



AGH

AKADEMIA GÓRNICZO-HUTNICZA IM. STANISŁAWA STASZICA W KRAKOWIE

Wydział Elektrotechniki, Automatyki, Informatyki i Inżynierii Biomedycznej

Praca dyplomowa

Gaussowskie aproksymacje filtrów optymalnych

Gaussian approximations of optimal filters

Autor:

Jakub Kłosiński

Kierunek studiów:

Automatyka i Robotyka

Opiekun pracy:

dr hab. inż. Piotr Bania

Kraków, 2022

Serdecznie dziękuję...

Streszczenie

W pracy zaprezentowano ...

Słowa kluczowe:

Abstract

The project presents ...

Keywords:

Spis treści

1. Wprowadzenie	9
2. Algorytmy filtracji	11
2.1. Filtr Kalmana	13
2.2. Rozszerzony filtr Kalmana	15
2.3. Filtr Kalmana Gaussa-Hermite'a	17
3. Eksperymenty numeryczne	23
3.1. Systemy teoretyczne	23
3.2. Problemy praktyczne	26
3.2.1. Śledzenie pocisku balistycznego	26
3.2.2. Dwuwymiarowe śledzenie ruchu obiektu	28
3.2.3. Śledzenie wyłącznie z wykorzystaniem namiaru	29
4. Eksperyment z wykorzystaniem wahadła reakcyjnego	33
5. Podsumowanie	35

1. Wprowadzenie

Wstęp w [1], zastosowania filtracji w [2]

2. Algorytmy filtracji

Termin *filtracja optymalna* odnosi się do zestawu metod, które mogą być używane do estymacji stanu systemu dynamicznego zmiennego w czasie. Stan systemu odnosi się do zbioru zmiennych, takich jak położenie, prędkość lub orientacja, które w pełni opisują badany system. Stan ten może być pośrednio obserwowany poprzez pomiary obciążone szumem, którego obecność oznacza, że obserwacje są niepewne; nawet w przypadku znajomości prawdziwego stanu systemu nie byłyby one jego deterministycznymi funkcjami, ale posiadałyby jedynie pewien rozkład możliwych wartości. Zmienność systemu w czasie jest modelowana jako system dynamiczny, który jest zaburzany pewnym szumem procesu. Szum ten jest używany do modelowania niepewności w dynamice systemu. W zdecydowanej większości przypadków wykorzystanie teorii prawdopodobieństwa wynika z braku pełnej wiedzy o systemie dynamicznym, a nie z jego wewnętrznej stochastycznej natury. [2, s. 1]

Zadanie filtracji optymalnej można zaklasyfikować jako problem inwersji statystycznej, gdzie nieznaną wielkością jest wektorowy szereg czasowy $\{x_0, x_1, x_2, \dots\}$ obserwowany poprzez zbiór zaszumionych pomiarów $\{z_1, z_2, z_3, \dots\}$, przy obecności sygnałów sterujących $\{u_1, u_2, u_3, \dots\}$. Celem wspomnianej inwersji statystycznej jest oszacowanie ukrytych stanów $x_{0:T} = \{x_0, \dots, x_T\}$ na podstawie pomiarów $z_{1:T} = \{z_1, \dots, z_T\}$ i dostarczanych sygnałów sterujących $u_{1:T} = \{u_1, \dots, u_T\}$. W sensie statystyki bayesowskiej celem jest obliczenie rozkładu łącznego a posteriori wszystkich stanów przy znajomości wszystkich pomiarów i sygnałów sterujących. Zasadniczo jest to możliwe poprzez proste zastosowanie twierdzenia Bayesa (2.1).

$$p(x_{0:T}|z_{1:T}, u_{1:T}) = \frac{p(z_{1:T}|x_{0:T}, u_{1:T})p(x_{0:T}|u_{1:T})}{p(z_{1:T}|u_{1:T})} \quad (2.1)$$

Gdzie:

- $p(x_{0:T}|u_{1:T})$ to rozkład a priori zdefiniowany przez model dynamiczny,
- $p(z_{1:T}|x_{0:T}, u_{1:T})$ to wiarygodność (prawdopodobieństwo otrzymania danych wartości pomiarów pod warunkiem wartości stanu i sterowania)
- $p(z_{1:T}|u_{1:T})$ to stała normalizacyjna zdefiniowana jako $\int p(z_{1:T}|x_{0:T}, u_{1:T})p(x_{0:T}|u_{1:T}) dx_{0:T}$

Takie sformułowanie pełnego rozkładu a posteriori ma jednak poważną wadę w postaci konieczności ponownego obliczania całego rozkładu, kiedy tylko pojawi się nowy pomiar. Problem ten jest szczególnie widoczny przy dynamicznej estymacji stanu, kiedy pomiary są otrzymywane po kolei i celem jest

uzyskanie możliwie najlepszej estymaty po każdej takiej aktualizacji wartości mierzonej. Przy wzroście liczby kroków czasowych, wymiarowość pełnego rozkładu a posteriori również wzrasta, co z kolei pociąga za sobą wzrost potrzebnej mocy obliczeniowej.

Obliczenia stają się jednak znacznie prostsze, jeśli zamiast pełnego rozkładu a posteriori, obliczane są jedynie wybrane rozkłady brzegowe. [2, s. 10] Uproszczonym celem obliczeń może być zatem rozkład brzegowy stanu w kroku t przy założeniu znajomości historii pomiarów i wartości sterowania. Możliwe jest również zastosowanie twierdzenia Bayesa do wspomnianego rozkładu (2.2).

$$p(\mathbf{x}_t | \mathbf{z}_{1:t}, \mathbf{u}_{1:t}) = \eta p(\mathbf{z}_t | \mathbf{x}_t, \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) p(\mathbf{x}_t | \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) \quad (2.2)$$

Gdzie η jest stałą normalizującą:

$$\eta = \frac{1}{p(\mathbf{z}_t | \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t})}$$

Możliwe jest przyjęcie szeregu założeń upraszczających: [3, s. 28–29]

- Żadne wartości pomiarów i sterowań przed krokiem t nie wpływają na przewidywanie pomiaru w kroku t przy założeniu znajomości stanu w kroku t (założenie Markowa):

$$p(\mathbf{z}_t | \mathbf{x}_t, \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) = p(\mathbf{z}_t | \mathbf{x}_t) \quad (2.3)$$

- Wprowadzenie zależności stanu w kroku t od stanu w kroku $t-1$ na podstawie twierdzenia o prawdopodobieństwie całkowitym:

$$p(\mathbf{x}_t | \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) = \int p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) p(\mathbf{x}_{t-1} | \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) d\mathbf{x}_{t-1} \quad (2.4)$$

- Jedynie znajomość sterowania w kroku t może wpłynąć na przewidywanie stanu w kroku t przy założeniu znajomości stanu w kroku $t-1$. Żadne wartości pomiarów i sterowań przed krokiem t nie wpływają na to przewidywanie (założenie Markowa):

$$p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) = p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{u}_t) \quad (2.5)$$

- Znajomość wartości sterowania w kroku t nie wpływa na przewidywanie stanu w kroku $t-1$:

$$p(\mathbf{x}_{t-1} | \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) = p(\mathbf{x}_{t-1} | \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t-1}) \quad (2.6)$$

Wykorzystanie powyższych założeń pozwala na rekurencyjne obliczanie rozkładu z równania 2.2. W otrzymanym w ten sposób rekurencyjnym algorytmie filtru Bayesa można wyróżnić dwa zasadnicze kroki:

- Predykcję, polegającą na znajdowaniu przewidywanego rozkładu stanu systemu w kroku t na podstawie sterowania w kroku t i poprzedniego stanu (z kroku $t-1$). Rozkład szukany w kroku predykcji to $p(\mathbf{x}_t | \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t})$,

- Korekcję, uwzględniającą pomiary z kroku t . Rozkład otrzymywany w tym kroku to $p(\mathbf{x}_t | \mathbf{z}_{1:t}, \mathbf{u}_{1:t})$.

$$p(\mathbf{x}_t | \mathbf{z}_{1:t}, \mathbf{u}_{1:t}) = \underbrace{\eta p(\mathbf{z}_t | \mathbf{x}_t)}_{\text{Korekcja}} \underbrace{\int \underbrace{p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{u}_t) p(\mathbf{x}_{t-1} | \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t-1})}_{\text{Predykcja}} d\mathbf{x}_{t-1}}_{\text{Korekcja}} \quad (2.7)$$

gdzie η to stała normalizacyjna zdefiniowana jako $\int p(\mathbf{z}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) d\mathbf{x}_t$.

Równanie 2.7 jest fundamentem dla wielu algorytmów filtracji, wykorzystujących rekurencyjne wyznaczanie wspomnianych rozkładów stanu systemu w krokach predykcji i korekcji. Takie podejście wymaga zdefiniowania pierwotnego rozkładu obrazującego początkowe przekonanie o wartości stanu, a także dwóch modeli - jednego opisującego ewolucję systemu w czasie (model dynamiczny: $\mathbf{x}_t \sim p(\mathbf{x}_t | \mathbf{x}_{t-1})$) oraz drugiego pokazującego rozkład wartości pomiarów dla danego stanu systemu (model obserwacyjny: $\mathbf{z}_t \sim p(\mathbf{z}_t | \mathbf{x}_t)$). [2, s. 10]

2.1. Filtr Kalmana

Przy założeniu liniowości modeli dynamicznego oraz obserwacyjnego, a także addytywności szumów i normalnego charakteru ich rozkładów, można znaleźć rozwiązanie równania 2.7 w zwartej formie. Wspomniane modele wyglądają zatem następująco: [2, s. 56–57]

$$\begin{aligned} \mathbf{x}_t &= \mathbf{A}_{t-1} \mathbf{x}_{t-1} + \mathbf{B}_t \mathbf{u}_t + \mathbf{q}_{t-1} \\ \mathbf{z}_t &= \mathbf{C}_t \mathbf{x}_t + \mathbf{r}_t \end{aligned} \quad (2.8)$$

$\mathbf{q}_{t-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{t-1})$ to szum procesu, natomiast $\mathbf{r}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_t)$ jest szumem pomiaru. Macierz \mathbf{A}_{t-1} jest macierzą przejścia modelu dynamicznego, zaś przez \mathbf{C}_t została oznaczona macierz modelu obserwacji. Modele można również przedstawić w notacji probabilistycznej:

$$\begin{aligned} p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{u}_t) &= \mathcal{N}(\mathbf{x}_t | \mathbf{A}_{t-1} \mathbf{x}_{t-1} + \mathbf{B}_t \mathbf{u}_t, \mathbf{Q}_{t-1}) \\ p(\mathbf{z}_t | \mathbf{x}_t) &= \mathcal{N}(\mathbf{z}_t | \mathbf{C}_t \mathbf{x}_t, \mathbf{R}_t) \end{aligned} \quad (2.9)$$

Działania wykonywane w krokach predykcji i korekcji nie powodują zmiany rodzaju rozkładu - wszystkie otrzymywane rozkłady są normalne:

$$\begin{aligned} p(\mathbf{x}_t | \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) &= \mathcal{N}(\mathbf{x}_t | \bar{\mathbf{m}}_t, \bar{\mathbf{P}}_t) \\ p(\mathbf{x}_t | \mathbf{z}_{1:t}, \mathbf{u}_{1:t}) &= \mathcal{N}(\mathbf{x}_t | \mathbf{m}_t, \mathbf{P}_t) \\ p(\mathbf{z}_t | \mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}) &= \mathcal{N}(\mathbf{z}_t | \mathbf{C}_t \bar{\mathbf{m}}_t, \mathbf{S}_t) \end{aligned} \quad (2.10)$$

Parametry powyższych rozkładów mogą zostać obliczone w krokach predykcji i korekcji filtru Kalmana:

- Krok predykcji:

$$\begin{aligned}\bar{\mathbf{m}}_t &= \mathbf{A}_{t-1} \mathbf{m}_{t-1} + \mathbf{B}_t \mathbf{u}_t \\ \bar{\mathbf{P}}_t &= \mathbf{A}_{t-1} \mathbf{P}_{t-1} \mathbf{A}_{t-1}^T + \mathbf{Q}_{t-1}\end{aligned}\quad (2.11)$$

- Krok korekcji:

$$\begin{aligned}\mathbf{v}_t &= \mathbf{z}_t - \mathbf{C}_t \bar{\mathbf{m}}_t \\ \mathbf{S}_t &= \mathbf{C}_t \bar{\mathbf{P}}_t \mathbf{C}_t^T + \mathbf{R}_t \\ \mathbf{K}_t &= \bar{\mathbf{P}}_t \mathbf{C}_t^T \mathbf{S}_t^{-1} \\ \mathbf{m}_t &= \bar{\mathbf{m}}_t + \mathbf{K}_t \mathbf{v}_t \\ \mathbf{P}_t &= (\mathbf{I} - \mathbf{K}_t \mathbf{C}_t) \bar{\mathbf{P}}_t\end{aligned}\quad (2.12)$$

Przewidywane parametry rozkładu $\bar{\mathbf{m}}_t$ i $\bar{\mathbf{P}}_t$ są obliczane przy użyciu modelu dynamicznego oraz sterowania dostarczanego do systemu. Sposób predykcji macierzy kowariancji $\bar{\mathbf{P}}_t$ bierze się z faktu, że zależność przyszłego stanu od stanu poprzedniego jest wyrażana poprzez macierz \mathbf{A}_{t-1} . W ten sposób przy obliczaniu niepewności uwzględniana jest również korelacja między zmiennymi stanu, wynikająca z modelu dynamicznego systemu. Do wyniku dodawana jest też macierz \mathbf{Q}_{t-1} , zatem po wykonaniu kroku predykcji wzrasta niepewność estymaty stanu.

Przewidywany stan jest korygowany poprzez uwzględnienie pomiarów w drugim etapie działania algorytmu. W zależności od podanej dokładności modelu dynamicznego oraz pomiarowego, algorytm podaje ostateczną estymatę bliższą przewidywaniom albo pomiarom. Macierz \mathbf{K}_t , nazywana macierzą wzmocnień Kalmana, precyzuje stopień zaufania do pomiarów i to na jej podstawie korygowane jest przewidywanie stanu. Uwzględnienie obserwacji jako kolejnego źródła informacji zmniejsza również niepewność oszacowania stanu. [3, s. 36–37]

Rozkład Gaussa jest w pełni określony przez wektor wartości średnich oraz macierz kowariancji, zatem obliczenia prowadzą do znalezienia tych dwóch charakterystyk rozkładu. Wektor wartości średnich zawiera optymalną estymatę stanu, natomiast diagonalne elementy macierzy kowariancji przedstawiają niepewność estymacji zmiennych stanu. Otrzymana estymata jest optymalna w każdym z najczęściej przyjmowanych sposobów, to znaczy *MAP* (*maximum a posteriori*), *MMSE* (*minimum mean squared error*) oraz przyjmując wartość bezwzględną błędu w funkcji kosztu (*Absolute error loss*). Wynika to z faktu, że moda, średnia arytmetyczna oraz mediana rozkładu normalnego pokrywają się. [2, s. 21]

Filtr Kalmana jest dość wydajnym obliczeniowo algorytmem. Dla najlepszych obecnie znanych algorytmów, złożoność obliczeniowa operacji odwracania macierzy jest w przybliżeniu rzędu $O(d^{2,8})$ dla macierzy rozmiaru $d \times d$. Każda iteracja algorytmu filtru Kalmana jest zatem ograniczona od dołu przez w przybliżeniu $O(k^{2,8})$, gdzie k jest rozmiarem wektora pomiarów \mathbf{z}_t . Wynika to z obserwacji, że każda iteracja algorytmu wiąże się z odwracaniem macierzy \mathbf{S}_t , rozmiaru $k \times k$. Kolejnym dolnym ograniczeniem złożoności filtru Kalmana jest $O(n^2)$, gdzie n to liczba zmiennych stanu, ze względu na mnożenie w ostatniej linii algorytmu. W wielu praktycznych zastosowaniach, wymiarowość przestrzeni pomiarów

jest znacznie mniejsza od przestrzeni stanu, i algorytm jest zdominowany przez operacje o złożoności $O(n^2)$. [3, s. 37]

2.2. Rozszerzony filtr Kalmana

Założenia o liniowych modelach dynamicznym oraz pomiarowym są rzadko spełnione w praktyce. Ten fakt, wraz z drugim założeniem o rozkładach jedynie normalnych, powoduje, że zwyczajny filtr Kalmana nadaje się tylko do najbardziej trywialnych rzeczywistych problemów. Istnieje kilka rozwiązań pozwalających na przewyższenie jednego z tych ograniczeń: założenia o liniowości. W tym wypadku zakładane jest jedynie, że wartością następnego stanu oraz pomiarami rządzą pewne (w ogólności nieliniowe) funkcje g_t i h_t :

$$\begin{aligned} \mathbf{x}_t &= \mathbf{g}_t(\mathbf{u}_t, \mathbf{x}_{t-1}) + \mathbf{q}_{t-1} \\ \mathbf{z}_t &= \mathbf{h}_t(\mathbf{x}_t) + \mathbf{r}_t \end{aligned} \quad (2.13)$$

Model przedstawiony w równaniu 2.13 uogólnia liniowy gaussowski model z równania 2.8, wykorzystywany w filtrze Kalmana. Funkcja g_t zastępuje macierze A_{t-1} oraz B_t , natomiast h_t występuje w miejsce macierzy C_t . W tym przypadku, przy dowolnych funkcjach g_t i h_t , otrzymywany rozkład nie jest już normalny. Wykonanie dokładnej aktualizacji estymaty stanu jest niemożliwe dla nieliniowych funkcji g_t i h_t , ponieważ algorytm filtru Bayesa z równania 2.7 nie posiada rozwiązania w zamkniętej formie.

Możliwe jest jednak poszukiwanie aproksymacji prawdziwego rozkładu stanu systemu. Jednym z pierwszych, podstawowych i najczęściej używanych rozwiązań jest rozszerzony filtr Kalmana (ang. *Extended Kalman Filter*, EKF). Algorytm ten również zakłada prostą reprezentację przekonania o stanie systemu za pomocą rozkładu normalnego, jednak w tym wypadku przekonanie to jest tylko przybliżeniem.

Główną ideą rozszerzonego filtru Kalmana jest linearyzacja, która przybliża g_t funkcją liniową styczną do g_t w miejscu średniej wartości rozkładu Gaussa. Poprzez projekcję rozkładu normalnego przez taką liniową aproksymację, wynikowy rozkład staje się normalny. W tym momencie cały mechanizm aktualizacji przekonania staje się taki sam jak w przypadku filtru Kalmana. Ten sam sposób może być zastosowany w przypadku funkcji h_t , zachowując w ten sposób gaussowską naturę rozkładu.

EKF wykorzystuje do linearyzacji metodę rozwinięcia Taylora pierwszego rzędu, która konstruuje liniowe przybliżenie funkcji poprzez jej wartość i pochodną cząstkową (równanie 2.14).

$$\mathbf{g}'_t(\mathbf{u}_t, \mathbf{x}_{t-1}) := \frac{\partial \mathbf{g}_t(\mathbf{u}_t, \mathbf{x}_{t-1})}{\partial \mathbf{x}_{t-1}} \quad (2.14)$$

Zarówno wartość funkcji g_t , jak i jej pochodna zależą od wartości argumentu funkcji. W rozszerzonym filtrze Kalmana jako argument wybiera się wartość stanu uznawaną za najbardziej prawdopodobną,

zatem funkcja g_t jest aproksymowana wokół \mathbf{m}_{t-1} (oraz \mathbf{u}_t):

$$\begin{aligned} g_t(\mathbf{u}_t, \mathbf{x}_{t-1}) &\approx g_t(\mathbf{u}_t, \mathbf{m}_{t-1}) + g'_t(\mathbf{u}_t, \mathbf{m}_{t-1})(\mathbf{x}_{t-1} - \mathbf{m}_{t-1}) \\ &= g_t(\mathbf{u}_t, \mathbf{m}_{t-1}) + \mathbf{G}_t(\mathbf{x}_{t-1} - \mathbf{m}_{t-1}) \end{aligned} \quad (2.15)$$

Macierz \mathbf{G}_t , często nazywana Jakobianem, jest macierzą rozmiaru $n \times n$, gdzie n to rozmiar wektora zmiennych stanu. Wartość Jakobianu zależy od \mathbf{u}_t oraz \mathbf{m}_{t-1} , zatem zmienia się dla różnych punktów w czasie.

EKF stosuje taką samą linearyzację dla funkcji h :

$$h'_t(\mathbf{x}_t) := \frac{\partial h_t(\mathbf{x}_t)}{\partial \mathbf{x}_t}$$

W tym przypadku rozwinięcie Taylora następuje w punkcie $\bar{\mathbf{m}}_t$, jako wartości najbardziej prawdopodobnej w momencie linearyzacji h :

$$\begin{aligned} h_t(\mathbf{x}_t) &\approx h_t(\bar{\mathbf{m}}_t) + h'_t(\bar{\mathbf{m}}_t)(\mathbf{x}_t - \bar{\mathbf{m}}_t) \\ &= h_t(\bar{\mathbf{m}}_t) + \mathbf{H}_t(\mathbf{x}_t - \bar{\mathbf{m}}_t) \end{aligned} \quad (2.16)$$

Modele przedstawione w notacji probabilistycznej wyglądają następująco:

$$\begin{aligned} p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{u}_t) &= \mathcal{N}(\mathbf{x}_t | g_t(\mathbf{u}_t, \mathbf{m}_{t-1}) + \mathbf{G}_t(\mathbf{x}_{t-1} - \mathbf{m}_{t-1}), \mathbf{Q}_{t-1}) \\ p(\mathbf{z}_t | \mathbf{x}_t) &= \mathcal{N}(\mathbf{z}_t | h_t(\bar{\mathbf{m}}_t) + \mathbf{H}_t(\mathbf{x}_t - \bar{\mathbf{m}}_t), \mathbf{R}_t) \end{aligned} \quad (2.17)$$

Podobnie jak w przypadku zwykłego filtra Kalmana, algorytm EKF, przedstawiony w równaniach 2.18 i 2.19, wyznacza potrzebne parametry w krokach predykcji oraz korekcji. [3, s. 48–51]

◦ Krok predykcji:

$$\begin{aligned} \bar{\mathbf{m}}_t &= g_t(\mathbf{u}_t, \mathbf{m}_{t-1}) \\ \bar{\mathbf{P}}_t &= \mathbf{G}_t \mathbf{P}_{t-1} \mathbf{G}_t^T + \mathbf{Q}_{t-1} \end{aligned} \quad (2.18)$$

◦ Krok korekcji:

$$\begin{aligned} \mathbf{v}_t &= \mathbf{z}_t - h_t(\bar{\mathbf{m}}_t) \\ \mathbf{S}_t &= \mathbf{H}_t \bar{\mathbf{P}}_t \mathbf{H}_t^T + \mathbf{R}_t \\ \mathbf{K}_t &= \bar{\mathbf{P}}_t \mathbf{H}_t^T \mathbf{S}_t^{-1} \\ \mathbf{m}_t &= \bar{\mathbf{m}}_t + \mathbf{K}_t \mathbf{v}_t \\ \mathbf{P}_t &= (\mathbf{I} - \mathbf{K}_t \mathbf{H}_t) \bar{\mathbf{P}}_t \end{aligned} \quad (2.19)$$

Bardziej ogólną sytuacją jest brak addytywności szumu w modelach dynamicznym i obserwacyjnym: [2, s. 71]

$$\begin{aligned} \mathbf{x}_t &= g_t(\mathbf{u}_t, \mathbf{x}_{t-1}, \mathbf{q}_{t-1}) \\ \mathbf{z}_t &= h_t(\mathbf{x}_t, \mathbf{r}_t) \end{aligned} \quad (2.20)$$

W takim przypadku możliwe jest obliczenie Jakobianów po zmiennych stanu oraz składowych szumu, oznaczonych odpowiednio G_{x_t} i G_{q_t} dla modelu dynamicznego, a także H_{x_t} i H_{r_t} dla obserwacji. Aproksymacja następuje, podobnie jak dla przypadku szumu addytywnego, wokół m_{t-1} i u_t (macierze G_{x_t} i G_{q_t}) oraz \bar{m}_t (dla macierzy H_{x_t} i H_{r_t}), a także wokół zerowych wartości składowych szumów. Algorytm rozszerzonego filtru Kalmana przyjmuje wówczas postać:

- Krok predykcji:

$$\begin{aligned}\bar{m}_t &= g_t(u_t, m_{t-1}, 0) \\ \bar{P}_t &= G_{x_t} P_{t-1} G_{x_t}^T + G_{q_t} Q_{t-1} G_{q_t}^T\end{aligned}\quad (2.21)$$

- Krok korekcji:

$$\begin{aligned}v_t &= z_t - h_t(\bar{m}_t, 0) \\ S_t &= H_{x_t} \bar{P}_t H_{x_t}^T + H_{r_t} \bar{R}_t H_{r_t}^T \\ K_t &= \bar{P}_t H_{x_t}^T S_t^{-1} \\ m_t &= \bar{m}_t + K_t v_t \\ P_t &= (I - K_t H_{x_t}) \bar{P}_t\end{aligned}\quad (2.22)$$

Algorytm rozszerzonego filtru Kalmana stał się najbardziej popularnym narzędziem wykorzystywanym do estymacji stanu systemów. Siłą tego rozwiązania leży w jego prostocie oraz efektywności obliczeniowej. Tak jak w przypadku filtru Kalmana, każda iteracja potrzebuje czasu $O(k^{2,8} + n^2)$, gdzie k jest rozmiarem wektora pomiarów z_t , a n jest wymiarem przestrzeni stanów.

Ważnym ograniczeniem algorytmu EKF jest fakt, że korzysta on z linearyzacji ewolucji stanu oraz pomiarów przy pomocy metody Taylora rozwinięcia funkcji w szereg. Dokładność uzyskanej w ten sposób aproksymacji zależy od dwóch głównych czynników. Po pierwsze, jest to stopień nieliniowości funkcji, która jest linearyzowana. Jeśli funkcja ta jest w przybliżeniu liniowa, aproksymacja algorytmu będzie dobra, co przełoży się na odwzorowanie wynikowego rozkładu z wystarczającą dokładnością. Drugim czynnikiem wpływającym na skuteczność takiego sposobu linearyzacji jest stopień niepewności estymaty stanu. Jeśli niepewność jest duża, gęstość rozkładu jest mniej skupiona wokół średniej, a przez to bardziej wpływają na nią nieliniowości funkcji. [3, s. 53–54]

2.3. Filtr Kalmana Gaussa-Hermite'a

Innym sposobem otrzymania rozkładu normalnego jest metoda dopasowania rozkładów za pomocą momentów. Jeśli zmienna losowa $x \sim \mathcal{N}(m, P)$ jest transformowana na zmienną losową $y = g(x) + q$, $q \sim \mathcal{N}(0, Q)$, to gaussowska aproksymacja bazująca na momentach rozkładu łącznego x i y jest dana wzorem: [2, s. 96–99]

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m \\ \mu_M \end{bmatrix}, \begin{bmatrix} P & C_M \\ C_M^T & S_M \end{bmatrix}\right)\quad (2.23)$$

Gdzie:

$$\begin{aligned}\mu_M &= \int g(x) \mathcal{N}(x|m, P) dx \\ S_M &= \int (g(x) - \mu_M)(g(x) - \mu_M)^T \mathcal{N}(x|m, P) dx + Q \\ C_M &= \int (x - m)(g(x) - \mu_M)^T \mathcal{N}(x|m, P) dx\end{aligned}\quad (2.24)$$

Dopasowanie rozkładów za pomocą momentów jest także możliwe w przypadku szumu nieaddytywnego, czyli jeśli $y = g(x, q)$:

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m \\ \mu_M \end{bmatrix}, \begin{bmatrix} P & C_M \\ C_M^T & S_M \end{bmatrix}\right)\quad (2.25)$$

Gdzie:

$$\begin{aligned}\mu_M &= \int g(x, q) \mathcal{N}(x|m, P) \mathcal{N}(q|0, Q) dx \\ S_M &= \int (g(x, q) - \mu_M)(g(x, q) - \mu_M)^T \mathcal{N}(x|m, P) \mathcal{N}(q|0, Q) dx dq \\ C_M &= \int (x - m)(g(x, q) - \mu_M)^T \mathcal{N}(x|m, P) \mathcal{N}(q|0, Q) dx dq\end{aligned}\quad (2.26)$$

W ten sposób możliwe jest aproksymowanie wynikowych rozkładów pojawiających się po nieliniowych transformacjach rozkładów normalnych poprzez rozkład Gaussa. Średnia m_t oraz kowariancja P_t rozkładu $p(x_t|z_{1:t}, u_{1:t}) \simeq \mathcal{N}(x|m_t, P_t)$ jest przybliżana przy użyciu metody dopasowania momentów. Dla przypadku szumu addytywnego uzyskany filtr Gaussa ma postać:

- Krok predykcji:

$$\begin{aligned}\bar{m}_t &= \int g(u_t, x_{t-1}) \mathcal{N}(x_{t-1}|m_{t-1}, P_{t-1}) dx_{t-1} \\ \bar{P}_t &= \int (g(u_t, x_{t-1}) - \bar{m}_t)(g(u_t, x_{t-1}) - \bar{m}_t)^T \mathcal{N}(x_{t-1}|m_{t-1}, P_{t-1}) dx_{t-1} + Q_{t-1}\end{aligned}\quad (2.27)$$

- Krok korekcji:

$$\begin{aligned}\mu_t &= \int h(x_t) \mathcal{N}(x_t, \bar{m}_t, \bar{P}_t) dx_t, \\ S_t &= \int (h(x_t) - \mu_t)(h(x_t) - \mu_t)^T \mathcal{N}(x_t, \bar{m}_t, \bar{P}_t) dx_t + R_t, \\ C_t &= \int (x_t - \bar{m}_t)((h(x_t) - \mu_t))^T \mathcal{N}(x_t, \bar{m}_t, \bar{P}_t) dx_t\end{aligned}\quad (2.28)$$

$$\begin{aligned}K_t &= C_t S_t^{-1} \\ P_t &= \bar{P}_t - K_t S_t K_t^T \\ m_t &= \bar{m}_t + K_t(z_t - \mu_t)\end{aligned}\quad (2.29)$$

Możliwe jest rozszerzenie algorytmu na przypadek szumu nieaddytywnego. Równania filtru przyjmą wówczas postać:

○ Krok predykcji:

$$\begin{aligned}\bar{\mathbf{m}}_t &= \int \mathbf{g}(\mathbf{u}_t, \mathbf{x}_{t-1}, \mathbf{q}_{t-1}) \mathcal{N}(\mathbf{x}_{t-1} | \mathbf{m}_{t-1}, \mathbf{P}_{t-1}) \mathcal{N}(\mathbf{q}_{t-1} | \mathbf{0}, \mathbf{Q}_{t-1}) d\mathbf{x}_{t-1} d\mathbf{q}_{t-1} \\ \bar{\mathbf{P}}_t &= \int (\mathbf{g}(\mathbf{u}_t, \mathbf{x}_{t-1}, \mathbf{q}_{t-1}) - \bar{\mathbf{m}}_t)(\mathbf{g}(\mathbf{u}_t, \mathbf{x}_{t-1}, \mathbf{q}_{t-1}) - \bar{\mathbf{m}}_t)^T \\ &\quad \times \mathcal{N}(\mathbf{x}_{t-1} | \mathbf{m}_{t-1}, \mathbf{P}_{t-1}) \mathcal{N}(\mathbf{q}_{t-1} | \mathbf{0}, \mathbf{Q}_{t-1}) d\mathbf{x}_{t-1} d\mathbf{q}_{t-1}\end{aligned}\quad (2.30)$$

○ Krok korekcji:

$$\begin{aligned}\boldsymbol{\mu}_t &= \int \mathbf{h}(\mathbf{x}_t, \mathbf{r}_t) \mathcal{N}(\mathbf{x}_t | \bar{\mathbf{m}}_t, \bar{\mathbf{P}}_t) \mathcal{N}(\mathbf{r}_t | \mathbf{0}, \mathbf{R}_t) d\mathbf{x}_t d\mathbf{r}_t \\ \mathbf{S}_t &= \int (\mathbf{h}(\mathbf{x}_t, \mathbf{r}_t) - \boldsymbol{\mu}_t)(\mathbf{h}(\mathbf{x}_t, \mathbf{r}_t) - \boldsymbol{\mu}_t)^T \mathcal{N}(\mathbf{x}_t, \bar{\mathbf{m}}_t, \bar{\mathbf{P}}_t) \mathcal{N}(\mathbf{r}_t | \mathbf{0}, \mathbf{R}_t) d\mathbf{x}_t d\mathbf{r}_t \\ \mathbf{C}_t &= \int (\mathbf{x}_t - \bar{\mathbf{m}}_t)(\mathbf{h}(\mathbf{x}_t, \mathbf{r}_t) - \boldsymbol{\mu}_t)^T \mathcal{N}(\mathbf{x}_t, \bar{\mathbf{m}}_t, \bar{\mathbf{P}}_t) \mathcal{N}(\mathbf{r}_t | \mathbf{0}, \mathbf{R}_t) d\mathbf{x}_t d\mathbf{r}_t \\ \mathbf{K}_t &= \mathbf{C}_t \mathbf{S}_t^{-1} \\ \mathbf{P}_t &= \bar{\mathbf{P}}_t - \mathbf{K}_t \mathbf{S}_t \mathbf{K}_t^T \\ \mathbf{m}_t &= \bar{\mathbf{m}}_t + \mathbf{K}_t (\mathbf{z}_t - \boldsymbol{\mu}_t)\end{aligned}\quad (2.31)$$

Powyższe ogólne równania filtru Gaussa są raczej teoretycznymi konstrukcjami, a nie praktycznymi algorytmami filtracji. Należy przyjąć pewną metodę rozwiązywania potrzebnych całek występujących w formie 2.32, aby uzyskać funkcjonalny algorytm.

$$\begin{aligned}\int \mathbf{g}(\mathbf{x}) \mathcal{N}(\mathbf{x} | \mathbf{m}, \mathbf{P}) d\mathbf{x} \\ = \frac{1}{(2\pi)^{n/2} (\det \mathbf{P})^{1/2}} \int \mathbf{g}(\mathbf{x}) \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m})^T \mathbf{P}^{-1}(\mathbf{x} - \mathbf{m})\right) d\mathbf{x}\end{aligned}\quad (2.32)$$

Jedną z takich numerycznych metod jest algorytm Gaussa-Hermite'a, który w swojej najprostszej formie odnosi się do przypadku jednowymiarowego ze standardową funkcją gęstości. Aproksymacja wygląda wówczas następująco:

$$\int g(x) \mathcal{N}(x | 0, 1) dx \approx \sum_{i=1}^p W_i g(\xi_i) \quad (2.33)$$

W_i to wagi, natomiast punkty ξ_i nazywane są węzłami lub punktami sigma. Jest nieskończenie wiele sposobów wyboru wag oraz węzłów. Przy rozwiązywaniu całek metodą Gaussa-Hermite'a, tak jak w przypadku innych kwadratur, są one wybierane w ten sposób, że dla funkcji podcałkowych będących wielomianami określonego stopnia, wynik jest dokładny. Okazuje się, że stopień ten jest maksymalizowany przy wyborze węzłów jako pierwiastków wielomianu Hermite'a. Dla wielomianu Hermite'a stopnia p , całkowanie jest dokładne dla wielomianów rzędu $2p - 1$ lub niższego.

Wielomian Hermite'a stopnia p jest definiowany następująco:

$$H_p(x) = (-1)^p e^{x^2/2} \frac{d^p}{dx^p} e^{-x^2/2} \quad (2.34)$$

Pierwsze wielomiany Hermite'a to:

$$\begin{aligned} H_0(x) &= 1, \\ H_1(x) &= x, \\ H_2(x) &= x^2 - 1, \\ H_3(x) &= x^3 - 3x, \\ H_4(x) &= x^4 - 6x^2 + 3, \end{aligned} \quad (2.35)$$

a kolejne mogą być obliczone z zależności rekurencyjnej:

$$H_{p+1}(x) = xH_p(x) - pH_{p-1}(x) \quad (2.36)$$

Dla każdego punktu sigma ξ_i można obliczyć odpowiadającą mu wagę W_i , korzystając z następującej zależności:

$$W_i = \frac{p!}{p^2 [H_{p-1}(\xi_i)]^2} \quad (2.37)$$

Całki z niestandardową funkcją gęstości $\mathcal{N}(x|m, P)$ mogą być obliczone poprzez zwykłą zmianę zmiennej:

$$\int g(x) \mathcal{N}(x|m, P) dx = \int g(P^{1/2}\xi + m) \mathcal{N}(\xi|0, 1) d\xi \quad (2.38)$$

W takim przypadku przybliżenie całki wygląda następująco:

$$\int g(x) \mathcal{N}(x|m, P) dx \approx \sum_{i=1}^p W_i g(P^{1/2}\xi_i + m) \quad (2.39)$$

Równanie 2.39 może być dalej uogólnione na przypadek wielowymiarowy, poprzez zdefiniowanie wektora nowych zmiennych ξ oraz zastosowanie rozkładu Choleskiego do macierzy kowariancji ($P = \sqrt{P}\sqrt{P}^T$):

$$x = m + \sqrt{P}\xi \quad (2.40)$$

Otrzymana w ten sposób całka nad wielowymiarową funkcją ze standardowym rozkładem normalnym jako funkcją wagową to:

$$\int g(x) \mathcal{N}(x|m, P) dx = \int g(m + \sqrt{P}\xi) \mathcal{N}(x|0, I) d\xi \quad (2.41)$$

Całka otrzymana w równaniu 2.41 może być przedstawiona jako całka iterowana i każdą z całek wchodzących w skład całki iterowanej można aproksymować z wykorzystaniem kwadratury Gaussa-Hermite'a:

$$\begin{aligned} & \int g(m + \sqrt{P}\xi) \mathcal{N}(x|0, I) d\xi \\ &= \int \cdots \int g(m + \sqrt{P}\xi) \mathcal{N}(\xi_1, 0, 1) \times \cdots \times \mathcal{N}(\xi_n, 0, 1) d\xi_1 \cdots d\xi_n \\ &\approx \sum_{i_1, \dots, i_n} W_{(i_1, \dots, i_n)} g(m + \sqrt{P}\xi_{(i_1, \dots, i_n)}) \end{aligned} \quad (2.42)$$

W tym wypadku węzły powstają jako iloczyn kartezjański jednowymiarowych punktów sigma, $\xi_{(i_1, \dots, i_n)} = (\xi_1 \dots \xi_n)^T$, natomiast wielowymiarowe wagi są tworzone poprzez pomnożenie jednowymiarowych wag odpowiadających węzłom: $\mathbf{W}_{(i_1, \dots, i_n)} = \mathbf{W}_{i_1} \times \dots \times \mathbf{W}_{i_n}$.

Całkowanie metodą Gaussa-Hermite'a jest dokładne dla jednomianów $x_1^{d_1} x_2^{d_2} \dots x_n^{d_n}$ i ich dowolnej kombinacji liniowej, gdzie każda potęga $d_i \leq 2p - 1$. Liczba węzłów (rozmiaru n) oraz wag potrzebnych do obliczenia całki n -wymiarowej przy zastosowaniu p węzłów jednowymiarowych jest równa p^n , zatem złożoność aproksymacji Gaussa-Hermite'a rośnie bardzo szybko wraz ze wzrostem wymiarowości i liczby p . [2, s. 103]

Zastosowanie metody Gaussa-Hermite'a do obliczenia całek z równań 2.27 i 2.28 daje w wyniku algorytm filtru Kalmana Gaussa-Hermite'a (ang. *Gauss-Hermite Kalman Filter*, GHKF) dla przypadku szumu addytywnego: [4]

o Krok predykcji:

$$\begin{aligned} \chi_{t-1}^{(i_1, \dots, i_n)} &= \mathbf{m}_{t-1} + \sqrt{\mathbf{P}_{t-1}} \xi_{(i_1, \dots, i_n)} \\ \hat{\chi}_t^{(i_1, \dots, i_n)} &= \mathbf{g}(\mathbf{u}_t, \chi_{t-1}^{(i_1, \dots, i_n)}) \\ \bar{\mathbf{m}}_t &= \sum_{i_1, \dots, i_n} \mathbf{W}_{(i_1, \dots, i_n)} \hat{\chi}_t^{(i_1, \dots, i_n)} \\ \bar{\mathbf{P}}_t &= \sum_{i_1, \dots, i_n} \mathbf{W}_{(i_1, \dots, i_n)} (\hat{\chi}_t^{(i_1, \dots, i_n)} - \bar{\mathbf{m}}_t)(\hat{\chi}_t^{(i_1, \dots, i_n)} - \bar{\mathbf{m}}_t)^T + \mathbf{Q}_{t-1} \end{aligned} \quad (2.43)$$

o Krok korekcji:

$$\begin{aligned} \bar{\chi}_t^{(i_1, \dots, i_n)} &= \bar{\mathbf{m}}_t + \sqrt{\bar{\mathbf{P}}_t} \xi_{(i_1, \dots, i_n)} \\ \hat{\mathbf{y}}_t^{(i_1, \dots, i_n)} &= \mathbf{h}(\bar{\chi}_t^{(i_1, \dots, i_n)}) \\ \boldsymbol{\mu}_t &= \sum_{i_1, \dots, i_n} \mathbf{W}_{(i_1, \dots, i_n)} \hat{\mathbf{y}}_t^{(i_1, \dots, i_n)} \\ \mathbf{S}_t &= \sum_{i_1, \dots, i_n} \mathbf{W}_{(i_1, \dots, i_n)} (\hat{\mathbf{y}}_t^{(i_1, \dots, i_n)} - \boldsymbol{\mu}_t)(\hat{\mathbf{y}}_t^{(i_1, \dots, i_n)} - \boldsymbol{\mu}_t)^T + \mathbf{R}_t \\ \mathbf{C}_t &= \sum_{i_1, \dots, i_n} \mathbf{W}_{(i_1, \dots, i_n)} (\bar{\chi}_t^{(i_1, \dots, i_n)} - \bar{\mathbf{m}}_t)(\hat{\mathbf{y}}_t^{(i_1, \dots, i_n)} - \boldsymbol{\mu}_t)^T \\ \mathbf{K}_t &= \mathbf{C}_t \mathbf{S}_t^{-1} \\ \mathbf{P}_t &= \bar{\mathbf{P}}_t - \mathbf{K}_t \mathbf{S}_t \mathbf{K}_t^T \\ \mathbf{m}_t &= \bar{\mathbf{m}}_t + \mathbf{K}_t (\mathbf{z}_t - \boldsymbol{\mu}_t) \end{aligned} \quad (2.44)$$

Przypadek szumu nieaddytywnego (2.20) wymaga tworzenia węzłów $\xi_{(i_1, \dots, i_n)}$ w oparciu o zwiększoną liczbę wymiarów (suma wymiarów zmiennych stanu i wektora szumu \mathbf{q}). Zwiększeniu ulegnie również wymiarowość zmodyfikowanego wektora wartości średnich oraz macierzy kowariancji:

$$\mathbf{m}'_{t-1} = \begin{bmatrix} \mathbf{m}_{t-1} \\ \mathbf{0} \end{bmatrix} \quad (2.45)$$

$$\mathbf{P}'_{t-1} = \begin{bmatrix} \mathbf{P}_{t-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{t-1} \end{bmatrix} \quad (2.46)$$

$$\bar{\mathbf{m}}'_t = \begin{bmatrix} \bar{\mathbf{m}}_t \\ \mathbf{0} \end{bmatrix} \quad (2.47)$$

$$\bar{\mathbf{P}}'_t = \begin{bmatrix} \bar{\mathbf{P}}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_t \end{bmatrix} \quad (2.48)$$

Stosując powyższe modyfikacje, filtr Kalmana Gaussa-Hermite'a dla szumu nieaddytywnego będzie wyglądał następująco:

- Krok predykcji:

$$\begin{aligned} \chi_{t-1}^{(i_1, \dots, i_n)} &= \mathbf{m}'_{t-1} + \sqrt{\mathbf{P}'_{t-1}} \boldsymbol{\xi}_{(i_1, \dots, i_n)} \\ \hat{\chi}_t^{(i_1, \dots, i_n)} &= \mathbf{g}(\mathbf{u}_t, \chi_{t-1}^{(i_1, \dots, i_n)}) \\ \bar{\mathbf{m}}_t &= \sum_{i_1, \dots, i_n} \mathbf{W}_{(i_1, \dots, i_n)} \hat{\chi}_t^{(i_1, \dots, i_n)} \\ \bar{\mathbf{P}}_t &= \sum_{i_1, \dots, i_n} \mathbf{W}_{(i_1, \dots, i_n)} (\hat{\chi}_t^{(i_1, \dots, i_n)} - \bar{\mathbf{m}}_t)(\hat{\chi}_t^{(i_1, \dots, i_n)} - \bar{\mathbf{m}}_t)^T \end{aligned} \quad (2.49)$$

- Krok korekcji:

$$\begin{aligned} \bar{\chi}_t^{(i_1, \dots, i_n)} &= \bar{\mathbf{m}}'_t + \sqrt{\bar{\mathbf{P}}'_t} \boldsymbol{\xi}_{(i_1, \dots, i_n)} \\ \hat{\mathbf{y}}_t^{(i_1, \dots, i_n)} &= \mathbf{h}(\bar{\chi}_t^{(i_1, \dots, i_n)}) \\ \boldsymbol{\mu}_t &= \sum_{i_1, \dots, i_n} \mathbf{W}_{(i_1, \dots, i_n)} \hat{\mathbf{y}}_t^{(i_1, \dots, i_n)} \\ \mathbf{S}_t &= \sum_{i_1, \dots, i_n} \mathbf{W}_{(i_1, \dots, i_n)} (\hat{\mathbf{y}}_t^{(i_1, \dots, i_n)} - \boldsymbol{\mu}_t)(\hat{\mathbf{y}}_t^{(i_1, \dots, i_n)} - \boldsymbol{\mu}_t)^T \\ \mathbf{C}_t &= \sum_{i_1, \dots, i_n} \mathbf{W}_{(i_1, \dots, i_n)} (\bar{\chi}_t^{(i_1, \dots, i_n)} - \bar{\mathbf{m}}_t)(\hat{\mathbf{y}}_t^{(i_1, \dots, i_n)} - \boldsymbol{\mu}_t)^T \\ \mathbf{K}_t &= \mathbf{C}_t \mathbf{S}_t^{-1} \\ \mathbf{P}_t &= \bar{\mathbf{P}}_t - \mathbf{K}_t \mathbf{S}_t \mathbf{K}_t^T \\ \mathbf{m}_t &= \bar{\mathbf{m}}_t + \mathbf{K}_t (\mathbf{z}_t - \boldsymbol{\mu}_t) \end{aligned} \quad (2.50)$$

3. Eksperymenty numeryczne

nieaddytywne? testy?

W celu przeprowadzenia eksperymentów numerycznych zaimplementowano rozszerzony filtr Kalmana oraz filtr Kalmana Gaussa-Hermite'a. Implementację przeprowadzono w języku Python z wykorzystaniem otwartoźródłowej biblioteki NumPy, przeznaczonej do szybkiego wykonywania obliczeń macierzowych. Interfejs wykonanych modułów został stworzony zgodnie z paradygmatem programowania obiektowego.

Funkcje teoretyczne: 2 funkcje Liu, moje funkcje: rmse jako funkcja parametru a funkcji, EKF, GHKF-2, GHKF-3, GHKF-5?, czasy działania algorytmów, 2 warianty: model procesu lub model pomiarów nieliniowy, 2 wymiary

3.1. Systemy teoretyczne

Do porównania działania filtrów użyto początkowo dyskretnego nieliniowego modelu systemu dynamicznego, wykorzystywanego wcześniej w [5] i [6]. Przeprowadzono 50 iteracji dla czasów 1-50 i wykorzystano EKF oraz GHKF trzeciego stopnia do estymacji stanu systemów.

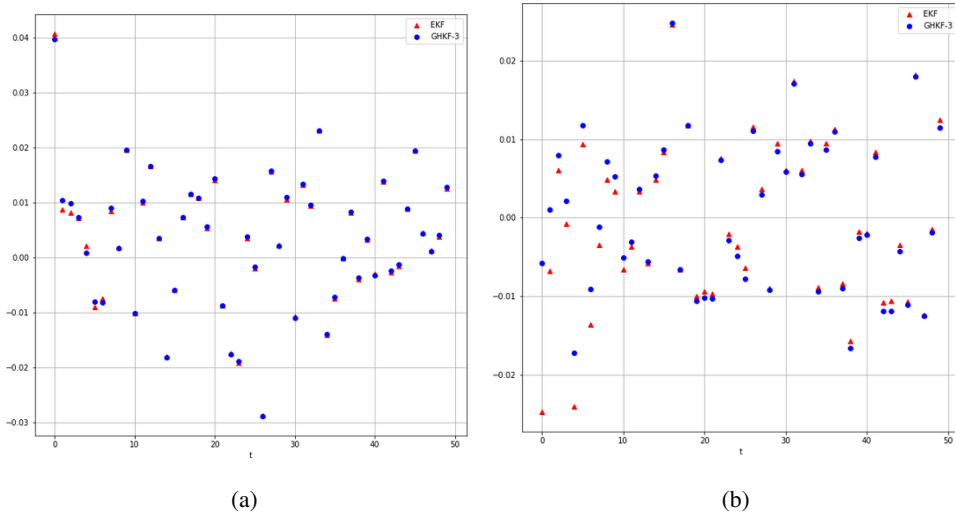
Model systemu wyglądał następująco:

$$\begin{cases} x_1(t+1) = 0.8x_1(t) + x_1(t)x_2(t) + 0.1 + 0.01w_1(t) \\ x_2(t+1) = 1.5x_2(t) - x_1(t)x_2(t) + 0.1 + 0.01w_1(t) \\ z_1(t+1) = x_1(t+1) + 0.04v_1(t+1) \\ z_2(t+1) = x_2(t+1) + 0.04v_2(t+1) \end{cases} \quad (3.1)$$

gdzie szum procesu oraz szum pomiaru są nieskorelowanymi białymi szumami gaussowskimi z następującymi rozkładami: $w(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$, $v(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$, $\mathbf{Q} = \text{diag}(0.1, 0.2)$, $\mathbf{R} = \text{diag}(0.1, 0.2)$; początkowy stan systemu to: $\mathbf{x}(0) = [1, 1]^T$, $\mathbf{P}(0|0) = \mathbf{I}_{2 \times 2}$.

Jako miarę skuteczności filtracji użyto pierwiastka z błędu średniokwadratowego (ang. *root-mean-square error*, RMSE), który dla T estymat \hat{x}_t rzeczywistych wartości x_t jest definiowany następująco:

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^T (\hat{x}_t - x_t)^2}{T}} \quad (3.2)$$



Rys. 3.1. Różnice między wartością estymowaną, a rzeczywistą dla x_1 (3.1a) oraz x_2 (3.1b).

Wykorzystanie EKF wymagało obliczenia macierzy G i H :

$$G = \begin{bmatrix} x_2 + 0.8 & x_1 \\ -x_2 & 1.5 - x_1 \end{bmatrix}$$

$$H = I_{2 \times 2} \quad (3.3)$$

Dla EKF wartość RMSE wyniosła, odpowiednio dla x_1 i x_2 , 0.01244 i 0.01042, natomiast dla GHKF 0.01246 oraz 0.00961. Rysunek 3.1 przedstawia wykresy różnic pomiędzy wartością estymowaną, a rzeczywistą dla zmiennych stanu badanego systemu. Oba algorytmy dobrze poradziły sobie z postawionym problemem, wartości błędów były bardzo niewielkie, a różnice między wartościami wyjściowymi algorytmów minimalne.

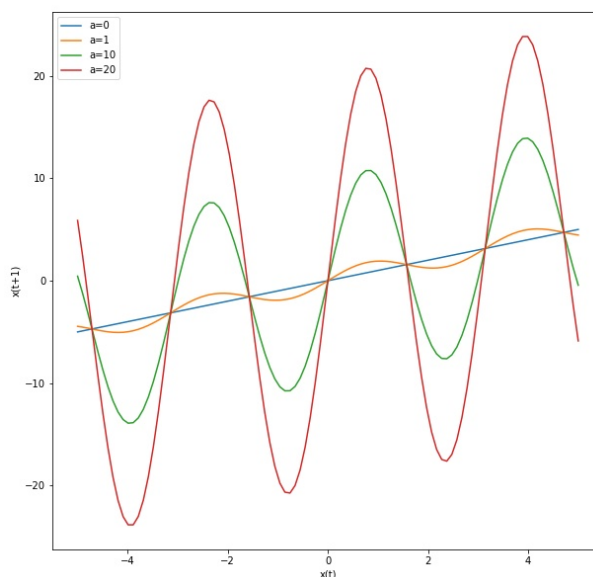
W celu dokładniejszego zbadania, jak nieliniowość funkcji wpływa na działanie obu rodzajów filtrów, przeprowadzono testy dla następującej rodziny systemów:

$$x(t+1) = x(t) + a \sin(2x(t)) + w(t)$$

$$z(t+1) = x(t+1) + v(t+1) \quad (3.4)$$

$a \in 0..20$, $w(t) \sim \mathcal{N}(0, Q)$, $v(t) \sim \mathcal{N}(0, R)$. Model procesu staje się bardziej nieliniowy dla większych wartości parametru a (Rysunek 3.2).

Dla każdej wartości parametru a wykonano 100 iteracji algorytmów rozszerzonego filtru Kalmana oraz filtru Kalmana Gaussa-Hermite'a stopnia 2, 3 i 5. Każda iteracja zakładała wykonanie 100 kroków predykcji oraz korekcji przy zastosowaniu danego algorytmu, którego działanie było oceniane za pomocą wartości wskaźnika RMSE. Zbadano również średni czas potrzebny na wykonanie zadania dla poszczególnych filtrów. W celu uzyskania takich samych wartości z generatora liczb losowych dla każdego filtra, przed uruchomieniem symulacji generator był inicjowany tym samym zarodkiem. Przyjęto następujące wartości parametrów symulacji i filtracji: $Q = 10$, $R = 10$, $x(0) = 1$, $P(0|0) = 1$.



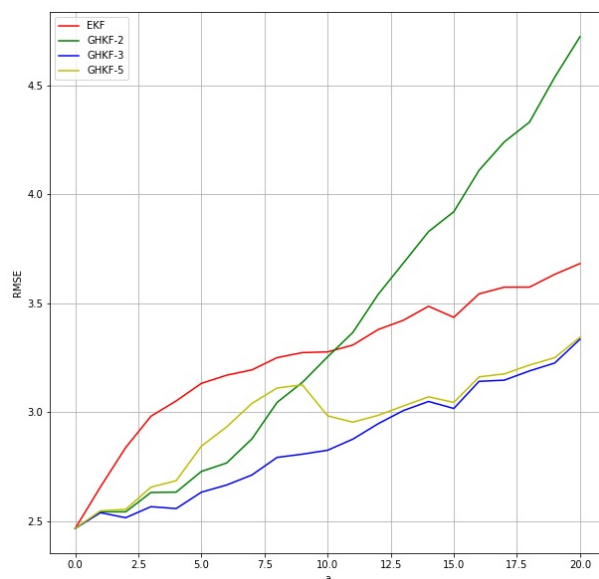
Rys. 3.2. Funkcje w modelu procesu użyte do przetestowania działania filtrów dla kilku wartości parametru a

Tabela 3.1. Średnie czasy działania algorytmów

EKF	GHKF-2	GHKF-3	GHKF-5
0.3212 s	0.3331 s	0.3372 s	0.3469 s

Na rysunku 3.3 pokazano średnie wartości RMSE dla różnych rodzajów filtrów. Dla parametru $a = 0$ modele procesu oraz pomiarów są modelami liniowymi i rezultaty filtracji były takie same. Dla innych wartości a w działaniu filtrów można zaobserwować różnice. Z postawionym zadaniem najlepiej poradził sobie filtr Kalmana Gaussa-Hermite’a stopnia 3, uzyskując najniższe średnie wartości RMSE dla każdego a . Rozszerzony filtr Kalmana uzyskiwał RMSE gorsze o około 0.45 dla bardziej nieliniowych modeli. Dwupunktowa aproksymacja wykorzystywana przez GHKF-2 okazała się mało dokładna i dla $a > 10$ uzyskane wyniki są zdecydowanie najslabsze. Zwiększenie liczby węzłów do 5 również nie przyniosło poprawy wyników - rezultaty uzyskane przez GHKF-5 są zbliżone do GHKF-3 dla $a > 12$, natomiast dla mniejszych a są nawet gorsze.

Średni czas na wykonanie postawionego zadania nieco różnił się dla poszczególnych filtrów (Tabela 3.1). Najszybszy był rozszerzony filtr Kalmana, natomiast w przypadku filtru Kalmana Gaussa-Hermite’a im więcej węzłów było wykorzystywanych, tym średni czas był większy. Największa różnica, między EKF i GHKF-5, nie przekroczyła 0.03 s, czyli 9.4%.



Rys. 3.3. Średnie wartości RMSE dla testowanych filtrów dla różnych wartości parametru a

3.2. Problemy praktyczne

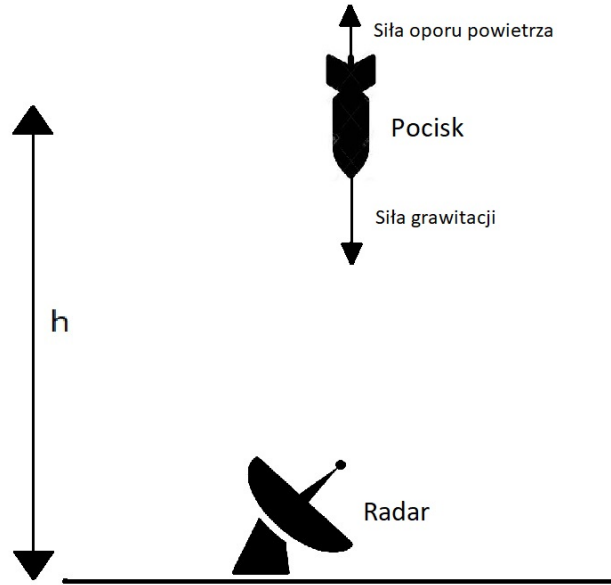
Problemy praktyczne: śledzenie pocisku balistycznego, dwuwymiarowe śledzenie ruchu obiektu, śledzenie wyłącznie z wykorzystaniem namiaru

3.2.1. Śledzenie pocisku balistycznego

Problem śledzenia pocisku balistycznego jest kluczowy dla skutecznego przechwycania pocisku, co ma duże znaczenie dla kwestii bezpieczeństwa i obronności. Rozważany jest scenariusz, kiedy pocisk balistyczny ponownie wchodzi w atmosferę Ziemi po przemierzeniu dużego dystansu, jego prędkość jest bardzo duża, a czas do uderzenia w ziemię stosunkowo niewielki. Problem ten jest uznawany za trudny ze względu na: szum pomiaru, nieliniowości w modelu procesu lub pomiaru oraz brak informacji o kształcie i wielkości pocisku. [7][8].

Założono, że pocisk spada pionowo na ziemię, tak jak pokazano na rysunku 3.4, opór powietrza i grawitacja działają w linii prostej, siła nośna działająca na pocisk jest pomijalnie mała, ziemia jest płaska i stacjonarna, a grawitacja nie zależy od wysokości. Bazując na powyższych założeniach, ruch obiektu można opisać następującymi równaniami [9]:

$$\begin{aligned}
 \dot{h} &= -v \\
 \dot{v} &= -\frac{\rho(h)gv^2}{2\beta} + g \\
 \dot{\beta} &= 0
 \end{aligned}
 \tag{3.5}$$



Rys. 3.4. Scenariusz śledzenia pocisku balistycznego przez radar umieszczony na ziemi

gdzie g to przyspieszenie ziemskie (9.81 m/s^2), h to wysokość, na jakiej znajduje się pocisk (w m), v to prędkość pocisku (m/s), natomiast jako β oznaczony jest współczynnik balistyczny. Gęstość powietrza $\rho(h) = a_1 e^{-a_2 h}$ jest wykładniczą funkcją wysokości, gdzie $a_1 = 1.754$ i $a_2 = 1.49 \cdot 10^{-4}$. Po dyskretyzacji i po uwzględnieniu szumu model procesu dla wektora zmiennych stanu $\mathbf{x} = [h \ v \ \beta]^T$ wygląda następująco:

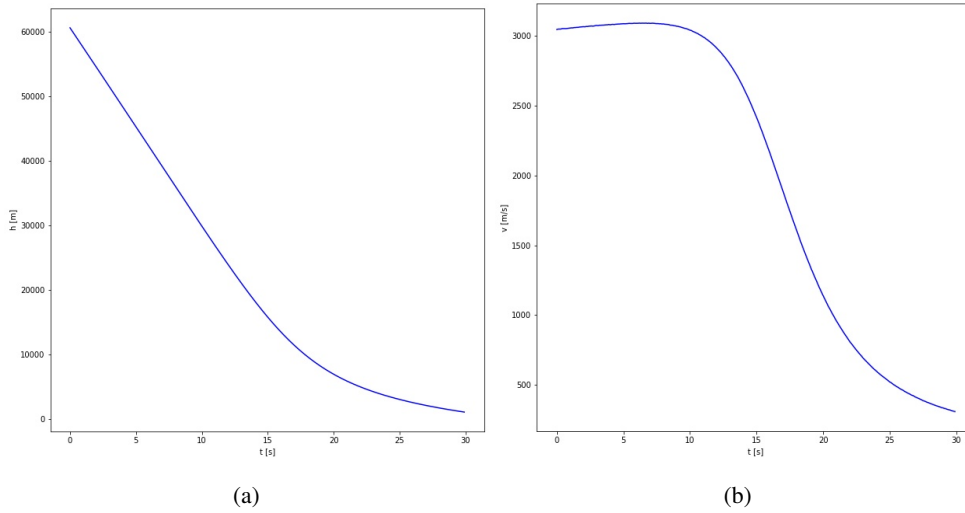
$$\begin{aligned} h(t+1) &= h(t) - v(t)T + w_1(t) \\ v(t+1) &= v(t) - \frac{\rho(h(t))g v(t)^2}{2\beta(t)}T + gT + w_2(t) \\ \beta(t+1) &= \beta(t) + w_3(t) \end{aligned} \quad (3.6)$$

Ze względu na siłę oporu powietrza, dynamika obiektu jest silnie nieliniowa. Szum procesu $\mathbf{w}(t)$ jest szumem gaussowskim o średniej $\mathbf{0}$ i macierzy kowariancji \mathbf{Q} ,

$$\mathbf{Q} = \begin{bmatrix} q_3 T^3/3 & q_3 T^2/2 & 0 \\ q_3 T^2/2 & q_3 T & 0 \\ 0 & 0 & q_4 T \end{bmatrix} \quad (3.7)$$

gdzie parametry q_3 (w m^2/s^3) oraz q_4 (w $kg^2 m^{-2} s^{-5}$) należy dostroić dla konkretnego systemu.

Na rysunku 3.5 przedstawiono typowe przebiegi wysokości oraz prędkości pocisku niezakłócone szumem procesu ($q_3 = q_4 = 0$). Przez kilka początkowych sekund prędkość nie zmienia się znacząco. Potem jednak gęstość powietrza zwiększa się i siła oporu spowalnia spadający obiekt. W końcu pocisk uzyskuje stałą prędkość, kiedy siła grawitacji i opór powietrza równoważą się. Początkowe wartości



Rys. 3.5. Typowe przebiegi wysokości (3.5a) oraz prędkości (3.5b) pocisku

wysokości, prędkości i współczynnika balistycznego ustawiono odpowiednio na 60960 m , 3048 m/s i 19161 kg/ms^2 . Czas próbkowania T wynosił 0.1 s .

Pozycja pocisku jest mierzona przy użyciu radaru umieszczonego na ziemi. Model pomiaru jest modelem liniowym, zakłócanym przez szum gaussowski o średniej 0 i wariancji R .

$$z(t+1) = \mathbf{H}\mathbf{x}(t+1) + \nu(t+1) \quad (3.8)$$

gdzie $\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$ Powyższy problem został rozwiązany przy użyciu rozszerzonego filtru Kalmana oraz filtru Kalmana Gaussa-Hermite'a. Parametry szumu procesu dla filtrów oraz symulacji ustalono podobnie jak w [7] na $q_3 = q_4 = 5$. Wartość wariancji szumu pomiarowego przyjęto jako $R = 200^2$. TODO: wyniki symulacji (takie same dla EKF i GHKF), do zastanowienia: niepewność beta

3.2.2. Dwuwymiarowe śledzenie ruchu obiektu

W przypadku problemu dwuwymiarowego śledzenia ruchu obiektu pozycja celu jest opisywana przy użyciu kartezjańskiego układu współrzędnych, natomiast pomiary są uzyskiwane w układzie współrzędnych biegunowych. Przyjmując wektor stanu $\mathbf{x} = \begin{bmatrix} x & v_x & a_x & y & v_y & a_y \end{bmatrix}^T$ model procesu wygląda następująco [10]:

$$\begin{aligned} x(t+1) &= x(t) + Tv_x(t) + 0.5T^2a_x(t) + w_1(t) \\ v_x(t+1) &= v_x(t) + Ta_x(t) + w_2(t) \\ a_x(t+1) &= a_x(t) + w_3(t) \\ y(t+1) &= y(t) + Tv_y(t) + 0.5T^2a_y(t) + w_4(t) \\ v_y(t+1) &= v_y(t) + Ta_y(t) + w_5(t) \\ a_y(t+1) &= a_y(t) + w_6(t) \end{aligned} \quad (3.9)$$

gdzie $w_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$, T - okres próbkowania.

Tabela 3.2. Czasy działania algorytmów dla problemu dwuwymiarowego śledzenia obiektu

EKF	GHKF-2	GHKF-3	GHKF-5
0.28 s	0.66 s	4.15 s	92.44 s

Model pomiarowy przyjmuje ogólną postać

$$\mathbf{z}(t+1) = \mathbf{h}(\mathbf{x}(t+1)) + \mathbf{v}(t+1) \quad (3.10)$$

gdzie $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$, natomiast nieliniowa funkcja $\mathbf{h}(\mathbf{x}(t))$ wygląda następująco:

$$\mathbf{h}(\mathbf{x}(t)) = \begin{bmatrix} \rho(t) & \theta(t) \end{bmatrix}^T = \begin{bmatrix} \sqrt{x(t)^2 + y(t)^2} & \arctan(\frac{x(t)}{y(t)}) \end{bmatrix}^T \quad (3.11)$$

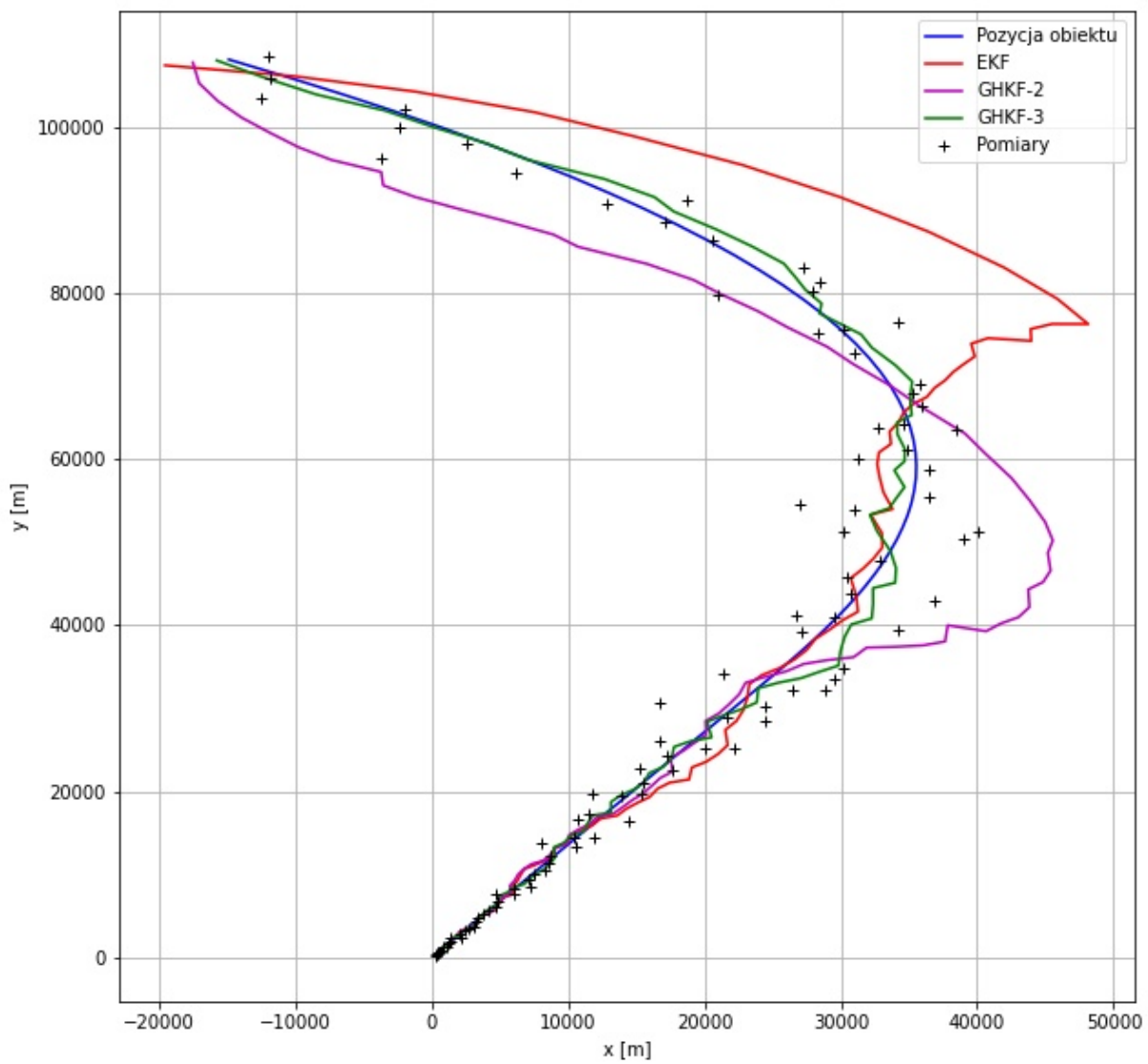
dla mierzonych wartości dystansu $\rho(t)$ oraz azymutu $\theta(t)$.

Podczas testów symulacyjnych wykonanych dla 100 s przyjęto następujące wartości parametrów: $\mathbf{P}(0|0) = \mathbf{I}_{6 \times 6}$, $\mathbf{x}(0|0) = \begin{bmatrix} 200 \text{ m} & 50 \frac{\text{m}}{\text{s}} & 15 \frac{\text{m}}{\text{s}^2} & 100 \text{ m} & 80 \frac{\text{m}}{\text{s}} & 20 \frac{\text{m}}{\text{s}^2} \end{bmatrix}^T$, $\mathbf{R} = \text{diag}(10, 0.03)$, $T = 1 \text{ s}$. Macierz kowariancji szumu procesu $\mathbf{Q} = \text{diag}(0, 0, 1000, 0, 0, 1000)$ wykorzystywana przez filtry została podana w formie uproszczonej [11, s. 247]. Podobnie jak w [10] w symulacji ruchu obiektu przyjęto, że śledzony cel wykonuje skręt w lewo, odejmując w każdym kroku od składowej x przyspieszenia $5 \frac{\text{m}}{\text{s}^2}$ dla $t < 50 \text{ s}$.

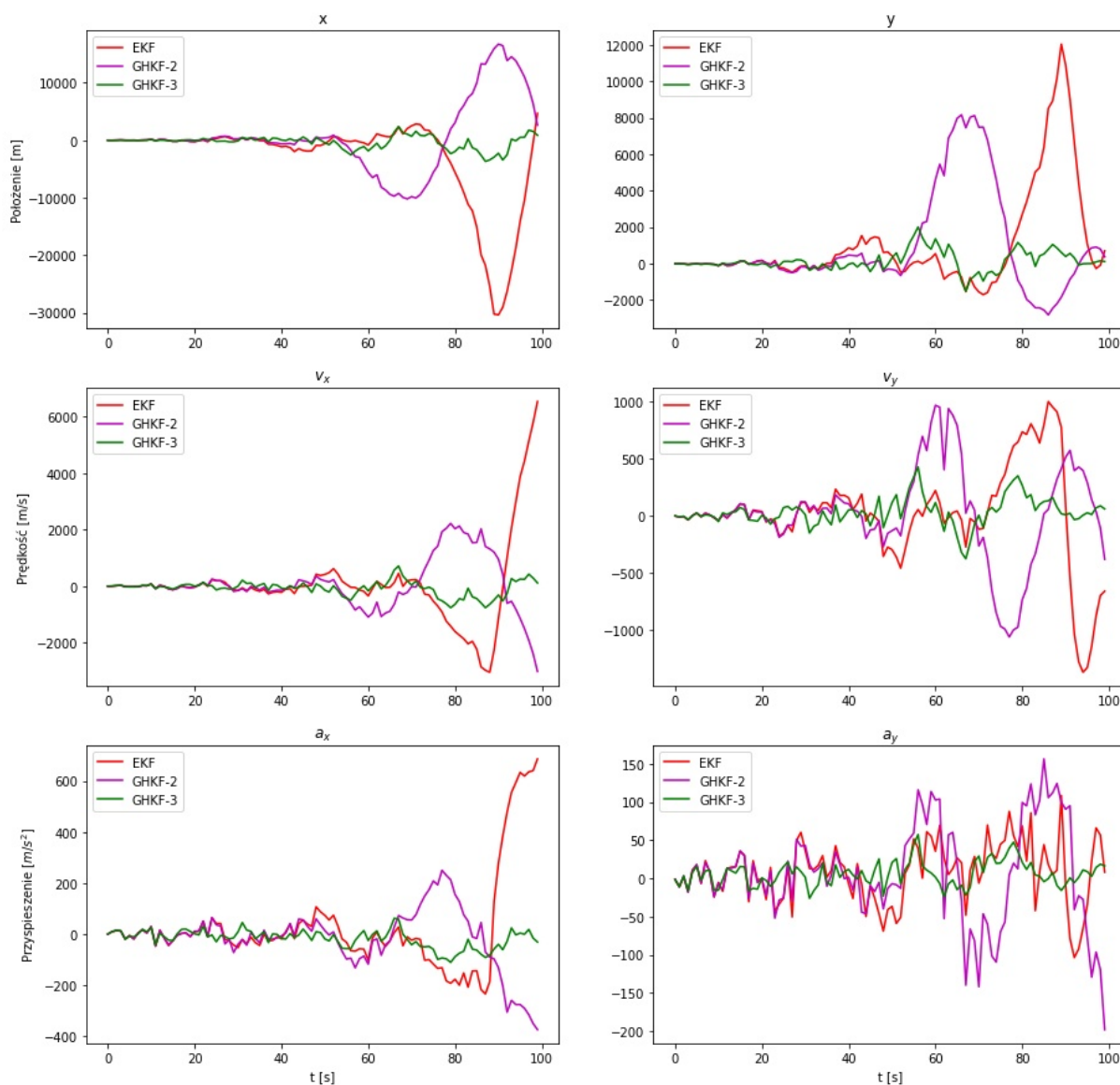
Na rysunku 3.6 przedstawiono położenie śledzonego obiektu oraz wyniki estymacji położenia wykonane przez rozszerzony filtr Kalmana oraz filtr Kalmana Gaussa-Hermite'a stopnia 2 i 3. Rysunek 3.7 przedstawia z kolei różnice pomiędzy rzeczywistymi wartościami zmiennych stanu, a estymatami wyznaczonymi przez filtry. Najlepiej z postawionym zadaniem poradził sobie filtr Kalmana Gaussa-Hermite'a stopnia trzeciego, estymując wszystkie zmienne stanu z niewielkim błędem przez cały czas trwania symulacji. Rozszerzony filtr Kalmana słabo poradził sobie z estymacją położenia podczas zmiany przyspieszenia, która nie była uwzględniona w modelu procesu. Estymowane wartości składowych x prędkości i przyspieszenia na koniec symulacji mocno odbiegały od wartości rzeczywistych. Dwupunktowa aproksymacja rozkładów używana przez GHKF-2 również okazała się mało dokładna i, podobnie jak w przypadku EKF, różnice pomiędzy wartościami rzeczywistymi, a estymowanymi są znaczące. Dla postawionego problemu badano również filtr Kalmana Gaussa-Hermite'a stopnia 5, jednak uzyskiwane wyniki były niemal identyczne jak w przypadku GHKF-3.

Czasy potrzebne do estymacji stanu przy użyciu różnych filtrów przedstawiono w tabeli 3.2. Pomiedzy algorytmami występowały znaczące różnice wynikające z dużej liczby zmiennych stanu w badanym problemie. Najkrótszy czas wystąpił dla rozszerzonego filtru Kalmana, natomiast dla algorytmów Gaussa-Hermite'a wzrastał bardzo szybko wraz ze wzrostem stopnia algorytmu i ostatecznie GHKF-5 okazał się ponad 300 razy wolniejszy od EKF.

3.2.3. Śledzenie wyłącznie z wykorzystaniem namiaru



Rys. 3.6. Położenie obiektu estymowane przez filtry



Rys. 3.7. Różnice pomiędzy rzeczywistymi wartościami zmiennych stanu, a estymowanymi przez filtry

4. Eksperyment z wykorzystaniem wahadła reakcyjnego

Opis problemu - sprawozdanie; problemy: wykrywanie końca wahadła, przeliczenie współrzędnych na ekranie na współrzędne w rzeczywistości, synchronizacja klatek filmu z pomiarami wartości sterowania i rzeczywistego wychylenia; zdjęcia z eksperymentu, prezentacja i omówienie wyników.

5. Podsumowanie

Bibliografia

- [1] Dan Crisan. „The stochastic filtering problem: a brief historical account”. W: *Journal of Applied Probability* 51.A (2014), 13–22. DOI: [10.1239/jap/1417528463](https://doi.org/10.1239/jap/1417528463).
- [2] Simo Sarka. „Bayesian Filtering and Smoothing”. W: Cambridge University Press, 2013.
- [3] Wolfram Burgard Sebastian Thrun Dieter Fox. „Probabilistic Robotics”. W: 1999.
- [4] Ienkarán Arasaratnam, Simon Haykin i Robert J. Elliott. „Discrete-Time Nonlinear Filtering Algorithms Using Gauss–Hermite Quadrature”. W: *Proceedings of the IEEE* 95.5 (2007), s. 953–977. DOI: [10.1109/JPROC.2007.894705](https://doi.org/10.1109/JPROC.2007.894705).
- [5] Xiaohan Liu, Chenglin Wen i Xiaohui Sun. „Design Method of High-Order Kalman Filter for Strong Nonlinear System Based on Kronecker Product Transform”. W: *Sensors* 22 (sty. 2022), s. 653. DOI: [10.3390/s22020653](https://doi.org/10.3390/s22020653).
- [6] A. Germani, C. Manes i P. Palumbo. „Polynomial extended Kalman filter”. W: *IEEE Transactions on Automatic Control* 50.12 (2005), s. 2059–2064. DOI: [10.1109/TAC.2005.860256](https://doi.org/10.1109/TAC.2005.860256).
- [7] Shovan Bhaumik, Nikhil Singh i Samar Bhattacharya. „A comparison of several nonlinear filters for ballistic missile tracking on re-entry”. W: sty. 2016. DOI: [10.1109/CMI.2016.7413790](https://doi.org/10.1109/CMI.2016.7413790).
- [8] Dai Hong-de i in. „Performance Comparison of EKF/UKF/CKF for the Tracking of Ballistic Target”. W: *TELKOMNIKA Indonesian Journal of Electrical Engineering* 10 (list. 2012). DOI: [10.11591/telkomnika.v10i7.1564](https://doi.org/10.11591/telkomnika.v10i7.1564).
- [9] Nitish Kumar Singh, Shovan Bhaumik i Samar Bhattacharya. „Tracking of ballistic target on re-entry using ensemble Kalman filter”. W: *2012 Annual IEEE India Conference (INDICON)*. 2012, s. 508–513. DOI: [10.1109/INDCON.2012.6420671](https://doi.org/10.1109/INDCON.2012.6420671).
- [10] S. Konatowski, P. Kaniewski i J. Matuszewski. „Comparison of estimation accuracy of EKF, UKF and PF filters”. W: *Annual of Navigation* No. 23 (2016), 69–87. ISSN: 1640-8632.
- [11] Roger Labbe. *Kalman and Bayesian Filters in Python*. <https://github.com/rlabbe/Kalman-and-Bayesian-Filters-in-Python>. 2014.