidden state). Der Zustandsvektor hängt ausschließlich vom letzten Zustand ab und repräsentiert die minimal notwendige Information über Gegenwart und Vergangenheit des Systems, die für die Beschreibung des zukünftigen Systemverhaltens notwendig ist [3]. Durch diesen Markov-Charakter² bilden die stochastischen Differentialgleichungen des Zustandsraummodells (2) zwei Markov-Prozesse eines Hidden Markov Models (HMM). Die Prozessgleichung entspricht einer Markov-Kette, welche durch eine initiale Zustandsverteilung $p(\mathbf{x}_0)$ und die Zustandstransitionswahrscheinlichkeit $p(\mathbf{x}_k | \mathbf{x}_{k-1})$ mit $k \geq 1$ gekennzeichnet ist. Die Beobachtungsgleichung stellt die zweite Markov-Kette dar, welche zu jedem Zeitpunkt beobachtbare Ausgangsgrößen gemäß der zustandsabhängigen

¹In den nachfolgend betrachteten Filterproblemen sind eventuelle exogene Größen \mathbf{u}_k nicht bekannt und werden daher hier nicht explizit betrachtet. Sämtliche Ausführungen können direkt auf dynamische Systeme mit exogenen Eingangsgrößen übertragen werden.

²Formal ausgedrückt ist ein Markov-Prozess ein stochastischer Prozess $\{\mathbf{x}_k, k \in \mathbb{Z}\}$ für den gilt $p(\mathbf{x}_{k+1} | \mathbf{x}_k, \mathbf{x}_{k-1}, \mathbf{x}_{k-2}, \dots, \mathbf{x}_0) = p(\mathbf{x}_{k+1} | \mathbf{x}_k)$ [1].

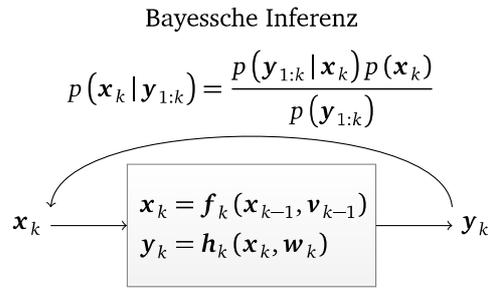


Abbildung 1: Darstellung des bayesschen Inferenzproblems für nichtlineare dynamische Systeme/Prozesse. Das System wird als zeitdiskretes, stochastisches Zustandsraummodell beschrieben.

Beobachtungswahrscheinlichkeit $p(\mathbf{y}_k | \mathbf{x}_k)$ mit $k \geq 0$ generiert. Damit definiert das stochastische Zustandsraummodell mit dem Initialzustand $p(\mathbf{x}_0)$ ein HMM erster Ordnung. Zustandsraummodell und HMM sind äquivalente Darstellungen eines dynamischen Systems und beschreiben probabilistisch wie sich das dynamische System zeitlich entwickelt und wie ungenau die Evolution der verborgenen Zustände beobachtet werden kann.

Aus bayesscher Perspektive ist die vollständige Lösung dieses Schätzproblems durch die bedingte a posteriori WDF des Zustandes $p(\mathbf{x}_k | \mathbf{y}_{1:k})$ gegeben, welche alle Messungen $\mathbf{y}_{1:k} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_k\}$ bis zum aktuellen Zeitpunkt k umfasst (siehe Abbildung 1). Da die a posteriori WDF alle vom Prozess verfügbaren Informationen beinhaltet, ermöglicht ihre Kenntnis die Berechnung optimaler Zustandsschätzungen für beliebige Optimalitätskriterien. So ist die MMSE-Schätzung der bedingte Erwartungswert $\hat{\mathbf{x}}_k = \mathbb{E}[\mathbf{x}_k | \mathbf{y}_{1:k}] = \int \mathbf{x}_k p(\mathbf{x}_k | \mathbf{y}_{1:k}) d\mathbf{x}_k$ oder die MAP-Schätzung $\hat{\mathbf{x}}_k = \arg \max p(\mathbf{x}_k | \mathbf{y}_{1:k})$ [4]. Schätzungen für weitere Parameter, wie Median, Moden, Konfidenzintervalle etc., können ebenso aus der a posteriori Dichte abgeleitet werden, was eine GUM-gemäße Bestimmung der Messunsicherheit ermöglicht.

Die Berechnung der bayesschen a posteriori Zustandsdichte $p(\mathbf{x}_k | \mathbf{y}_{1:k})$ erfordert in jedem Zeitschritt k die Speicherung und Verarbeitung aller zurückliegenden Messungen $\mathbf{y}_{1:k}$ (batch processing), was eine Echtzeitverarbeitung verhindert. Das Ziel ist daher die optimale Schätzung des verborgenen Zustandes durch sequentielle Verarbeitung fortlaufend eintreffender Messdaten, was einer iterativen Aktualisierung der vorherigen Zustandsschätzung entspricht und als rekursive bayessche Schätzung oder Filterung bezeichnet wird.

Die rekursive bayessche Schätzung erlaubt eine effiziente und echtzeitfähige Messdatenverarbeitung mit geringem Speicherbedarf und schneller Signalanpassung. Für zeitdiskrete Systeme kann sie aus dem Bayes-Theorem (1) wie nachstehend angegeben hergeleitet werden.

$$\begin{aligned}
 p(\mathbf{x}_k | \mathbf{y}_{1:k}) &= \frac{p(\mathbf{y}_{1:k} | \mathbf{x}_k) p(\mathbf{x}_k)}{p(\mathbf{y}_{1:k})} \\
 &= \frac{p(\mathbf{y}_k, \mathbf{y}_{1:k-1} | \mathbf{x}_k) p(\mathbf{x}_k)}{p(\mathbf{y}_k, \mathbf{y}_{1:k-1})} \\
 &= \frac{p(\mathbf{y}_k | \mathbf{y}_{1:k-1}, \mathbf{x}_k) p(\mathbf{y}_{1:k-1} | \mathbf{x}_k) p(\mathbf{x}_k)}{p(\mathbf{y}_k | \mathbf{y}_{1:k-1}) p(\mathbf{y}_{1:k-1})} \\
 &= \frac{p(\mathbf{y}_k | \mathbf{y}_{1:k-1}, \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) p(\mathbf{y}_{1:k-1}) p(\mathbf{x}_k)}{p(\mathbf{y}_k | \mathbf{y}_{1:k-1}) p(\mathbf{y}_{1:k-1}) p(\mathbf{x}_k)} \\
 &= \frac{p(\mathbf{y}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{y}_{1:k-1})}{p(\mathbf{y}_k | \mathbf{y}_{1:k-1})}
 \end{aligned} \tag{3}$$

Die rekursive Berechnung der bayesschen a posteriori WDF $p(\mathbf{x}_k | \mathbf{y}_{1:k})$ in Gleichung (3) erfolgt nun für jeden Zeitpunkt k in zwei Schritten: einem Prädiktionsschritt $p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}) \rightarrow p(\mathbf{x}_k | \mathbf{y}_{1:k-1})$ und einem Korrekturschritt $p(\mathbf{x}_k | \mathbf{y}_{1:k-1}, \mathbf{y}_k) \rightarrow p(\mathbf{x}_k | \mathbf{y}_{1:k})$. Im Prädiktionsschritt wird zunächst unter Verwendung der letzten a posteriori Zustandsdichte $p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1})$ und der Prozessgleichung f die aktuelle Zustandsdichte $p(\mathbf{x}_k | \mathbf{y}_{1:k-1})$ vorhergesagt.

Diese a priori Dichte des Zustandes \mathbf{x}_k zum Zeitpunkt k berechnet sich entsprechend der Chapman–Kolmogorov–Gleichung [1]

$$p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_k | \mathbf{x}_{k-1}) p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}) d\mathbf{x}_{k-1}, \quad (4)$$

mit der Zustandstransitionsdichte

$$p(\mathbf{x}_k | \mathbf{x}_{k-1}) = \int \delta(\mathbf{x}_k - \mathbf{f}_k(\mathbf{x}_{k-1}, \mathbf{v}_k)) p(\mathbf{v}_k) d\mathbf{v}_k \quad (5)$$

und berücksichtigt nicht die aktuelle Messung \mathbf{y}_k sondern basiert nur auf zurückliegenden Messungen. Da die Zustandsevolution allerdings unbekanntem Störungen unterliegt (modelliert durch das Prozessrauschen) und darüber hinaus zumeist imperfekt modelliert ist, erfolgt hierbei eine verfälschte Vorhersage. In dem anschließenden Korrekturschritt wird daher die prognostizierte a priori Zustandsdichte (4) entsprechend der Likelihood–Dichte der aktuellen Messung

$$p(\mathbf{y}_k | \mathbf{x}_k) = \int \delta(\mathbf{y}_k - \mathbf{h}_k(\mathbf{x}_k, \mathbf{w}_k)) p(\mathbf{w}_k) d\mathbf{w}_k \quad (6)$$

korrigiert³ und letztendlich mit der Normalisierungskonstante

$$p(\mathbf{y}_k | \mathbf{y}_{1:k-1}) = \int p(\mathbf{y}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) d\mathbf{x}_k \quad (7)$$

die aktuelle a posteriori Dichte $p(\mathbf{x}_k | \mathbf{y}_{1:k})$ des Zustandes entsprechend dem Bayes–Theorem (3) berechnet.

3 Kalman–Filter

Für den Fall eines linearen, zeitdiskreten Systems

$$\mathbf{x}_k = \mathbf{F}_{k-1} \mathbf{x}_{k-1} + \mathbf{v}_{k-1} \quad (8)$$

mit der Systemmatrix \mathbf{F}_{k-1} und dem normalverteilten Prozessrauschen \mathbf{v}_{k-1} und der linearen Messgleichung

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{w}_k \quad (9)$$

mit der Beobachtungsmatrix \mathbf{H}_k und dem vom Prozessrauschen \mathbf{v}_k unabhängigen, normalverteilten Rauschen \mathbf{w}_k lässt sich auf analytischem Weg eine optimale Lösung des bayesschen Filterungsproblems finden. Diese Lösung ist das bereits 1960 von R.E. Kalman vorgestellte Kalman–Filter [5].

Für ein lineares System mit normalverteiltem weißem Rauschen ist die Wahrscheinlichkeitsdichte des Systemzustands zu jeder Zeit normalverteilt. Sie kann daher mit den beiden Parametern Mittelwert \mathbf{m} und Kovarianzmatrix \mathbf{P} vollständig beschrieben werden.

Ausgehend von den Gleichungen (3) und (4) gilt der rekursive Zusammenhang

$$\begin{aligned} p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}) &= \mathcal{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1}) \\ p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) &= \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}) \\ p(\mathbf{x}_k | \mathbf{y}_{1:k}) &= \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}). \end{aligned} \quad (10)$$

$\mathcal{N}(x; \mu, \sigma^2)$ bzw. $\mathcal{N}(\mathbf{x}; \mathbf{m}, \mathbf{P})$ beschreibt eine Normalverteilung, die im eindimensionalen Fall als

$$\mathcal{N}(x; \mu, \sigma^2) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2}\right) \quad (11)$$

³Zumeist wird die prädierte Dichte $p(\mathbf{x}_k | \mathbf{y}_{1:k-1})$ im Zustandsraum konzentriert und verschoben, wodurch insgesamt eine fortlaufende Nachbesserung der zuvor prädierten Dichte erreicht wird.

und im mehrdimensionalen Fall als

$$\mathcal{N}(\mathbf{x}; \mathbf{m}, \mathbf{P}) := |\mathbf{P}|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m})^T \mathbf{P}^{-1}(\mathbf{x} - \mathbf{m})\right) \quad (12)$$

definiert ist, mit dem Argument \mathbf{x} , dem Mittelwert μ bzw. \mathbf{m} und der Varianz σ^2 bzw. der Kovarianzmatrix \mathbf{P} .

Das Kalman-Filter berechnet rekursiv den Mittelwert und die Kovarianz der normalverteilten a posteriori Dichte $p(\mathbf{x}_k | \mathbf{y}_{1:k})$ anhand der nachstehenden Gleichungen⁴.

Prädiktion

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}_{k-1} \hat{\mathbf{x}}_{k-1|k-1} \quad a \text{ priori Schätzung} \quad (13)$$

$$\mathbf{P}_{k|k-1} = \mathbf{F}_{k-1} \mathbf{P}_{k-1|k-1} \mathbf{F}_{k-1}^T + \mathbf{Q}_{k-1} \quad a \text{ priori Kovarianzmatrix} \quad (14)$$

Korrektur

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \quad \text{Kalman-Verstärkung} \quad (15)$$

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k (\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}) \quad a \text{ posteriori Schätzung} \quad (16)$$

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1} \quad a \text{ posteriori Kovarianzmatrix} \quad (17)$$

Die Matrizen \mathbf{Q} und \mathbf{R} sind dabei die Kovarianzmatrizen des mittelwertfreien normalverteilten System- bzw. Messrauschens. Der Vektor $\hat{\mathbf{x}}$ ist der Mittelwert des Systemzustands und \mathbf{P} die dazugehörige a posteriori Kovarianzmatrix. Sie ist ein Maß für die Varianz der Schätzung des Kalman-Filter. Mit \mathbf{K} wird die Kalman-Verstärkung bezeichnet.

Bei Einhaltung der restriktiven Voraussetzungen stellt das Kalman-Filter die optimale Lösung der bayesschen Schätzung dar. Für lineare Systeme mit additivem weißen gaußschen Rauschen existiert kein besserer Algorithmus zur Zustandsschätzung als das Kalman-Filter.

4 Partikel-Filter

Die Gleichungen (4), (5), (6) und (7) formulieren die optimale bayessche Lösung zur rekursiven Schätzung der Zustände dynamischer Systeme aus abweichungsbehafteten Beobachtungen. Jedoch ist diese bayessche Inferenz nur eine konzeptionelle Lösung, da die Integralfunktionen im Allgemeinen nicht analytisch gelöst werden können. Des Weiteren würde die algorithmische Implementierung der bayesschen Schätzung eine Speicherung der a posteriori WDF als unendlich dimensionale Matrix erfordern, da die bei nichtlinearen Systemen entstehenden a posteriori Dichten im Allgemeinen nicht vollständig durch endlichdimensionale suffiziente Statistiken (wie z. B. Erwartungswert und Varianz im Fall der Normalverteilung) beschrieben werden können.

Im Allgemeinen ist die Prozess- und/oder Messgleichung jedoch nichtlinear mit nicht gaußscher Charakteristik. Dadurch können die Integrale der bayesschen Rekursionsgleichungen nicht exakt gelöst werden und müssen mittels approximativer Lösungsverfahren berechnet werden. Von allen hierfür entwickelten Verfahren, haben sich für praktische Anwendungen die PF als sehr geeignet erwiesen. Der Erfolg der PF liegt in ihrer Fähigkeit nicht lineare und/oder nicht gaußsche Systeme zu behandeln sowie in ihrer programmierlich sehr einfachen Umsetzbarkeit. PF sind Realisierungen stichprobenbasierter sequentieller Monte-Carlo-Methoden, welche eine Menge gewichteter Stichproben zur rekursiven Berechnung von bayesschen a posteriori WDF verwenden.

Sequentielle Monte-Carlo-Methoden sind stochastische stichprobenbasierte Methoden, welche die Monte-Carlo-Integration mit sequentiellen Stichprobenverfahren nutzen, um die hochdimensionalen Integrale der bayesschen Rekursionsgleichungen zu lösen [6]. Diese Methoden erfordern keine explizite Annahme über die Gestalt der a posteriori Dichte und approximieren die Integrale durch endliche Summen diskreter, gewichteter Stichproben (auch Sample oder Partikel genannt). Dadurch wird es möglich, auch Zustände mit multimodalen oder sehr schiefen Verteilungen zu schätzen, die sich bei nichtlinearen, nicht gaußschen Systemen ergeben. Ein weiterer wesentlicher Vorteil dieser Methoden ist die weitgehende Unabhängigkeit von der Dimension der zu integrierenden

⁴Die Filtergleichungen berücksichtigen keine deterministischen Eingangsgrößen \mathbf{u}_k , da diese in reinen Tracking-Anwendungen nicht vorliegen. Für gesteuerte Systeme d. h. $\mathbf{x}_k = \mathbf{F}_{k-1} \mathbf{x}_{k-1} + \mathbf{B}_{k-1} \mathbf{u}_{k-1} + \mathbf{v}_{k-1}$ ändert sich die Zustandsprädiktion des Kalman-Filter zu $\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}_{k-1} \hat{\mathbf{x}}_{k-1|k-1} + \mathbf{B}_{k-1} \mathbf{u}_{k-1}$.

Funktion. Sie unterliegen nicht dem sogenannten »Fluch der Dimensionalität«⁵. Im Gegensatz zu klassischen Integrationsverfahren, wie der Trapezregel oder der Simpsonregel, wächst bei SMCM der Berechnungsaufwand nicht exponentiell mit der Anzahl der Dimensionen des Integrationsproblems. Aus Platzgründen wird für die mathematische Herleitung der PF auf die Fachliteratur verwiesen [6, 8, 9].

5 Dynamische Koordinatenmessungen

In den vorherigen Abschnitten wurde dargelegt, dass alle Bayes-Filter (BF) und damit auch PF sowie PF modellbasierte algorithmische Umsetzungen der bayesschen rekursiven Schätzung sind, welche durch die Gleichungen (4), (5) und (6) gegeben ist. Die essentielle Voraussetzung und der Schlüssel zu ihrer erfolgreichen Anwendung ist daher die adäquate Modellierung des betrachteten dynamischen Systems sowie die effiziente Extraktion der Informationen über den Systemzustand aus abweichungsbehafteten Messungen. Zunächst werden mathematischen Modelle zur Beschreibung derartiger Messaufgaben präsentiert.

In der Koordinatenmesstechnik wird unter dynamischen Koordinatenmessungen die zeitlich kontinuierliche Erfassung der Koordinaten bewegter oder unbewegter Objekte verstanden. Dies kann entweder mit dynamischen Koordinatenmessgeräten, wie Lasertrackern oder photogrammetrischen Messsystemen, oder mit taktil oder optisch scannenden Sensoren erfolgen. Typische Messaufgaben sind Bewegungsmessungen (d. h. zurückgelegte Raumbahnen und höhere kinematische Momente), Trajektorienfolgereglungen von mechatronischen Antriebssystemen, Kalibrierungsmessungen von Fertigungsmaschinen oder Robotern sowie die dimensionelle Analyse von Bauteilen, Strukturen oder Oberflächen.

Derartige Messaufgaben können sensorunabhängig und allgemeingültig als Trackingproblem eines Massepunktes⁶ formuliert werden, welcher sich zeitlich in beliebigen Koordinaten bewegt. Die Dynamik des Massepunktes entspricht dann einer Trajektorie, welche entweder die zurückgelegte Raumbahn oder das Gestaltprofil, im Fall der Erfassung von Oberflächenpunkten eines Objektes, beschreibt. In der Koordinatenmesstechnik können prinzipiell zwei Messaufgaben unterschieden werden. Die Messung von a priori bekannten oder erwarteten Trajektorien und die Messung von unbekanntem Trajektorien. Der erste Fall tritt stets bei der Erfassung von Objekten oder Bewegungen auf, welche durch idealgeometrische Ersatzelemente (z. B. Standardgeometrien in der Fertigungsmesstechnik), Splines (z. B. Freiformkurven industrieller Produkte) oder Lookup-Tabellen (z. B. programmierte Robotertrajektorien) beschrieben werden können und deren Ausprägung bereits vor der eigentlichen Messung bekannt ist. Für die häufige Aufgabe der Messung von idealgeometrischen Bahnen sind in Tabelle 1 geeignete Dynamikmodelle aufgeführt. Die hierin fehlenden Regelgeometrien Zylinder, Ebene und Torus können aus den dargestellten Modellen zusammengesetzt werden. So kann die Erfassung zylindrischer Geometrien auf die Antastung entlang kreisförmiger Bahnen in verschiedenen Ebenen oder einer Helix zurückgeführt werden.

Das in der Praxis dynamischer Koordinatenmessungen, wie dem optischen Tracking oder dem Abtasten (Scanning) von Freiformflächen, relevantere und eigentliche Problem ist die Beschreibung von freien Bewegungen und deren Verfolgung. In diesen Fällen können, in Ermangelung vorab verfügbarer Informationen über die zu erwartende Dynamik oder Trajektorie des Messobjektes, nur allgemeingültige Polynommodelle für die Beschreibung der Dynamik verwendet werden. Die hierbei gebräuchlichsten Ansätze zur Formulierung von Prozessfunktionen sind in Tabelle 6 zusammengefasst.

Auf die Modellierung der Messgleichung wird nicht weiter explizit eingegangen, da diese a priori aus Kalibrierungen oder experimentellen Analysen des Messsystems ermittelt werden kann. Zumeist handelt es sich hierbei um einfache Abbildungsfunktionen zwischen Mess- und Zustandsraum. Bei Koordinatenmessungen ergibt sich für gewöhnlich eine direkte Abbildung oder eine einfache Koordinatentransformation zwischen Inertial- oder Sensorkoordinatensystem und dem körperfesten Koordinatensystem des Massepunktes (z. B. Kugelkoordinaten bei richtungs- und entfernungsmessenden Sensoren).

⁵Entgegen der häufig veröffentlichten Meinung sind sie nicht vollständig frei von diesem Effekt. Allerdings sind die sequentielle Monte-Carlo-Methoden (SMCM) weitaus weniger betroffen als die klassischen numerischen Integrationsverfahren, wodurch dieser Effekt für Dimensionen $n_x \leq 100$ vernachlässigt werden kann[7].

⁶In den folgenden Ausführungen wird primär von Einkoordinatenmessungen ausgegangen, da diese in der dynamischen Koordinatenmesstechnik überwiegen. Die Verallgemeinerung der hier betrachteten Modelle auf Mehrkoordinatenmessungen ist ohne Weiteres möglich.

Tabelle 1: Beschreibung idealgeometrischer Trajektorien im euklidischen Raum mit dem Zustandsvektor $\mathbf{x}_k = [x, y, z]^T$

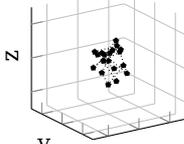
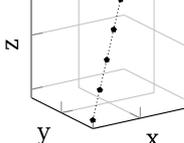
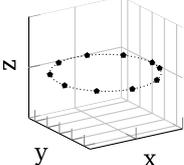
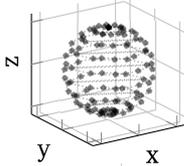
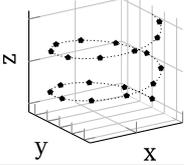
Trajektorie	Visualisierung	Dynamikmodell
Punkt		$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{v}_k$
Gerade		$\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda \cdot \begin{bmatrix} e_x \\ e_y \\ e_z \end{bmatrix} + \mathbf{v}_k$
Kreis (Polarkoordinaten)		$\mathbf{x}_k = \begin{bmatrix} x_k \\ y_k \end{bmatrix} = \begin{bmatrix} r_k \cos(\varphi_k) \\ r_k \sin(\varphi_k) \end{bmatrix} + \mathbf{v}_k$
Kugel (Kugelkoordinaten)		$\mathbf{x}_k = \begin{bmatrix} r_k \cos(\theta_k) \cos(\varphi_k) \\ r_k \cos(\theta_k) \sin(\varphi_k) \\ r_k \sin(\theta_k) \end{bmatrix} + \mathbf{v}_k$
Helix (Zylinderkoordinaten)		$\mathbf{x}_k = \begin{bmatrix} r_k \cos(\varphi_k) \\ r_k \sin(\varphi_k) \\ z_k \end{bmatrix} + \mathbf{v}_k$

Tabelle 2: Beschreibung beliebiger nichtlinearer Trajektorien im euklidischen Raum

Bezeichnung	Dynamikmodell
Polynommodell	$\begin{bmatrix} x(t) \\ y(t) \\ z(t) \end{bmatrix} = \begin{bmatrix} a_0 & \dots & a_n \\ b_0 & \dots & b_n \\ c_0 & \dots & c_n \end{bmatrix} \begin{bmatrix} 1 \\ t \\ \vdots \\ \frac{t^n}{n!} \end{bmatrix} + \begin{bmatrix} v_x(t) \\ v_y(t) \\ v_z(t) \end{bmatrix}$
Kinematikmodelle	<p>CWNA $\begin{bmatrix} \dot{s}(t) \\ \dot{v}(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} s(t) \\ v(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \cdot v(t)$</p> <p>CWPA $\begin{bmatrix} \dot{s}(t) \\ \dot{v}(t) \\ \dot{a}(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} s(t) \\ v(t) \\ a(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \cdot v(t)$</p>
Zeitreihenmodelle	ARMA(m, n) $x_t = \varepsilon_t + \sum_{i=1}^m a_i x_{t-i} + \sum_{j=1}^n b_j \varepsilon_{t-j}$

6 Bayes–Filter für dynamische Koordinatenmessungen

Bayes–Filter können auf Grundlage der in Tabelle dargestellten Kinematikmodelle für Massepunkte direkt zur kontinuierlichen Messung (Verfolgung) der Koordinaten bewegter Objekte verwendet werden. Bei adäquater Modellparametrisierung können abweichungsverursachende Einflüsse des Messprozesses kompensiert werden und eine bessere Bestimmung der tatsächlichen Objektkoordinaten erzielt werden. Neben dieser Verbesserung der Messgenauigkeit kann gleichzeitig erfolgreich eine Ermittlung der Messunsicherheit erfolgen. Dies soll im Folgenden für eine bewusst einfach gehaltene Messaufgabe demonstriert werden. Als Anwendung hierfür werden dynamische, d. h. kontinuierlich scannende, Koordinatenmessungen eines Einstellringes mit einem Tastsystem der Fertigungsmesstechnik betrachtet.

Der Taster ist vom Hersteller gemäß ISO 10360 mit einer Antastunsicherheit $1,5\ \mu\text{m}$ für scannende Messungen spezifiziert. Alle Messungen wurden auf einem Multisensor–Koordinatenmessgerät mit maximal zulässiger Längenmessabweichung von $E2 = (0.5 + L/600)\ \mu\text{m}$, mit der Messlänge L in mm bei konstanter Temperatur $\vartheta = 20\ \text{°C} \pm 0,1\ \text{K}$ durchgeführt. Das Werkstückkoordinatensystem wurde durch iteratives Einmessen so bestimmt, dass der Nullpunkt und die Messebene dem Mittelpunkt bzw. der x – y –Ebene des Einstellringes entsprechen. Anschließend wurde mit einer Scangeschwindigkeit $v \approx 80\ \mu\text{m/s}$ die Kreiskontur des Einstellringes im mathematisch positiven Drehsinn erfasst. Die Unsicherheit der Verfahrgeschwindigkeit wurde experimentell mit $2\ \mu\text{m/s}$ ermittelt. Für die Beschreibung des Messprozesses wurde das nachfolgende Zustandsraummodell verwendet (vgl. Abschnitt 5).

$$\mathbf{x}_k = \begin{bmatrix} x_k \\ y_k \end{bmatrix} = \begin{bmatrix} r \cos(\varphi_k) \\ r \sin(\varphi_k) \end{bmatrix} + \mathbf{v}_k \quad (18)$$

$$\mathbf{y}_k = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{x}_k + \mathbf{w}_k \quad (19)$$

Der Drehwinkel φ_k wird im Filter in jedem Zeitschritt aus der Scangeschwindigkeit berechnet. Das Messrauschen in jeder Koordinate \mathbf{w}_k wurde entsprechend der spezifizierten Messunsicherheit des Koordinatenmessgeräts, der Antastunsicherheit des Tasters, der Unsicherheit der Verfahrgeschwindigkeit sowie umgebungsbedingter Abweichungsursachen als gleichverteiltes Rauschen $\mathcal{U}(0; 2,9\ \mu\text{m})$ parametrisiert. Das Prozessrauschen wurde unter Berücksichtigung der Kalibrierunsicherheit, der Unsicherheit des Koordinatenursprungs, der Rundheitsabweichung und der Rauheit des Einstellringes als normalverteiltes Rauschen mit der Standardabweichung $\mathcal{N}(0; 0,5\ \mu\text{m})$ definiert.

Die Messunsicherheit ist gemäß GUM als Standardabweichung oder als Überdeckungswahrscheinlichkeit anzugeben [10]. Für die Ermittlung der hierfür erforderlichen Messabweichungen wird eine inverse Filterung mit Adaption des Prozessrauschens umgesetzt. Mit inverser Filterung ist hierbei der Umstand gemeint, dass nicht von der Beobachtung auf den Zustand sondern vom Zustand auf die Beobachtung geschlossen wird. Aufgrund der sehr genauen a priori Kenntnis der Prozessdynamik wird die Zustandsvorhersage des nächsten Antastpunktes nach Gleichung 18 als Referenz für den nachgelagerten Korrekturschritt verwendet. Dadurch wird die aktuelle Koordinatenmessung mit dem berechneten Zustand verglichen und gewichtet. Es erfolgt somit keine beobachtungs-basierte sondern eine prozessmodellbasierte Abweichungsbestimmung, woraus sich die Notwendigkeit einer genauen a priori Kenntnis der Prozessdynamik ergibt. Für die Schätzung der Standardabweichung wird zusätzlich eine dynamische Anpassung des Prozessrauschens in Abhängigkeit der Abweichung zwischen Schätz- und Messwert im letzten Zeitschritt umgesetzt. In den folgenden Untersuchungen wird ein PF mit 200 Partikeln verwendet. Die Anzahl der Partikel kann aufgrund der hohen Güte der Prozessmodellierung und des hier gegebenen Prozessrauschens deutlich niedriger als sonst üblich gewählt werden.

Für die Messunsicherheitsermittlung dynamischer Koordinatenmessungen ist die sequentielle Bestimmung der Unsicherheit jeder Einzelantastung erforderlich. Diese Unsicherheitsaussage kann nun sehr einfach aus der bayesschen a posteriori Dichteschätzung des PF abgeleitet werden. In Abbildung 2 ist beispielhaft für einen Zeitschritt die PF-basierte Approximation und Berechnung der Kovarianzmatrix visualisiert. Weitere stochastische Momente, wie Schiefe und Wölbung, können ebenfalls abgeleitet werden. Es ist zu beachten, dass die bayessche Inferenz nicht immer zu einer GUM– bzw. GS1–identischen Messunsicherheit führt. Sowohl im GS1 als auch bei Bayes–Filtern wird die Messunsicherheit als Unkenntnis über den tatsächlichen Wert einer Messgröße aufgefasst und diese durch WDF beschrieben. Jedoch wird im GS1 die WDF einer Messgröße im Allgemeinen nicht aus einer bayesschen Inferenz ermittelt. Die bayessche Schätzung basiert auf einem Beobachtungsmodell während im GS1 ein Messmodell der Auswertung zugrunde gelegt wird. Dadurch können sich in Abhängigkeit der Nichtlinearität des Modells und der

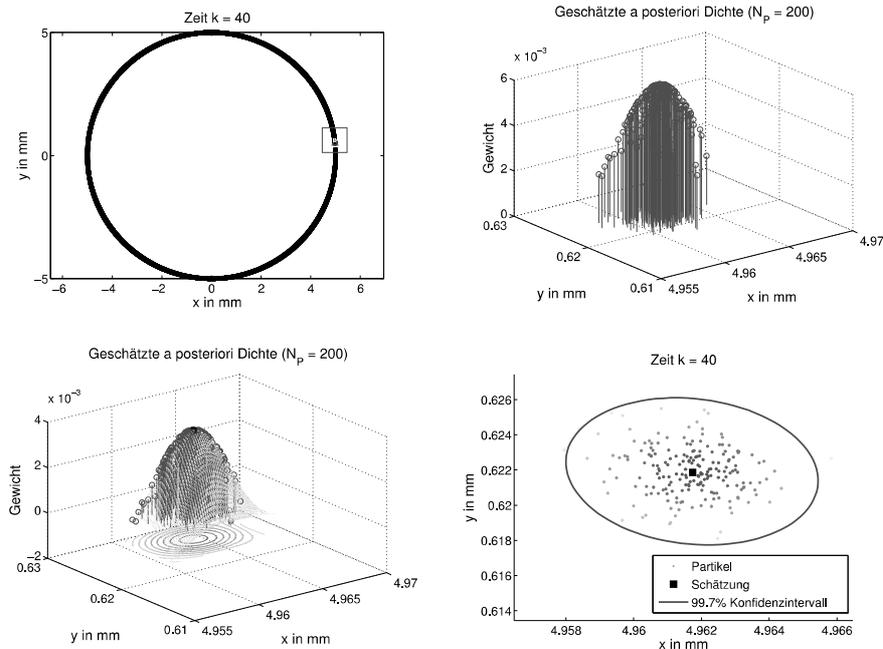


Abbildung 2: Darstellung der PF-basierten Schätzung der a posteriori Dichte und des 99,7%-Konfidenzintervalls für scannende Tastermessungen der Kreiskontur eines Einstellringes im Zeitschritt $k = 40$.

verwendeten a priori WDF die ermittelten Messunsicherheiten beider Ansätze unterscheiden [11]. Da allerdings weder GUM noch GS1 dynamische Messungen berücksichtigen, ist die bayessche Inferenz die einzige Möglichkeit einer GS1-gemäßen Messunsicherheitsermittlung für derartige Messaufgaben.

Abschließend ist für das bewusst einfach gewählte Beispiel anzumerken, dass aufgrund der präzisen Dynamikkenntnis und der geringen Schrittweite durch Linearisierung des Zustandsraummodell aus Gleichung 18 und Anwendung des KF eine annähernd identische Schätzgenauigkeit wie das PF erzielt werden kann. Eine Zunahme der Abweichungen ist erst bei größerer Nichtlinearität der Prozess- und/oder Messgleichung sowie bei nicht gaußschem Rauschen zu erwarten. Auch die prinzipbedingte Verwendung von normalverteiltem Messrauschen mit der zuvor spezifizierten Standardabweichung zeigt keinen signifikanten Einfluss auf das Schätzergebnis. Prinzipiell sind sowohl KF als auch PF geeignet, die Messunsicherheit dynamischer Koordinatenmessungen zu schätzen und auch zu reduzieren. Der Vorteil der PF liegt vornehmlich in ihrer Fähigkeit, auch nicht gaußsche Verteilungen erfassen und abbilden zu können. Dies sollte aufgabenspezifisch genutzt und Antastabweichungen entsprechend modelliert werden.

7 Zusammenfassung und Ausblick

In diesem Beitrag wurde eine anwendungsorientierte Einführung in das Gebiet der Bayes-Filter zur Genauigkeitsverbesserung und Unsicherheitsermittlung von dynamischen Koordinatenmessungen gegeben. Dazu wurden zunächst die Grundlagen der bayesschen Estimations- und Filtertheorie dieser modellbasierten Filterverfahren präsentiert. Anschließend erfolgte die Erarbeitung von Modellen für den erfolgreichen Einsatz der Bayes-Filter im Bereich der Koordinatenmesstechnik. Abschließend wurden die Leistungsfähigkeit der Bayes-Filter zur Genauigkeitsverbesserung und Unsicherheitsermittlung anhand einfacher dynamischer Koordinatenmessaufgaben demonstriert.

In zukünftigen Arbeiten müssen stochastische Eigenschaften der Bayes-Filter intensiver untersucht werden. Daneben sind differenzierte Systemmodelle für weitere Messaufgaben sowie programmiertechnische Implementierungen zu erarbeiten.

Danksagung

Die Autoren danken der Deutschen Forschungsgemeinschaft für die Förderung der Arbeiten im Rahmen des Forschungsprojektes WE 918/34-1 »Datenfusion multisensoriell ausgeführter Koordinatenmessungen«.

Literatur

- [1] I. N. Bronstein u. a. *Taschenbuch der Mathematik*. 8., überarb. Aufl. Deutsch (Harri), Aug. 2012.
- [2] R. Unbehauen. *Systemtheorie 1: Allgemeine Grundlagen, Signale und lineare Systeme im Zeit- und Frequenzbereich. Mit 148 Aufgaben*. Oldenbourg Verlag, 2002.
- [3] R. Schlittgen und B. H. J. Streitberg. *Zeitreihenanalyse*. München, Wien: Oldenbourg Verlag, 2001.
- [4] B. Ristic u. a. *Beyond the Kalman filter: particle filters for tracking applications*. Artech House, Feb. 2004.
- [5] R. E. Kalman. “A New Approach to Linear Filtering and Prediction Problems”. In: *Transactions of the ASME–Journal of Basic Engineering* 82.Series D (1960), S. 35–45.
- [6] A. Doucet u. a. *Sequential Monte Carlo Methods in Practice*. Hrsg. von A. Doucet u. a. Statistics for engineering and information science. New York, NY: Springer, 2001.
- [7] F. Daum und J. Huang. “Curse of dimensionality and particle filters”. In: *Proceedings IEEE Aerospace Conference, 2003*. Bd. 4. März 2003, S. 1979–1993.
- [8] M. S. Arulampalam u. a. “A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking”. In: *IEEE Transactions on Signal Processing* 50 (2002), S. 174–188. URL: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.117.1144>.
- [9] B. D. O. Anderson und J. B. Moore. *Optimal Filtering*. Dover Publications, Jan. 2005.
- [10] ISO/IEC Guide 98-3:2008. *Uncertainty of measurement – Part 3: Guide to the expression of uncertainty in measurement*. ISO, Genf, Schweiz, 2008. URL: http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf.
- [11] C. Elster und B. Toman. “Bayesian uncertainty analysis for a regression model versus application of GUM Supplement 1 to the least-squares estimate”. In: *Metrologia* 48.5 (Okt. 2011), S. 233. URL: <http://iopscience.iop.org/0026-1394/48/5/001>.