AKADEMIA GÓRNICZO-HUTNICZA IM. STANISŁAWA STASZICA W KRAKOWIE

Wydział Elektrotechniki, Automatyki, Informatyki i Inżynierii Biomedycznej

Praca dyplomowa

Gaussowskie aproksymacje filtrów optymalnych
English title

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. Algorytmy filtracji

Termin *filtracja optymalna* odnosi się do zestawu metod, które mogą być używane do estymacji stanu systemu 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 obarczone szumem, którego obecność oznacza, że obserwacje są niepewne; nawet w przypadku znajomości prawdziwego stanu systemu nie byłyby one jego deterministycznymi funkcjami, ale posiadały 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 zachowanie obiektu nie jest prawdziwie, wewnętrznie losowe, ale w celu przedstawienia niepewności modelu używany jest aparat matematyczny teorii prawdopodobieństwa.

Zadanie filtracji optymalnej można zaklasyfikować jako problem inwersji statystycznej, gdzie nieznaną wielkością jest wektorowy szereg czasowy $\{\boldsymbol{x_0}, \boldsymbol{x_1}, \boldsymbol{x_2}, \dots\}$ obserwowany poprzez zbiór zaszumionych pomiarów $\{\boldsymbol{z_1}, \boldsymbol{z_2}, \boldsymbol{z_3}, \dots\}$, przy obecności sygnałów sterujących $\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}, \dots\}$ Celem wspomnianej inwersji statystycznej jest oszacowanie ukrytych stanów $\boldsymbol{x_{0:T}} = \{\boldsymbol{x_0}, \dots, \boldsymbol{x_T}\}$ na podstawie pomiarów $\boldsymbol{z_{1:T}} = \{\boldsymbol{z_1}, \dots, \boldsymbol{z_T}\}$ i dostarczanych sygnałów sterujących $\boldsymbol{u_{1:T}} = \{\boldsymbol{u_1}, \dots, \boldsymbol{u_T}\}$. 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 (1.1).

$$p(\boldsymbol{x}_{0:T}|\boldsymbol{z}_{1:T}, \boldsymbol{u}_{1:T}) = \frac{p(\boldsymbol{z}_{1:T}|\boldsymbol{x}_{0:T}, \boldsymbol{u}_{1:T})p(\boldsymbol{x}_{0:T}|\boldsymbol{u}_{1:T})}{p(\boldsymbol{z}_{1:T}|\boldsymbol{u}_{1:T})} \tag{1.1}$$

Gdzie:

- $p(\boldsymbol{x}_{0:T}|\boldsymbol{u}_{1:T})$ to rozkład a priori zdefiniowany przez model dynamiczny,
- $p(\boldsymbol{z}_{1:T}|\boldsymbol{x}_{0:T}, \boldsymbol{u}_{1:T})$ to wiarygodność (prawdopodobieństwo otrzymania danych wartości pomiarów pod warunkiem wartości stanu i sterowania)
- $p(\boldsymbol{z}_{1:T}|\boldsymbol{u}_{1:T})$ to stała normalizacyjna zdefiniowana jako $\int p(\boldsymbol{z}_{1:T}|\boldsymbol{x}_{0:T}, \boldsymbol{u}_{1:T})p(\boldsymbol{x}_{0:T}|\boldsymbol{u}_{1:T})\, d\boldsymbol{x}_{0:T}$

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. Uproszczonym celem obliczeń może być zatem rozkład brzegowy stanu w kroku k przy założeniu znajomości historii pomiarów i wartości sterowania. Możliwe jest również zastosowanie twierdzenia Bayesa do wspomnianego rozkładu (1.2).

$$p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t}, \boldsymbol{u}_{1:t}) = \eta p(\boldsymbol{z}_t|\boldsymbol{x}_t, \boldsymbol{z}_{1:t-1}, \boldsymbol{u}_{1:t}) p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t-1}, \boldsymbol{u}_{1:t}) \tag{1.2}$$

Gdzie η jest stałą normalizującą:

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

Możliwe jest przyjęcie szeregu założeń upraszczających:

– 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):

$$p(\boldsymbol{z}_t|\boldsymbol{x}_t, \boldsymbol{z}_{1:t-1}, \boldsymbol{u}_{1:t}) = p(\boldsymbol{z}_t|\boldsymbol{x}_t) \tag{1.3}$$

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

$$p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t-1}, \boldsymbol{u}_{1:t}) = \int p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{z}_{1:t-1}, \boldsymbol{u}_{1:t}) p(\boldsymbol{x}_{t-1}|\boldsymbol{z}_{1:t-1}, \boldsymbol{u}_{1:t})\, d\boldsymbol{x}_{t-1} \tag{1.4}$$

– 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):

$$p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{z}_{1:t-1}, \boldsymbol{u}_{1:t}) = p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{u}_t) \tag{1.5}$$

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

$$p(\boldsymbol{x}_{t-1}|\boldsymbol{z}_{1:t-1}, \boldsymbol{u}_{1:t}) = p(\boldsymbol{x}_{t-1}|\boldsymbol{z}_{1:t-1}, \boldsymbol{u}_{1:t-1}) \tag{1.6}$$

Wykorzystanie powyższych założeń pozwala na rekurencyjne obliczanie rozkładu z równiania 1.2. 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 $p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t-1}, \boldsymbol{u}_{1:t})$,

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

$$p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t}, \boldsymbol{u}_{1:t}) = \underbrace{\eta p(\boldsymbol{z}_t|\boldsymbol{x}_t) \underbrace{\int p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{u}_t) p(\boldsymbol{x}_{t-1}|\boldsymbol{z}_{1:t-1}, \boldsymbol{u}_{1:t-1})\, d\boldsymbol{x}_{t-1}}_{\text{Predykcja}}}_{\text{Korekcja}} \tag{1.7}$$

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

Równanie 1.7 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: $\boldsymbol{x}_t \sim p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$) oraz drugiego pokazującego rozkład wartości pomiarów dla danego stanu systemu (model obserwacyjny: $\boldsymbol{z}_t \sim p(\boldsymbol{z}_t|\boldsymbol{x}_t)$).

1.1. Filtr Kalmana

Przy założeniu liniowości modeli dynamicznego oraz obserwacyjnego, a także addytywności szumów i normalnego charakteru ich rozkładów, można znaleźć rozwiązanie równiania 1.7 w zwartej formie. Wspomniane modele wyglądają zatem następująco:

$$\begin{aligned} \boldsymbol{x}_t &= \boldsymbol{A}_{t-1}\boldsymbol{x}_{t-1} + \boldsymbol{B}_t\boldsymbol{u}_t + \boldsymbol{q}_{t-1} \\ \boldsymbol{z}_t &= \boldsymbol{H}_t\boldsymbol{x}_t + \boldsymbol{r}_t \end{aligned} \tag{1.8}$$

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

$$\begin{aligned} p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{u}_t) &= \mathcal{N}(\boldsymbol{x}_t|\boldsymbol{A}_{t-1}\boldsymbol{x}_{t-1} + \boldsymbol{B}_t\boldsymbol{u}_t, \boldsymbol{Q}_{t-1}) \\ p(\boldsymbol{z}_t|\boldsymbol{x}_t) &= \mathcal{N}(\boldsymbol{z}_t|\boldsymbol{H}_t\boldsymbol{x}_t, \boldsymbol{R}_t) \end{aligned} \tag{1.9}$$

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

$$\begin{aligned} p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t-1}, \boldsymbol{u}_{1:t}) &= \mathcal{N}(\boldsymbol{x}_t|\bar{\boldsymbol{m}}_t, \bar{\boldsymbol{P}}_t) \\ p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t}, \boldsymbol{u}_{1:t}) &= \mathcal{N}(\boldsymbol{x}_t|\boldsymbol{m}_t, \boldsymbol{P}_t) \\ p(\boldsymbol{z}_t|\boldsymbol{z}_{1:t-1}, \boldsymbol{u}_{1:t}) &= \mathcal{N}(\boldsymbol{z}_t|\boldsymbol{H}_t\bar{\boldsymbol{m}}_t, \boldsymbol{S}_t) \end{aligned} \tag{1.10}$$

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

- Krok predykcji:

$$\begin{aligned}\bar{\boldsymbol{m}}_t =& \mathbf{A}_{t-1}\boldsymbol{m}_{t-1} + \mathbf{B}_t\boldsymbol{u}_t \\ \bar{\mathbf{P}}_t =& \mathbf{A}_{t-1}\mathbf{P}_{t-1}\mathbf{A}_{t-1}^T + \mathbf{Q}_{t-1}\end{aligned} \tag{1.11}$$

- Krok korekcji:

$$\begin{aligned}\mathbf{v}_t =& \mathbf{z}_t - \mathbf{H}_t\bar{\mathbf{m}}_t \\ \mathbf{S}_t =& \mathbf{H}_t\bar{\mathbf{P}}_t\mathbf{H}_t^T + \mathbf{R}_t \\ \mathbf{K}_t =& \bar{\mathbf{P}}_t\mathbf{H}_t^T\mathbf{S}_t^{-1} \\ \mathbf{m}_t =& \bar{\mathbf{m}}_t + \mathbf{K}_t\mathbf{v}_t \\ \mathbf{P}_t =& (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\bar{\mathbf{P}}_t\end{aligned} \tag{1.12}$$

Przewidywane parametry rozkładu $\bar{\boldsymbol{m}}_t$ 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.

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.

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 $O(n^2)$.

1.2. 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 przezwyciężenie 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 $\boldsymbol{g}_t$ i $\boldsymbol{h}_t$:

$$\begin{aligned} \boldsymbol{x}_t &= \boldsymbol{g}_t(\boldsymbol{u}_t, \boldsymbol{x}_{t-1}) + \boldsymbol{q}_{t-1} \\ \boldsymbol{z}_t &= \boldsymbol{h}_t(\boldsymbol{x}_t) + \boldsymbol{r}_t \end{aligned} \tag{1.13}$$

Model przedstawiony w równaniu 1.13 uogólnia liniowy gaussowski model z równania 1.8, wykorzystywany w filtrze Kalmana. Funkcja $\boldsymbol{g}_t$ zastępuje macierze $\boldsymbol{A}_{t-1}$ oraz $\boldsymbol{B}_t$, natomiast $\boldsymbol{h}_t$ występuje w miejsce macierzy $\boldsymbol{H}_t$. W tym przypadku, przy dowolnych funkcjach $\boldsymbol{g}_t$ i $\boldsymbol{h}_t$, otrzymywany rozkład nie jest już normalny. Wykonanie dokładnej aktualizacji estymaty stanu jest niemożliwe dla nieliniowych funkcji $\boldsymbol{g}_t$ i $\boldsymbol{h}_t$, ponieważ algorytm filtru Bayesa z równania 1.7 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 $\boldsymbol{g}_t$ funkcją liniową styczną do $\boldsymbol{g}_t$ 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 $\boldsymbol{h}_t$, 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 1.14).

$$\boldsymbol{g}'_t(\boldsymbol{u}_t, \boldsymbol{x}_{t-1}) := \frac{\partial \boldsymbol{g}_t(\boldsymbol{u}_t, \boldsymbol{x}_{t-1})}{\partial \boldsymbol{x}_{t-1}} \tag{1.14}$$

Zarówno wartość funkcji $\boldsymbol{g}_t$, jak i jej pochodna zależą od wartości argumentu funkcji. W rozszerzonym filtrze Kalmana jako argument wybiera się wartość stanu uznawaną za najbardziej prawdopodobną,

zatem funcja $\boldsymbol{g}_t$ jest aproksymowana wokół $\boldsymbol{m}_{t-1}$ (oraz $\boldsymbol{u}_t$):

$$\begin{aligned}\boldsymbol{g}_t(\boldsymbol{u}_t, \boldsymbol{x}_{t-1}) \approx& \boldsymbol{g}_t(\boldsymbol{u}_t, \boldsymbol{m}_{t-1}) + \boldsymbol{g}'_t(\boldsymbol{u}_t, \boldsymbol{m}_{t-1})(\boldsymbol{x}_{t-1} - \boldsymbol{m}_{t-1}) \\ =& \boldsymbol{g}_t(\boldsymbol{u}_t, \boldsymbol{m}_{t-1}) + \boldsymbol{G}_t(\boldsymbol{x}_{t-1} - \boldsymbol{m}_{t-1})\end{aligned} \tag{1.15}$$

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

EKF stosuje taką samą linearyzację dla funkcji $\boldsymbol{h}$:

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

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

$$\begin{aligned}\boldsymbol{h}_t(\boldsymbol{x}_t) \approx& \boldsymbol{h}_t(\bar{\boldsymbol{m}}_t) + \boldsymbol{h}'_t(\bar{\boldsymbol{m}}_t)(\boldsymbol{x}_t - \bar{\boldsymbol{m}}_t) \\ =& \boldsymbol{h}_t(\bar{\boldsymbol{m}}_t) + \boldsymbol{H}_t(\boldsymbol{x}_t - \bar{\boldsymbol{m}}_t)\end{aligned} \tag{1.16}$$

Modele przedstawione w notacji probabilistycznej wyglądają następująco:

$$\begin{aligned}p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{u}_t) =& \mathcal{N}(\boldsymbol{x}_t|\boldsymbol{g}_t(\boldsymbol{u}_t, \boldsymbol{m}_{t-1}) + \boldsymbol{G}_t(\boldsymbol{x}_{t-1} - \boldsymbol{m}_{t-1}), \boldsymbol{Q}_{t-1}) \\ p(\boldsymbol{z}_t|\boldsymbol{x}_t) =& \mathcal{N}(\boldsymbol{z}_t|\boldsymbol{h}_t(\bar{\boldsymbol{m}}_t) + \boldsymbol{H}_t(\boldsymbol{x}_t - \bar{\boldsymbol{m}}_t), \boldsymbol{R}_t)\end{aligned} \tag{1.17}$$

Podobnie jak w przypadku zwykłego filtru Kalmana, algorytm EKF, przedstawiony w równaniach 1.18 i 1.19, wyznacza potrzebne parametry w krokach predykcji oraz korekcji.

- Krok predykcji:

$$\begin{aligned}\bar{\boldsymbol{m}}_t =& \boldsymbol{g}_t(\boldsymbol{u}_t, \boldsymbol{m}_{t-1}) \\ \bar{\mathbf{P}}_t =& \mathbf{G}_t\mathbf{P}_{t-1}\mathbf{G}_t^T + \mathbf{Q}_{t-1}\end{aligned} \tag{1.18}$$

- Krok korekcji:

$$\begin{aligned}\mathbf{v}_t =& \mathbf{z}_t - \boldsymbol{h}_t(\bar{\boldsymbol{m}}_t) \\ \mathbf{S}_t =& \mathbf{H}_t\bar{\mathbf{P}}_t\mathbf{H}_t^T + \mathbf{R}_t \\ \mathbf{K}_t =& \bar{\mathbf{P}}_t\mathbf{H}_t^T\mathbf{S}_t^{-1} \\ \mathbf{m}_t =& \bar{\mathbf{m}}_t + \mathbf{K}_t\mathbf{v}_t \\ \mathbf{P}_t =& (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\bar{\mathbf{P}}_t\end{aligned} \tag{1.19}$$

Bardziej ogólną sytuacją jest brak addytywności szumu w modelach dynamicznym i obserwacyjnym:

$$\begin{aligned}\boldsymbol{x}_t &= \boldsymbol{g}_t(\boldsymbol{u}_t, \boldsymbol{x}_{t-1}, \boldsymbol{q}_{t-1}) \\ \boldsymbol{z}_t &= \boldsymbol{h}_t(\boldsymbol{x}_t, \boldsymbol{r}_t)\end{aligned} \tag{1.20}$$

W takim przypadku możliwe jest obliczenie Jakobianów po zmiennych stanu oraz składowych szumu, oznaczonych odpowiednio $\boldsymbol{G}_{\boldsymbol{x}t}$ i $\boldsymbol{G}_{\boldsymbol{q}_t}$ dla modelu dynamicznego, a także $\boldsymbol{H}_{\boldsymbol{x}t}$ i $\boldsymbol{H}_{\boldsymbol{r}t}$ dla obserwacji. Aproksymacja następuje, podobnie jak dla przypadku szumu addytywnego, wokół $\boldsymbol{m}_{t-1}$ i $\boldsymbol{u}_t$ (macierze $\mathbf{G}_{\mathbf{x}t}$ i $\mathbf{G}_{\mathbf{q}_t}$) oraz $\bar{\boldsymbol{m}}_t$ (dla macierzy $\mathbf{H}_{\mathbf{x}t}$ i $\mathbf{H}_{\mathbf{r}t}$), a także wokół zerowych wartości składowych szumów. Algorytm rozszerzonego filtru Kalmana przyjmuje wówczas postać:

- Krok predykcji:

$$\begin{aligned}\bar{\boldsymbol{m}}_t =& \boldsymbol{g}_t(\boldsymbol{u}_t, \boldsymbol{m}_{t-1}, \mathbf{0})\\ \bar{\mathbf{P}}_t =& \mathbf{G}_{\mathbf{x}t}\mathbf{P}_{t-1}\mathbf{G}_{\mathbf{x}t}^T + \mathbf{G}_{\mathbf{q}_t}\mathbf{Q}_{t-1}\mathbf{G}_{\mathbf{q}_t}^T\end{aligned} \tag{1.21}$$

- Krok korekcji:

$$\begin{aligned}\mathbf{v}_t =& \mathbf{z}_t - \boldsymbol{h}_t(\bar{\boldsymbol{m}}_t, \mathbf{0})\\ \mathbf{S}_t =& \mathbf{H}_{\mathbf{x}t}\bar{\mathbf{P}}_t\mathbf{H}_{\mathbf{x}t}^T + \mathbf{H}_{\mathbf{r}t}\bar{\mathbf{R}}_t\mathbf{H}_{\mathbf{r}t}^T\\ \mathbf{K}_t =& \bar{\mathbf{P}}_t\mathbf{H}_{\mathbf{x}t}^T\mathbf{S}_t^{-1}\\ \mathbf{m}_t =& \bar{\mathbf{m}}_t + \mathbf{K}_t\mathbf{v}_t\\ \mathbf{P}_t =& (\mathbf{I} - \mathbf{K}_t\mathbf{H}_{\mathbf{x}t})\bar{\mathbf{P}}_t\end{aligned} \tag{1.22}$$

Algorytm rozszerzonego filtru Kalmana stał się najbardziej popularnym narzędziem wykorzystywanym do estymacji stanu systemów. Siła 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 $\boldsymbol{z}_t$, 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.

1.3. Filtr Kalmana Gaussa-Hermite'a

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

$$\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\boldsymbol{m}\\ \boldsymbol{\mu}_{\boldsymbol{M}}\end{bmatrix}, \begin{bmatrix}\boldsymbol{P} & \boldsymbol{C}_{\boldsymbol{M}}\\ \boldsymbol{C}_{\boldsymbol{M}}{}^T & \boldsymbol{S}_{\boldsymbol{M}}\end{bmatrix}\right) \tag{1.23}$$

Gdzie:

$$\begin{aligned}
\boldsymbol{\mu_M} &= \int \boldsymbol{g}(\boldsymbol{x})\mathcal{N}(\boldsymbol{x}|\boldsymbol{m},\boldsymbol{P})\,d\boldsymbol{x} \\
\boldsymbol{S_M} &= \int (\boldsymbol{g}(\boldsymbol{x})-\boldsymbol{\mu_M})(\boldsymbol{g}(\boldsymbol{x})-\boldsymbol{\mu_M})^T\mathcal{N}(\boldsymbol{x}|\boldsymbol{m},\boldsymbol{P})\,d\boldsymbol{x}+\boldsymbol{Q} \\
\boldsymbol{C_M} &= \int (\boldsymbol{x}-\boldsymbol{m})(\boldsymbol{g}(\boldsymbol{x})-\boldsymbol{\mu_M})^T\mathcal{N}(\boldsymbol{x}|\boldsymbol{m},\boldsymbol{P})\,d\boldsymbol{x}
\end{aligned} \tag{1.24}$$

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

$$\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{bmatrix} \sim \mathcal{N}(\begin{bmatrix}\boldsymbol{m}\\ \boldsymbol{\mu_M}\end{bmatrix},\begin{bmatrix}\boldsymbol{P} & \boldsymbol{C_M}\\ \boldsymbol{C_M}^T & \boldsymbol{S_M}\end{bmatrix}) \tag{1.25}$$

Gdzie:

$$\begin{aligned}
\boldsymbol{\mu_M} &= \int \boldsymbol{g}(\boldsymbol{x},\boldsymbol{q})\mathcal{N}(\boldsymbol{x}|\boldsymbol{m},\boldsymbol{P})\mathcal{N}(\boldsymbol{q}|\boldsymbol{0},\boldsymbol{Q})\,d\boldsymbol{x} \\
\boldsymbol{S_M} &= \int (\boldsymbol{g}(\boldsymbol{x},\boldsymbol{q})-\boldsymbol{\mu_M})(\boldsymbol{g}(\boldsymbol{x},\boldsymbol{q})-\boldsymbol{\mu_M})^T\mathcal{N}(\boldsymbol{x}|\boldsymbol{m},\boldsymbol{P})\mathcal{N}(\boldsymbol{q}|\boldsymbol{0},\boldsymbol{Q})\,d\boldsymbol{x}\,d\boldsymbol{q} \\
\boldsymbol{C_M} &= \int (\boldsymbol{x}-\boldsymbol{m})(\boldsymbol{g}(\boldsymbol{x},\boldsymbol{q})-\boldsymbol{\mu_M})^T\mathcal{N}(\boldsymbol{x}|\boldsymbol{m},\boldsymbol{P})\mathcal{N}(\boldsymbol{q}|\boldsymbol{0},\boldsymbol{Q})\,d\boldsymbol{x}\,d\boldsymbol{q}
\end{aligned} \tag{1.26}$$

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

- Krok predykcji:

$$\begin{aligned}
\bar{\boldsymbol{m}}_t &= \int \boldsymbol{f}(\boldsymbol{x}_{t-1})\mathcal{N}(\boldsymbol{x}_{t-1}|\boldsymbol{m}_{t-1},\boldsymbol{P}_{t-1})\,d\boldsymbol{x}_{t-1} \\
\bar{\mathbf{P}}_t &= \int (\boldsymbol{f}(\boldsymbol{x}_{t-1})-\bar{\boldsymbol{m}}_t)(\boldsymbol{f}(\boldsymbol{x}_{t-1})-\bar{\boldsymbol{m}}_t)^T\mathcal{N}(\boldsymbol{x}_{t-1}|\boldsymbol{m}_{t-1},\boldsymbol{P}_{t-1})\,d\boldsymbol{x}_{t-1}+\boldsymbol{Q}_{t-1}
\end{aligned} \tag{1.27}$$

- Krok korekcji:

$$\begin{aligned}
\boldsymbol{\mu}_t &= \int \boldsymbol{h}(\boldsymbol{x}_t)\mathcal{N}(\boldsymbol{x}_t,\bar{\boldsymbol{m}}_t,\bar{\boldsymbol{P}_t})d\boldsymbol{x}_t, \\
\boldsymbol{S}_t &= \int (\boldsymbol{h}(\boldsymbol{x}_t)-\boldsymbol{\mu}_t)(\boldsymbol{h}(\boldsymbol{x}_t)-\boldsymbol{\mu}_t)^T\mathcal{N}(\boldsymbol{x}_t,\bar{\boldsymbol{m}}_t,\bar{P}_t)d\boldsymbol{x}_t+\boldsymbol{R}_t, \\
\boldsymbol{C}_t &= \int (\boldsymbol{x}_t-\bar{\boldsymbol{m}}_t)((\boldsymbol{h}(\boldsymbol{x}_t)-\boldsymbol{\mu}_t))^T\mathcal{N}(\boldsymbol{x}_t,\bar{\boldsymbol{m}}_t,\bar{P}_t)d\boldsymbol{x}_t
\end{aligned} \tag{1.28}$$

$$\begin{aligned}
\boldsymbol{K}_t &= \boldsymbol{C}_t\boldsymbol{S}_t^{-1} \\
\boldsymbol{P}_t &= \bar{\mathbf{P}}_t-\boldsymbol{K}_t\boldsymbol{S}_t\boldsymbol{K}_t^T \\
\boldsymbol{m}_t &= \bar{\boldsymbol{m}}_t+\boldsymbol{K}_t(\boldsymbol{z}_t-\boldsymbol{\mu}_t)
\end{aligned} \tag{1.29}$$

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

- Krok predykcji:

$$\begin{aligned}
\bar{\boldsymbol{m}}_t &= \int \boldsymbol{f}(\boldsymbol{x}_{t-1}, \boldsymbol{q}_{t-1}) \mathcal{N}(\boldsymbol{x}_{t-1} | \boldsymbol{m}_{t-1}, \boldsymbol{P}_{t-1}) \mathcal{N}(\boldsymbol{q}_{t-1} | \boldsymbol{0}, \boldsymbol{Q}_{t-1})\, d\boldsymbol{x}_{t-1}\, d\boldsymbol{q}_{t-1} \\
\bar{\mathbf{P}}_t &= \int (\boldsymbol{f}(\boldsymbol{x}_{t-1}, \boldsymbol{q}_{t-1}) - \bar{\boldsymbol{m}}_t)(\boldsymbol{f}(\boldsymbol{x}_{t-1}, \boldsymbol{q}_{t-1}) - \bar{\boldsymbol{m}}_t)^T \\
&\quad \times \mathcal{N}(\boldsymbol{x}_{t-1} | \boldsymbol{m}_{t-1}, \boldsymbol{P}_{t-1}) \mathcal{N}(\boldsymbol{q}_{t-1} | \boldsymbol{0}, \boldsymbol{Q}_{t-1})\, d\boldsymbol{x}_{t-1}\, d\boldsymbol{q}_{t-1}
\end{aligned} \tag{1.30}$$

- Krok korekcji:

$$\begin{aligned}
\boldsymbol{\mu}_t &= \int \boldsymbol{h}(\boldsymbol{x}_t, \boldsymbol{r}_t) \mathcal{N}(\boldsymbol{x}_t | \bar{\boldsymbol{m}}_t, \bar{\boldsymbol{P}_t}) \mathcal{N}(\boldsymbol{r}_t | \boldsymbol{0}, \boldsymbol{R}_t) d\boldsymbol{x}_t d\boldsymbol{r}_t \\
\boldsymbol{S}_t &= \int (\boldsymbol{h}(\boldsymbol{x}_t, \boldsymbol{r}_t) - \boldsymbol{\mu}_t)(\boldsymbol{h}(\boldsymbol{x}_t, \boldsymbol{r}_t) - \boldsymbol{\mu}_t)^T \mathcal{N}(\boldsymbol{x}_t, \bar{\boldsymbol{m}}_t, \bar{P}_t) \mathcal{N}(\boldsymbol{r}_t | \boldsymbol{0}, \boldsymbol{R}_t) d\boldsymbol{x}_t d\boldsymbol{r}_t \\
\boldsymbol{C}_t &= \int (\boldsymbol{x}_t, -\bar{\boldsymbol{m}}_t)(\boldsymbol{h}(\boldsymbol{x}_t, \boldsymbol{r}_t) - \boldsymbol{\mu}_t)^T \mathcal{N}(\boldsymbol{x}_t, \bar{\boldsymbol{m}}_t, \bar{P}_t) \mathcal{N}(\boldsymbol{r}_t | \boldsymbol{0}, \boldsymbol{R}_t) d\boldsymbol{x}_t d\boldsymbol{r}_t \\
\boldsymbol{K}_t &= \boldsymbol{C}_t \boldsymbol{S}_t^{-1} \\
\boldsymbol{P}_t &= \bar{\mathbf{P}}_t - \boldsymbol{K}_t \boldsymbol{S}_t \boldsymbol{K}_t^T \\
\boldsymbol{m}_t &= \bar{\boldsymbol{m}}_t + \boldsymbol{K}_t(\boldsymbol{z}_t - \boldsymbol{\mu}_t)
\end{aligned} \tag{1.31}$$

Powyższe ogólne równania filtru Gaussa są raczej teoretycznymi konstrukcjami, a nie praktycznymi algorytmami filtracji. Należy przyjąć pewną metodę rozwiązywania potrzebnych całek występujących w formie 1.32, aby uzyskać funkcjonalny algorytm.

$$\begin{aligned}
&\int \boldsymbol{g}(\boldsymbol{x}) \mathcal{N}(\boldsymbol{x} | \boldsymbol{m}, \boldsymbol{P}))d\boldsymbol{x} \\
&\quad = \frac{1}{(2\pi)^{n/2} (\det \boldsymbol{P})^{1/2}} \int \boldsymbol{g}(\boldsymbol{x}) \exp(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{m})^T \boldsymbol{P}^{-1} (\boldsymbol{x} - \boldsymbol{m}))d\boldsymbol{x}
\end{aligned} \tag{1.32}$$

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:

$$\int g(x) \mathcal{N}(x | 0, 1) dx \approx \sum_{i=1}^{p} W_i g(\xi_i) \tag{1.33}$$

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 bedących 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:

$$H_p(x) = (-1)^p e^{x^2/2} \frac{d^p}{dx^p} e^{-x^2/2} \tag{1.34}$$

Pierwsze wielomiany Hermite'a to:

$$\begin{aligned} H_0(x) =& 1, \\ H_1(x) =& x, \\ H_2(x) =& x^2 - 1, \\ H_3(x) =& x^3 - 3x, \\ H_4(x) =& x^4 - 6x^2 + 3, \end{aligned} \tag{1.35}$$

a kolejne mogą być obliczone z zależności rekurencyjnej:

$$H_{p+1}(x) = xH_p(x) - pH_{p-1}(x) \tag{1.36}$$

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

$$W_i = \frac{p!}{p^2[H_{p-1}(\xi_i)]^2} \tag{1.37}$$

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

$$\int g(x)\mathcal{N}(x|m, P)dx = \int g(P^{1/2}\xi + m)\mathcal{N}(\xi|0, 1)d\xi \tag{1.38}$$

W takim przypadku przybliżenie całki wygląda następująco:

$$\int g(x)\mathcal{N}(x|m, P)dx \approx \sum_{i=1}^{p} W_i g(P^{1/2}\xi_i + m) \tag{1.39}$$

Równianie 1.39 może być dalej uogólnione na przypadek wielowymiarowy, poprzez zdefiniowanie wektora nowych zmiennych $\boldsymbol{\xi}$ oraz zastosowanie rozkładu Choleskiego do macierzy kowariancji ($\boldsymbol{P} = \sqrt{\boldsymbol{P}}\sqrt{\boldsymbol{P}}^T$):

$$\boldsymbol{x} = \boldsymbol{m} + \sqrt{\boldsymbol{P}}\boldsymbol{\xi} \tag{1.40}$$

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

$$\int \boldsymbol{g}(\boldsymbol{x})\mathcal{N}(\boldsymbol{x}|\boldsymbol{m}, \boldsymbol{P})d\boldsymbol{x} = \int \boldsymbol{g}(\boldsymbol{m} + \sqrt{\boldsymbol{P}}\boldsymbol{\xi})\mathcal{N}(\boldsymbol{x}|\boldsymbol{0}, \boldsymbol{I})d\boldsymbol{\xi} \tag{1.41}$$

Całka otrzymana w równianiu 1.41 może być przedstawiona jako całka iterowana i każdą z całek wchodzących w skład całki iterowanej można aproksymować z wykorzystaniem kwadratury Gaussa-Hermite'a:

$$\begin{aligned} &\int \boldsymbol{g}(\boldsymbol{m} + \sqrt{\boldsymbol{P}}\boldsymbol{\xi})\mathcal{N}(\boldsymbol{x}|\boldsymbol{0}, \boldsymbol{I})d\boldsymbol{\xi} \\ &= \int \cdots \int \boldsymbol{g}(\boldsymbol{m} + \sqrt{\boldsymbol{P}}\boldsymbol{\xi})\mathcal{N}(\xi_1, 0, 1) \times \cdots \times \mathcal{N}(\xi_n, 0, 1)d\xi_1 \ldots d\xi_n \\ &\approx \sum_{i_1,\ldots,i_n} \boldsymbol{W}_{(i_1,\ldots,i_n)} \boldsymbol{g}(\boldsymbol{m} + \sqrt{\boldsymbol{P}}\xi_{(i_1,\ldots,i_n)}) \end{aligned} \tag{1.42}$$

Węzły powstają jako iloczyn kartezjański jednowymiarowych punktów sigma, $\xi_{(i_1,\ldots,i_n)} = \begin{pmatrix}\xi_1 & \ldots & \xi_n\end{pmatrix}^T$, natomiast wielowymiarowe wagi są tworzone poprzez pomnożenie jednowymiarowych wag odpowiadających węzłom: $\boldsymbol{W}_{(i_1,\ldots,i_n)} = \boldsymbol{W}_{i_1} \times \cdots \times \boldsymbol{W}_{i_n}$.

Całkowanie metodą Gaussa-Hermite-a jest dokładne dla jednomianów $x_1^{d_1} x_2^{d_2} \ldots x_n^{d_n}$ 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.

Zastosowanie metody Gaussa-Hermite-a do obliczenia całek z równań 1.27 i 1.28 daje w wyniku algorytm filtru Kalmana Gaussa-Hermite-a (ang. *Gauss-Hermite Kalman Filter*, GHKF) dla przypadku szumu addytywnego:

- Krok predykcji:

$$\begin{aligned}
\boldsymbol{\chi}_{t-1}^{(i_1,\ldots,i_n)} &= \boldsymbol{m}_{t-1} + \sqrt{\boldsymbol{P}_{t-1}}\boldsymbol{\xi}_{(i_1,\ldots,i_n)} \\
\hat{\boldsymbol{\chi}}_t^{(i_1,\ldots,i_n)} &= \boldsymbol{f}(\boldsymbol{\chi}_{t-1}^{(i_1,\ldots,i_n)}) \\
\bar{\boldsymbol{m}}_t &= \sum_{i_1,\ldots,i_n} \boldsymbol{W}_{(i_1,\ldots,i_n)} \hat{\boldsymbol{\chi}}_t^{(i_1,\ldots,i_n)} \\
\bar{\boldsymbol{P}}_t &= \sum_{i_1,\ldots,i_n} \boldsymbol{W}_{(i_1,\ldots,i_n)} (\hat{\boldsymbol{\chi}}_t^{(i_1,\ldots,i_n)} - \bar{\boldsymbol{m}}_t)(\hat{\boldsymbol{\chi}}_t^{(i_1,\ldots,i_n)} - \bar{\boldsymbol{m}}_t)^T + \boldsymbol{Q}_{t-1}
\end{aligned} \tag{1.43}$$

- Krok korekcji:

$$\begin{aligned}
\bar{\boldsymbol{\chi}}_t^{(i_1,\ldots,i_n)} &= \bar{\boldsymbol{m}}_t + \sqrt{\bar{\boldsymbol{P}}_t}\boldsymbol{\xi}_{(i_1,\ldots,i_n)} \\
\hat{\boldsymbol{\Psi}}_t^{(i_1,\ldots,i_n)} &= \boldsymbol{h}(\bar{\boldsymbol{\chi}}_t^{(i_1,\ldots,i_n)}) \\
\boldsymbol{\mu}_t &= \sum_{i_1,\ldots,i_n} \boldsymbol{W}_{(i_1,\ldots,i_n)} \hat{\boldsymbol{\Psi}}_t^{(i_1,\ldots,i_n)} \\
\boldsymbol{S}_t &= \sum_{i_1,\ldots,i_n} \boldsymbol{W}_{(i_1,\ldots,i_n)} (\hat{\boldsymbol{\Psi}}_t^{(i_1,\ldots,i_n)} - \boldsymbol{\mu}_t)(\hat{\boldsymbol{\Psi}}_t^{(i_1,\ldots,i_n)} - \boldsymbol{\mu}_t)^T + \boldsymbol{R}_t \\
\boldsymbol{C}_t &= \sum_{i_1,\ldots,i_n} \boldsymbol{W}_{(i_1,\ldots,i_n)} (\bar{\boldsymbol{\chi}}_t^{(i_1,\ldots,i_n)} - \bar{\boldsymbol{m}}_t)(\hat{\boldsymbol{\Psi}}_t^{(i_1,\ldots,i_n)} - \boldsymbol{\mu}_t)^T \\
\boldsymbol{K}_t &= \boldsymbol{C}_t \boldsymbol{S}_t^{-1} \\
\boldsymbol{P}_t &= \bar{\mathbf{P}}_t - \boldsymbol{K}_t \boldsymbol{S}_t \boldsymbol{K}_t^T \\
\boldsymbol{m}_t &= \bar{\boldsymbol{m}}_t + \boldsymbol{K}_t(\boldsymbol{z}_t - \boldsymbol{\mu}_t)
\end{aligned} \tag{1.44}$$

Filtr dla szumu nieaddytywnego (do zweryfikowania):

- Krok predykcji:

  $\boldsymbol{\xi}_{(i_1,\ldots,i_n)}$ będzie tworzone na podstawie zwiększonej liczby wymiarów (suma wymiarów zmiennych stanu i wektora szumu $\boldsymbol{q}$). $\boldsymbol{m}_{t-1}$ będzie miało także zwiększony wymiar, dla zmiennych

szumu będą tam zera. Macierz $\boldsymbol{P}_{t-1}$ będzie miała dołożoną na przekątnej macierz $\boldsymbol{Q}_{t-1}$, a pozostałe miejsca macierzy (odpowiadające kowariancji zmiennych stanu i zmiennych szumu) będą uzupełnione zerami. Funkcja $\boldsymbol{f}$ będzie przyjmować współrzędne węzłów odpowiadające zmiennym stanu i szumom:

$$\begin{aligned}
&\boldsymbol{\chi}_{t-1}^{(i_1,\ldots,i_n)} = \boldsymbol{m}_{t-1} + \sqrt{\boldsymbol{P}_{t-1}}\boldsymbol{\xi}_{(i_1,\ldots,i_n)} \\
&\hat{\boldsymbol{\chi}}_{t}^{(i_1,\ldots,i_n)} = \boldsymbol{f}(\boldsymbol{\chi}_{t-1}^{(i_1,\ldots,i_n)}) \\
&\bar{\boldsymbol{m}}_t = \sum_{i_1,\ldots,i_n} \boldsymbol{W}_{(i_1,\ldots,i_n)} \hat{\boldsymbol{\chi}}_{t}^{(i_1,\ldots,i_n)}
\end{aligned}$$

Nie będzie już dodawana macierz $\boldsymbol{Q}$:

$$\bar{\boldsymbol{P}}_t = \sum_{i_1,\ldots,i_n} \boldsymbol{W}_{(i_1,\ldots,i_n)} (\hat{\boldsymbol{\chi}}_{t}^{(i_1,\ldots,i_n)} - \bar{\boldsymbol{m}}_t)(\hat{\boldsymbol{\chi}}_{t}^{(i_1,\ldots,i_n)} - \bar{\boldsymbol{m}}_t)^T \tag{1.45}$$

- Krok korekcji:

  Tutaj podobnie jak w kroku predykcji, $\boldsymbol{\xi}_{(i_1,\ldots,i_n)}$ tworzone będzie na podstawie zwiększonej liczby wymiarów (suma wymiarów zmiennych stanu i wektora szumu $\boldsymbol{r}$). $\bar{\boldsymbol{P}}_t$ będzie miało dołożoną macierz $\boldsymbol{R}_t$. $\bar{\boldsymbol{m}}_t$ będzie też uzupełnione zerami. Funkcja $\boldsymbol{h}$ przyjmie współrzędne węzłów odpowiadające i zmiennym stanu, i szumom:

$$\begin{aligned}
&\bar{\boldsymbol{\chi}}_{t}^{(i_1,\ldots,i_n)} = \bar{\boldsymbol{m}}_t + \sqrt{\bar{\boldsymbol{P}}_t}\boldsymbol{\xi}_{(i_1,\ldots,i_n)} \\
&\hat{\boldsymbol{\Psi}}_{t}^{(i_1,\ldots,i_n)} = \boldsymbol{h}(\bar{\boldsymbol{\chi}}_{t}^{(i_1,\ldots,i_n)}) \\
&\boldsymbol{\mu}_t = \sum_{i_1,\ldots,i_n} \boldsymbol{W}_{(i_1,\ldots,i_n)} \hat{\boldsymbol{\Psi}}_{t}^{(i_1,\ldots,i_n)}
\end{aligned}$$

Bez dodawania macierzy $\boldsymbol{R}$:

$$\begin{aligned}
&\boldsymbol{S}_t = \sum_{i_1,\ldots,i_n} \boldsymbol{W}_{(i_1,\ldots,i_n)} (\hat{\boldsymbol{\Psi}}_{t}^{(i_1,\ldots,i_n)} - \boldsymbol{\mu}_t)(\hat{\boldsymbol{\Psi}}_{t}^{(i_1,\ldots,i_n)} - \boldsymbol{\mu}_t)^T \\
&\boldsymbol{C}_t = \sum_{i_1,\ldots,i_n} \boldsymbol{W}_{(i_1,\ldots,i_n)} (\bar{\boldsymbol{\chi}}_{t}^{(i_1,\ldots,i_n)} - \bar{\boldsymbol{m}}_t)(\hat{\boldsymbol{\Psi}}_{t}^{(i_1,\ldots,i_n)} - \boldsymbol{\mu}_t)^T \\
&\boldsymbol{K}_t = \boldsymbol{C}_t \boldsymbol{S}_t^{-1} \\
&\boldsymbol{P}_t = \bar{\mathbf{P}}_t - \boldsymbol{K}_t \boldsymbol{S}_t \boldsymbol{K}_t^T \\
&\boldsymbol{m}_t = \bar{\boldsymbol{m}}_t + \boldsymbol{K}_t(\boldsymbol{z}_t - \boldsymbol{\mu}_t)
\end{aligned} \tag{1.46}$$