



Universität Karlsruhe (TH), Fakultät für Informatik
Institut für Technische Informatik
Lehrstuhl für Intelligente Sensor-Aktor-Systeme (ISAS)
Prof. Dr.-Ing. Uwe D. Hanebeck



Studienarbeit

Mehrstufiges Clustering-Verfahren zur
Komponentenreduktion von Gaußmischdichten

Dominik Itte

17. Juni 2009

Referent: Prof. Dr.-Ing. Uwe D. Hanebeck

Betreuer: Dipl.-Inform. Marco Huber
Dipl.-Phys, Dipl.-Inform. Dennis Schieferdecker

Eidesstattliche Erklärung

Hiermit erkläre ich, die vorliegende Studienarbeit selbstständig angefertigt zu haben. Die verwendeten Quellen sind im Text gekennzeichnet und im Literaturverzeichnis aufgeführt.

Karlsruhe, 17. Juni 2009

Dominik Itte

Inhaltsverzeichnis

Abbildungsverzeichnis	III
Tabellenverzeichnis	V
Nomenklatur	VII
1 Einleitung	1
1.1 Motivation	1
1.2 Aufbau	2
2 Grundlagen	3
2.1 Multivariate Normalverteilung	3
2.2 Multivariate Gaußmischdichte	6
3 Reduktionsverfahren	9
3.1 Gütemaße	11
3.2 Verfahren von West	12
3.3 Verfahren von Salmond	13
3.4 Verfahren von Runnalls	13
3.5 Verfahren von Williams	15
3.6 PGMR Verfahren	16
4 Komponentenreduktion durch Clusterbildung	17
4.1 Clusterverfahren	17
4.1.1 Kmeans	18
4.1.2 Kmeans++	18

4.2	Reduktionsverfahren	22
4.2.1	Initialisierung	23
4.2.2	Clustering	23
4.2.3	Optimierung	25
4.2.4	Gewichtsoptimierung	25
4.2.5	Laufzeiteigenschaften	26
5	Minimierung der Gütefunktion mittels Gradientenabstieg	29
5.1	Gradientenabstiegsverfahren	29
5.2	Schrittweitenbestimmung	30
5.3	Gradient des Gütemaßes	31
5.3.1	Beispiel	32
6	Auswertung und Simulation	35
6.1	Beispielablauf	35
6.2	Simulation 1	41
6.3	Simulation 2	43
6.4	Anwendung	45
7	Schlussfolgerung	49
7.1	Ausblick	49
A	Herleitung des Gradienten	51
	Literaturverzeichnis	63

Abbildungsverzeichnis

1.1	Gaußmischdichte mit 10 Komponenten.	1
2.1	Dichte und Wahrscheinlichkeitsverteilung für die Standardnormalverteilung für eine (oben) und zwei Dimensionen (unten)	5
2.2	Zweidimensionale Gaußmischdichte mit 3 Komponenten.	6
4.1	Datenpunkte bzw. Komponenten der wahren Dichte.	20
4.2	Zufällige Auswahl des ersten Zentrums (rot).	20
4.3	Zuordnung der Komponenten zu ihrem nächstgelegenen Cluster.	20
4.4	Wahl des zweiten Clusters mit max-min-Distanz und erneute Zuordnung.	20
4.5	Wahl des dritten Clusters und neue Zuordnung der Komponenten zum nächsten Zentrum.	20
4.6	Wahl des letzten Zentrums.	20
4.7	Initiale Zentren aus initialReduction mit Kmeans++.	24
4.8	Zuordnung der Datenpunkte zum jeweils nächsten Zentrum mit initialClustering.	24
4.9	Neuberechnung der Zentren mit computeCenters.	24
4.10	Wechsel des Clusters.	24
4.11	Neuberechnete Zentren, mit computeCenters.	24
5.1	Wahre Dichte bestehend aus 2 Komponenten mit Approximation.	33
5.2	Wahre Dichte mit optimierter Approximation durch Gradientenabstieg.	33
6.1	Wahre Dichte bestehend aus 10 Komponenten.	37
6.2	Wahre Dichte bestehend aus 5 Komponenten.	37
6.3	Reduktion mit 6 Komponenten nach Phase 1 mit Abweichung von 2%.	38
6.4	Reduktion mit 3 Komponenten nach Phase 1 mit Abweichung von 15,1%.	38

6.5	Reduktion nach Phase 2 mit Abweichung von 1,5%.	39
6.6	Reduktion nach Phase 2 mit Abweichung von 9,7%.	39
6.7	Reduktion nach Phase 3 mit Abweichung von 0,1%.	40
6.8	Reduktion nach Phase 3 mit Abweichung von 4,7%.	40
6.9	Komponentenreduktion der Dichte mit $D=4$, $G=50$	44
6.10	Komponentenreduktion der Dichte mit $D=8$, $G=50$	44
6.11	Ablaufdiagramm des Filters mit und ohne Komponentenreduktion. In jedem Zeitschritt fallen 4 Messungen an.	46
6.12	Wahre Trajektorien und Hypothese der Reduktionsverfahren im Vergleich.	47
6.13	Erfolgreiche Trackingdauer der Reduktionsverfahren im Vergleich.	48
6.14	Anzahl der Trackverluste der Reduktionsverfahren im Vergleich.	48

Tabellenverzeichnis

3.1	Matrix \mathbf{B} vor der Fusion.	14
3.2	Matrix \mathbf{B} nach der Fusion.	14
3.3	Matrix J_{hh} : Bewertet Selbstähnlichkeit der wahren Dichte.	15
3.4	Matrix J_{hr} : Bewertet Ähnlichkeit zwischen wahrer und reduzierter Dichte.	15
3.5	Matrix J_{rr} : Bewertet Selbstähnlichkeit der reduzierten Dichte.	15
6.1	Abweichung in % für Dichte der Dimension 1.	41
6.2	Abweichung in % für Dichte der Dimension 2.	42
6.3	Abweichung in % für Dichte der Dimension 4.	42
6.4	Abweichung in % für Dichte der Dimension 8.	42
6.5	Zeit in Sekunden für Dichte der Dimension 2.	42

Nomenklatur

\mathbb{N} Menge der natürlichen Zahlen

\mathbb{R} Menge der reellen Zahlen

Skalare

x Skalar

X Zufallsvariable

\tilde{w} Gewicht der wahren Dichtefunktion

w Gewicht der reduzierten Dichtefunktion

μ Erwartungswert

σ Standardabweichung

M_0^k Nulltes Moment der k-ten Dichtefunktion

α_t Schrittweitenparameter zum Zeitschritt t

Vektoren

\underline{x} Spaltenvektor

\underline{x}^T Zeilenvektor

\underline{X} Zufallsvektor

$\underline{\tilde{\mu}}$ Erwartungswertvektor der wahren Dichtefunktion

$\underline{\mu}$ Erwartungswertvektor der reduzierten Dichtefunktion

$\underline{M_1^k}$ Erstes Moment der k-ten Dichtefunktion

$\underline{\tilde{\eta}}$ Parametervektor der wahren Dichtefunktion

$\underline{\eta}$ Parametervektor der reduzierten Dichtefunktion

Matrizen

\mathbf{A}	Matrix
\mathbf{A}^T	Transponierte Matrix
$ A $	Determinante der Matrix \mathbf{A}
$\text{tr}(\mathbf{A})$	Spur der Matrix \mathbf{A}
$\tilde{\Sigma}$	Symmetrische, positiv semidefinite Kovarianzmatrix einer Dichtefunktion
$\tilde{\mathbf{C}}$	Symmetrische, positiv semidefinite Kovarianzmatrix der wahren Dichtefunktion
\mathbf{C}	Symmetrische, positiv semidefinite Kovarianzmatrix der reduzierten Dichtefunktion
\mathbf{M}_2^k	Zweites Moment der k-ten Dichtefunktion

Funktionen

$\tilde{f}(\underline{x})$	Wahre Dichtefunktion
$f(\underline{x})$	Reduzierte Dichtefunktion
$\mathcal{N}_N(\underline{x}, \underline{\mu}, \underline{\Sigma})$	Normalverteilung der N-ten Dimension
G	Gütefunktion
$D(x)$	Kleinster Abstand von x zum nächsten Clusterzentrum
$\text{chol}(\mathbf{A})$	Cholesky-Zerlegung
$\text{sqrt}(x)$	Wurzel
$\text{vec}(\mathbf{A})$	Vektorisierung

Abkürzungen

NISD	<u>N</u> ormalized <u>I</u> ntegral <u>S</u> quare <u>D</u> ifferenz
GMRC	<u>G</u> aussian <u>M</u> ixture <u>R</u> eduction via <u>C</u> lustering

Einleitung

1.1 Motivation

Die Gaußmischfunktion entspricht einer gewichteten Summe von einzelnen Gaußdichten und wird vollständig durch die statistischen Kenngrößen Erwartungswert und Kovarianz der einzelnen Dichtefunktionen, sowie deren Gewichte beschrieben. Häufig wird die Gaußmischfunktion zur Modellierung verschiedenster Funktionen verwendet, dabei ist die Gaußmischfunktion ein universeller Approximator [1]. Ein wichtiger Anwendungsbereich für Gaußmischfunktionen ist die Schätztheorie. In diesem Falle spricht man von einer Gaußmischdichte. In Abbildung 1.1 ist eine Gaußmischdichte mit 10 Komponenten abgebildet.

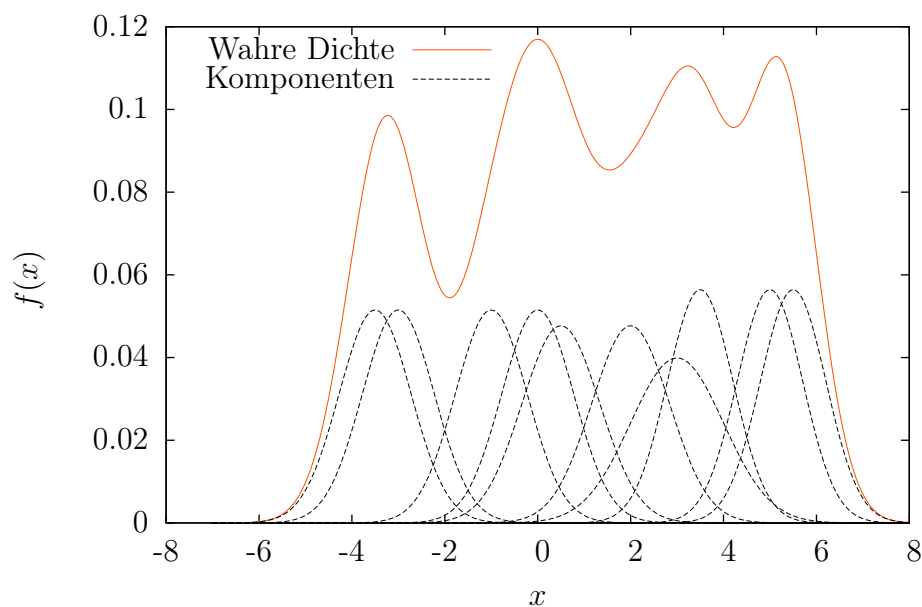


Abbildung 1.1: Gaußmischdichte mit 10 Komponenten.

Hier wird die Gaußmischdichte verwendet um auftretende Dichtefunktionen beliebig genau zu repräsentieren. Dabei ergibt sich ein wesentliches Problem. Wird die Gaußmischdichte bei der Zustandsschätzung verwendet und werden sowohl Likelihood, als auch die priore Wahrscheinlichkeitsdichte durch eine Gaußmischdichte beschrieben, kommt es in jedem Filterschritt zu einer Multiplikation zweier Gaußmischdichten. Als Ergebnis erhält man wiederum eine Gaußmischdichte. In diesem Fall folgt aus der rekursiven Verarbeitung, dass die Anzahl der Komponenten der verwendeten Gaußmischdichte und damit sowohl Speicher-, als auch Rechenaufwand exponentiell wächst.

Daher ist es zwingend notwendig die Anzahl an Komponenten zu reduzieren. Verfahren die dies leisten werden als Gaußmischreduktionsverfahren bezeichnet. Im Rahmen dieser Arbeit wurde ein neuartiges Verfahren [1] zur Reduktion univariater Gaußmischdichten für den multivariaten Fall weiterentwickelt.

Die Grundidee dieses Verfahrens besteht darin unter Verwendung der Clusteranalyse Ähnlichkeiten zwischen den Komponenten der Gaußmischdichte zu finden.

1.2 Aufbau

In Kapitel 2 wird auf die Grundlagen eingegangen, die für das Verständnis der Arbeit notwendig sind. Dazu werden die Eigenschaften der multivariaten Normalverteilung und der multivariaten Gaußmischdichte näher vorgestellt. Es werden die Vorteile der Gaußmischdichte, deren Anwendungsgebiete, sowie deren Nachteile angesprochen. Dem Nachteil des exponentiellen Anstiegs bei rekursiver Verarbeitung gilt es mit Reduktionsverfahren zu begegnen.

In Kapitel 3 werden gängige Reduktionsverfahren näher beschrieben.

In Kapitel 4 wird im Hinblick auf das mehrstufige Clustering-Verfahren zur Komponentenreduktion zuerst auf die Clusteranalyse eingegangen. Im weiteren Verlauf des Kapitels werden die wichtigsten Aspekte der einzelnen Phasen des Verfahrens erklärt.

Kapitel 5 beschäftigt sich mit dem Gradientenabstiegsverfahren, dies kommt in der finalen Phase zum Einsatz und dient der Optimierung. Zusätzlich werden die notwendigen Schritte für die Herleitung des Gradienten aufgeführt.

In Kapitel 6.1 wird anhand eines Beispiels der Ablauf des Verfahrens veranschaulicht. Im Anschluss werden mehrere Simulationen vorgestellt. Dabei werden die Ergebnisse der bestehende Reduktionsverfahren mit den Ergebnissen des mehrstufigen Clustering-Verfahren verglichen. Abschließend werden in Kapitel 7 noch einmal die wichtigsten Aspekte zusammengefasst und ein Ausblick für mögliche Verbesserungen gegeben.

Grundlagen

Wir beginnen mit den Grundlagen, die für das Verständnis dieser Arbeit notwendig sind. Einführend wird sowohl die multivariate Normalverteilung als auch die multivariate Gaußmischdichte vorgestellt. Die universelle Approximationsfähigkeit der Gaußmischdichte ist ebenso Thema, wie ihre Anwendungsgebiete und den damit verbundenen Problemen. Der exponentiell ansteigende Ressourcenverbrauch soll den Einsatz von Reduktionsverfahren motivieren und eine Überleitung zu Kapitel 3 bilden.

2.1 Multivariate Normalverteilung

Viele naturwissenschaftliche Prozesse lassen sich mit Hilfe der Normal-/ Gaußverteilung beschreiben. Dabei liefert der *zentrale Grenzwertsatz* [2] eine mathematische Begründung für dieses Phänomen. Er besagt, dass die normierte Summe von n unabhängig und identisch verteilten Zufallsvariablen mit Erwartungswertvektor $\underline{\mu}$ und Kovarianzmatrix $\underline{\Sigma}$ für $n \rightarrow \infty$ gegen die Normalverteilung $\mathcal{N}(x, \underline{\mu}, \frac{\underline{\Sigma}}{n})$ konvergiert.

Das heißt, alle stetigen Zufallsgrößen, die aus einer Vielzahl von im Einzelbeitrag kleinen und unabhängigen Einflüssen resultieren, normalverteilt sind. Ein Beispiel dafür sind Messfehler.

Sei X eine eindimensionalen und normalverteilte Zufallsvariable, dann ist ihre Dichtefunktion definiert als

$$f(x) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma^2}} \cdot \exp\left\{-\frac{(x - \mu)^2}{2 \cdot \sigma^2}\right\}, \quad (2.1)$$

wobei μ dem Erwartungswert $E\{X\} = \mu$ und $E\{X^2\} - E\{X\}^2 = \sigma^2$ der Varianz entspricht.

Die Dichte der *Normalverteilung* ist symmetrisch und hat die Form einer Glocke, daher wird sie auch Glockenkurve genannt. Lage und Verlauf der Funktion werden durch die beiden statistischen Kenngrößen Erwartungswert μ und Standardabweichung σ festgelegt.

Die Dichte der multivariaten Normalverteilung $\mathcal{N}_N(\underline{x}, \underline{\mu}, \underline{\Sigma})$ ist definiert als

$$f(\underline{x}) = \frac{1}{\sqrt{(2\pi)^N \cdot |\underline{\Sigma}|}} \cdot \exp\left\{-\frac{1}{2} \cdot (\underline{x} - \underline{\mu})^T \cdot \underline{\Sigma}^{-1} \cdot (\underline{x} - \underline{\mu})\right\}, \quad (2.2)$$

wobei $\underline{\mu}$ dem Erwartungswertvektor $E\{\underline{X}\} = \underline{\mu}$ und $E\{(\underline{X} - E\{\underline{X}\}) \cdot (\underline{X} - E\{\underline{X}\})^T\} = \underline{\Sigma}$ der positiv, semidefiniten Kovarianzmatrix entspricht. Dabei steht N für die Dimension der Zufallsvariable \underline{X} .

Die Verteilungsfunktion von $\mathcal{N}_N(\underline{x}, \underline{\mu}, \underline{\Sigma})$ ist definiert als

$$F(\underline{x}) = \frac{1}{\sqrt{(2\pi)^N \cdot |\underline{\Sigma}|}} \cdot \int_{-\infty}^{\underline{x}} \exp\left\{-\frac{1}{2} \cdot (\underline{t} - \underline{\mu})^T \cdot \underline{\Sigma}^{-1} \cdot (\underline{t} - \underline{\mu})\right\} d\underline{t}. \quad (2.3)$$

Dichte und Verteilungsfunktion für $\mathcal{N}_1(x, 0, 1)$ und $\mathcal{N}_2(x, \underline{\mu}, \underline{\Sigma})$ mit

$$\underline{\mu} = \begin{bmatrix} 0 & 0 \end{bmatrix}^T, \quad \underline{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

sind in Abbildung 2.1 dargestellt.

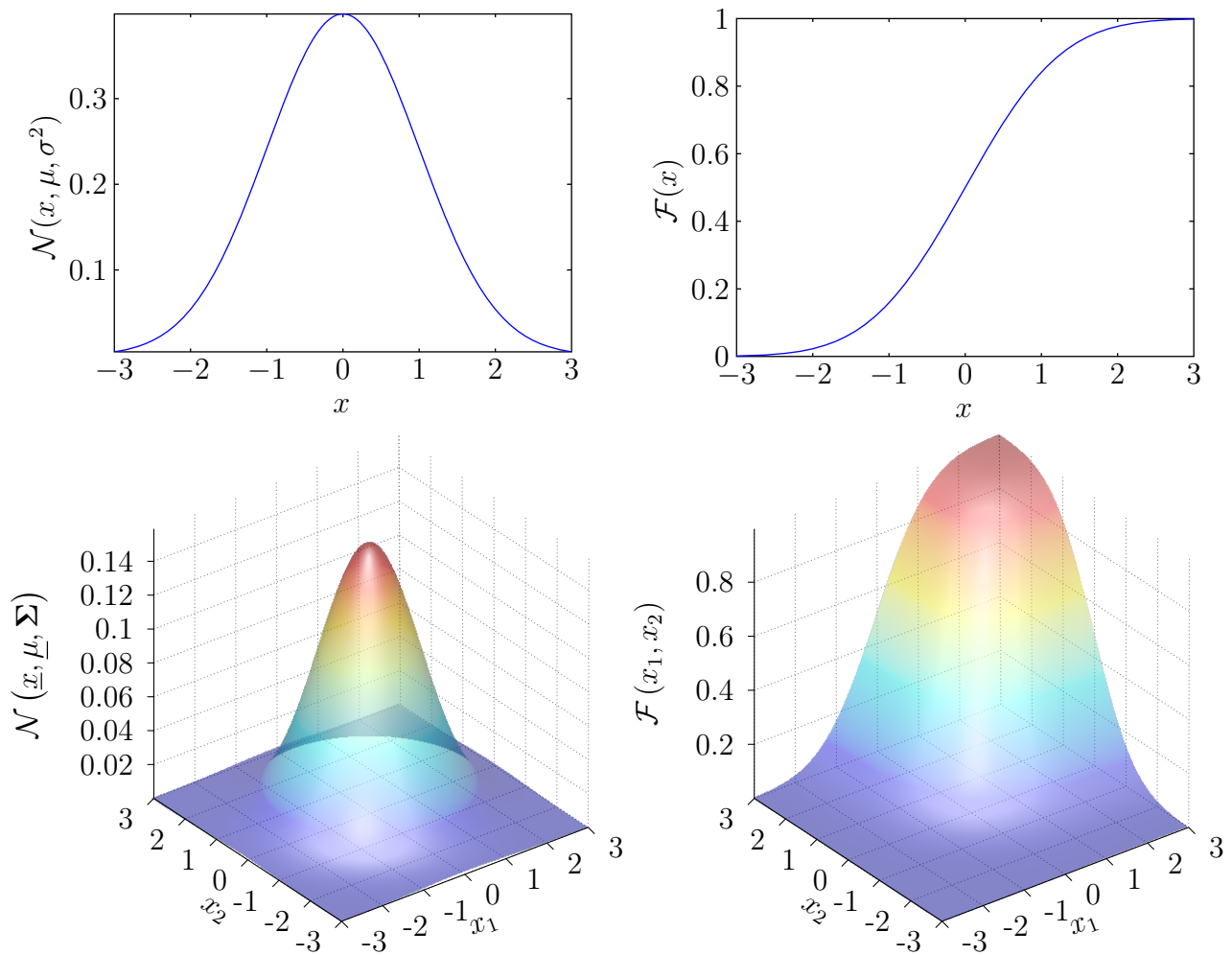


Abbildung 2.1: Dichte und Wahrscheinlichkeitsverteilung für die Standardnormalverteilung für eine (oben) und zwei Dimensionen (unten) .

2.2 Multivariate Gaußmischdichte

Die *Gaußmischdichte* ist eine Linearkombination einzelner Gaußdichten. Eine Gaußmischdichte kann durch die Funktion

$$f(\underline{x}) = \sum_{i=1}^K w_i \cdot \mathcal{N}_N(\underline{x}, \underline{\mu}_i, \underline{\Sigma}_i)$$

beschrieben werden.

Für die Parameter Gewicht w_i , Erwartungswert $\underline{\mu}_i$ und Kovarianz $\underline{\Sigma}_i$ einer speziellen Komponente i , wird in den folgenden Kapiteln abkürzend $\underline{\eta}_i = [w_i, \underline{\mu}_i, \text{vec}(\underline{\Sigma}_i)]^T$ verwendet. Eine Gaußmischdichte der Dimension N mit K Komponenten wird durch $K \cdot (N^2 + N + 1)$ Parameter repräsentiert. Genauer gesagt benötigt man pro Summand jeweils einen Parameter für das Gewicht, N für den Erwartungswert und N^2 für die Kovarianzmatrix. Ein Spezialfall ist $K = 1$ mit $\underline{\eta} = [1, \underline{\mu}, \underline{\Sigma}]^T$. Dies entspricht einer einzelnen Gaußdichte. In Abbildung 2.2 ist ein Beispiel einer Gaußmischdichte für den zweidimensionalen Fall zu sehen.

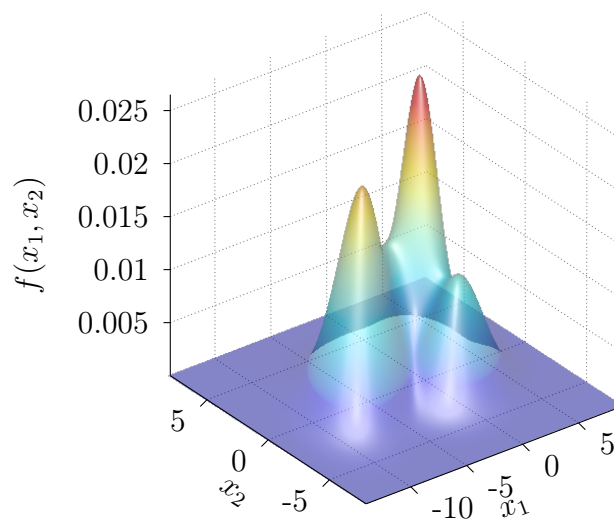


Abbildung 2.2: Zweidimensionale Gaußmischdichte mit 3 Komponenten.

Gilt für die Gewichte w_i zusätzlich

$$\sum_{i=1}^K w_i = 1, \quad w_i > 0,$$

so spricht man von einer *normierten Gaußmischdichte*. In diesem Fall gilt

$$\int_{-\infty}^{\infty} f(\underline{x}) = 1.$$

Die Gaußmischdichte besitzt die Eigenschaft eines universellen Approximators [1], wodurch jede beliebige kontinuierliche Dichtefunktion beliebig genau durch eine Gaußmischdichte mit beliebiger Komponentenanzahl approximiert werden kann. Durch diese Eigenschaft finden Gaußmischdichten insbesondere bei der (approximativen) Zustandsschätzung ihre Anwendung. Im Falle der Zustandsschätzung tritt dabei häufig der Fall auf, dass zwei Gaußmischdichten multipliziert werden müssen. Seien also zwei Gaußmischdichten f_1 und f_2 gegeben mit L bzw. K Komponenten. Für f_g gilt

$$\underbrace{\sum_{i=1}^L w_i \cdot \mathcal{N}_N(\underline{x}, \underline{\mu}_i, \underline{\Sigma}_i)}_{f_1} \cdot \underbrace{\sum_{j=1}^K w_j \cdot \mathcal{N}_N(\underline{x}, \underline{\mu}_j, \underline{\Sigma}_j)}_{f_2} = \sum_{i=1}^L \sum_{j=1}^K w_i \cdot w_j \cdot \mathcal{N}_N(\underline{x}, \underline{\mu}_i, \underline{\Sigma}_i) \cdot \mathcal{N}_N(\underline{x}, \underline{\mu}_j, \underline{\Sigma}_j) \\ = \underbrace{\sum_{k=1}^{L \cdot K} w_k \cdot \mathcal{N}_N(\underline{x}, \underline{\mu}_k, \underline{\Sigma}_k)}_{f_g} .$$

Dabei gilt für $w_k = w_i \cdot w_j$ und für $\underline{\mu}_k, \underline{\Sigma}_k$ werden die Formeln (A.10) und (A.11) aus dem Anhang A verwendet.

Wird die resultierende Gaußmischdichte f_g weiteren Verarbeitungsschritten unterworfen, steigt die Anzahl der Komponenten exponentiell an, wodurch der Einsatz von Reduktionsverfahren erforderlich ist, um den Rechen- und Speicheraufwand begrenzt zu halten.

Reduktionsverfahren

Dieses Kapitel bietet eine Einführung in viele der gängigsten Reduktionsverfahren. Es werden sowohl Unterschiede, als auch Gemeinsamkeiten der verschiedenen Reduktionsverfahren aufgezeigt. Dabei wird der jetzige Stand der Technik näher beschrieben. Zudem wird auf die wichtigsten Gütemaße, die zum Vergleich von Dichtefunktionen verwendet werden, eingegangen. Ebenso werden die wichtigsten Prinzipien, die zur Reduktion verwendet werden können vorgestellt.

Die Aufgabe von Reduktionsverfahren ist es, zu einer gegebenen Gaußmischdichte eine möglichst gute Approximation zu finden. Dabei wird eine optimale Lösung meist nicht erreicht. Vielmehr versuchen diese Verfahren so nah, wie möglich an die optimale Lösung heranzukommen. Gesucht ist dabei eine Dichte mit $L \ll K$. Die Bezeichner L, K stehen für die Anzahl der Komponenten der approximierten und der wahren Dichte.

Die Grundidee dieser Verfahren besteht darin Redundanzen in der wahren Dichte zu eliminieren. Oft wird eine einzelne Mode der Gaußmischverteilung durch mehrere Komponenten repräsentiert, obwohl eine Komponente ausreichen würde. Dies geschieht alles unter der Devise, dass die resultierende reduzierte Dichte so geringfügig wie möglich von der wahren Dichte abweicht. Dies kann durch verschiedene Operationen erreicht werden. Eine Möglichkeit ist es, Komponenten aus der Gaußmischdichte zu entfernen, die einen verhältnismäßig geringen Anteil an der Dichte haben. So werden Komponenten entfernt, deren Gewicht w sehr klein ist. Eine weitere Möglichkeit besteht darin, Komponenten zusammenzufassen, die eine besonders große Ähnlichkeit aufweisen.

Grundsätzlich werden Verfahren anhand ihrer Arbeitsweise unterschieden. Wird mit der gesamten Dichtefunktion begonnen und sukzessive die Anzahl der Komponenten verringert, so spricht man von einem *top-down* Ansatz. Beginnen die Verfahren mit einer Dichte, die aus einer Komponente besteht und fügen nach und nach Komponenten hinzu, spricht man von einem *bottom-up* Ansatz.

Für die Fusion zweier Komponenten $\underline{\eta}_i, \underline{\eta}_j$ mit den Parametern Gewicht, Erwartungswert und Kovarianz zu einer Komponente $\underline{\eta}_{i,j} = [w_{i,j}, \underline{\mu}_{i,j}, \text{vec}(\underline{\Sigma}_{i,j})]$ gilt

$$\begin{aligned} w_{i,j} &= w_i + w_j , \\ \underline{\mu}_{i,j} &= \frac{1}{w_{i,j}} \cdot (w_i \cdot \underline{\mu}_i + w_j \cdot \underline{\mu}_j) , \\ \underline{\Sigma}_{i,j} &= \frac{1}{w_{i,j}} \cdot (w_i \cdot \underline{\Sigma}_i + w_j \cdot \underline{\Sigma}_j + w_i \cdot w_j \cdot (\underline{\mu}_i - \underline{\mu}_j) \cdot (\underline{\mu}_i - \underline{\mu}_j)^T) . \end{aligned}$$

Es werden so lange Komponenten entfernt oder zusammengefasst, bis die reduzierte Dichte so viele Komponenten enthält, wie vorher festgelegt.

Die Anzahl der Komponenten der reduzierten Dichte ist entscheidend für die Genauigkeit der Approximation. Es kann beispielsweise vorkommen, dass aufgrund dieser Anzahl Komponenten, die stark an der Form der Gaußmischdichte beteiligt sind, entfernt oder zusammengefasst werden. Dadurch werden im schlimmsten Fall keine redundanten Informationen zusammengefasst, sondern es geht Information verloren. Daher kann schon eine Reduktion mit unwesentlich mehr Komponenten eine bessere Approximation der wahren Gaußmischdichte liefern.

Zur Bewertung der Abweichung zwischen Approximation und der wahren Gaußmischdichte wird ein Maß benötigt, das festlegt, in wie weit sich diese ähneln oder unterscheiden. Der Wert des Gütemaßes wird für die Auswahl von Komponenten zur Fusion betrachtet.

In Abhängigkeit von dem verwendeten Gütemaß lassen sich die Reduktionsverfahren neben ihrer Arbeitsweise (top-down oder bottom-up) auch wie folgt unterscheiden:

- Lokale Verfahren

- Globale Verfahren

Lokale Verfahren, zu denen das von Salmond (siehe Abschnitt 3.3) oder das von West (siehe Abschnitt 3.2) zählen, berücksichtigen nur statistische Kenngrößen wie Erwartungswert und Kovarianz, um auszuwerten in wie weit sich Komponenten ähneln.

Globale Verfahren, wie das von Williams (siehe Abschnitt 3.5) und das PGMR Verfahren, betrachten hingegen zusätzlich noch die Form der Gaußmischdichte, um geeignete Kandidaten für die Fusion zu ermitteln. Komplexität und Rechenaufwand sind für globale Verfahren wesentlich höher als bei lokalen Verfahren, dafür ist auch die Abweichung zwischen wahrer und approximierter Dichte wesentlich geringer. Bei der Anwendung der verschiedenen Verfahren gilt es daher abzuwägen, ob sich das Verhältnis von Aufwand zu Nutzen in einem vernünftigen Rahmen bewegt.

3.1 Gütemaße

Es existieren verschiedene Gütemaße zur Bewertung des Unterschieds zwischen zwei beliebigen Funktionen. Im folgenden werden die allgemeinen Formulierungen der Gütefunktionen speziell für Gaußmischdichten bzw. deren Komponenten betrachtet.

Um zwei Kandidaten für die paarweise Fusion zu ermitteln, verwendet Williams [3] die *Integrale Quadratische Differenz* (ISD). Ausgangspunkt ist die *Kolmogorov Distanz*

$$G_{Kol}(\underline{\eta}) = \int_{\mathbb{R}} |\tilde{f}(\underline{x}) - f(\underline{x}, \underline{\eta})| \, d\underline{x} ,$$

die den Unterschied in der Wahrscheinlichkeitsmasse vergleicht. Für die Kolmogorov Distanz gilt $G_{Kol}(\underline{\eta}) \in [0, 1]$. Der Ausdruck wird zu 0 ausgewertet, falls die Dichten identisch sind und entspricht 1, wenn die Dichten vollkommen unterschiedlich sind. Der Vorteil der ISD

$$G_{ISD}(f, \tilde{f}) = \frac{1}{2} \cdot \int_{\mathbb{R}} (\tilde{f}(\underline{x}) - f(\underline{x}, \underline{\eta}))^2 \, d\underline{x} \quad (3.1)$$

ist, dass sie im Gegensatz zur Kolmogorov Distanz geschlossen berechnet werden kann. Es ist also kein numerisches Verfahren notwendig um das Integral auszuwerten. Hierbei lässt sich ausnutzen, dass das Produkt zweier Gaußdichten wieder einer (unnormierten) Gaußdichte ergibt. Ausmultiplizieren von (3.1) ergibt

$$G_{ISD}(f, \tilde{f}) = J_{hh} - 2 \cdot J_{hr} + J_{rr} ,$$

wobei

$$J_{hh} = \int_{\mathbb{R}} \tilde{f}(\underline{x})^2 \, d\underline{x} , \quad J_{hr} = \int_{\mathbb{R}} \tilde{f}(\underline{x}) \cdot f(\underline{x}) \, d\underline{x} , \quad J_{rr} = \int_{\mathbb{R}} f(\underline{x})^2 \, d\underline{x} . \quad (3.2)$$

Die einzelnen Produkte in (3.2) können auf Produkte zweier Gaußdichten reduziert werden, wodurch sich diese Ausdrücke sehr einfach berechnen lassen.

Für die *Normierte Integralen Quadratischen Differenz* gilt

$$G_{NISD}(f, \tilde{f}) = \frac{J_{hh} - 2 \cdot J_{hr} + J_{rr}}{J_{hh} + J_{rr}} . \quad (3.3)$$

Diese liegt immer im Intervall $[0, 1]$. Dieses Distanzmaß wird zu 0 ausgewertet, wenn wahre und reduzierte Dichte identisch sind. Sind diese aber komplett verschieden erhält man 1.

Ein weiteres Gütemaß ist die *Kullback-Leibler Divergenz* [4]

$$G_{KL}(f, \tilde{f}) = \int_{\mathbb{R}} \tilde{f} \cdot \log\left(\frac{\tilde{f}}{f}\right) d\underline{x}. \quad (3.4)$$

Die Formel enthält den Logarithmus über Gaußmischdichten und kann daher nicht geschlossen berechnet werden.

Die *Mahalanobis-Distanz* [5] ist mit der euklidischen Distanz verwandt. Für unkorrelierte Daten ergibt sie das gleiche Ergebnis wie die euklidische Distanz. Dieses Distanzmaß berücksichtigt sowohl Unterschiede bei der Skalierung der Daten, als auch die Korrelation zwischen den Daten. Der Unterschied zwischen der Mahalanobis-Distanz und der Integralen Quadratischen Differenz besteht darin, dass nicht der Unterschied zwischen der wahren und der reduzierten Dichtefunktion, sondern der Unterschied zwischen zwei Komponenten einer Gaußmischdichte bestimmt wird. Bei der gewichtete Mahalanobis-Distanz

$$G_{MAH}(\mathcal{N}_N^i, \mathcal{N}_N^j) = \sqrt{\frac{w_i \cdot w_j}{w_i + w_j} \cdot (\underline{\mu}_i - \underline{\mu}_j)^T \cdot \underline{\Sigma}_i^{-1} \cdot (\underline{\mu}_i - \underline{\mu}_j)}, \quad (3.5)$$

werden noch die Unterschiede in der Gewichtung der einzelnen Komponenten berücksichtigt. Im nächsten Abschnitt werden ein paar der gängigsten Reduktionsverfahren näher beschrieben, um einen Überblick über den jetzigen Stand der Technik zu erhalten.

3.2 Verfahren von West

West verwendet für sein Verfahren [6] die gewichtete Mahalanobis-Distanz (3.5) als Gütemaß. Zu Beginn werden alle Komponenten nach ihren Gewichten sortiert. Solange noch mehr Komponenten enthalten sind als gewünscht, wird immer die Komponente i mit dem kleinsten Gewicht ausgewählt. Tritt der Fall ein, dass für das Minimum mehrere Komponenten zur Auswahl stehen, kann zufällig zwischen ihnen gewählt werden.

Danach wird zu der gewählten Komponente i , mit Hilfe der Gütefunktion, die am nächsten liegende Komponente j ermittelt. Im nächsten Schritt werden die beiden Komponenten fusioniert, aus der wahren Dichte entfernt und das Ergebnis der Fusion sowohl der wahren Dichte, als auch der aktuellen Reduktion hinzugefügt.

Bildlich gesprochen werden zu den Komponenten mit dem geringsten Anteil an der gesamten Gaußmischdichte, mit Hilfe des Distanzmaßes geeignete Fusionspartner ermittelt und fusioniert. Dieses Verfahren ist wesentlich schneller als die Verfahren von Runnalls und Williams, da in jedem Reduktionsschritt nur $n - 1$ Distanzen statt n^2 zwischen der Komponente mit dem kleinsten Gewicht und allen verbleibenden Komponenten berechnet werden müssen, wobei n für die Anzahl an Komponenten der aktuellen Reduktion steht. Die Berechnungersparnis gegenüber Williams und Runnalls erkaufte man sich mit teils deutlich schlechteren Reduktionsergebnissen.

3.3 Verfahren von Salmond

Das Verfahren von Salmond [7] verwendet, wie auch das Verfahren von West, als Distanzmaß die gewichtete Mahalanobis-Distanz (3.5). Beide Verfahren verfolgen einen top-down Ansatz. Im Gegensatz zu den anderen Verfahren werden in der Regel in jedem Reduktionsschritt mehr als nur zwei Komponenten fusioniert.

Wie auch bei den anderen Verfahren werden solange Komponenten fusioniert, bis die Reduktion die gewünschte Anzahl an Komponenten beinhaltet. Zu Beginn eines jeden Reduktionsschrittes wird diejenige Komponente gewählt die den stärksten Beitrag (größtes Gewicht) zur gesamten Gaußmischdichte liefert. Diese Komponente wird im Folgenden als Hauptkomponente bezeichnet. Nun wird ein sogenanntes Cluster C gebildet. Es werden alle Komponenten zu diesem Cluster hinzugefügt, die weniger als ein vorher definierter Schwellwert von der Hauptkomponente entfernt sind. Es ist zu beachten, dass für die Berechnung des Distanzmaßes die Kovarianz der Hauptkomponente verwendet wird. Schließlich werden alle Komponenten des Clusters zu einer einzigen Dichte fusioniert. Alle Komponenten die zuvor fusioniert wurden, werden aus der wahren Gaußmischdichte entfernt und das Ergebnis der Fusion wird der reduzierten und der wahren Dichte hinzugefügt. Danach wird der Vorgang für die verbleibende Gaußmischdichte wiederholt.

Tritt der Fall ein, dass in einem Reduktionsschritt alle Distanzen oberhalb des Schwellwertes liegen, wird der Schwellwert um einen gewissen Anteil inkrementiert. Nachdem der Schwellwert angepasst ist, wird die Reduktion von vorne begonnen.

Ein wesentliches Problem an diesem Verfahren kann die Wahl des Schwellwertes darstellen. Ist er zu gering, muss die Reduktion mehrmals von vorne begonnen werden, da kein passender Fusionskandidat gefunden wird. Ist er zu groß, können auch ungeeignete Komponenten zu einem Cluster hinzugefügt werden, was zu einer schlechteren Approximation führt.

3.4 Verfahren von Runnalls

Runnalls verwendet für sein Reduktionsverfahren [8] eine obere Schranke der Kullback-Leibler Distanz zwischen zwei Komponenten

$$G_{KL}(\mathcal{N}_N^i, \mathcal{N}_N^j) = \frac{1}{2} \cdot \left[(w_i + w_j) \cdot \ln |\boldsymbol{\Sigma}_{ij}| - w_i \cdot \ln |\boldsymbol{\Sigma}_i| - w_j \cdot \ln |\boldsymbol{\Sigma}_j| \right] \quad (3.6)$$

als Gütefunktion für das Finden von Fusionspartnern. Dies kann in geschlossener Form berechnet werden. Die Gütefunktion wird zwischen allen Komponenten der wahren Gaußmischdichte berechnet. Für eine Dichte mit K Komponenten müssen also $\frac{1}{2} \cdot (K \cdot (K - 1))$ Distanzen berechnet werden. Solange die gewünschte Anzahl an Komponenten L noch nicht erreicht ist, werden in jedem Schritt jeweils immer die zwei Dichten i und j aus der wahren Dichte fusioniert, für die die obere Schranke der Kullback-Leibler Distanz (3.6) am Kleinsten ist. Das Ergebnis der

Fusion wird der reduzierten Dichte hinzugefügt und die beiden Komponenten werden aus der wahren Dichte gelöscht. Das Ergebnis der Fusion wird zusätzlich noch zur wahren Dichte hinzugefügt. Dies wird solange wiederholt, bis die gewünschte Anzahl an Komponenten erreicht ist.

Da sich bei der Fusion der Komponenten i und j , die Distanzen zwischen allen anderen Komponenten nicht verändern, werden, um unnötigen Rechenaufwand zu vermeiden, alle Distanzen vorberechnet und dann nur die Distanzen zur fusionierten Dichte aktualisiert. Zu diesem Zweck werden bei der effizienten Umsetzung des Verfahrens von Runnalls die Distanzen zwischen den einzelnen Komponenten in einer oberen Dreiecksmatrix \mathbf{B} gespeichert. Aus Matrix \mathbf{B} können nach jeder Fusion jeweils eine Spalte und eine Zeile gelöscht und das Ergebnis der Fusion hinzugefügt werden. Zusätzlich müssen die Distanzen zu allen verbliebenen Komponenten und der neuen Komponente aktualisiert werden. In Tabelle 3.1 ist Matrix \mathbf{B} vor der Fusion von den Komponenten $i = 3$ und $j = 4$ abgebildet. Die rot markierten Elemente der Matrix müssen neu berechnet werden. In Tabelle 3.2 sind die neu berechneten Elemente in Zeile und Spalte i gespeichert und blau markiert, wobei die Zeile und Spalte j gelöscht wurde. In Listing 3.1 ist ein Pseudocode für das Verfahren von Runnalls beschrieben.

1, 2	1, 3	1, j	...	1, L
	$i - 1, 3$	\vdots
		i, j	...	i, L
			\ddots	\vdots
				$L - 1, L$

L

Tabelle 3.1: Matrix \mathbf{B} vor der Fusion.

1, 2	1, 3	1, 5	...	1, L
	2, 3	2, 5	...	\vdots
		3, 5	...	3, L
			\ddots	\vdots
				$L - 1, L$

$L-1$

Tabelle 3.2: Matrix \mathbf{B} nach der Fusion.

```

1  $B \leftarrow G_{KL}(\tilde{\eta}_1 \dots \tilde{\eta}_n, \tilde{\eta}_1 \dots \tilde{\eta}_n)$ 
2  $k \leftarrow 0$ 
3 repeat
4      $i, j \leftarrow \arg \min_{i,j} (\mathbf{B})$ 
5      $\tilde{\eta}_i \leftarrow \text{Fusion}(\tilde{\eta}_i, \tilde{\eta}_j)$ 
6     entferne  $\tilde{\eta}_j$  aus  $\tilde{f}(\underline{x})$ 
7      $B_{i,1 \dots n} \leftarrow G_{KL}(\tilde{\eta}_i, \tilde{\eta}_{i+1} \dots \tilde{\eta}_n)$  // Aktualisierung der Zeile  $i$ 
8      $B_{1 \dots n, i} \leftarrow G_{KL}(\tilde{\eta}_1 \dots \tilde{\eta}_{i-1}, \tilde{\eta}_i)$  // Aktualisierung der Spalte  $i$ 
9     lösche Zeile und Spalte  $j$ 
10     $k++$ 
11 until  $k = L$ 

```

Listing 3.1: Pseudocode für das Verfahren von Runnalls.

3.5 Verfahren von Williams

Das Verfahren von Williams [3] kann in zwei Phasen unterteilt werden. In Phase 1 werden analog zu Runnalls solange, bis die wahre Gaußmischdichte auf die gewünschte Anzahl von Komponenten reduziert ist, jeweils zwei Komponenten miteinander fusioniert. Zusätzlich verwendet Williams das Löschen von Komponenten. Um die Komponenten auszuwählen, verwendet Williams für sein Reduktionsverfahren die Integrale Quadratische Distanz (3.1) als Gütefunktion und wählt die Komponente aus, für die sich die ISD am wenigsten verändert.

Da sich in jedem Schritt des Verfahrens nur die Distanzen zur hinzugefügten Dichte ändern, werden wie bei Runnalls auch nur diese Neuberechnet. Zudem muss auch nicht das gesamte Distanzmaß neu ausgewertet werden, sondern nur die Summanden J_{hr} und J_{rr} , wie in Formel (3.2) beschrieben. Daher reicht es, J_{hh} aus Tabelle 3.3 zu Beginn vorzuberechnen. In den Tabellen 3.4 und 3.5 sieht man die Situation nach der Fusion von zwei Komponenten. Die neu ausgewerteten Elemente sind blau markiert. Das Ergebnis der Fusion wurde an der Stelle L eingefügt. Listing 3.2 stellt einen Pseudocode für die erste Phase des Verfahrens von Williams vor.

Ist die gewünschte Anzahl an Komponenten erreicht, dient die berechnete Reduktion als Startlösung für ein numerisches Optimierungsverfahren zur Verbesserung der Approximation. Als Optimierungsverfahren verwendet Williams einen Gradientenabstieg.

1, 2	...	1, j	...	1, K
	\ddots	\vdots
		i, j	...	i, K
			\ddots	\vdots
				$K - 1, K$

Tabelle 3.3: Matrix J_{hh} : Bewertet Selbstähnlichkeit der wahren Dichte.

1, 1	1, j	...	1, L
\vdots	2, 2	\vdots
$i, 1$...	\ddots	i, j	...	i, L
\vdots	\ddots	...	\vdots
$K, 1$	K, j	$K, L - 1$	K, L

Tabelle 3.4: Matrix J_{hr} : Bewertet Ähnlichkeit zwischen wahrer und reduzierter Dichte.

1, 1	1, j	...	1, L
\vdots	2, 2	\vdots
$i, 1$...	\ddots	i, j	...	i, L
\vdots	\ddots	...	\vdots
$L, 1$	L, j	\ddots	L, L

Tabelle 3.5: Matrix J_{rr} : Bewertet Selbstähnlichkeit der reduzierten Dichte.

```

1  $B_{hh} \leftarrow J_{hh}(\tilde{\eta}_1 \dots \tilde{\eta}_k)$ 
2  $k \leftarrow 0$ 
3 repeat
4      $B_{hr} \leftarrow J_{hr}(\tilde{\eta}_1 \dots \tilde{\eta}_k, \eta_1 \dots \eta_n)$ 
5      $B_{rr} \leftarrow J_{rr}(\eta_1 \dots \eta_n)$ 
6      $i, j \leftarrow \arg \min_{\tilde{\eta}_i, \tilde{\eta}_j} (B_{hh} - 2 \cdot B_{hr} + B_{rr})$ 
7      $\eta_k \leftarrow \text{Fusion}(\tilde{\eta}_i, \tilde{\eta}_j)$ 
8     füge  $\eta_k$  zu  $f(\underline{x})$  hinzu
9     entferne  $\tilde{\eta}_i, \tilde{\eta}_j$  aus  $\tilde{f}(\underline{x})$ 
10     $k \leftarrow k + 1$ 
11 until  $k = L$ 

```

Listing 3.2: Pseudocode für Phase 1.

3.6 PGMR Verfahren

Bei dem PGMR Verfahren handelt es sich um ein progressives Verfahren [9]. Es verwendet die normierte Integrale Quadratische Differenz als Distanzmaß (3.3). Zusätzlich zu der Anzahl an Komponenten der approximierten Dichte kann die maximale Abweichung vorgegeben werden. Dieses Verfahren verwendet eine Prädiktor-Korrektor Struktur und nimmt eine Sonderstellung gegenüber den bereits genannten Verfahren ein, da sich die Berechnung der Approximation vollkommen unterscheidet. Dieses Verfahren verfolgt einen bottom-up Ansatz. Zuerst wird eine Startlösung gewählt, dafür wird eine einzelne Gaußmischdichte verwendet. Diese Startlösung wird nun im Prädiktionsschritt schrittweise zur wahren Dichte hintransformiert. Dazu werden die Stellen identifiziert, an denen die Abweichung zwischen aktueller Reduktion und wahrer Dichte am größten sind. An diesen Stellen wird die Reduktion mit Hilfe der Operationen Splitting und Adding verfeinert. Beim Splitting wird eine Gaußdichte in zwei einzelne Dichten aufgeteilt, während beim Adding eine vollkommen neue Dichte hinzugefügt wird. Beim Löschen wird eine Dichte entfernt. Dazu wird im Laufe des Verfahrens stets überprüft, ob aus der aktuellen Reduktion eine Komponente entfernt werden kann.

Auftretende Abweichungen während des Prädiktionsschritts werden im Korrekturschritt korrigiert. Dazu kann das Newtonverfahren oder ein Gradientenabstieg verwendet werden.

Komponentenreduktion durch Clusterbildung

In diesem Abschnitt wird das von Dennis Schieferdecker entwickelte *Mehrstufiges Clustering-Verfahren* [1] (GMRC) vorgestellt. Dazu wird in einem ersten Schritt näher auf das Clusterverfahren eingegangen. Ablauf und Einsatzgebiete der Clusteralgorithmen werden ebenso präsentiert, wie auch der Zusammenhang zum Reduktionsverfahren.

Bisher können nur eindimensionale Zufallsvariablen betrachtet werden. Das eigentliche Reduktionsverfahren unterteilt sich in drei Phasen, deren Abläufe genau betrachtet werden. Vorteil dieses Reduktionsverfahrens ist die Unabhängigkeit der einzelnen Phasen untereinander. Daher können für diese Phasen unterschiedlichste Verfahren verwendet werden. Zu beachten ist lediglich, dass die Berechnungen der einzelnen Phasen korrekt und in gewünschter Form übergeben werden.

4.1 Clusterverfahren

Die Clusterverfahren dienen allgemein als multivariate Methode zur Segmentierung bzw. Klassifikation von Daten. Ziel der Clusterbildung ist es eine gegebene Menge von Objekten in einzelne Gruppen/Cluster zu unterteilen. Weiter sei vereinbart, dass jedes Cluster auch mindestens ein Element enthält. Für die Cluster und deren Elemente sollte folgendes gelten:

- Homogenität im Cluster
- Heterogenität zwischen den Clustern

Mit Homogenität im Cluster ist gemeint, dass alle Elemente eines Clusters eine gewisse Ähnlichkeit aufweisen. Die Zentren untereinander sollen möglichst verschieden sein. Die Ähnlichkeit selbst wird durch die Abstände der Daten untereinander beschrieben. Um also ein Maß für die Ähnlichkeit zu erhalten, wird der Abstand dieser Punkt-zu-Punkt-Verbindung berechnet. Als Abstandsmaß kann beispielsweise die euklidische oder Mahalanobis Distanz verwendet werden. Anwendung finden diese Clusterverfahren bei der automatischen Datenklassifikation [10], Data Mining [11] oder auch in der Bildverarbeitung [12].

Grundsätzlich können die Clusterverfahren in zwei Gruppen gegliedert werden:

- *Hierachische Verfahren*
- *Partitionierende Verfahren*

Zu den partitionierenden Verfahren zählt beispielsweise die *Kmeans Methode* [13]. Diese Verfahren bestimmen für die Initialisierung eine beliebige Anzahl von Clusterzentren. Danach wird jedes Objekt ihrem nächsten Zentrum zugeordnet und danach die Zentren neu berechnet. Dies wird solange wiederholt, bis sich an der Zuordnung der Objekte nichts mehr ändert.

Bei den hierarchischen Verfahren unterscheidet man zwischen agglomerativen und divisiven Verfahren. Agglomerative Verfahren beginnen mit einem Cluster für jedes einzelne Datum und fassen dann sukzessive ähnliche Cluster zu einem Neuen zusammen. Divisive Verfahren hingegen beginnen mit einem einzigen Cluster, suchen Daten, die unähnlich zueinander sind und teilen diese in einzelne Cluster auf.

4.1.1 Kmeans

Das Kmeans Verfahren ermittelt aus einer Menge von Datenpunkten eine bestimmte Anzahl von Repräsentanten, so dass ein Abstandsmaß ϕ minimiert wird.

Gegeben ist eine Menge X von Datenpunkten x mit $|X| = n$. Gesucht ist die Menge \mathcal{C} der Zentren von X mit $|\mathcal{C}| = l$. Die Repräsentanten werden so gewählt, dass die Abstände zwischen ihnen und den Datenpunkten minimal ist. Dabei gilt also

$$\phi = \sum_{x \in X} \min_{\mathcal{C}} \|x - \mathcal{C}\|^2 .$$

Um nun eine gültige Zuordnung zu berechnen werden folgende Schritte verwendet.

1. Wähle Clusterzentren $\mathcal{C} = \{c_1, \dots, c_l\}$ zufällig
2. Ordne alle Datenpunkte ihrem nächsten Clusterzentrum zu
3. Berechne alle Zentren $c_i = \frac{1}{|c_i|} \cdot \sum_{x \in c_i} x$ neu
4. Wiederhole Schritt 2 und 3 solange bis sich die Zentren nicht mehr verändern.

4.1.2 Kmeans++

Das *Kmeans++ Verfahren* [14] unterscheidet sich nur geringfügig vom Kmeans Verfahren. Es wird als Initialisierung des eigentlichen Kmeans Verfahren verwendet. Für die Initialisierung sind l Clusterzentren gesucht. Man versucht Zentren zu finden, die möglichst weit voneinander entfernt sind. Der Zustandsraum wird also durch die initialen Zentren großflächig abgedeckt. $D(x)$ bestimmt die minimale Distanz zwischen einem Datenpunkt x und dem am nächsten liegenden Zentrum.

1. Wähle Clusterzentren c_1 zufällig
2. Wähle das nächste Zentrum $c_i = x' \in X$ mit der Wahrscheinlichkeit $\frac{D(x')^2}{\sum_{x \in C_i} D(x)^2}$
3. Wiederhole Schritt 2 solange bis genügend Zentren gewählt sind
4. Fahre mit dem Kmeans Verfahren fort

Zum besseren Verständnis des Verfahrens wird im Folgenden noch einmal auf die einzelnen Schritte eingegangen und der Ablauf in den Abbildungen 4.1-4.6 dargestellt:

- Abbildung 4.1 zeigt die Ausgangssituation. In diesem Fall wird von sieben dieser Datenpunkte ausgegangen. Aus den Datenpunkten werden im Folgenden schrittweise 4 Clusterzentren ermittelt.
- Dazu wird als erstes zufällig einer von den Datenpunkten ausgewählt (siehe Abbildung 4.2). Dieser wird als erstes Clusterzentrum (rot) verwendet.
- Alle Datenpunkte werden nun in Abbildung 4.3 diesem Zentrum zugeordnet. Als zweites Zentrum wird dann derjenige Datenpunkt gewählt, der am Weitesten vom ersten Clusterzentrum entfernt ist.
- Danach geschieht eine neue Zuordnung der Datenpunkte zu ihrem jeweils nächstgelegenen Clusterzentrum. Dies ist in Abbildung 4.4 zu sehen, wobei das zweite Clusterzentrum grün markiert wurde. Das dritte Clusterzentrum (blau) entspricht wieder der Zuordnung mit maximaler Distanz.
- In Abbildung 4.5 ist die Zuordnung nach der Wahl des dritten Clusterzentrums dargestellt.
- Der Ablauf findet nun in Abbildung 4.6 durch Wahl des vierten und letzten Clusterzentrums (gelb) sein Ende.

Wie man sieht werden durch das Verfahren vier Clusterzentren gefunden, die den Raum der Datenpunkte großflächig abdecken.

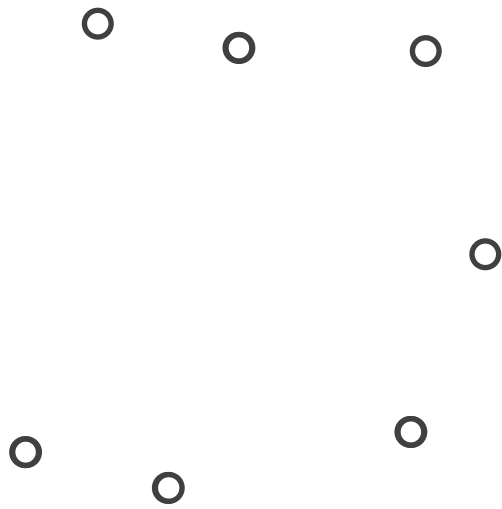


Abbildung 4.1: Datenpunkte bzw. Komponenten der wahren Dichte.

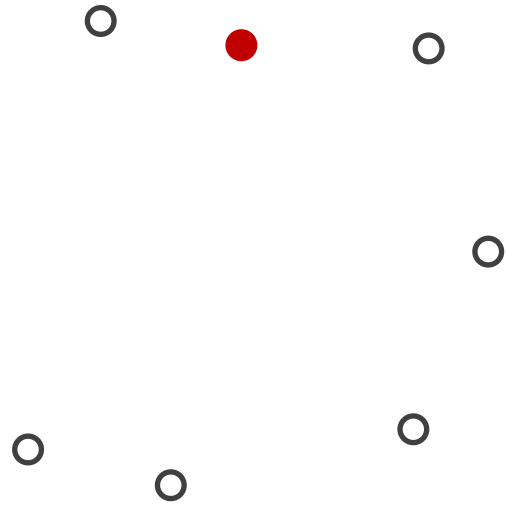


Abbildung 4.2: Zufällige Auswahl des ersten Zentrums (rot).

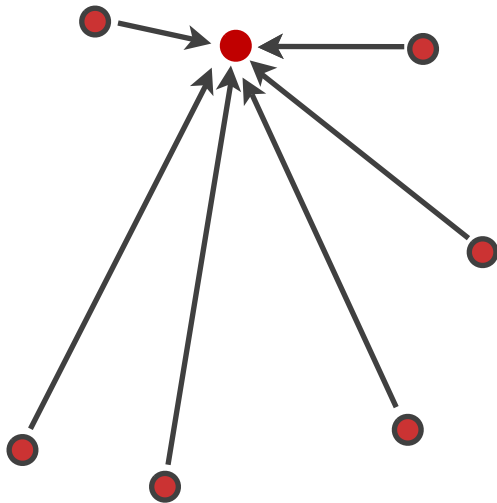


Abbildung 4.3: Zuordnung der Komponenten zu ihrem nächstgelegenen Cluster.

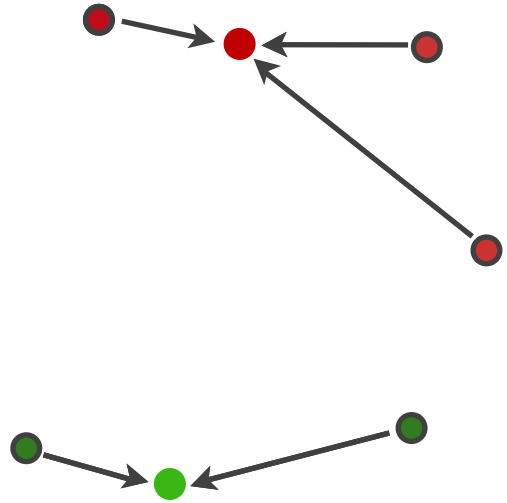


Abbildung 4.4: Wahl des zweiten Clusters mit max-min-Distanz und erneute Zuordnung.

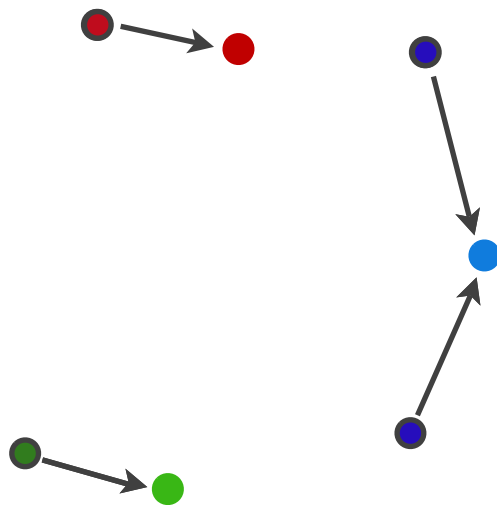


Abbildung 4.5: Wahl des dritten Clusters und neue Zuordnung der Komponenten zum nächsten Zentrum.

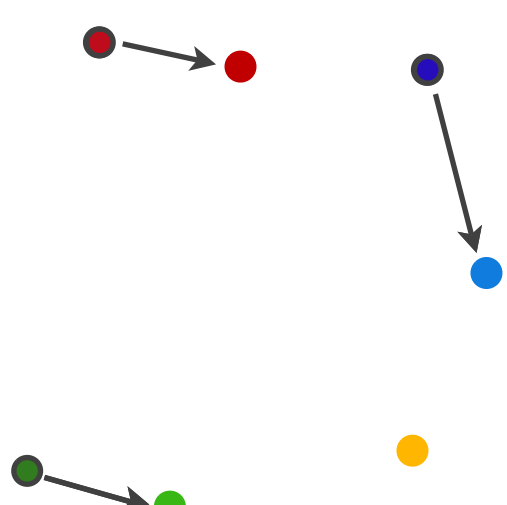


Abbildung 4.6: Wahl des letzten Zentrums.

Überträgt man die Gaußmischdichten-Reduktion auf Kmeans, gelten folgende Beziehungen:

1. Komponenten der wahren Gaußmischdichte $\hat{=}$ Datenpunkte
2. Komponenten der reduzierten Gaußmische $\hat{=}$ Clusterzentren
3. Abstandsmaß $\hat{=}$ Distanz zwischen zwei Komponenten
4. Gütemaß $\hat{=}$ Abweichung zwischen wahrer und reduzierter Gaußmischdichte

4.2 Reduktionsverfahren

Mit Hilfe der Clusterbildung wird nun zu einer wahren Gaußmischdichte eine reduzierte Gaußmischdichte berechnet. Ziel ist dabei die Abweichung zwischen der wahren Dichte und der reduzierten Gaußmischdichte anhand der Integrale Quadratische Differenz zu minimieren. Das Reduktionsverfahren (*GMRC*) kann in drei Phasen unterteilt werden:

1. Bestimmung einer Startlösung
2. Verbesserung der Startlösung durch Clusterverfahren
3. Optimierung der Approximation

Listing 4.1 enthält einen Pseudo-Code für das Clusteringverfahren und fasst alle wichtigen Schritte des Verfahren zusammen. Im Folgenden werden die einzelnen Phasen des Reduktionsverfahrens näher betrachtet.

```

1 /* Phase 1/*
2  $\underline{\eta}_{init} \leftarrow \text{initialReduction}(\tilde{\eta})$ 
3  $\mathcal{C} \leftarrow \text{initialClustering}(\eta, \tilde{\eta})$ 
4  $\underline{\eta} \leftarrow \text{computeCenters}(\mathcal{C}, \tilde{\eta})$ 
5  $\underline{\eta} \leftarrow \text{correctCenters}(\underline{\eta}, \underline{\eta}_{init})$ 
6 /* Phase 2/*
7 repeat
8   for  $i = 1; i \leq n; i ++$ ,
9     for  $j = 1; j \leq L; j ++$ 
10        $\mathcal{C}_{tmp} \leftarrow \text{Zuordnung von } \underline{\eta}_i \text{ zu } \mathcal{C}_j$ 
11        $\underline{\eta}_{tmp} \leftarrow \text{computeCenters}(\mathcal{C}_{tmp}, \tilde{\eta})$ 
12        $\mathbf{D}_j \leftarrow G_{NSID}(\tilde{\eta}, \underline{\eta}_{tmp})$ 
13     end
14      $k \leftarrow \arg \min_{1..j} (\mathbf{D})$ 
15      $\mathcal{C}_j \leftarrow \underline{\eta}_k$ 
16      $\underline{\eta} \leftarrow \text{computeCenters}(\mathcal{C}, \tilde{\eta})$ 
17   end
18 until Abbruchbedingung erfüllt
19  $\underline{\eta} \leftarrow \text{correctCenters}(\underline{\eta}, \underline{\eta}_{init})$ 
20 /* Phase 3/*
21  $\underline{\eta} \leftarrow \text{optimizeReduction}(\underline{\eta}, \tilde{\eta})$ 

```

Listing 4.1: Pseudocode für das mehrstufige Clustering.

4.2.1 Initialisierung

Die Berechnung einer initialen Reduktion und deren Parametervektor $\underline{\eta}$ wird im folgenden *initialReduction* genannt werden. Für eine gültige Startlösung werden zuerst L Clusterzentren benötigt. Dies entspricht der Berechnung einer Approximation mit L Komponenten. Daher kann entweder eines der bereits genannten Reduktionsverfahren aus Abschnitt 3.4-3.3 verwendet werden, oder z.B. die Initialisierung der Kmeans++ Methode, wie in Abschnitt 4.1.2 beschrieben. Ausgehend von diesen Zentren, wird jede Komponente der wahren Gaußmischdichte ihrem nächsten Clusterzentrum zugeordnet. Als Abstandsmaß wird dabei die KL-Divergenz verwendet. Dies entspricht der Funktion *InitialClustering*. Nach der Zuordnung werden alle Komponenten desselben Clusters zu einer Komponente fusioniert. Dies ist gleichbedeutend mit der Neuberechnung aller Clusterzentren. In der bestehenden Implementierung entspricht dies der Funktion *ComputeCenters*.

Als letztes muss noch sichergestellt werden, dass kein Zentrum leer ist. Dafür wird die Funktion *CorrectCenters* verwendet. Sollte ein Cluster leer sein, wird ihm zufällig eine Komponente zugewiesen.

4.2.2 Clustering

In der zweiten Phase wird versucht die Startlösung schrittweise zu verbessern. Im Verlauf des Clusterings werden folgende Operationen verwendet:

- Zuordnung der Komponente
- Neuberechnung des Cluster
- Berechnung der Abweichung

In jedem Schritt des Clusterings wechselt jeweils nur maximal eine Komponente der wahren Gaußmischdichte ihren Cluster, d.h. diese Komponente wird einem anderen Cluster zugeordnet. Ausgehend von der initialen Zuordnung einer Komponente der wahren Dichte erfolgt schrittweise eine neue Zuordnung zu einem der L Clusterzentren der aktuellen Approximation. Gleichzeitig wird das betroffene Cluster neu berechnet, dabei werden alle Komponenten dieses Clusters fusioniert.

Im Anschluss an die Neuberechnung des Clusters erfolgt die Berechnung der Abweichung mithilfe der normierten Integralen Quadratischen Differenz (3.3). Genau wie in Abschnitt 3.5 beschrieben, wird nicht die komplette Formel des Gütemaßes neu berechnet, sondern nur die Stellen der Formel an denen sich Änderungen ergeben. Ist eine Verbesserung eingetreten, entspricht diese Zuordnung der aktuellen Komponente zu einem Zentrum der aktuell besten Approximation. Verbessert sich hingegen die Abweichung nicht, wird mit der nächsten Komponente fortgefahren. Die Berechnung der verbleibenden Cluster erfolgt analog.

Im folgenden wird wie bei der Initialisierung der Verlauf des Clusterings anhand eines einfachen Beispiels in den Abbildungen 4.7-4.11 grafisch dargestellt.

- Ausgehend von den initialen Zentren (rot,blau,grün) wird ein erstes Clustering vorgenommen (siehe Abbildung 4.7). Dazu werden alle Komponenten K der Gaußmischdichte ihrem jeweils nächsten Zentrum zugeordnet 4.8, siehe Abbildung 4.9. Als Abstandsmaß kommt hierfür die Kullback-Leibler Divergenz (3.4) zum Einsatz.
- Neuberechnung der Zentren, siehe Abbildung 4.9.
- Komponente des roten Clusters wird dem Blauen zugewiesen, siehe Abbildung 4.10.

Um festzustellen, ob sich durch diese Neuordnung eine Verbesserung ergeben hat, werden die Zentren Neuberechnet (siehe Abbildung 4.11) und die Abweichung bestimmt.

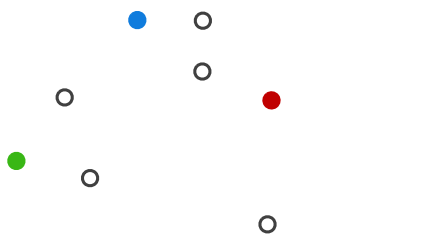


Abbildung 4.7: Initiale Zentren aus initialReduction mit Kmeans++.

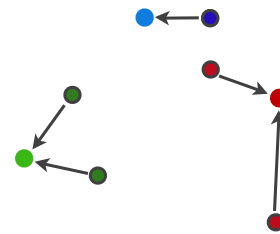


Abbildung 4.8: Zuordnung der Datenpunkte zum jeweils nächsten Zentrum mit initialClustering.

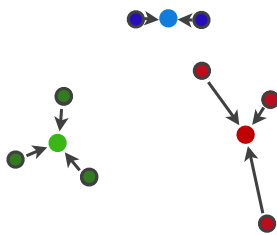


Abbildung 4.9: Neuberechnung der Zentren mit computeCenters.

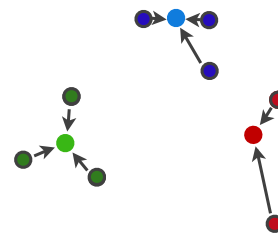


Abbildung 4.10: Wechsel des Clusters.

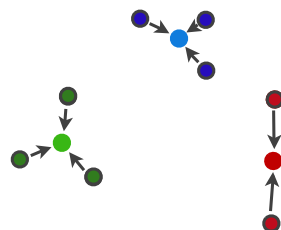


Abbildung 4.11: Neuberechnete Zentren, mit computeCenters.

4.2.3 Optimierung

In dieser Phase (*optimizeReduction*) wird die Approximation durch Einsatz numerischer Methoden weiter verbessert. Dies entspricht der Lösung eines nichtlinearen Optimierungsproblems mit

$$\underline{\eta}_{min} = \arg \min_{\underline{\eta}} G(\tilde{f}(\underline{x}), f(\underline{x}, \underline{\eta})) .$$

Im eindimensionalen Fall wird das Newtonverfahren verwendet, um eine Lösung des Optimierungsproblems zu berechnen. Im höherdimensionalen Fall kommt üblicherweise ein Gradientenabstiegsverfahren zum Einsatz. Dadurch vermeidet man die für das Newtonverfahren notwendige Berechnung der zweiten Ableitung. Auf den Gradientenabstieg wird in Kapitel 5 näher eingegangen.

Das Newtonverfahren berechnet die Nullstelle einer Funktion, genau wie das Regula-Falsi-, Bisektions- oder das Sekantenverfahren. Die Nullstelle wird nicht direkt berechnet. Es erfolgt lediglich eine näherungsweise Berechnung. Voraussetzung dafür ist, dass die vorliegende Funktion zweifach stetig differenzierbar ist, da sonst die benötigten partiellen Ableitungen nicht berechnet werden können. Hier wird die Nullstelle von $\nabla G = \frac{\partial G}{\partial \eta}$ berechnet. Das Verfahren benötigt eine erste Näherung $\underline{\eta}_0$ für die Nullstelle. An die Stelle $\nabla G(\tilde{f}(\underline{x}), f(\underline{x}, \underline{\eta}_0))$ wird eine Tangente angelegt. Die Nullstelle der Tangente wird dann für die nächste Iteration verwendet und entspricht dabei der aktuellen Näherung. Zum Zeitschritt k entspricht die k - Näherung

$$\underline{\eta}_k = \underline{\eta}_{k-1} - \mathbf{H}^{-1} \cdot \nabla G(\tilde{f}(\underline{x}), f(\underline{x}, \underline{\eta}_{k-1})) , \quad (4.1)$$

wobei \mathbf{H} der Hessematrix entspricht.

4.2.4 Gewichtsoptimierung

Zusätzlich zu den genannten Verfahren kann eine Gewichtsoptimierung vorgenommen werden. Hierbei wird versucht die Gütefunktion (3.1) zwischen wahrer und approximierter Dichte ausschließlich durch Variation der Gewichte w_1, \dots, w_k der approximierten Dichte zu minimieren. Dabei sind die Nebenbedingungen (4.2) und (4.3) einzuhalten. Bei der Gewichtsoptimierung wird das folgende quadratische Optimierungsproblem mit den Nebenbedingungen gelöst.

Zielfunktion :

$$Z(w_1, \dots, w_k) = \arg \min_{\underline{w}} G(f(x), \tilde{f}(x))$$

unter den Nebenbedingungen :

$$[w_1, \dots, w_k]^T \geq \underline{0} \quad (4.2)$$

$$\sum_{j=1}^k w_j = 1 . \quad (4.3)$$

Als Lösung des Optimierungsproblems erhält man die Gewichte w_1 bis w_k . Durch Nebenbedingung (4.2) wird sichergestellt, dass die berechneten Gewichte nicht negativ sind. Die Nebenbedingung (4.3) repräsentiert die Forderung nach normierten Gewichten.

Einzelne Gewichte können durch die Optimierung auf 0 gesetzt werden. Dies hat zur Folge, dass die zugehörige Komponente aus der Dichte entfernt werden kann. In diesem Fall kann die reduzierte Dichte auch weniger Komponenten haben als vorher festgelegt. Die Gewichtsoptimierung kann sowohl vor, als auch nach dem Newton-/Gradientenabstiegsverfahren angewendet werden. Ohne die Nebenbedingungen kann das Optimierungsproblem geschlossen durch folgendes lineares Gleichungssystem

$$\begin{bmatrix} \int_{\mathbb{R}} \tilde{\mathcal{N}}_1^2 dx & & \\ & \ddots & \\ & & \int_{\mathbb{R}} \tilde{\mathcal{N}}_L^2 dx \end{bmatrix} \cdot \underline{w} = \sum_{j=1}^K \tilde{w}_j \cdot \begin{bmatrix} \int_{\mathbb{R}} \tilde{\mathcal{N}}_j \cdot \tilde{\mathcal{N}}_1 dx \\ \vdots \\ \int_{\mathbb{R}} \tilde{\mathcal{N}}_j \cdot \tilde{\mathcal{N}}_L dx \end{bmatrix}$$

gelöst werden. Die Lösung des Gleichungssystem dient als Startlösung für das quadratische Optimierungsproblem verwendet. Zur Lösung von quadratischen Optimierungsproblemen können Verfahren des Operations Research, wie z.B. das Verfahren von Kuhn-Tucker verwendet werden.

4.2.5 Laufzeiteigenschaften

Initialisierung

Um eine gültige Startlösung zu erhalten, kann eines der bereits genannten Reduktionsverfahren verwendet werden. Problematisch ist, dass die Laufzeit des gesamten Verfahrens direkt von dem verwendeten Initialisierung abhängt. Wichtig für die Wahl des Verfahrens ist das Verhältnis aus Zeitaufwand und Genauigkeit. Es ist ratsam ein Verfahren zu verwenden, das eine gute Startlösung ermittelt, da dadurch die zweite Phase, die Clusterbildung, abgekürzt oder gar übersprungen und direkt mit der Optimierung begonnen werden kann.

Wird das Kmeans++ Verfahren verwendet, hängt die Güte der Startlösung wesentlich von der zufälligen Wahl des ersten Clusterzentrums ab. Man kann für die Initialisierung eine obere Schranke für die Abweichung definieren. Sollte diese überschritten werden, d.h. die Abweichung zwischen der Startlösung und der wahren Dichte ist zu groß, wird erneut eine Startlösung berechnet. Damit sich das initiale Clustering auch unterscheidet, muss sichergestellt werden, dass sich die zu Anfang zufällig gewählten Clusterzentren nicht ähneln. Es ist zu beachten, dass bei einer Initialisierung mit Kmeans++ die Abweichungen der einzelnen Startlösungen unterschiedlich sind, da sich sonst die Güte der Initialisierung nur marginal ändert.

Der zusätzliche Zeit- und Rechenaufwand für die mehrfache Berechnung der Startlösung lässt

sich durch eine verringerte Laufzeit des eigentlichen Clusterings rechtfertigt. Im besten Fall kann eine Startlösung gefunden werden, die so gering von der wahren Dichte abweicht, dass die zweite Phase übersprungen und direkt mit der Optimierung begonnen werden kann. Außerdem wird für die Berechnung einer Startlösung mit Hilfe der Kmeans++ Methode viel weniger Zeit und Rechenaufwand benötigt als für einen Schritt des Clusterings. Da die erste Komponente zufällig gewählt wird, müssen $L - 1$ Cluster berechnet werden. Dafür müssen in jedem Schritt $L \cdot K$ Distanzen berechnet werden.

Clustering

Es werden zwei Abbruchkriterien verwendet. Anhand dieser wird entschieden, ob es sich lohnt eine neue Zuordnung und deren Abweichung zur wahren Dichte zu berechnen:

- unterschreiten der Abweichungsschranke
- Zuordnungen konstant

Konkret bedeutet dies, dass sobald sich für alle K Komponenten der wahren Dichte keine Verbesserung des Gütemaßes ergibt, die Iteration vorzeitig abgebrochen werden kann. Genau wie die obere Schranke für die Abweichung, kann ebenfalls die maximale Anzahl an Iterationen festgelegt werden.

Wird diese Phase nicht vorzeitig unterbrochen, wird jede der K Komponenten im Verlauf der Clusterbildung einmal jedem der L Clusterzentren, ausschließlich des initialen Zentrums zugeordnet. Folglich müssen die vorher beschriebenen Operationen $K \cdot (L - 1)$ mal für jeden Iterationsschritt durchgeführt werden.

Minimierung der Gütefunktion mittels Gradientenabstieg

Ziel der Studienarbeit ist es das univariate Verfahren beschrieben in [1] auf den multivariaten Fall zu erweitern. Dabei gestaltet sich die Erweiterung von Phase 1 und 2 einfach, da Clustering-Verfahren grundsätzlich für den multivariaten Fall ausgelegt sind. Die Erweiterung von Phase 3 ist schwieriger, da darauf geachtet werden muss, dass die Eigenschaften der Kovarianzmatrix erhalten bleiben. Ebenfalls ist die Berechnung der zweiten Ableitung, die für die Schrittweite des Newtonverfahrens benötigt wird, sehr aufwendig. Daher wird die Schrittweite durch andere Methoden berechnet. Dieses Kapitel stellt im Wesentlichen das für die Optimierung verwendete *Gradientenabstiegsverfahren* vor. In einem ersten Schritt wird der Ablauf des Gradientenabstiegs beschrieben. Dafür werden die mathematischen Grundlagen und Ansätze näher betrachtet. Außerdem wird auf die Herleitung des Gradienten selbst eingegangen. Hierfür sind teilweise komplexe Umformungen und auch fundiertes mathematisches Wissen nötig. Daher wird versucht auf die Herleitung schrittweise hinzuführen. Da ein anderes Vorgehen als bei Williams [3] eingeschlagen wurde, wird der Unterschied zu seiner Herleitung erklärt. In Anhang A befinden sich ausführliche Umformungen zwischen den beiden Herleitungen. Dadurch wird die Korrektheit der Herleitung sichergestellt.

5.1 Gradientenabstiegsverfahren

Die berechnete Approximation der Gaußmischdichte durch Phase 1 und 2 kann mit Hilfe des Gradientenabstiegsverfahren weiter verbessert werden. Dazu wird die Funktion des verwendeten Gütemaßes (3.1) in Abhängigkeit von den Parametern der approximierten Dichte minimiert. Genau wie im eindimensionalen Fall, gilt für multivariate Funktionen, dass die Ableitung nach $\underline{\eta}$ im Minimum $\underline{0}$ sein muss. Die Methode des steilsten Abstiegs startet mit dem Parametervektor $\underline{\eta}_0$ der berechneten Approximation und tastet sich über eine Folge von Iterationsschritten zum tiefsten erreichbaren Punkt des quadratischen Gütemaßes hinab. Es kann aber nicht gewährleistet werden, dass ausgehend von dem Startvektor $\underline{\eta}_0$ auch das globale Minimum gefunden wird. Vielmehr ist es wahrscheinlich, dass der Gradientenabstieg in einem lokalen Minimum endet.

Dies ist darauf zurückzuführen, dass die Gütefunktion (3.1) nicht konvex ist. Für das Gradientenabstiegsverfahren werden zuerst die partiellen Ableitungen des Gütemaßes $G(\underline{\eta})$ nach dem Parametervektor $\underline{\eta}$ berechnet. Der Gradient der Gütefunktion zeigt stets in Richtung des steilsten Anstiegs. Folglich zeigt der negative Gradient in Richtung des steilsten Abstiegs und es gilt zum Zeitschritt t bei Wahl einer geeigneter Schrittweite $\alpha_t > 0$ der Zusammenhang

$$G(\underline{\eta}_{t+1}) = G(\underline{\eta}_t - \alpha_{t+1} \cdot \nabla G(\underline{\eta})_{t+1}) \leq G(\underline{\eta}_t) .$$

Die Parameter der approximierten Dichte werden iterativ verbessert. Dabei ist

$$\underline{\eta}_{t+1} = \underline{\eta}_t - \alpha_{t+1} \cdot \nabla G(\underline{\eta})_{t+1} .$$

Listing 5.1 gibt den Pseudo-Code für das Verfahren an. Im folgenden wird auf zwei wichtige Punkte des Verfahrens näher eingegangen.

```

1 /*  $\underline{\eta}_0$  aus Phase 2 /*
2  $dist_0 \leftarrow G_{NSID}(\tilde{\eta}, \underline{\eta}_0)$ 
3  $w_{1\dots k} \leftarrow \text{sqrt}(w_{1\dots k})$ 
4  $\mathbf{P}_{1\dots k} \leftarrow \text{chol}(\mathbf{\Sigma}_{1\dots k})$ 
5  $t \leftarrow 0$ 
6 repeat
7    $\left[ \frac{\partial G(\underline{\eta}_t)}{\partial w}, \frac{\partial G(\underline{\eta}_t)}{\partial \underline{\mu}}, \frac{\partial G(\underline{\eta}_t)}{\partial \mathbf{P}} \right] \leftarrow \text{calcGrad}(\tilde{\eta}, \underline{\eta}_t)$ 
8    $\nabla G(\underline{\eta})_t \leftarrow \text{vectorize} \left( \frac{\partial G(\underline{\eta})}{\partial w}, \frac{\partial G(\underline{\eta})}{\partial \underline{\mu}}, \frac{\partial G(\underline{\eta})}{\partial \mathbf{P}} \right)$ 
9    $\alpha_t \leftarrow \text{lineSearch}(\tilde{\eta}, \underline{\eta}_{t-1}, \nabla G(\underline{\eta})_0, \nabla G(\underline{\eta})_t, dist_{t-1})$ 
10   $\underline{\eta}_t \leftarrow \underline{\eta}_{t-1} + \alpha_t \cdot \nabla G(\underline{\eta})_t$ 
11   $dist_t \leftarrow G_{NSID}(\tilde{\eta}, \underline{\eta}_t)$ 
12   $t++$ 
13 until  $|dist_t - dist_{t+1}| \leq 10^{-5}$ 
14  $w_{1\dots k} \leftarrow w_{1\dots k}^2$ 
15  $\mathbf{\Sigma}_{1\dots k} \leftarrow \mathbf{P}_{1\dots k}^T \cdot \mathbf{P}_{1\dots k}$ 

```

Listing 5.1: Pseudocode für Gradientenabstieg

5.2 Schrittweitenbestimmung

Die Berechnung einer geeigneten Schrittweite gestaltet sich nicht einfach. Für ihre Berechnung wird die *Linien suche* [15] in Richtung des Gradienten verwendet. Dabei gilt es die Lösung des folgenden eindimensionalen Optimierungsproblem

$$\alpha_t = \min_{\alpha_t} G(\underline{\eta}_t - \alpha_{t+1} \cdot \nabla G(\underline{\eta})_{t+1}) , \quad \alpha_t \in [0, 1]$$

zu finden. Berechnet wird, wie weit in Richtung des Gradienten abgestiegen werden kann, ohne das Minimum zu überschreiten und noch eine möglichst große Verbesserung der Parameter zu erzielen. Ist die Schrittweite zu groß verpasst man das Minimum. Während eine zu kurze Schrittweite ineffizient ist. Die Liniensuche versucht zuerst die Position des Minimums zu schätzen und dann zu interpolieren. Meistens wird die zu minimierende Funktion durch eine quadratische oder kubische Funktion interpoliert. Mit Hilfe der Interpolation kann die optimale Schrittweite ermittelt werden. Es wird überprüft, ob das Minimum verpasst wurde. Ist dies der Fall wird eine geringere Schrittweite berechnet. Ist die Veränderung des Funktionswert unterhalb einer Grenze $0 < \Delta G \ll 1$, wird der Gradientenabstieg vorzeitig abgebrochen.

5.3 Gradient des Gütemaßes

Im Folgenden wird kurz auf die einzelnen Schritte eingegangen, die zur Herleitung des Gradienten des integralen quadratischen Gütemaßes nötig sind. Eine ausführliche Herleitung und Umformung auf den von Williams [16] berechneten Gradienten ist in Anhang A zu finden.

Grundsätzlich existieren zwei unterschiedliche Wege den Gradienten herzuleiten. Zum einen kann man zuerst die Ableitung des Gütemaßes berechnen und löst dann die Integrale auf, oder man berechnet zuerst das Integral und danach die Ableitung. Dies ist in diesem speziellen Fall möglich, da die Variablen nach denen differenziert wird und die Variablen über die integriert wird unabhängig voneinander sind. Im Gegensatz zur Herleitung von Williams wurde hier zuerst differenziert und danach integriert.

Für das Gewicht der reduzierten Dichte wird w_j^2 verwendet um sicherzustellen, dass das Gewicht während des Gradientenabstiegs immer positiv bleibt. Für die Kovarianzmatrix \mathbf{C}_k einer Komponente der reduzierten Dichte wird ihre Wurzel verwendet $\mathbf{C}_k = (\mathbf{P}_k \cdot \mathbf{P}_k^T) = \mathbf{C}_k^T = (\mathbf{P}_k^T \cdot \mathbf{P}_k)$, damit die Kovarianzmatrix stets positiv semidefinit bleibt. Nachfolgend wird nur der Gradient für ein einzelnes $\underline{\eta}_k$ berechnet. Die Bestimmung des gesamten Gradienten für $\underline{\eta}$ erfolgt analog. Der gesamte Gradient besteht aus $3K$ partiellen Ableitungen und enthält $K \cdot (N + N^2 + 1)$ Einträge. K entspricht der Anzahl an Komponenten und N der Dimension der Gaußmischdichte.

Für die Ableitung des Gütemaßes nach $\underline{\eta}_k$ gilt

$$\nabla G(\underline{\eta}_k) = \int_{\mathbb{R}^N} f(\underline{x}, \underline{\eta}) \cdot \frac{\partial f(\underline{x}, \underline{\eta})}{\partial \underline{\eta}_k} d\underline{x} - \int_{\mathbb{R}^N} \tilde{f}(\underline{x}, \tilde{\eta}) \cdot \frac{\partial f(\underline{x}, \underline{\eta})}{\partial \underline{\eta}_k} d\underline{x} . \quad (5.1)$$

Zusätzlich benötigt man die Ableitung der Gaußmischdichte nach Gewicht $\frac{\partial f(\underline{x}, \underline{\eta})}{\partial w_k}$, Erwartungswert $\frac{\partial f(\underline{x}, \underline{\eta})}{\partial \underline{\mu}_k}$ und Wurzel der Kovarianz $\frac{\partial f(\underline{x}, \underline{\eta})}{\partial \mathbf{P}_k}$ der k -ten Komponente benötigt, welche sich zu

$$\frac{\partial f(\underline{x}, \underline{\eta})}{\partial w_k} = 2w_k \cdot \mathcal{N}(\underline{x}, \underline{\mu}_k, \mathbf{C}_k) , \quad (5.2)$$

$$\frac{\partial f(\underline{x}, \underline{\eta})}{\partial \underline{\mu}_k} = w_k^2 \cdot \mathbf{C}_k^{-1} \cdot (\underline{x} - \underline{\mu}_k) \cdot \mathcal{N}(\underline{x}, \underline{\mu}_k, \mathbf{C}_k) , \quad (5.3)$$

$$\frac{\partial f(\underline{x}, \underline{\eta})}{\partial \mathbf{P}_k} = w_k^2 \cdot [\mathbf{C}_k^{-1} \cdot (\underline{x} - \underline{\mu}_k) \cdot (\underline{x} - \underline{\mu}_k)^T - \mathbf{I}] \cdot \mathbf{C}_k^{-1} \cdot \mathbf{P}_k \cdot \mathcal{N}(\underline{x}, \underline{\mu}_k, \mathbf{C}_k) \quad (5.4)$$

ergeben. Setzt man die partiellen Ableitungen (5.2), (5.3), (5.4) in die Ableitung des Gütemaßes (5.1) ein, müssen als nächstes die Integrale berechnet werden. Dazu benötigt man die Rechenregeln für das Produkt zweier Gaußmischdichten [3] und erhält die partiellen Ableitungen

$$\frac{\partial G(\underline{\eta})}{\partial w_k} = 2 \cdot w_k \cdot \left[\sum_{j=1}^m w_j^2 \cdot \mathcal{N}(\underline{\mu}_j, \underline{\mu}_k, \mathbf{C}_j + \mathbf{C}_k) - \sum_{i=1}^n \tilde{w}_i \cdot \mathcal{N}(\tilde{\underline{\mu}}_i, \underline{\mu}_k, \tilde{\mathbf{C}}_i + \mathbf{C}_k) \right] , \quad (5.5)$$

$$\frac{\partial G(\underline{\eta})}{\partial \underline{\mu}_k} = \text{Vec} \left(w_k^2 \cdot \left[\sum_{i=1}^n \tilde{w}_i \cdot \mathcal{N}(\tilde{\underline{\mu}}_i, \underline{\mu}_k, \tilde{\mathbf{C}}_i + \mathbf{C}_k) \cdot \mathbf{C}_k^{-1} \cdot (\underline{M}_1^s - \underline{\mu}_k) \right. \right. \\ \left. \left. - \sum_{j=1}^m w_j^2 \cdot \mathcal{N}(\underline{\mu}_j, \underline{\mu}_k, \mathbf{C}_j + \mathbf{C}_k) \cdot \mathbf{C}_k^{-1} \cdot (\underline{M}_1^t - \underline{\mu}_k) \right] \right) ,$$

$$\frac{\partial G(\underline{\eta})}{\partial \mathbf{P}_k} = \text{Vec} \left(w_k^2 \cdot \mathbf{C}_k^{-1} \cdot \left[\sum_{j=1}^m w_j^2 \cdot \mathcal{N}(\underline{\mu}_j, \underline{\mu}_k, \mathbf{C}_j + \mathbf{C}_k) \right. \right. \\ \cdot [\mathbf{M}_2^t + \underline{M}_1^t (\underline{M}_1^t)^T - \underline{M}_1^t \underline{\mu}_k^T - \underline{\mu}_k (\underline{M}_1^t)^T + \underline{\mu}_k \underline{\mu}_k^T - \mathbf{C}_k] \\ \left. \left. - \sum_{i=1}^n \tilde{w}_i \cdot \mathcal{N}(\tilde{\underline{\mu}}_i, \underline{\mu}_k, \tilde{\mathbf{C}}_i + \mathbf{C}_k) \right. \right. \\ \left. \left. \cdot [\mathbf{M}_2^s + \underline{M}_1^s (\underline{M}_1^s)^T - \underline{M}_1^s \underline{\mu}_k^T - \underline{\mu}_k (\underline{M}_1^s)^T + \underline{\mu}_k \underline{\mu}_k^T - \mathbf{C}_k] \right] \cdot \mathbf{C}_k^{-1} \cdot \mathbf{P}_k \right) .$$

Hierbei stehen $\underline{M}_1^s, \mathbf{M}_2^s$ bzw. $\underline{M}_1^t, \mathbf{M}_2^t$ mit $s \hat{=} k, i$ und $t \hat{=} k, j$ für das erste und zweite Moment, also Erwartungswert und Kovarianz stehen, die mit den Gleichungen (A.8) und (A.9) errechnet werden.

5.3.1 Beispiel

Im folgenden Beispiel wird veranschaulicht, inwiefern der Gradientenabstieg eine Verbesserung erzielt. Dazu wird eine eindimensionale Gaußmischdichte mit zwei sehr ähnlichen Komponenten $\tilde{\eta} = [[0.5 \ 0.5], [0.5 \ 0.4], [0.5 \ 0.5]]$ gewählt (siehe Abbildung 5.1). Es wird bewusst eine Reduktion dieser Dichte verwendet, die sehr weit vom Optimum entfernt ist. Das diese Reduktion gegen die Regeln der Fusion verstößt, spielt für dieses Beispiel keine Rolle. Ausgehend von dieser

Reduktion $\underline{\eta}_0 = [1, 1.5, 0.5]$ wurde der Gradientenabstieg verwendet um die Reduktion zu optimieren.

Nach 32 Schritten erhält man $\underline{\eta}_{32} = [1, 0.45, 0.5025]$. Wie man in Abbildung 5.2 sieht, gelingt dies sehr gut, der Verlauf der Funktionen ist deckungsgleich. Anfänglich bestand eine Abweichung von 65% , die nun auf 0.001% verringert werden konnte.

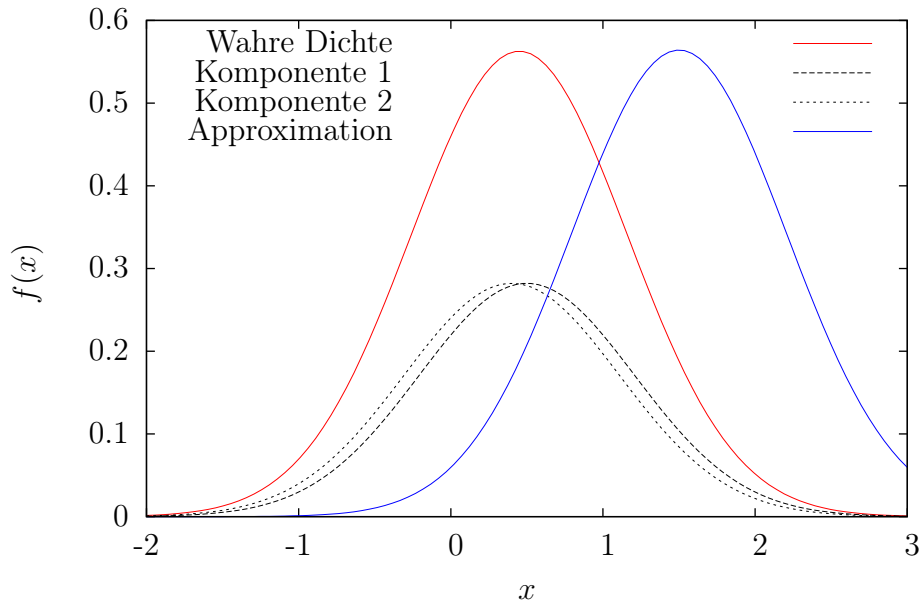


Abbildung 5.1: Wahre Dichte bestehend aus 2 Komponenten mit Approximation.

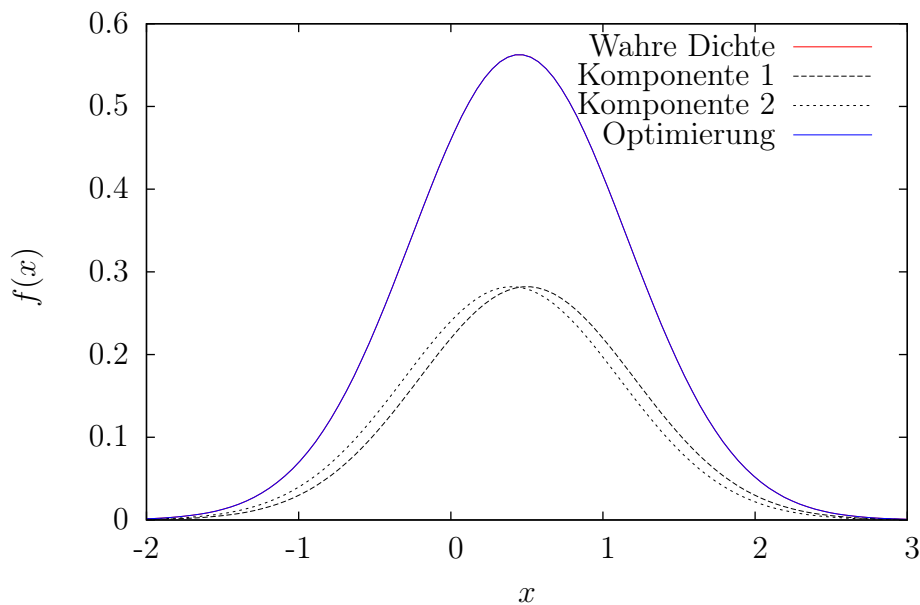


Abbildung 5.2: Wahre Dichte mit optimierter Approximation durch Gradientenabstieg.

Auswertung und Simulation

In diesem Kapitel werden die Simulationsergebnisse für das vorgestellte GMRC-Verfahren präsentiert. Ziel ist es sowohl die Stärken, als auch die Schwächen des mehrstufigen Clustering-Verfahrens aufzuzeigen. Dazu wird gezielt Aufwand und Nutzen der einzelnen Phasen genauer untersucht.

Die Simulationen erfolgten auf einem System mit Intel(R) Core(TM)2 Quad Processor. Der Prozessor besitzt vier Kerne mit jeweils 2.4 GHz. Der Hauptspeicher besteht aus 2 GB. Als Betriebssystem wurde openSUSE Linux Version 11.1 verwendet.

6.1 Beispielablauf

Zunächst wird kurz anhand eines eindimensionalen und zweidimensionalen Beispiels gezeigt werden, wie die einzelnen Phasen die Reduktion bilden und sie ggf. verbessern. Dafür wird eine eindimensionale Dichte mit den Parametern

$$\begin{aligned}\tilde{w} &= [0, 1 \ 0, 1 \ 0, 1 \ 0, 1 \ 0, 1 \ 0, 1 \ 0, 1 \ 0, 1 \ 0, 1 \ 0, 1 \ 0, 1]^T, \\ \tilde{\mu} &= [-3, 5 \ -3 \ -1 \ 0 \ 0, 5 \ 2 \ 3 \ 3, 5 \ 5 \ 5, 5]^T, \\ \tilde{\sigma}^2 &= [0, 6 \ 0, 6 \ 0, 6 \ 0, 6 \ 0, 7 \ 0, 7 \ 1 \ 0, 5 \ 0, 5 \ 0, 5],\end{aligned}$$

verwendet [9], die auf 6 Komponenten reduziert wird.

Für den zweidimensionalen Fall wird eine Dichte mit den Parametern

$$\begin{aligned}\tilde{w} &= \left[\frac{1}{5} \ \frac{1}{5} \ \frac{1}{5} \ \frac{1}{5} \ \frac{1}{5} \right]^T, \\ \tilde{\mu} &= \left[[2 \ 2] \ [1 \ -2] \ [-1 \ 2] \ [-5 \ 0] \ [-7 \ 0] \right]^T, \\ \tilde{\Sigma} &= \text{vec} \left(\Sigma_1 \ \Sigma_2 \ \Sigma_3 \ \Sigma_4 \ \Sigma_5 \right),\end{aligned}$$

$$\Sigma_1 = \begin{bmatrix} 5 & 0,4 \\ 0,4 & 1,5 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 5 & 1 \\ 1 & 2 \end{bmatrix}, \Sigma_3 = \begin{bmatrix} 5 & 0,1 \\ 0,1 & 2 \end{bmatrix}, \Sigma_4 = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}, \Sigma_5 = \begin{bmatrix} 4 & 1,7 \\ 1,7 & 4 \end{bmatrix}$$

verwendet, die auf 3 Komponenten reduziert wird. In Abbildung 6.1 und 6.2 sind die wahren Dichten mit ihren einzelnen Komponenten abgebildet. In Phase 1 wird das Kmeans++ Verfahren für die Initialisierung verwendet. Die Abweichung zwischen wahrer und reduzierter Dichte wurde mit der NISD bewertet. Nach der Initialisierung beträgt die Abweichung 2% bzw. 15,1%. Die initiale Reduktion ist in Abbildung 6.3 und 6.4 veranschaulicht.

Im Anschluss an die Initialisierung wird die Abweichung durch das Clustering auf 1,5% bzw. 9,7% verbessert. In Abbildung 6.5 und 6.6 sind die reduzierten Dichten und ihre Komponenten nach dem eigentlichen Clustering zu sehen.

Nach der dritten und letzten Phase beträgt die Abweichung lediglich 0,1% bzw. 4,7%. Die endgültige reduzierte Dichte und deren Komponenten nach Abschluss der Optimierung ist in Abbildung 6.7 und 6.8 dargestellt. Runnalls und West liefern für das zweidimensionale Beispiel eine Reduktion mit einer Abweichung von 15,1%, dies entspricht der Abweichung nach der Initialisierung mit Hilfe von Kmeans++. Für das eindimensionale Beispiel findet eine Verbesserung um den Faktor 20 statt.

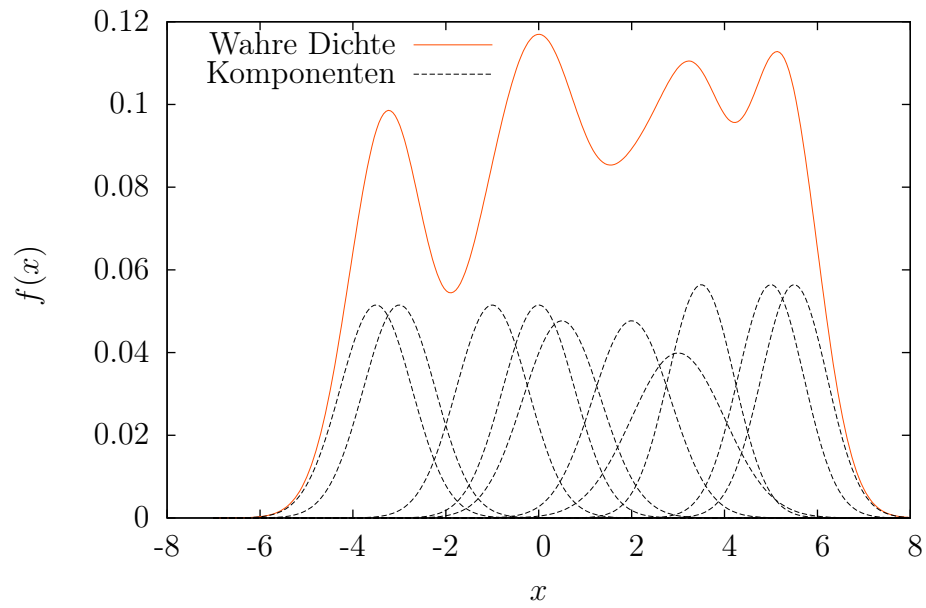


Abbildung 6.1: Wahre Dichte bestehend aus 10 Komponenten.

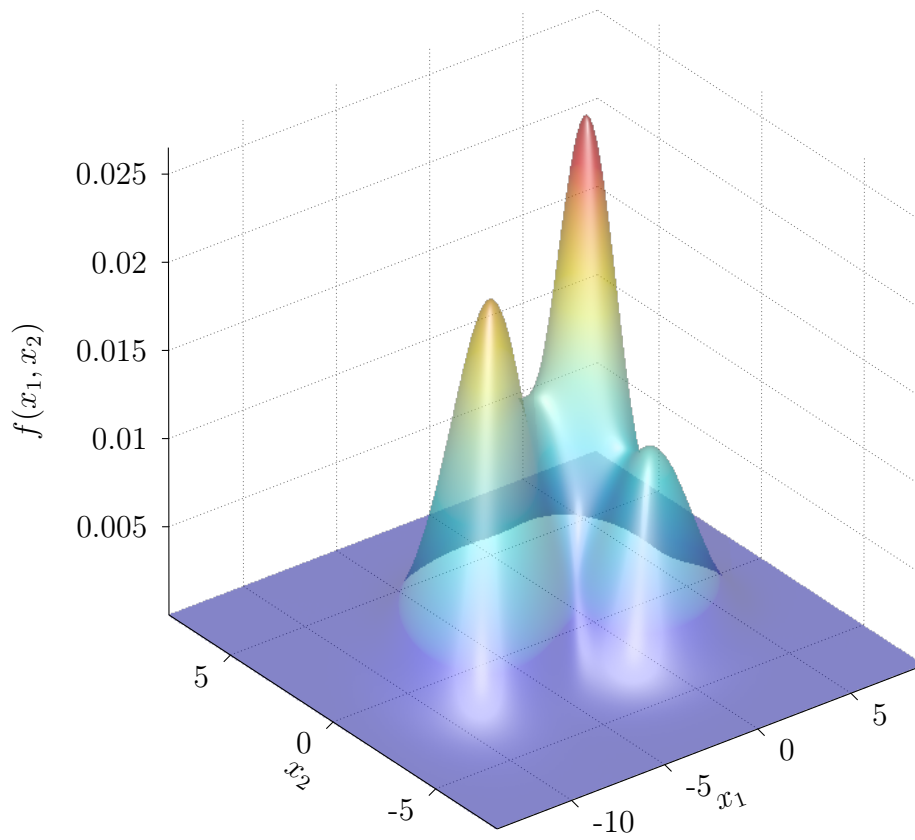


Abbildung 6.2: Wahre Dichte bestehend aus 5 Komponenten.

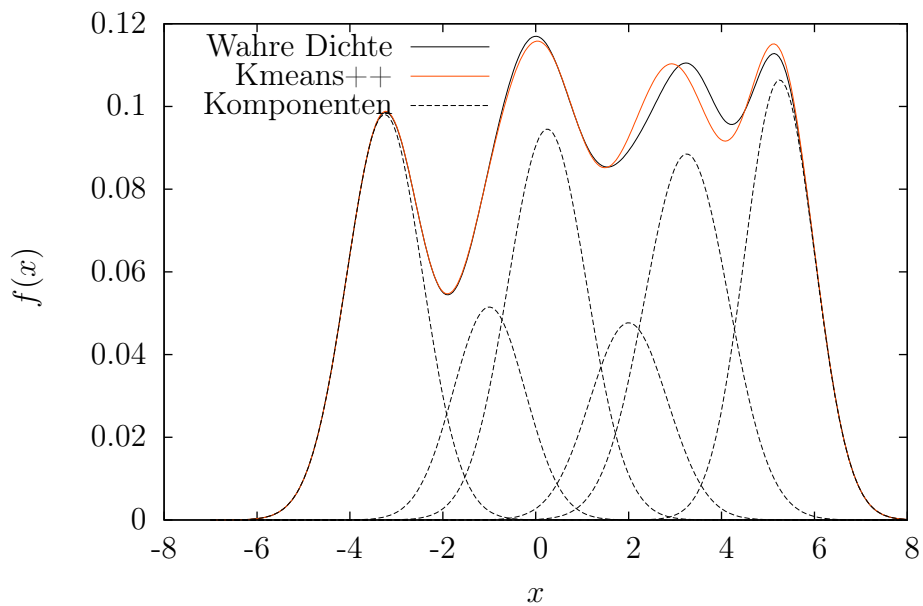


Abbildung 6.3: Reduktion mit 6 Komponenten nach Phase 1 mit Abweichung von 2%.

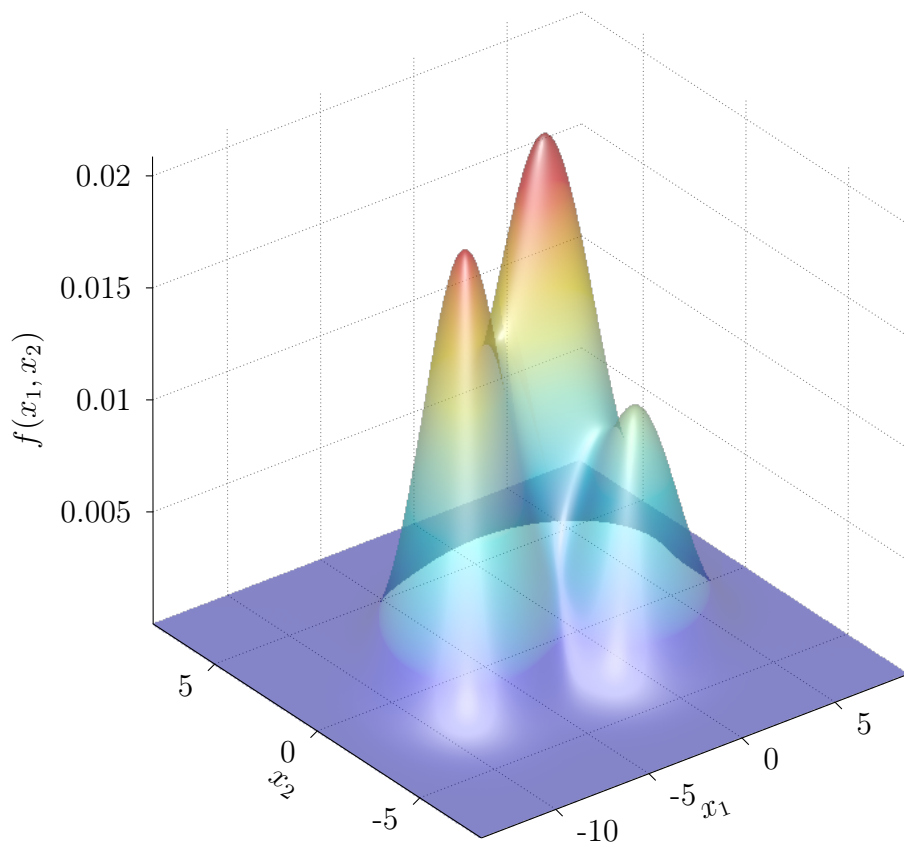


Abbildung 6.4: Reduktion mit 3 Komponenten nach Phase 1 mit Abweichung von 15,1%.

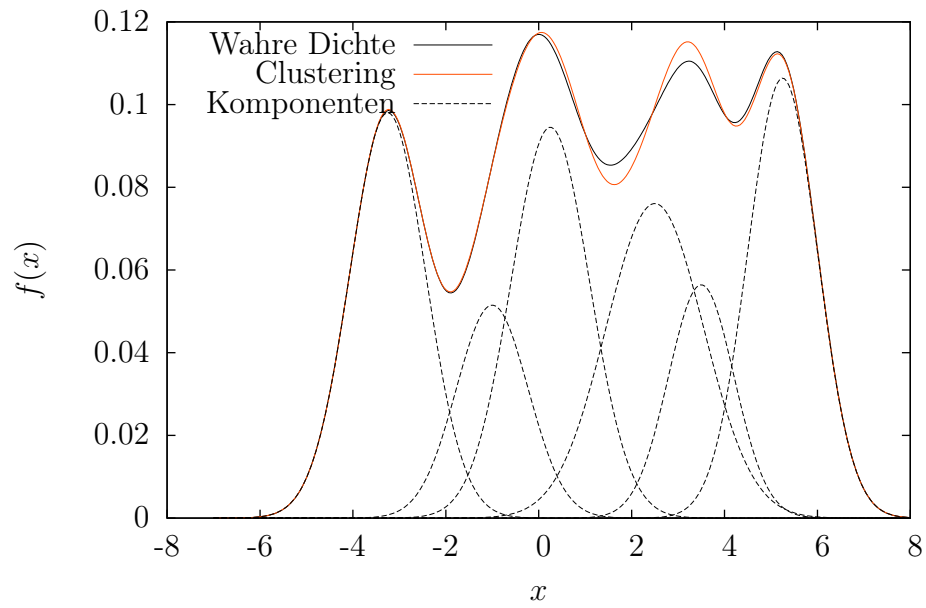


Abbildung 6.5: Reduktion nach Phase 2 mit Abweichung von 1,5%.

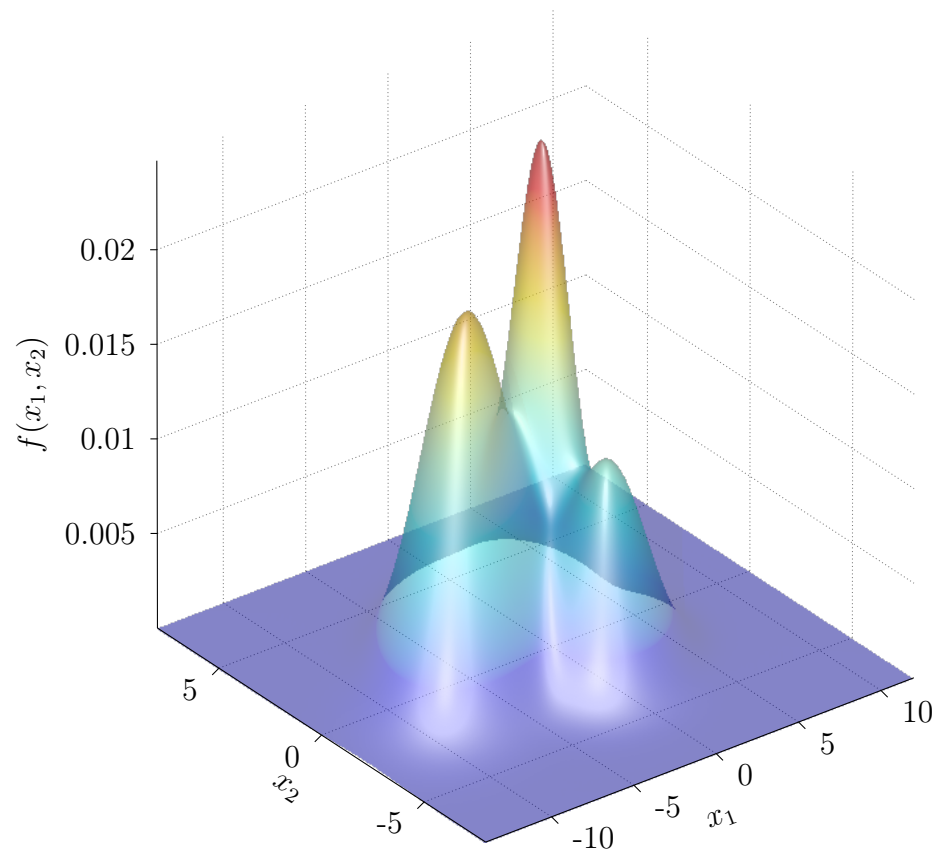


Abbildung 6.6: Reduktion nach Phase 2 mit Abweichung von 9,7%.

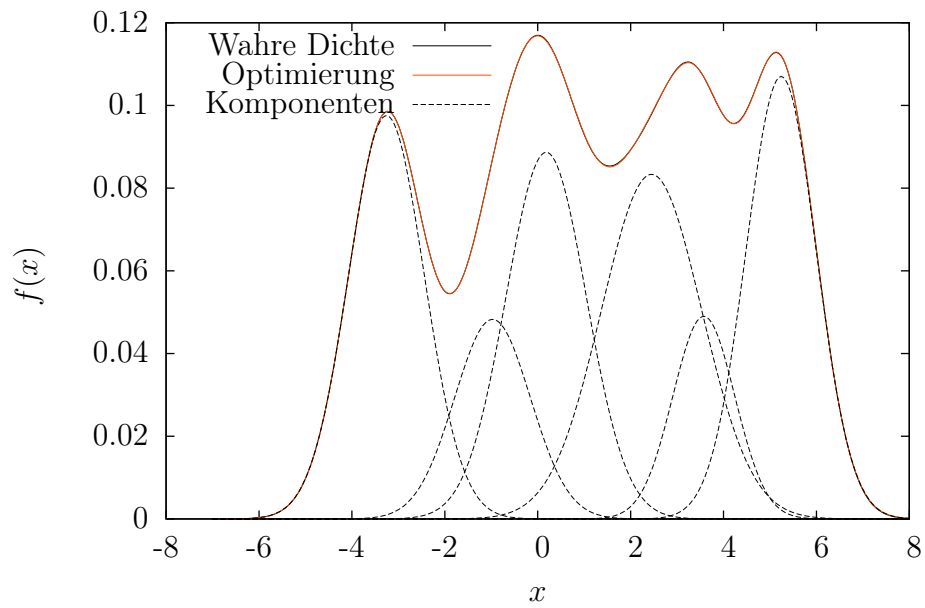


Abbildung 6.7: Reduktion nach Phase 3 mit Abweichung von 0,1%.

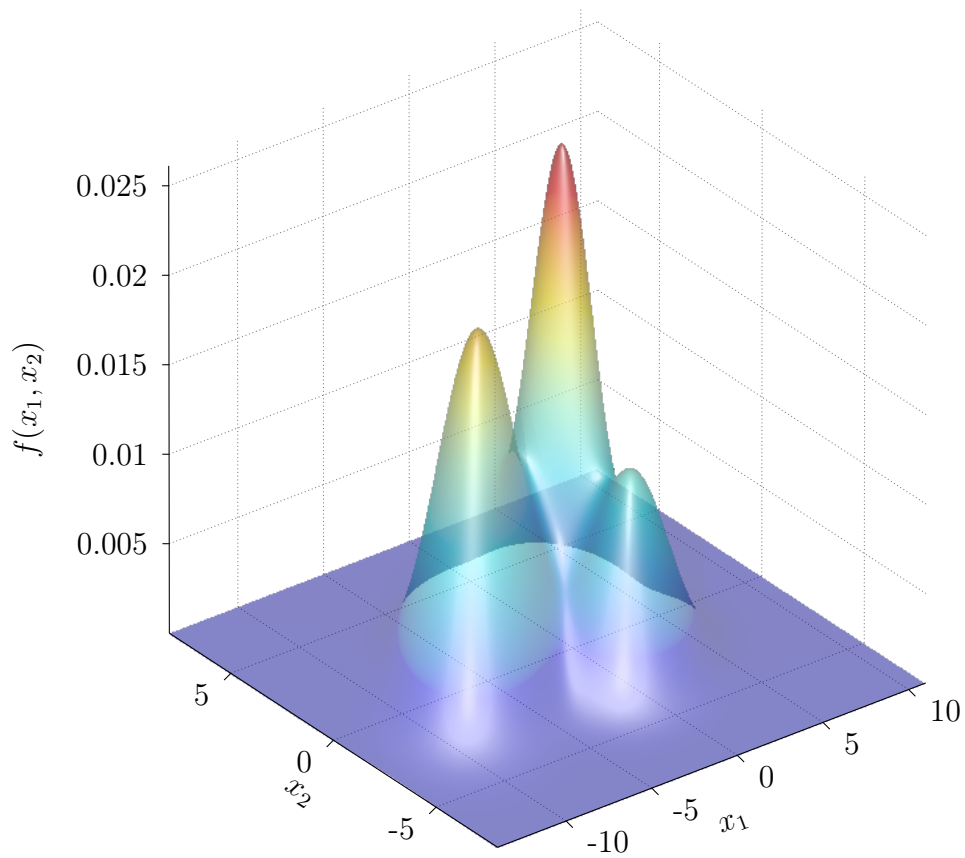


Abbildung 6.8: Reduktion nach Phase 3 mit Abweichung von 4,7%.

6.2 Simulation 1

Im folgenden werden zwei Monte Carlo Simulationen mit jeweils 20 Iterationen durchgeführt. Dadurch beeinflussen Ausreißer die Ergebnisse nur marginal und es erfolgt eine Mittelung der Ergebnisse. Die wahren Gaußmischdichten werden hierfür zufällig erzeugt.

Die Werte für die Gewichte liegen in den Intervallen $w \in (0,05, 0,55)$, die Elemente des Erwartungswertvektors $\underline{\mu}$ und der Kovarianzmatrix Σ liegen in den Intervallen $(0, 64)$ und $(0, 17)$. Die erste Simulation vergleicht die Auswirkungen der verschiedenen Initialisierungen. In den Tabellen 6.1 bis 6.4 sind die Ergebnisse für verschiedene Dimensionen veranschaulicht. Die Abweichung wurde mit der Normierten Integralen Quadratischen Distanz in Prozent berechnet. Zu beachten ist hierbei, dass die einer Phase zugeordnete Abweichung einschließlich aller vorigen Phasen angegeben wird.

Es ist deutlich zu erkennen das mit zunehmender Dimension die Abweichung steigt. Für die Initialisierung liefern Runnalls und Kmeans++ mit wiederholter Initialisierung sehr gute Ergebnisse, wobei mit West die ungenauesten Ergebnisse berechnet werden.

Verwendet man die Ergebnisse von West und Runnalls für die Initialisierung als Referenzwerte, kann man deutlich sehen, wie diese mit Hilfe der Clusteranalyse und der Optimierung weiter verbessert werden können. Gerade höherdimensionale Fälle wie in Tabelle 6.4 aufgeführt sind hierfür bezeichnend. In allen Fällen wurde durch das eigentliche Clustering die größte Verbesserung erzielt.

Da die wahren Gaußmischdichten jeweils nur auf die Hälfte ihrer Komponenten reduziert werden, ist die Verbesserung durch die Optimierung gegenüber dem Clustering eher gering.

Eine stärkere Verbesserung durch die Optimierung ist zu erwarten, wenn auf weniger als die Hälfte der Komponenten reduziert wird.

Die kumulierten Laufzeiten nach den einzelnen Phasen des Verfahrens für Dichten der Dimension 2 können der Tabelle 6.5 entnommen werden. Wie man sieht, hängt die Laufzeit der Initialisierung vom verwendeten Verfahren ab. Damit hängt auch die Laufzeit des gesamten Verfahrens von der verwendeten Initialisierung ab. Vergleicht man die Laufzeiten der unterschiedlichen Phasen, benötigt Phase 1 am wenigsten Zeit. Für die Berechnungen in Phase 2 wird die meiste Zeit benötigt, denn vorallem die Operationen für die Fusion der Komponenten sind multivariaten Fall sehr aufwendig. Phase 3 ist deshalb auch im Vergleich zu Phase 2 eher unkritisch.

Tabelle 6.1: Abweichung in % für Dichte der Dimension 1.

#Komp.	nach Initialisierung				nach Clustering				nach Optimierung			
	Runnalls	West	Kmeans	Kmeans.R	Runnalls	West	Kmeans	Kmeans.R	Runnalls	West	Kmeans	Kmeans.R
10 → 5	11,5	18,1	14,5	12,4	9,9	14,2	11,1	10,5	8,1	12,7	9,5	8,8
20 → 10	1,8	9,2	2,7	2,0	1,3	6,5	1,5	1,3	0,7	5,3	0,7	0,6
30 → 15	1,1	5,6	1,5	1,2	0,7	5,6	0,8	0,7	0,4	2,8	0,3	0,3
40 → 20	1,1	6,6	1,4	1,1	0,8	9,6	0,7	0,7	0,4	3,1	0,3	0,3
50 → 25	0,3	4,7	0,6	0,4	0,2	11,5	0,2	0,2	0,1	1,8	0,1	0,1

Tabelle 6.2: Abweichung in % für Dichte der Dimension 2.

#Komp.	nach Initialisierung				nach Clustering				nach Optimierung			
	Runnalls	West	Kmeans	Kmeans.R	Runnalls	West	Kmeans	Kmeans.R	Runnalls	West	Kmeans	Kmeans.R
10→5	14,5	23,8	16,1	13,2	12,6	16,2	12,1	11,6	12,2	15,8	11,7	11,2
20→10	5,9	14,6	6,9	5,8	4,6	10,2	4,8	4,8	4,3	9,7	4,4	4,4
30→15	3,7	9,6	4,8	3,8	2,7	7,5	3,0	2,7	2,5	6,2	2,7	2,4
40→20	2,9	9,7	3,5	3,0	2,1	7,8	2,2	2,0	1,9	5,8	1,9	1,8
50→25	2,3	8,8	3,2	2,3	1,6	6,5	1,7	1,6	1,4	4,5	1,4	1,4

Tabelle 6.3: Abweichung in % für Dichte der Dimension 4.

#Komp.	nach Initialisierung				nach Clustering				nach Optimierung			
	Runnalls	West	Kmeans	Kmeans.R	Runnalls	West	Kmeans	Kmeans.R	Runnalls	West	Kmeans	Kmeans.R
10→5	31,2	42,9	37,2	32,5	21,1	25,4	23,9	21,7	20,6	24,7	23,4	21,3
20→10	29,1	48,0	34,6	30,8	22,0	26,7	25,7	22,6	21,4	26,3	25,1	22,2
30→15	22,3	37,2	28,9	24,8	16,0	20,9	18,6	17,6	15,5	20,4	18,1	17,0
40→20	19,4	35,5	25,4	21,0	14,6	19,4	16,9	15,7	14,2	19,1	16,4	15,3
50→25	16,1	31,8	21,3	18,4	12,1	17,6	13,9	12,9	11,8	17,3	13,4	12,5

Tabelle 6.4: Abweichung in % für Dichte der Dimension 8.

#Komp.	nach Initialisierung				nach Clustering				nach Optimierung			
	Runnalls	West	Kmeans	Kmeans.R	Runnalls	West	Kmeans	Kmeans.R	Runnalls	West	Kmeans	Kmeans.R
10→5	42,5	72,9	58,1	47,3	22,7	38,0	17,3	13,7	22,1	36,6	16,6	13,4
20→10	38,1	56,3	42,5	37,9	21,1	24,1	21,5	21,4	20,9	23,9	21,1	21,0
30→15	38,2	56,2	43,6	37,3	20,2	24,3	25,4	22,0	19,5	24,0	24,8	21,5
40→20	30,3	59,9	35,8	29,3	18,7	26,5	21,7	18,3	18,2	26,0	20,9	17,7
50→25	25,2	45,9	29,3	26,2	16,3	20,4	19,2	17,1	15,9	20,2	18,6	16,5

Tabelle 6.5: Zeit in Sekunden für Dichte der Dimension 2.

#Komp.	nach Initialisierung				nach Clustering				nach Optimierung			
	Runnalls	West	Kmeans	Kmeans.R	Runnalls	West	Kmeans	Kmeans.R	Runnalls	West	Kmeans	Kmeans.R
10→5	0,01	0,01	0,01	0,03	0,11	0,10	0,11	0,12	0,26	0,20	0,20	0,22
20→10	0,02	0,01	0,02	0,09	0,67	0,62	0,67	0,73	0,97	0,99	0,97	1,04
30→15	0,04	0,02	0,04	0,17	2,14	1,63	2,11	2,22	2,78	2,47	2,85	2,79
40→20	0,08	0,03	0,07	0,26	4,93	3,87	4,85	5,04	6,06	5,17	6,01	6,12
50→25	0,12	0,04	0,11	0,42	9,56	7,84	9,33	9,63	10,97	9,40	10,93	11,24

6.3 Simulation 2

In einer weiteren Simulation werden die Ergebnisse des mehrstufigen Verfahrens direkt mit den bestehenden Implementierungen für West und Runnalls verglichen. Hierfür wurde wieder eine Monte Carlo Simulation mit 20 Iterationen verwendet. Die Parameter der wahren Dichten liegen in den Intervallen $w \in (0.05, 0.55)$, $\mu \in (0, 64)$ und $\sigma^2 \in (0, 17)$. Alle zufällig erzeugten Gaußmischdichten haben dieselbe Anzahl an Komponenten G mit der Dimension D . Im Laufe der Simulation wird die Anzahl der Komponenten der approximierten Dichte schrittweise bis zu einer unteren Schranke (5 Komponenten) verringert.

Wie auch schon bei Simulation 1 wird die Abweichung mit der Normierten Integralen Quadratischen Distanz berechnet. Die Ergebnisse für unterschiedlich dimensionierte Gaußmischdichten sind in den Abbildungen 6.9 und 6.10 dokumentiert.

In den Abbildungen kann man deutlich sehen, dass je mehr Komponenten zusammengefasst werden, desto größer ist auch die Abweichung. Abbildung 6.10 ist hierfür ein gutes Beispiel. Wird hier auf weniger als 15 Komponenten reduziert, vergrößert sich die Abweichung drastisch. Außerdem sind insbesondere für sehr wenige Komponenten die signifikanten Stärken des mehrstufigen Clustering-Verfahrens zu erkennen. Vor allem in Abbildung 6.10 ist dieses Verhalten ersichtlich. Bei drastischer Reduktion der wahren Dichte werden hier im Vergleich zu Runnalls und vor allem auch zu West wesentlich bessere Ergebnisse erzielt. Nimmt allerdings die Anzahl der Komponenten zu, wird der Unterschied geringer. Beispielsweise für eine Reduktion auf 40 Komponenten ist die Verbesserung gegenüber Runnalls nur noch marginal. Im Vergleich zu West sind die Ergebnisse aber dennoch besser.

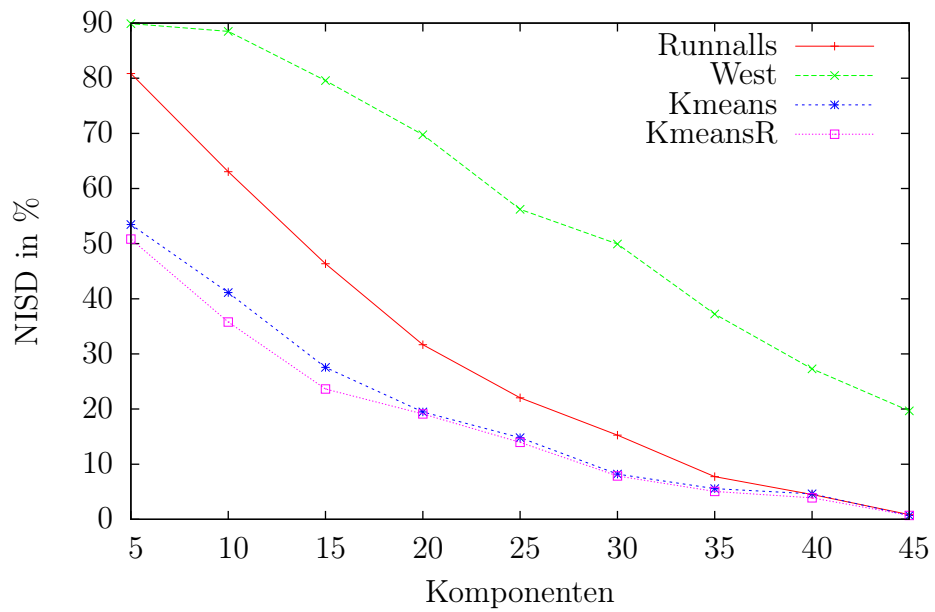


Abbildung 6.9: Komponentenreduktion der Dichte mit D=4, G=50.

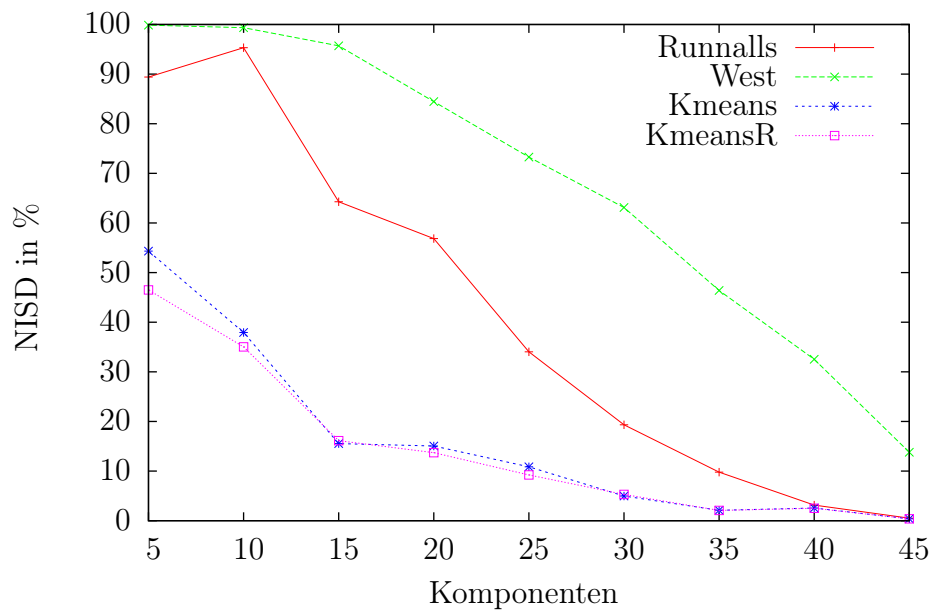


Abbildung 6.10: Komponentenreduktion der Dichte mit D=8, G=50.

6.4 Anwendung

Als letztes wird das Verfahren in einer konkreten Anwendung zum Einsatz gebracht. Grundlage dafür bildet das von Salmond [17] bzw. Williams [16] beschriebene Single-Target Multihypothesen-Tracking-Szenario. Dabei wird eine Flugobjekt lokalisiert, wobei neben Messungen der Objektposition und Störmessungen (*Clutter*) mitberücksichtigt werden.

Das Flugobjekt bewegt sich innerhalb eines zweidimensionalen Raumes. Der Zustand des Flugobjekts zum Zeitpunkt t entspricht

$$\underline{x}_t = [x_t, \dot{x}_t, y_t, \dot{y}_t],$$

wobei $[x, y]^T$ die Position und $[\dot{x}, \dot{y}]^T$ die Geschwindigkeit darstellt.

Als Bewegungsmodell wird

$$\underline{x}_{t+1} = \begin{bmatrix} 1 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta t \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \underline{x}_t + \begin{bmatrix} \frac{\Delta t^2}{2} & 0 \\ \Delta t & 0 \\ 0 & \frac{\Delta t^2}{2} \\ 0 & \Delta t \end{bmatrix} \cdot \underline{w}_t$$

verwendet, \underline{w}_t entspricht einem gaußverteilten Systemrauschen mit Erwartungswert $\underline{0}$ und konstanter Kovarianzmatrix $\Sigma = \mathbf{I}_2$, wobei \mathbf{I}_2 die 2×2 Einheitsmatrix darstellt, Δt gibt den Abstand zwischen zwei Messungen an. In diesem konkreten Fall wurde $\Delta t = 1$ gewählt.

Das Messmodell entspricht

$$\underline{z}_{t+1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \underline{x}_t + \underline{v}_t.$$

Gemessen wird dabei nur die Position. Die Geschwindigkeit fließt nicht in das Messmodell mit ein. Hier entspricht \underline{v}_t einem gaußverteilten Systemrauschen mit Erwartungswert $\underline{0}$ und konstanter Kovarianzmatrix $\Sigma = \mathbf{I}_2$, wobei \mathbf{I}_2 die 2×2 Einheitsmatrix darstellt. Wie zu Anfangs erwähnt, müssten zu jedem Zeitschritt alle Hypothesen der Objektposition und alle unmöglichen Messungen (inkl. Clutter) betrachtet werden. Damit würde sowohl der Aufwand für die Prädiktion, als auch der Aufwand für die Fusion exponentiell ansteigen. Um den Aufwand konstant zu halten, werden nun die Anzahl der Hypothesen, das heißt die Komponenten der Gaußmischdichte reduziert (siehe Abbildung 6.11).

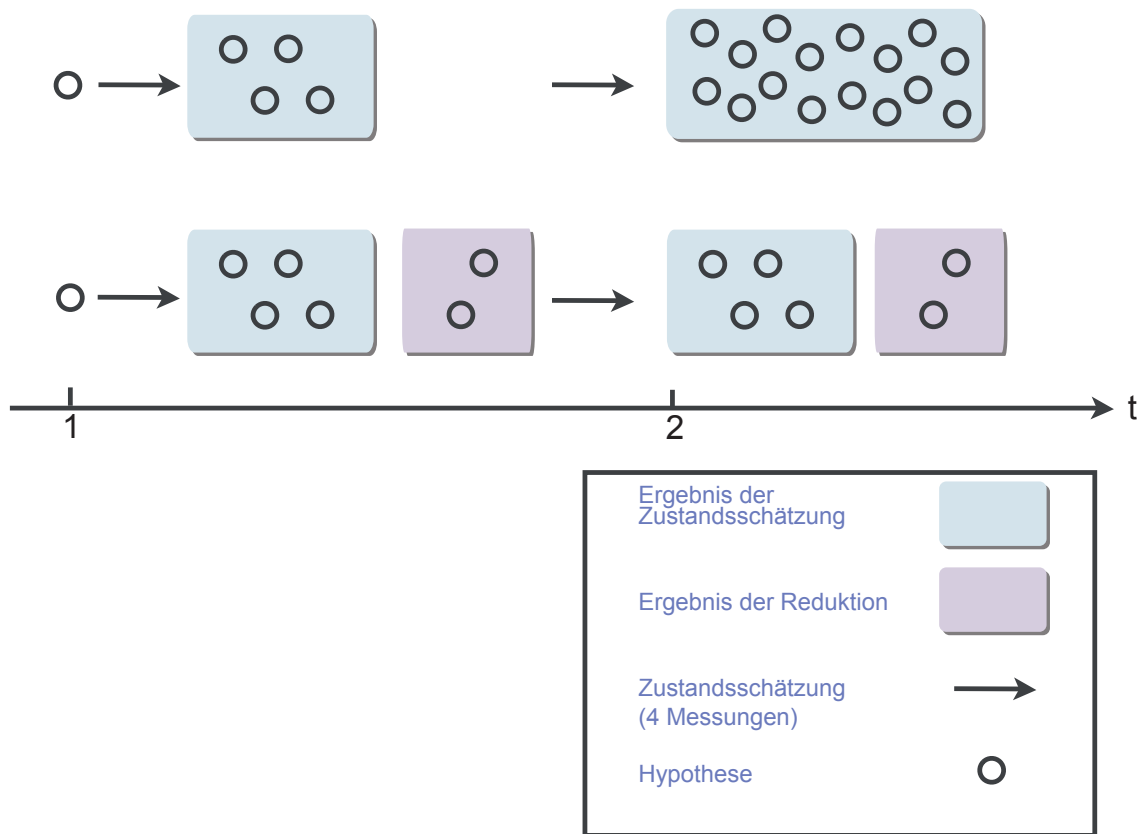


Abbildung 6.11: Ablaufdiagramm des Filters mit und ohne Komponentenreduktion. In jedem Zeitschritt fallen 4 Messungen an.

Das Lokalisierungsverfahren bricht ab, sobald die Trajektorie als verloren gilt. Eine Trajektorie gilt als verloren sobald 5 Zeitschritte hintereinander entweder die "wahre" Messung durch den Plausibilitätstests verworfen wurde, oder wenn die Abweichung zwischen der "wahren" Trajektorie und der aktuellen Hypothese zu groß ist.

Es wurde untersucht wie lange das Flugobjekt beim Einsatz der verschiedenen Reduktionsverfahren erfolgreich lokalisiert werden kann, dafür wurden 50 Iterationen der Monte-Carlo-Simulation verwendet. Die beim Einsatz der verschiedenen Reduktionsverfahren entstandenen Trajektorien sind in Abbildung 6.12 zu sehen. Im Zoom ist der Abbruch der Lokalisierung für das Verfahren von West gut zu erkennen.

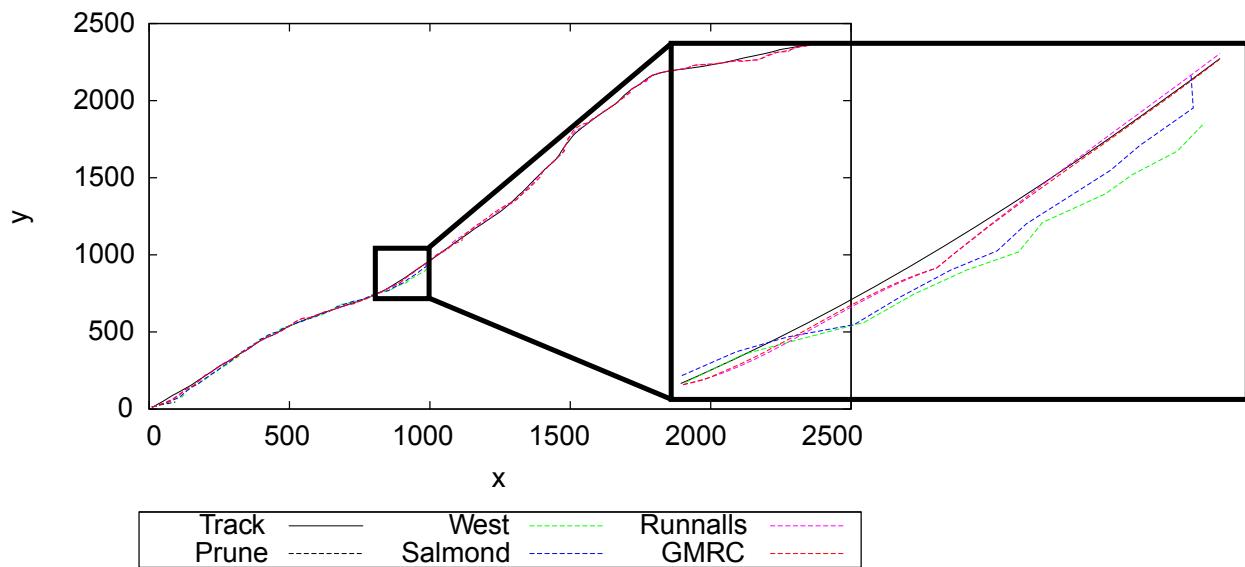


Abbildung 6.12: Wahre Trajektorien und Hypothese der Reduktionsverfahren im Vergleich.

Die Ergebnisse des Multihypothesen-Tracking-Szenarios für die Dauer bis zu einem Verlust der Trajektorie und die Anzahl der Verluste in Abhängigkeit zum verwendeten Reduktionsverfahren und der Anzahl der Komponenten der approximierten Dichte sind in den Abbildungen 6.13 und 6.14 dargestellt. Dabei liefert die Reduktion durch Pruning die mit Abstand schlechtesten Ergebnisse. Salmond und West liefern annähernd gleiche Ergebnisse. Die Reduktion mit Runnalls kann den Track schon entscheidend länger erfolgreich verfolgen. Unter Einsatz von GMRC mit Runnalls und Kmeans++ kann das Flugobjekt noch länger erfolgreich lokalisiert werden. In Abbildung 6.14 ist dargestellt bei wieviel der 50 verfolgten Tracks die Trajektorie verloren ging. Unter Einsatz von Pruning, West und Salmond gingen alle der 50 Tracks verloren. Wesentlich mehr Tracks konnten für Runnalls und das GMRC erfolgreich verfolgt werden, wobei GMRC bessere Ergebnisse liefert als Runnalls.

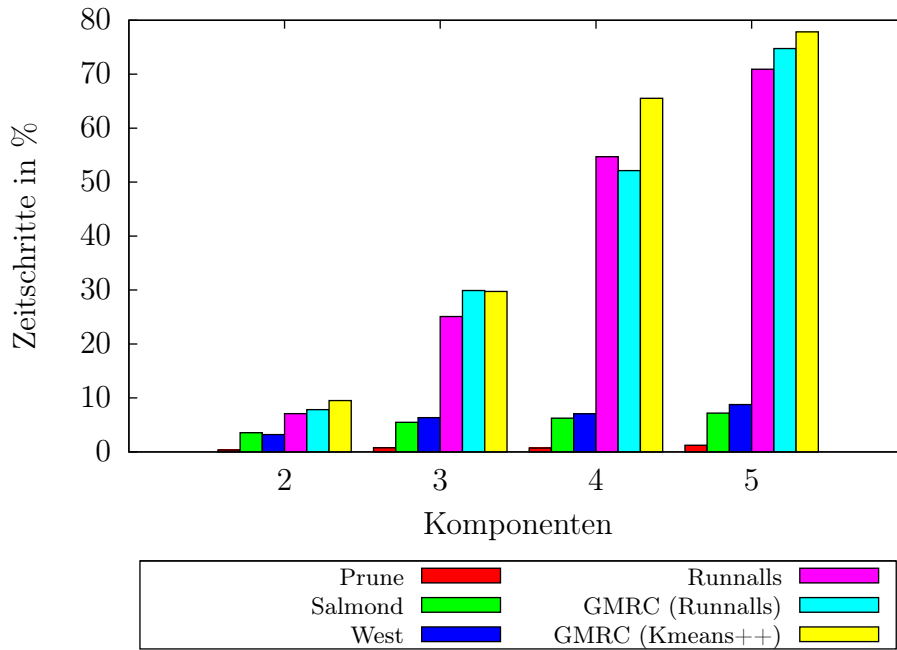


Abbildung 6.13: Erfolgreiche Trackingdauer der Reduktionsverfahren im Vergleich.

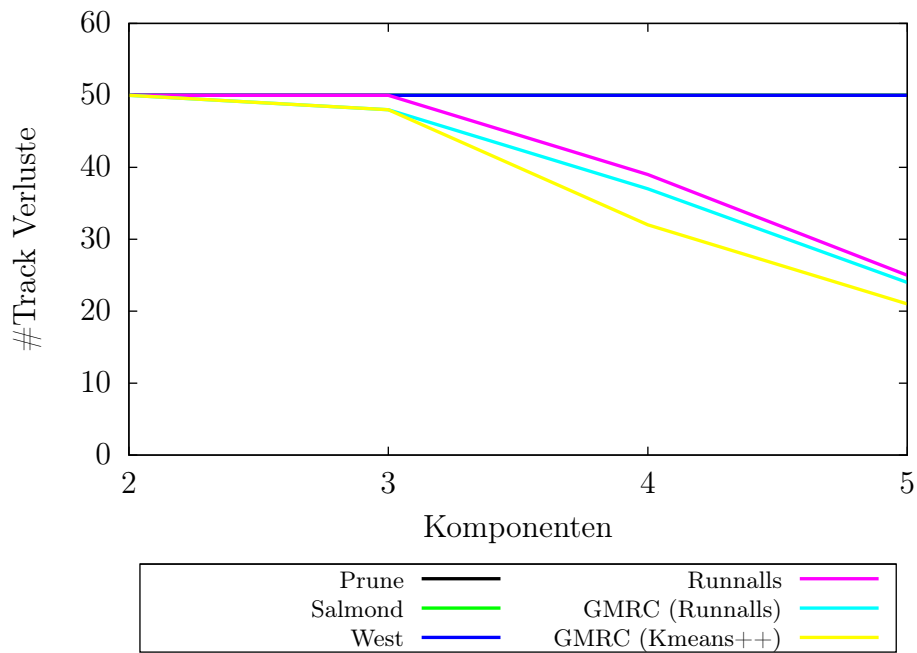


Abbildung 6.14: Anzahl der Trackverluste der Reduktionsverfahren im Vergleich.

Schlussfolgerung

Ziel dieser Studienarbeit war es, ein bestehendes Verfahren zur Reduktion von Gaußmischdichten zu erweitern. Das mehrstufige Clustering Verfahren zur Komponentenreduktion verfolgt einen Top-down Ansatz und besteht aus drei separaten Phasen: Initialisierung, Clustering und Optimierung.

In der Initialisierungsphase können verschiedene Verfahren verwendet werden, um eine erste Approximation der gegebenen Dichte mit der gewünschten Anzahl an Komponenten zu berechnen.

In der zweiten Phase wird diese Approximation verbessert. Dabei werden die Komponenten einzelnen Clustern zugeordnet. Nachdem die Zuordnung beendet ist, werden alle Komponenten aus demselben Cluster zusammengefasst. Schließlich wird der Grad der Abweichung zwischen der wahren Gaußmischdichte und der approximierten Dichte berechnet. Nach und nach können so verschiedene Zuordnungen geprüft werden, um damit eine gute Approximation zu finden.

Phase 3 dient der Verbesserung der berechneten Approximation. Die Lösung eines nichtlinearen Optimierungsproblems steht dabei im Vordergrund. Als Zielfunktion des Optimierungsproblem wird die NISD als Gütemaß verwendet. Zur Lösungsfindung werden numerische Lösungsverfahren verwendet. Dabei werden die Komponenten der approximierten Dichte schrittweise angepasst, bis letztendlich das verwendete Distanzmaß minimiert ist.

Die guten Ergebnisse des Verfahrens werden allerdings durch sehr hohe Laufzeiten erkauft. Vor allem im multivariaten Fall ergeben sich wesentlich höhere Laufzeiten. Phase 2 verfolgt eine Greedy Suche über die möglichen Reduktionen und benötigt insbesondere für die Auswertung des Gütemaßes einen hohen Rechenaufwand. Phase 3 verbessert die Ergebnisse der vorangegangenen Phasen besonders gut für Reduktionen auf sehr wenige Komponenten.

7.1 Ausblick

Erweiterungsspielraum bieten insbesondere die rechenintensiven Phasen 2 und 3. Denkbar ist beispielsweise diese Phasen in C++ zu implementieren. Zusätzlich könnte in Phase 3 anstatt des Gradientenabstiegs auch das Newtonverfahren verwendet werden. Dazu müsste aber zunächst die zweite Ableitung des Gütemaßes berechnet werden.

Unter Verwendung eines adaptiven Ansatzes, der schon für eindimensionale Dichten vielversprechende Ergebnisse geliefert hat, könnte ausgehend von einer festen Komponentenanzahl eine Reduktion mit unwesentlich mehr oder weniger Komponenten gebildet werden und jeweils die Reduktion mit geringster Abweichung gewählt werden. Dies ist vor allem von Vorteil, wenn aufgrund der festen Komponentenanzahl eine wichtige Komponente entfernt werden müsste und somit auch wesentliche Information verloren gehen würde.

Durch den Einsatz anderer Gütemaße könnten die Laufzeiteigenschaften verbessert werden. Zusätzlich könnte Phase 2 modifiziert werden um die Performance weiter zu verbessern, dazu könnte Neuronales Clustering verwendet werden.

Herleitung des Gradienten

Für die Herleitung werden folgende Ableitungsregeln aus [18],[19] verwendet. Der Gradient wird für ein bestimmtes $\underline{\eta}_k$ berechnet. Dieser besteht aus den partiellen Ableitungen nach Gewicht, Erwartungswert und Kovarianzmatrix. Außerdem werden folgende Funktionen und Konventionen benutzt:

- Eine Funktion die Elemente einer Matrix vektorisiert:

$$\text{Vec}(\mathbf{A}) = (a_{11}, \dots, a_{1n}, \dots, a_{1n}, \dots, a_{mn})^T$$

- Die wahre Gaußmischdichte ist durch $\tilde{f}(\underline{x}, \tilde{\underline{\eta}}) = \sum_{i=1}^n \tilde{w}_i \cdot \mathcal{N}(\underline{x}, \tilde{\underline{\mu}}_i, \tilde{\mathbf{C}}_i)$ gegeben mit

$$\mathcal{N}(\underline{x}, \tilde{\underline{\mu}}_i, \tilde{\mathbf{C}}_i) = \frac{1}{\sqrt{(2\pi)^N \cdot |\tilde{\mathbf{C}}_i|}} \cdot \exp\left\{-\frac{1}{2} \cdot (\underline{x} - \tilde{\underline{\mu}}_i)^T \cdot \tilde{\mathbf{C}}_i^{-1} \cdot (\underline{x} - \tilde{\underline{\mu}}_i)\right\} .$$

Ihre Komponenten Gewicht, Erwartungswert und Kovarianz entsprechen

$$\tilde{\underline{\eta}} = [\tilde{\underline{\eta}}_1, \dots, \tilde{\underline{\eta}}_n]^T, \quad \tilde{\underline{\eta}}_i = [\tilde{w}_i, \tilde{\underline{\mu}}_i, \tilde{\mathbf{C}}_i]^T .$$

- Die reduzierte Gaußmischdichte ist durch $f(\underline{x}, \underline{\eta}) = \sum_{j=1}^m w_j^2 \cdot \mathcal{N}(\underline{x}, \underline{\mu}_j, \mathbf{C}_j)$ gegeben, mit

$$\mathcal{N}(\underline{x}, \underline{\mu}_j, \mathbf{C}_j) = \frac{1}{\sqrt{(2\pi)^N \cdot |\mathbf{C}_j|}} \cdot \exp\left\{-\frac{1}{2} \cdot (\underline{x} - \underline{\mu}_j)^T \cdot \mathbf{C}_j^{-1} \cdot (\underline{x} - \underline{\mu}_j)\right\} .$$

Ihre Komponenten Gewicht, Erwartungswert und Kovarianz entsprechen

$$\underline{\eta} = [\underline{\eta}_1, \dots, \underline{\eta}_m]^T, \quad \underline{\eta}_j = [w_j, \underline{\mu}_j, \mathbf{C}_j]^T .$$

Für das Gewicht der reduzierten Dichte wird \underline{w}_j^2 verwendet um sicherzustellen, dass das Gewicht während des Gradientenabstiegs immer positiv bleibt. Für die Kovarianzmatrix

der reduzierten Dichte wird ihre Cholesky-Zerlegung verwendet $\mathbf{C}_x = (\mathbf{P}_x \cdot \mathbf{P}_x^T) = \mathbf{C}_x^T = (\mathbf{P}_x^T \cdot \mathbf{P}_x)$.

- Die Gütefunktion ist durch

$$G(\underline{\tilde{\eta}}, \underline{\eta}) = \frac{1}{2} \cdot \int_{\mathbb{R}} (\tilde{f}(\underline{x}, \underline{\tilde{\eta}}) - f(\underline{x}, \underline{\eta}))^2 d\underline{x}$$

definiert.

- Der Gradient besteht aus den partiellen Ableitungen nach den Komponenten der reduzierten Dichte, also Gewicht, Erwartungswert und Kovarianz:

$$\nabla G(\underline{\tilde{\eta}}, \underline{\eta}) = \left(\underbrace{\frac{\partial G(\underline{\tilde{\eta}}, \underline{\eta})}{\partial w_1}, \frac{\partial G(\underline{\tilde{\eta}}, \underline{\eta})}{\partial \underline{\mu}_1}, \frac{\partial G(\underline{\tilde{\eta}}, \underline{\eta})}{\partial \mathbf{P}_1}}_{\nabla G(\underline{\eta}_1)}, \dots, \frac{\partial G(\underline{\tilde{\eta}}, \underline{\eta})}{\partial w_m}, \frac{\partial G(\underline{\tilde{\eta}}, \underline{\eta})}{\partial \underline{\mu}_m}, \frac{\partial G(\underline{\tilde{\eta}}, \underline{\eta})}{\partial \mathbf{P}_m} \right)^T$$

Partielle Ableitung nach den k-ten Komponenten der reduzierten Dichte:

$$\frac{\partial f(\underline{x}, \underline{\eta})}{\partial \underline{\eta}_k} = \left(\frac{\partial f(\underline{x}, \underline{\eta})}{\partial w_k}, \frac{\partial f(\underline{x}, \underline{\eta})}{\partial \underline{\mu}_k}, \frac{\partial f(\underline{x}, \underline{\eta})}{\partial \mathbf{P}_k} \right)^T$$

$$\begin{aligned} \nabla G(\underline{\eta}_k) &= \frac{\partial}{\partial \underline{\eta}_k} \frac{1}{2} \cdot \int_{\mathbb{R}^N} (\tilde{f}(\underline{x}, \underline{\tilde{\eta}}) - f(\underline{x}, \underline{\eta}))^2 d\underline{x} \\ &= \frac{1}{2} \cdot \int_{\mathbb{R}^N} \left(\frac{\partial}{\partial \underline{\eta}_k} \tilde{f}(\underline{x}, \underline{\tilde{\eta}}) - f(\underline{x}, \underline{\eta}) \right)^2 d\underline{x} \\ &= - \int_{\mathbb{R}^N} \tilde{f}(\underline{x}, \underline{\tilde{\eta}}) - f(\underline{x}, \underline{\eta}) \cdot \frac{\partial f(\underline{x}, \underline{\eta})}{\partial \underline{\eta}_k} d\underline{x} \\ &= - \int_{\mathbb{R}^N} \tilde{f}(\underline{x}, \underline{\tilde{\eta}}) \cdot \frac{\partial f(\underline{x}, \underline{\eta})}{\partial \underline{\eta}_k} d\underline{x} + \int_{\mathbb{R}^N} f(\underline{x}, \underline{\eta}) \cdot \frac{\partial f(\underline{x}, \underline{\eta})}{\partial \underline{\eta}_k} d\underline{x} \end{aligned}$$

Ableitung nach dem Gewicht:

$$\frac{\partial f(\underline{x}, \eta)}{\partial w_k} = 2w_k \cdot \mathcal{N}(\underline{x}, \underline{\mu}_k, \mathbf{C}_k) \quad (\text{A.1})$$

Ableitung nach dem Erwartungswert:

$$\begin{aligned} \frac{\partial f(\underline{x}, \eta)}{\partial \underline{\mu}_k} &= -\frac{1}{2} \cdot w_k^2 \cdot \left[2 \cdot \mathbf{C}_k^{-1} \cdot \underline{\mu}_k - 2 \cdot \mathbf{C}_k^{-1} \cdot \underline{x} \right] \cdot \mathcal{N}(\underline{x}, \underline{\mu}_k, \mathbf{C}_k) \\ &= w_k^2 \cdot \mathbf{C}_k^{-1} \cdot (\underline{x} - \underline{\mu}_k) \cdot \mathcal{N}(\underline{x}, \underline{\mu}_k, \mathbf{C}_k) \end{aligned} \quad (\text{A.2})$$

Ableitung nach der Kovarianz:

$$\begin{aligned} \frac{\partial f(\underline{x}, \eta)}{\partial \mathbf{P}_k} &= -\frac{1}{2} \cdot [(2\pi)^N \cdot \mathbf{C}_k]^{-\frac{3}{2}} \cdot (2\pi)^N \cdot \det(\mathbf{C}_k) \cdot \mathbf{C}_k^{-\text{T}} \cdot 2 \cdot \mathbf{P}_k \\ &\quad \cdot w_k^2 \cdot \exp\left\{-\frac{1}{2} \cdot (\underline{x} - \underline{\mu}_k)^{\text{T}} \cdot \mathbf{C}_k^{-1} \cdot (\underline{x} - \underline{\mu}_k)\right\} \\ &\quad + [(2\pi)^N \cdot \det(\mathbf{C}_k)]^{-\frac{1}{2}} \cdot \frac{w_k^2}{2} \cdot \mathbf{C}_k^{-\text{T}} \cdot (\underline{x} - \underline{\mu}_k) \cdot (\underline{x} - \underline{\mu}_k)^{\text{T}} \cdot \mathbf{C}_k^{-\text{T}} \cdot 2 \cdot \mathbf{P}_k \\ &\quad \cdot \exp\left\{-\frac{1}{2} \cdot (\underline{x} - \underline{\mu}_k)^{\text{T}} \cdot \mathbf{C}_k^{-1} \cdot (\underline{x} - \underline{\mu}_k)\right\} \\ &= -w_k^2 \cdot [(2\pi)^N \cdot \mathbf{C}_k]^{-1} \cdot (2\pi)^N \cdot \det(\mathbf{C}_k) \cdot \mathbf{C}_k^{-1} \cdot 2 \cdot \mathbf{P}_k \cdot \mathcal{N}(\underline{x}, \underline{\mu}_k, \mathbf{C}_k) \\ &\quad + w_k^2 \cdot \mathbf{C}_k^{-1} \cdot (\underline{x} - \underline{\mu}_k) \cdot (\underline{x} - \underline{\mu}_k)^{\text{T}} \cdot \mathbf{C}_k^{-1} \cdot \mathbf{P}_k \cdot \mathcal{N}(\underline{x}, \underline{\mu}_k, \mathbf{C}_k) \\ &= w_k^2 \cdot [\mathbf{C}_k^{-1} \cdot (\underline{x} - \underline{\mu}_k) \cdot (\underline{x} - \underline{\mu}_k)^{\text{T}} - \mathbf{I}] \cdot \mathbf{C}_k^{-1} \cdot \mathbf{P}_k \cdot \mathcal{N}(\underline{x}, \underline{\mu}_k, \mathbf{C}_k) \end{aligned} \quad (\text{A.3})$$

Ableitungsregeln:

$$\frac{\partial \det(\mathbf{C})}{\partial \mathbf{C}} = \det(\mathbf{C}) \cdot (\mathbf{C}^{-1})^{\text{T}}, \quad \frac{\partial \underline{a}^{\text{T}} \cdot \mathbf{C}^{-1} \cdot \underline{b}}{\partial \mathbf{C}} = -\mathbf{C}^{-\text{T}} \cdot \underline{a} \underline{b}^{\text{T}} \cdot \mathbf{C}^{-\text{T}}, \quad \frac{\partial \underline{x}^{\text{T}} \cdot \underline{a}}{\partial \underline{x}} = \frac{\partial \underline{a} \cdot \underline{x}}{\partial \underline{x}} = \underline{a},$$

$$\frac{\partial \underline{x}^{\text{T}} \cdot \mathbf{C}}{\partial \underline{x}} = \frac{\partial \mathbf{C} \cdot \underline{x}}{\partial \underline{x}} = \mathbf{C}, \quad \frac{\partial \underline{x}^{\text{T}} \cdot \mathbf{C} \cdot \underline{x}}{\partial \underline{x}} = \underline{x}^{\text{T}} \cdot (\mathbf{C} + \mathbf{C}^{\text{T}}).$$

Partiellen Ableitungen:

$$\frac{\partial f(\underline{x}, \underline{\eta})}{\partial w_k} = \underbrace{2w_k}_{T_1} \cdot \mathcal{N}(\underline{x}, \underline{\mu}_k, \mathbf{C}_k) \quad (\text{A.4})$$

$$\frac{\partial f(\underline{x}, \underline{\eta})}{\partial \underline{\mu}_k} = \underbrace{w_k^2 \cdot \mathbf{C}_k^{-1} \cdot (\underline{x} - \underline{\mu}_k)}_{T_2} \cdot \mathcal{N}(\underline{x}, \underline{\mu}_k, \mathbf{C}_k) \quad (\text{A.5})$$

$$\frac{\partial f(\underline{x}, \underline{\eta})}{\partial \mathbf{P}_k} = \underbrace{w_k^2 \cdot [\mathbf{C}_k^{-1} \cdot (\underline{x} - \underline{\mu}_k) \cdot (\underline{x} - \underline{\mu}_k)^T - \mathbf{I}] \cdot \mathbf{C}_k^{-1} \cdot \mathbf{P}_k}_{T_3} \cdot \mathcal{N}(\underline{x}, \underline{\mu}_k, \mathbf{C}_k) \quad (\text{A.6})$$

$$\frac{\partial f(\underline{x}, \underline{\eta})}{\partial \underline{\eta}_k} = \left(\frac{\partial f(\underline{x}, \underline{\eta})}{\partial w_k}, \frac{\partial f(\underline{x}, \underline{\eta})}{\partial \underline{\mu}_k}, \frac{\partial f(\underline{x}, \underline{\eta})}{\partial \mathbf{P}_k} \right)^T = (T_1, T_2, T_3)^T \cdot \mathcal{N}(\underline{x}, \underline{\mu}_k, \mathbf{C}_k) \quad (\text{A.7})$$

Für das Produkt von 2 Gaußdichten gilt:

$$\mathcal{N}(\underline{x}, \underline{\mu}_i, \tilde{\mathbf{C}}_i) \cdot \mathcal{N}(\underline{x}, \underline{\mu}_k, \mathbf{C}_k) = \mathcal{N}(\underline{\mu}_i, \underline{\mu}_k, \tilde{\mathbf{C}}_i + \mathbf{C}_k) \cdot \mathcal{N}(\underline{x}, \underline{\mu}_s, \mathbf{C}_s) \quad (\text{A.8})$$

$$\mathcal{N}(\underline{x}, \underline{\mu}_j, \mathbf{C}_j) \cdot \mathcal{N}(\underline{x}, \underline{\mu}_k, \mathbf{C}_k) = \mathcal{N}(\underline{\mu}_j, \underline{\mu}_k, \mathbf{C}_j + \mathbf{C}_k) \cdot \mathcal{N}(\underline{x}, \underline{\mu}_t, \mathbf{C}_t) \quad (\text{A.9})$$

mit $\underline{\mu}_s, \underline{\mu}_t$ und $\mathbf{C}_s, \mathbf{C}_t$ ($s \hat{=} k, i$ und $t \hat{=} k, j$):

$$\underline{\mu}_s = \mathbf{C}_s \cdot (\tilde{\mathbf{C}}_i^{-1} \cdot \underline{\mu}_i + \mathbf{C}_k^{-1} \cdot \underline{\mu}_k), \quad (\text{A.10}) \quad \mathbf{C}_s = (\tilde{\mathbf{C}}_i^{-1} + \mathbf{C}_k^{-1})^{-1}, \quad (\text{A.11})$$

$$\underline{\mu}_t = \mathbf{C}_t \cdot (\mathbf{C}_j^{-1} \cdot \underline{\mu}_j + \mathbf{C}_k^{-1} \cdot \underline{\mu}_k), \quad (\text{A.12}) \quad \mathbf{C}_t = (\mathbf{C}_j^{-1} + \mathbf{C}_k^{-1})^{-1}. \quad (\text{A.13})$$

Erster Summand des Gradienten:

$$\begin{aligned} \int_{\mathbf{R}^N} \tilde{f}(\underline{x}, \underline{\eta}) \cdot \frac{\partial f(\underline{x}, \underline{\eta})}{\partial \underline{\eta}_k} \, d\underline{x} &\stackrel{(\text{A.7})}{=} \int_{\mathbf{R}^N} \sum_{i=1}^n \tilde{w}_i \cdot \mathcal{N}(\underline{x}, \underline{\mu}_i, \tilde{\mathbf{C}}_i) \cdot (T_1, T_2, T_3)^T \cdot \mathcal{N}(\underline{x}, \underline{\mu}_k, \mathbf{C}_k) \, d\underline{x} \\ &= \int_{\mathbf{R}^N} \sum_{i=1}^n \tilde{w}_i \cdot \mathcal{N}(\underline{x}, \underline{\mu}_i, \tilde{\mathbf{C}}_i) \cdot \mathcal{N}(\underline{x}, \underline{\mu}_k, \mathbf{C}_k) \cdot (T_1, T_2, T_3)^T \, d\underline{x} \\ &\stackrel{(\text{A.8})}{=} \int_{\mathbf{R}^N} \sum_{i=1}^n \tilde{w}_i \cdot \mathcal{N}(\underline{\mu}_i, \underline{\mu}_k, \tilde{\mathbf{C}}_i + \mathbf{C}_k) \cdot \mathcal{N}(\underline{x}, \underline{\mu}_s, \mathbf{C}_s) \cdot (T_1, T_2, T_3)^T \, d\underline{x} \\ &= \sum_{i=1}^n \tilde{w}_i \cdot \mathcal{N}(\underline{\mu}_i, \underline{\mu}_k, \tilde{\mathbf{C}}_i + \mathbf{C}_k) \cdot \int_{\mathbf{R}^N} (T_1, T_2, T_3)^T \cdot \mathcal{N}(\underline{x}, \underline{\mu}_s, \mathbf{C}_s) \, d\underline{x} \end{aligned} \quad (\text{A.14})$$

Zweiter Summand des Gradienten:

$$\begin{aligned}
\int_{\mathbf{R}^N} f(\underline{x}, \underline{\eta}) \frac{\partial f(\underline{x}, \underline{\eta})}{\partial \eta_k} d\underline{x} &\stackrel{(A.7)}{=} \sum_{j=1}^m w_j^2 \cdot \mathcal{N}(\underline{x}, \underline{\mu}_j, \mathbf{C}_j) \cdot \left(T_1, T_2, T_3\right)^T \mathcal{N}(\underline{x}, \underline{\mu}_k, \mathbf{C}_k) d\underline{x} \\
&= \int_{\mathbf{R}^N} \sum_{j=1}^m w_j^2 \cdot \mathcal{N}(\underline{x}, \underline{\mu}_j, \mathbf{C}_j) \cdot \mathcal{N}(\underline{x}, \underline{\mu}_k, \mathbf{C}_k) \cdot \left(T_1, T_2, T_3\right)^T d\underline{x} \\
&\stackrel{(A.9)}{=} \int_{\mathbf{R}^N} \sum_{j=1}^m w_j^2 \cdot \mathcal{N}(\underline{\mu}_j, \underline{\mu}_k, \mathbf{C}_j + \mathbf{C}_k) \cdot \mathcal{N}(\underline{x}, \underline{\mu}_k, \mathbf{C}_k) \cdot \left(T_1, T_2, T_3\right)^T d\underline{x} \\
&= \sum_{j=1}^m w_j^2 \cdot \mathcal{N}(\underline{\mu}_j, \underline{\mu}_k, \mathbf{C}_j + \mathbf{C}_k) \cdot \int_{\mathbf{R}^N} \left(T_1, T_2, T_3\right)^T \cdot \mathcal{N}(\underline{x}, \underline{\mu}_k, \mathbf{C}_k) d\underline{x}
\end{aligned} \tag{A.15}$$

Für das Integral über eine Verteilungsfunktion gilt:

$$\begin{aligned}
1 &= \int_{\mathbf{R}^N} \mathcal{N}(\underline{x}, \underline{\mu}, \mathbf{C}) d\underline{x}, & E(x^k) &= \int_{\mathbf{R}^N} x^k \cdot \mathcal{N}(\underline{x}, \underline{\mu}, \mathbf{C}) d\underline{x} = M_k \text{ (k-tes Moment)}, \\
\mathbf{C} &= \int_{\mathbf{R}^N} (\underline{x} - \underline{\mu}) \cdot (\underline{x} - \underline{\mu})^T \cdot \mathcal{N}(\underline{x}, \underline{\mu}, \mathbf{C}) d\underline{x} = \int_{\mathbf{R}^N} \underline{x} \underline{x}^T \cdot \mathcal{N}(\underline{x}, \underline{\mu}, \mathbf{C}) d\underline{x} - \underline{\mu}.
\end{aligned}$$

Integral über partielle Ableitung nach dem Gewicht:

$$\int_{\mathbf{R}^N} T_1 \cdot \mathcal{N}(\underline{x}, \underline{\mu}_s, \mathbf{C}_s) \stackrel{(A.4)}{=} \int_{\mathbf{R}^N} 2 \cdot w_k \cdot \mathcal{N}(\underline{x}, \underline{\mu}_s, \mathbf{C}_s) d\underline{x} = 2 \cdot w_k \cdot \int_{\mathbf{R}^N} \mathcal{N}(\underline{x}, \underline{\mu}_s, \mathbf{C}_s) d\underline{x} = 2 \cdot w_k \tag{A.16}$$

Integral über partielle Ableitung nach dem Erwartungswert:

$$\begin{aligned}
\int_{\mathbf{R}^N} T_2 \cdot \mathcal{N}(\underline{x}, \underline{\mu}_s, \mathbf{C}_s) &\stackrel{(A.5)}{=} \int_{\mathbf{R}^N} w_k^2 \cdot \mathbf{C}_k^{-1} \cdot (\underline{x} - \underline{\mu}_k) \cdot \mathcal{N}(\underline{x}, \underline{\mu}_s, \mathbf{C}_s) d\underline{x} \\
&= w_k^2 \cdot \mathbf{C}_k^{-1} \cdot \int_{\mathbf{R}^N} (\underline{x} - \underline{\mu}_k) \cdot \mathcal{N}(\underline{x}, \underline{\mu}_s, \mathbf{C}_s) d\underline{x} \\
&= w_k^2 \cdot \mathbf{C}_k^{-1} \cdot \left[\int_{\mathbf{R}^N} \underline{x} \cdot \mathcal{N}(\underline{x}, \underline{\mu}_s, \mathbf{C}_s) d\underline{x} - \underline{\mu}_k \cdot \int_{\mathbf{R}^N} \mathcal{N}(\underline{x}, \underline{\mu}_s, \mathbf{C}_s) d\underline{x} \right] \\
&= w_k^2 \cdot \mathbf{C}_k^{-1} \cdot (\underline{M}_1^s - \underline{\mu}_k)
\end{aligned} \tag{A.17}$$

Integral über partielle Ableitung nach der Kovarianz:

$$\begin{aligned}
 \int_{\mathbf{R}^N} T_3 \cdot \mathcal{N}(\underline{x}, \underline{\mu}_s, \mathbf{C}_s) &\stackrel{(A.6)}{=} \int_{\mathbf{R}^N} w_k^2 \cdot \left[\mathbf{C}_k^{-1} \cdot (\underline{x} - \underline{\mu}_k) \cdot (\underline{x} - \underline{\mu}_k)^T - \mathbf{I} \right] \cdot \mathbf{C}_k^{-1} \cdot \mathbf{P}_k \cdot \mathcal{N}(\underline{x}, \underline{\mu}_s, \mathbf{C}_s) \, d\underline{x} \\
 &= w_k^2 \cdot \int_{\mathbf{R}^N} (\mathbf{C}_k^{-1} \cdot (\underline{x} - \underline{\mu}_k) \cdot (\underline{x} - \underline{\mu}_k)^T - \mathbf{I}) \cdot \mathcal{N}(\underline{x}, \underline{\mu}_s, \mathbf{C}_s) \, d\underline{x} \cdot \mathbf{C}_k^{-1} \cdot \mathbf{P}_k \\
 &= w_k^2 \cdot \left[\mathbf{C}_k^{-1} \cdot \int_{\mathbf{R}^N} (\underline{x}\underline{x}^T - \underline{x}\underline{\mu}_k^T - \underline{\mu}_k\underline{x}^T + \underline{\mu}_k\underline{\mu}_k^T) \cdot \mathcal{N}(\underline{x}, \underline{\mu}_s, \mathbf{C}_s) \, d\underline{x} - \mathbf{I} \right] \\
 &\quad \cdot \mathbf{C}_k^{-1} \cdot \mathbf{P}_k \\
 &= w_k^2 \cdot \left[\mathbf{C}_k^{-1} \cdot (\mathbf{M}_2^s + \underline{M}_1^s(\underline{M}_1^s)^T - \underline{M}_1^s\underline{\mu}_k^T - \underline{\mu}_k(\underline{M}_1^s)^T + \underline{\mu}_k\underline{\mu}_k^T) - \mathbf{I} \right] \cdot \mathbf{C}_k^{-1} \cdot \mathbf{P}_k
 \end{aligned} \tag{A.18}$$

Erster Summand des Gradienten für das Gewicht:

$$\begin{aligned}
 \int_{\mathbf{R}^N} \tilde{f}(\underline{x}, \underline{\eta}) \cdot \frac{\partial f(\underline{x}, \underline{\eta})}{\partial w_k} \, d\underline{x} &\stackrel{(A.14)(A.16)}{=} \sum_{i=1}^n \tilde{w}_i \cdot \mathcal{N}(\underline{\mu}_i, \underline{\mu}_k, \tilde{\mathbf{C}}_i + \mathbf{C}_k) \cdot 2 \cdot w_k \\
 &= 2 \cdot w_k \cdot \sum_{i=1}^n \tilde{w}_i \cdot \mathcal{N}(\underline{\mu}_i, \underline{\mu}_k, \tilde{\mathbf{C}}_i + \mathbf{C}_k)
 \end{aligned} \tag{A.19}$$

Zweiter Summand des Gradienten für das Gewicht:

$$\begin{aligned}
 \int_{\mathbf{R}^N} f(\underline{x}, \underline{\eta}) \cdot \frac{\partial f(\underline{x}, \underline{\eta})}{\partial w_k} \, d\underline{x} &\stackrel{(A.15)(A.16)}{=} \sum_{j=1}^m w_j^2 \cdot \mathcal{N}(\underline{\mu}_j, \underline{\mu}_k, \mathbf{C}_j + \mathbf{C}_k) \cdot 2 \cdot w_k \\
 &= 2 \cdot w_k \cdot \sum_{j=1}^m w_j^2 \cdot \mathcal{N}(\underline{\mu}_j, \underline{\mu}_k, \mathbf{C}_j + \mathbf{C}_k)
 \end{aligned} \tag{A.20}$$

Damit ergibt sich der Gradient über das Gewicht:

$$\frac{\partial G(\underline{\eta})}{\partial w_k} \stackrel{(A.19)(A.20)}{=} 2 \cdot w_k \cdot \left[\sum_{j=1}^m w_j^2 \cdot \mathcal{N}(\underline{\mu}_j, \underline{\mu}_k, \mathbf{C}_j + \mathbf{C}_k) - \sum_{i=1}^n \tilde{w}_i \cdot \mathcal{N}(\underline{\mu}_i, \underline{\mu}_k, \tilde{\mathbf{C}}_i + \mathbf{C}_k) \right] \tag{A.21}$$

Erster Summand des Gradienten für den Erwartungswert:

$$\begin{aligned}
 \int_{\mathbf{R}^N} \tilde{f}(\underline{x}, \underline{\eta}) \cdot \frac{\partial f(\underline{x}, \underline{\eta})}{\partial \underline{\mu}_k} \, d\underline{x} &\stackrel{(A.14)(A.17)}{=} \sum_{i=1}^n \tilde{w}_i \cdot \mathcal{N}(\underline{\mu}_i, \underline{\mu}_k, \tilde{\mathbf{C}}_i + \mathbf{C}_k) \cdot w_k^2 \cdot \mathbf{C}_k^{-1} \cdot (\underline{M}_1^s - \underline{\mu}_k) \\
 &= w_k^2 \cdot \sum_{i=1}^n \tilde{w}_i \cdot \mathcal{N}(\underline{\mu}_i, \underline{\mu}_k, \tilde{\mathbf{C}}_i + \mathbf{C}_k) \cdot \mathbf{C}_k^{-1} \cdot (\underline{M}_1^s - \underline{\mu}_k) \\
 &\stackrel{(A.10)}{=} w_k^2 \cdot \sum_{i=1}^n \tilde{w}_i \cdot \mathcal{N}(\underline{\mu}_i, \underline{\mu}_k, \tilde{\mathbf{C}}_i + \mathbf{C}_k) \\
 &\quad \cdot \underbrace{\mathbf{C}_k^{-1} \cdot \left[(\tilde{\mathbf{C}}_i^{-1} + \mathbf{C}_k^{-1})^{-1} \cdot (\tilde{\mathbf{C}}_i^{-1} \cdot \underline{\mu}_i + \mathbf{C}_k^{-1} \cdot \underline{\mu}_k) - \underline{\mu}_k \right]}_S
 \end{aligned} \tag{A.22}$$

Dies lässt sich umformen auf die Herleitung von Williams [3]:

$$\begin{aligned}
S &= \mathbf{C}_k^{-1} \cdot \left[(\tilde{\mathbf{C}}_i^{-1} + \mathbf{C}_k^{-1})^{-1} \cdot (\tilde{\mathbf{C}}_i^{-1} \cdot \tilde{\underline{\mu}}_i + \mathbf{C}_k^{-1} \cdot \underline{\mu}_k) - (\tilde{\mathbf{C}}_i^{-1} + \mathbf{C}_k^{-1})^{-1} \cdot (\tilde{\mathbf{C}}_i^{-1} + \mathbf{C}_k^{-1}) \cdot \underline{\mu}_k \right] \\
&= \mathbf{C}_k^{-1} \cdot (\tilde{\mathbf{C}}_i^{-1} + \mathbf{C}_k^{-1})^{-1} \cdot \left[\tilde{\mathbf{C}}_i^{-1} \cdot \tilde{\underline{\mu}}_i + \mathbf{C}_k^{-1} \cdot \underline{\mu}_k - (\tilde{\mathbf{C}}_i^{-1} + \mathbf{C}_k^{-1}) \cdot \underline{\mu}_k \right] \\
&= \mathbf{C}_k^{-1} \cdot (\tilde{\mathbf{C}}_i^{-1} + \mathbf{C}_k^{-1})^{-1} \cdot \left[\tilde{\mathbf{C}}_i^{-1} \cdot \tilde{\underline{\mu}}_i + \mathbf{C}_k^{-1} \cdot \underline{\mu}_k - \tilde{\mathbf{C}}_i^{-1} \cdot \underline{\mu}_k - \mathbf{C}_k^{-1} \cdot \underline{\mu}_k \right] \\
&= \mathbf{C}_k^{-1} \cdot (\tilde{\mathbf{C}}_i^{-1} + \mathbf{C}_k^{-1})^{-1} \cdot \left[\tilde{\mathbf{C}}_i^{-1} \cdot \tilde{\underline{\mu}}_i - \tilde{\mathbf{C}}_i^{-1} \cdot \underline{\mu}_k \right] \\
&= \mathbf{C}_k^{-1} \cdot (\tilde{\mathbf{C}}_i^{-1} + \mathbf{C}_k^{-1})^{-1} \cdot \tilde{\mathbf{C}}_i^{-1} \cdot (\tilde{\underline{\mu}}_i - \underline{\mu}_k) \\
&= (\mathbf{C}_k \cdot \tilde{\mathbf{C}}_i^{-1} \cdot \tilde{\mathbf{C}}_i + \mathbf{C}_k \cdot \mathbf{C}_k^{-1} \cdot \tilde{\mathbf{C}}_i)^{-1} \cdot (\tilde{\underline{\mu}}_i - \underline{\mu}_k) \\
&= (\mathbf{C}_k + \tilde{\mathbf{C}}_i)^{-1} \cdot (\tilde{\underline{\mu}}_i - \underline{\mu}_k)
\end{aligned} \tag{A.23}$$

Damit ergibt sich die Herleitung nach Williams:

$$\int_{\mathbf{R}^N} \tilde{f}(\underline{x}, \tilde{\underline{\eta}}) \cdot \frac{\partial f(\underline{x}, \underline{\eta})}{\partial \underline{\mu}_k} \, d\underline{x} \stackrel{(A.22)(A.23)}{=} w_k^2 \cdot \sum_{i=1}^n \tilde{w}_i \cdot \mathcal{N}(\tilde{\underline{\mu}}_i, \underline{\mu}_k, \tilde{\mathbf{C}}_i + (\mathbf{P}_k \cdot \mathbf{P}_k^T)) \cdot (\tilde{\mathbf{C}}_i + (\mathbf{P}_k \cdot \mathbf{P}_k^T))^{-1} \cdot (\tilde{\underline{\mu}}_i - \underline{\mu}_k) \tag{A.24}$$

Zweiter Summand des Gradienten für den Erwartungswert:

$$\begin{aligned}
&\int_{\mathbf{R}^N} f(\underline{x}, \underline{\eta}) \cdot \frac{\partial f(\underline{x}, \underline{\eta})}{\partial \underline{\mu}_k} \, d\underline{x} \stackrel{(A.15)}{=} \sum_{j=1}^m w_j^2 \cdot \mathcal{N}(\underline{\mu}_j, \underline{\mu}_k, \mathbf{C}_j + \mathbf{C}_k) \cdot w_k^2 \cdot \mathbf{C}_k^{-1} \cdot (\underline{M}_1^t - \underline{\mu}_k) \\
&= w_k^2 \cdot \sum_{j=1}^m w_j^2 \cdot \mathcal{N}(\underline{\mu}_j, \underline{\mu}_k, \mathbf{C}_j + \mathbf{C}_k) \cdot \mathbf{C}_k^{-1} \cdot (\underline{M}_1^t - \underline{\mu}_k) \\
&\stackrel{(A.12)}{=} w_k^2 \cdot \sum_{j=1}^m w_j^2 \cdot \mathcal{N}(\underline{\mu}_j, \underline{\mu}_k, \mathbf{C}_j + \mathbf{C}_k) \\
&\quad \cdot \underbrace{\mathbf{C}_k^{-1} \cdot \left[(\mathbf{C}_j + \mathbf{C}_k^{-1})^{-1} \cdot (\mathbf{C}_j \cdot \tilde{\underline{\mu}}_j + \mathbf{C}_k^{-1} \cdot \underline{\mu}_k) - \underline{\mu}_k \right]}_T
\end{aligned} \tag{A.25}$$

Dieser lässt sich umformen auf die Herleitung von Williams:

$$\begin{aligned}
T &= \mathbf{C}_k^{-1} \cdot \left[(\mathbf{C}_j^{-1} + \mathbf{C}_k^{-1})^{-1} \cdot (\mathbf{C}_j^{-1} \cdot \tilde{\underline{\mu}}_j + \mathbf{C}_k^{-1} \cdot \underline{\mu}_k) - (\mathbf{C}_j^{-1} + \mathbf{C}_k^{-1})^{-1} \cdot (\mathbf{C}_j^{-1} + \mathbf{C}_k^{-1}) \cdot \underline{\mu}_k \right] \\
&= \mathbf{C}_k^{-1} \cdot (\mathbf{C}_j^{-1} + \mathbf{C}_k^{-1})^{-1} \cdot \left[\mathbf{C}_j^{-1} \cdot \tilde{\underline{\mu}}_j + \mathbf{C}_k^{-1} \cdot \underline{\mu}_k - (\mathbf{C}_j^{-1} + \mathbf{C}_k^{-1}) \cdot \underline{\mu}_k \right] \\
&= \mathbf{C}_k^{-1} \cdot (\mathbf{C}_j^{-1} + \mathbf{C}_k^{-1})^{-1} \cdot \left[\mathbf{C}_j^{-1} \cdot \tilde{\underline{\mu}}_j + \mathbf{C}_k^{-1} \cdot \underline{\mu}_k - \mathbf{C}_j^{-1} \cdot \underline{\mu}_k - \mathbf{C}_k^{-1} \cdot \underline{\mu}_k \right] \\
&= \mathbf{C}_k^{-1} \cdot (\mathbf{C}_j^{-1} + \mathbf{C}_k^{-1})^{-1} \cdot \left[\mathbf{C}_j^{-1} \cdot \tilde{\underline{\mu}}_j - \mathbf{C}_j^{-1} \cdot \underline{\mu}_k \right] \\
&= \mathbf{C}_k^{-1} \cdot (\mathbf{C}_j^{-1} + \mathbf{C}_k^{-1})^{-1} \cdot \mathbf{C}_j^{-1} \cdot (\tilde{\underline{\mu}}_j - \underline{\mu}_k) \\
&= (\mathbf{C}_k \cdot \mathbf{C}_j^{-1} \cdot \mathbf{C}_j + \mathbf{C}_k \cdot \mathbf{C}_k^{-1} \cdot \mathbf{C}_j)^{-1} \cdot (\tilde{\underline{\mu}}_j - \underline{\mu}_k) \\
&= (\mathbf{C}_k + \mathbf{C}_j)^{-1} \cdot (\tilde{\underline{\mu}}_j - \underline{\mu}_k)
\end{aligned} \tag{A.26}$$

Damit ergibt sich die Herleitung nach Williams:

$$\int_{\mathbb{R}^N} f(\underline{x}, \underline{\eta}) \cdot \frac{\partial f(\underline{x}, \underline{\eta})}{\partial \underline{\mu}_k} d\underline{x} \stackrel{(A.22)(A.26)}{=} w_k^2 \cdot \sum_{j=1}^m w_j \cdot \mathcal{N}(\underline{\mu}_j, \underline{\mu}_k, (\mathbf{P}_j \cdot \mathbf{P}_j^T) + (\mathbf{P}_k \cdot \mathbf{P}_k^T)) \cdot ((\mathbf{P}_j \cdot \mathbf{P}_j^T) + (\mathbf{P}_k \cdot \mathbf{P}_k^T)) \cdot (\underline{\mu}_j - \underline{\mu}_k) \quad (A.27)$$

Partielle Ableitung nach dem Erwartungswert:

$$\nabla G(\underline{\mu}_k) \stackrel{(A.22)(A.25)}{=} \text{Vec} \left(w_k^2 \cdot \left[\sum_{i=1}^n \tilde{w}_i \cdot \mathcal{N}(\tilde{\underline{\mu}}_i, \underline{\mu}_k, \tilde{\mathbf{C}}_i + \mathbf{C}_k) \cdot \mathbf{C}_k^{-1} \cdot (\underline{M}_1^s - \underline{\mu}_k) - \sum_{j=1}^m w_j^2 \cdot \mathcal{N}(\underline{\mu}_j, \underline{\mu}_k, \mathbf{C}_j + \mathbf{C}_k) \cdot \mathbf{C}_k^{-1} \cdot (\underline{M}_1^t - \underline{\mu}_k) \right] \right) \quad (A.28)$$

Diese ist identisch mit der Herleitung von Williams:

$$\nabla G(\underline{\mu}_k) \stackrel{(A.24)(A.27)}{=} w_k^2 \cdot \left[\sum_{i=1}^n \tilde{w}_i \cdot \mathcal{N}(\tilde{\underline{\mu}}_i, \underline{\mu}_k, \tilde{\mathbf{C}}_i + (\mathbf{P}_k \cdot \mathbf{P}_k^T)) \cdot (\tilde{\mathbf{C}}_i + (\mathbf{P}_k \cdot \mathbf{P}_k^T))^{-1} \cdot (\tilde{\underline{\mu}}_i - \underline{\mu}_k) - \sum_{j=1}^m w_j \cdot \mathcal{N}(\underline{\mu}_j, \underline{\mu}_k, (\mathbf{P}_j \cdot \mathbf{P}_j^T) + (\mathbf{P}_k \cdot \mathbf{P}_k^T)) \cdot ((\mathbf{P}_j \cdot \mathbf{P}_j^T) + (\mathbf{P}_k \cdot \mathbf{P}_k^T))^{-1} \cdot (\underline{\mu}_j - \underline{\mu}_k) \right] \quad (A.29)$$

Erster Summand des Gradienten für die Kovarianz:

$$\int_{\mathbb{R}^N} \tilde{f}(\underline{x}, \tilde{\underline{\eta}}) \cdot \frac{\partial f(\underline{x}, \underline{\eta})}{\partial \mathbf{C}_k} d\underline{x} \stackrel{(A.14)(A.18)}{=} w_k^2 \cdot \sum_{i=1}^n \tilde{w}_i \cdot \mathcal{N}(\tilde{\underline{\mu}}_i, \underline{\mu}_k, \tilde{\mathbf{C}}_i + \mathbf{C}_k) \cdot \underbrace{\mathbf{C}_k^{-1} \cdot [\mathbf{M}_2^s + \overbrace{\underline{M}_1^s (\underline{M}_1^s)^T - \underline{M}_1^s \underline{\mu}_k^T - \underline{\mu}_k (\underline{M}_1^s)^T + \underline{\mu}_k \underline{\mu}_k^T}_{S_2} - \mathbf{C}_k]}_{S_1} \cdot \mathbf{C}_k^{-1} \cdot \mathbf{P}_k \quad (A.30)$$

$$\mathbf{C}_s^{-1} \cdot \underline{\mu}_k = (\tilde{\mathbf{C}}_i^{-1} + \mathbf{C}_k^{-1}) \cdot \underline{\mu}_k = \tilde{\mathbf{C}}_i^{-1} \cdot \underline{\mu}_k + \mathbf{C}_k^{-1} \cdot \underline{\mu}_k \quad (A.31)$$

$$\mathbf{C}_k^{-1} \cdot \mathbf{C}_s \cdot \tilde{\mathbf{C}}_i^{-1} = \mathbf{C}_k^{-1} \cdot (\tilde{\mathbf{C}}_i^{-1} + \mathbf{C}_k^{-1})^{-1} \cdot \tilde{\mathbf{C}}_i^{-1} = \tilde{\mathbf{C}}_i^{-1} \cdot (\tilde{\mathbf{C}}_i^{-1} + \mathbf{C}_k^{-1})^{-1} \cdot \mathbf{C}_k^{-1} = \tilde{\mathbf{C}}_i^{-1} \cdot \mathbf{C}_s \cdot \mathbf{C}_k^{-1} = (\tilde{\mathbf{C}}_i + \mathbf{C}_k)^{-1} \quad (A.32)$$

$$\mathbf{C}_k \cdot \mathbf{C}_s^{-1} \cdot \tilde{\mathbf{C}}_i = \mathbf{C}_k \cdot (\tilde{\mathbf{C}}_i^{-1} + \mathbf{C}_k^{-1}) \cdot \tilde{\mathbf{C}}_i = \tilde{\mathbf{C}}_i \cdot (\tilde{\mathbf{C}}_i^{-1} + \mathbf{C}_k^{-1}) \cdot \mathbf{C}_k = \tilde{\mathbf{C}}_i \cdot \mathbf{C}_s^{-1} \cdot \mathbf{C}_k = \tilde{\mathbf{C}}_i + \mathbf{C}_k \quad (A.33)$$

$$\begin{aligned} \tilde{\mathbf{C}}_i \cdot \mathbf{C}_s^{-1} \cdot \tilde{\mathbf{C}}_i - \tilde{\mathbf{C}}_i \cdot \mathbf{C}_s^{-1} \cdot \mathbf{C}_k \cdot \mathbf{C}_s^{-1} \cdot \tilde{\mathbf{C}}_i &= \tilde{\mathbf{C}}_i \cdot \mathbf{C}_s^{-1} \cdot (\tilde{\mathbf{C}}_i - \mathbf{C}_k \cdot \mathbf{C}_s^{-1} \cdot \tilde{\mathbf{C}}_i) \\ &\stackrel{(A.33)}{=} \tilde{\mathbf{C}}_i \cdot \mathbf{C}_s^{-1} \cdot (\tilde{\mathbf{C}}_i - \tilde{\mathbf{C}}_i - \mathbf{C}_k) \\ &= -\tilde{\mathbf{C}}_i \cdot \mathbf{C}_s^{-1} \cdot \mathbf{C}_k \\ &\stackrel{(A.33)}{=} -(\tilde{\mathbf{C}}_i + \mathbf{C}_k) \end{aligned} \quad (A.34)$$

Dieser lässt sich umformen auf die Herleitung von Williams:

$$\begin{aligned}
S_1 &\stackrel{(A.30)(A.30)}{=} \underline{\mu}_s \underline{\mu}_s^T - \underline{\mu}_s \underline{\mu}_k^T - \underline{\mu}_k \underline{\mu}_k^T + \underline{\mu}_k \underline{\mu}_k^T = (\underline{\mu}_s - \underline{\mu}_k) \cdot (\underline{\mu}_s - \underline{\mu}_k)^T \\
&= (\mathbf{C}_s \cdot (\tilde{\mathbf{C}}_i^{-1} \cdot \tilde{\underline{\mu}}_i + \mathbf{C}_k^{-1} \cdot \underline{\mu}_k) - \underline{\mu}_k) \cdot (\mathbf{C}_s \cdot (\tilde{\mathbf{C}}_i^{-1} \cdot \tilde{\underline{\mu}}_i + \mathbf{C}_k^{-1} \cdot \underline{\mu}_k) - \underline{\mu}_k)^T \\
&= \mathbf{C}_s \cdot (\tilde{\mathbf{C}}_i^{-1} \cdot \tilde{\underline{\mu}}_i + \mathbf{C}_k^{-1} \cdot \underline{\mu}_k - \mathbf{C}_s^{-1} \cdot \underline{\mu}_k) \cdot (\tilde{\mathbf{C}}_i^{-1} \cdot \tilde{\underline{\mu}}_i + \mathbf{C}_k^{-1} \cdot \underline{\mu}_k - \mathbf{C}_s^{-1} \cdot \underline{\mu}_k)^T \cdot \mathbf{C}_s \\
&= \mathbf{C}_s \cdot (\tilde{\mathbf{C}}_i^{-1} \cdot \tilde{\underline{\mu}}_i + \mathbf{C}_k^{-1} \cdot \underline{\mu}_k - \mathbf{C}_s^{-1} \cdot \underline{\mu}_k) \cdot (\tilde{\mathbf{C}}_i^{-1} \cdot \tilde{\underline{\mu}}_i + \mathbf{C}_k^{-1} \cdot \underline{\mu}_k - \mathbf{C}_s^{-1} \cdot \underline{\mu}_k)^T \cdot \mathbf{C}_s \\
&\stackrel{(A.31)}{=} \mathbf{C}_s \cdot (\tilde{\mathbf{C}}_i^{-1} \cdot \tilde{\underline{\mu}}_i - \tilde{\mathbf{C}}_i^{-1} \cdot \underline{\mu}_k) \cdot (\tilde{\mathbf{C}}_i^{-1} \cdot \tilde{\underline{\mu}}_i - \tilde{\mathbf{C}}_i^{-1} \cdot \underline{\mu}_k)^T \cdot \mathbf{C}_s \\
&= \mathbf{C}_s \cdot \tilde{\mathbf{C}}_i^{-1} \cdot (\tilde{\underline{\mu}}_i - \underline{\mu}_k) \cdot (\tilde{\underline{\mu}}_i - \underline{\mu}_k)^T \cdot \tilde{\mathbf{C}}_i^{-1} \cdot \mathbf{C}_s
\end{aligned} \tag{A.35}$$

$$\begin{aligned}
S_2 &\stackrel{(A.32)}{=} (\tilde{\mathbf{C}}_i + \mathbf{C}_k)^{-1} \cdot [(\tilde{\underline{\mu}}_i - \underline{\mu}_k) \cdot (\tilde{\underline{\mu}}_i - \underline{\mu}_k)^T + \tilde{\mathbf{C}}_i \cdot \mathbf{C}_s^{-1} \cdot \tilde{\mathbf{C}}_i - \tilde{\mathbf{C}}_i \cdot \mathbf{C}_s^{-1} \cdot \mathbf{C}_k \cdot \mathbf{C}_s^{-1} \cdot \tilde{\mathbf{C}}_i] \cdot (\tilde{\mathbf{C}}_i + \mathbf{C}_k)^{-1} \\
&\stackrel{(A.34)}{=} (\tilde{\mathbf{C}}_i + \mathbf{C}_k)^{-1} \cdot [(\tilde{\underline{\mu}}_i - \underline{\mu}_k) \cdot (\tilde{\underline{\mu}}_i - \underline{\mu}_k)^T - (\tilde{\mathbf{C}}_i + \mathbf{C}_k)] \cdot (\tilde{\mathbf{C}}_i + \mathbf{C}_k)^{-1}
\end{aligned} \tag{A.36}$$

Damit ergibt sich die Herleitung nach Williams:

$$\begin{aligned}
\int_{\mathbb{R}^N} \tilde{f}(\underline{x}, \tilde{\underline{\eta}}) \cdot \frac{\partial f(\underline{x}, \underline{\eta})}{\partial \mathbf{C}_k} \, d\underline{x} &\stackrel{(A.36)(A.30)}{=} w_k^2 \cdot \sum_{i=1}^n \tilde{w}_i \cdot \mathcal{N}(\tilde{\underline{\mu}}_i, \underline{\mu}_k, \tilde{\mathbf{C}}_i + \mathbf{C}_k) \\
&\quad \cdot (\tilde{\mathbf{C}}_i + \mathbf{C}_k)^{-1} \cdot [(\tilde{\underline{\mu}}_i - \underline{\mu}_k) \cdot (\tilde{\underline{\mu}}_i - \underline{\mu}_k)^T - (\tilde{\mathbf{C}}_i + \mathbf{C}_k)] \cdot (\tilde{\mathbf{C}}_i + \mathbf{C}_k)^{-1} \cdot \mathbf{P}_k
\end{aligned} \tag{A.37}$$

Zweiter Summand des Gradienten für die Kovarianz:

$$\begin{aligned}
\int_{\mathbb{R}^N} f(\underline{x}, \underline{\eta}) \cdot \frac{\partial f(\underline{x}, \underline{\eta})}{\partial \mathbf{C}_k} \, d\underline{x} &\stackrel{(A.15)(A.18)}{=} w_k^2 \cdot \sum_{i=j}^m w_j^2 \cdot \mathcal{N}(\underline{\mu}_j, \underline{\mu}_k, \mathbf{C}_j + \mathbf{C}_k) \\
&\quad \cdot \underbrace{\mathbf{C}_k^{-1} \cdot [\mathbf{M}_2^t + \overbrace{\mathbf{M}_1^t (\mathbf{M}_1^t)^T - \mathbf{M}_1^t \underline{\mu}_k^T - \underline{\mu}_k (\mathbf{M}_1^t)^T + \underline{\mu}_k \underline{\mu}_k^T}_{T_2} - \mathbf{C}_k]}_{T_2} \cdot \mathbf{C}_k^{-1} \cdot \mathbf{P}_k
\end{aligned} \tag{A.38}$$

$$\mathbf{C}_t^{-1} \cdot \underline{\mu}_k = (\mathbf{C}_j^{-1} + \mathbf{C}_k^{-1}) \cdot \underline{\mu}_k = \mathbf{C}_j^{-1} \cdot \underline{\mu}_k + \mathbf{C}_k^{-1} \cdot \underline{\mu}_k \tag{A.39}$$

$$\begin{aligned}
\mathbf{C}_k^{-1} \cdot \mathbf{C}_t \cdot \mathbf{C}_j^{-1} &= \mathbf{C}_k^{-1} \cdot (\mathbf{C}_j^{-1} + \mathbf{C}_k^{-1})^{-1} \cdot \mathbf{C}_j^{-1} = \mathbf{C}_j^{-1} \cdot (\mathbf{C}_j^{-1} + \mathbf{C}_k^{-1})^{-1} \cdot \mathbf{C}_k^{-1} = \mathbf{C}_j^{-1} \cdot \mathbf{C}_t \cdot \mathbf{C}_k^{-1} \\
&= (\mathbf{C}_j + \mathbf{C}_k)^{-1}
\end{aligned} \tag{A.40}$$

$$\begin{aligned}
\mathbf{C}_k \cdot \mathbf{C}_t^{-1} \cdot \tilde{\mathbf{C}}_j &= \mathbf{C}_k \cdot (\mathbf{C}_j^{-1} + \mathbf{C}_k^{-1}) \cdot \mathbf{C}_j = \mathbf{C}_j \cdot (\mathbf{C}_j^{-1} + \mathbf{C}_k^{-1}) \cdot \mathbf{C}_k = \mathbf{C}_j \cdot \mathbf{C}_t^{-1} \cdot \mathbf{C}_k \\
&= \mathbf{C}_j + \mathbf{C}_k
\end{aligned} \tag{A.41}$$

$$\begin{aligned}
 \mathbf{C}_j \cdot \mathbf{C}_t^{-1} \cdot \mathbf{C}_j - \mathbf{C}_j \cdot \mathbf{C}_t^{-1} \cdot \mathbf{C}_k \cdot \mathbf{C}_t^{-1} \cdot \mathbf{C}_j &= \mathbf{C}_j \cdot \mathbf{C}_t^{-1} \cdot (\mathbf{C}_j - \mathbf{C}_k \cdot \mathbf{C}_t^{-1} \cdot \mathbf{C}_j) \\
 &\stackrel{(A.41)}{=} \mathbf{C}_j \cdot \mathbf{C}_t^{-1} \cdot (\mathbf{C}_j - \mathbf{C}_j - \mathbf{C}_k) \\
 &= -\mathbf{C}_j \cdot \mathbf{C}_t^{-1} \cdot \mathbf{C}_k \\
 &\stackrel{(A.41)}{=} -(\mathbf{C}_j + \mathbf{C}_k)
 \end{aligned} \tag{A.42}$$

Dieser lässt sich umformen auf die Herleitung von Williams:

$$\begin{aligned}
 T_1 &= \underline{\mu}_t \underline{\mu}_t^\top - \underline{\mu}_t \underline{\mu}_k^\top - \underline{\mu}_k \underline{\mu}_k^\top + \underline{\mu}_k \underline{\mu}_k^\top = (\underline{\mu}_t - \underline{\mu}_k) \cdot (\underline{\mu}_t - \underline{\mu}_k)^\top \\
 &= (\mathbf{C}_t \cdot (\mathbf{C}_j^{-1} \cdot \underline{\mu}_j + \mathbf{C}_k^{-1} \cdot \underline{\mu}_k) - \underline{\mu}_k) \cdot (\mathbf{C}_t \cdot (\mathbf{C}_j^{-1} \cdot \underline{\mu}_j + \mathbf{C}_k^{-1} \cdot \underline{\mu}_k) - \underline{\mu}_k)^\top \\
 &= \mathbf{C}_t \cdot (\mathbf{C}_j^{-1} \cdot \underline{\mu}_j + \mathbf{C}_k^{-1} \cdot \underline{\mu}_k - \mathbf{C}_t^{-1} \cdot \underline{\mu}_k) \cdot (\mathbf{C}_j^{-1} \cdot \underline{\mu}_j + \mathbf{C}_k^{-1} \cdot \underline{\mu}_k - \mathbf{C}_t^{-1} \cdot \underline{\mu}_k)^\top \cdot \mathbf{C}_t \\
 &= \mathbf{C}_t \cdot (\mathbf{C}_j^{-1} \cdot \underline{\mu}_j + \mathbf{C}_k^{-1} \cdot \underline{\mu}_k - \mathbf{C}_t^{-1} \cdot \underline{\mu}_k) \cdot (\mathbf{C}_j^{-1} \cdot \underline{\mu}_j + \mathbf{C}_k^{-1} \cdot \underline{\mu}_k - \mathbf{C}_t^{-1} \cdot \underline{\mu}_k)^\top \cdot \mathbf{C}_t \\
 &\stackrel{(A.39)}{=} \mathbf{C}_t \cdot (\mathbf{C}_j^{-1} \cdot \underline{\mu}_j - \mathbf{C}_j^{-1} \cdot \underline{\mu}_k) \cdot (\mathbf{C}_j^{-1} \cdot \underline{\mu}_j - \mathbf{C}_j^{-1} \cdot \underline{\mu}_k)^\top \cdot \mathbf{C}_t \\
 &= \mathbf{C}_t \cdot \mathbf{C}_j^{-1} \cdot (\underline{\mu}_j - \underline{\mu}_k) \cdot (\underline{\mu}_j - \underline{\mu}_k)^\top \cdot \mathbf{C}_j^{-1} \cdot \mathbf{C}_t
 \end{aligned} \tag{A.43}$$

$$\begin{aligned}
 T_2 &\stackrel{(A.40)}{=} (\mathbf{C}_j + \mathbf{C}_k)^{-1} \cdot [(\underline{\mu}_j - \underline{\mu}_k) \cdot (\underline{\mu}_j - \underline{\mu}_k)^\top + \mathbf{C}_j \cdot \mathbf{C}_t^{-1} \cdot \mathbf{C}_j - \mathbf{C}_j \cdot \mathbf{C}_t^{-1} \cdot \mathbf{C}_k \cdot \mathbf{C}_t^{-1} \cdot \mathbf{C}_j] \cdot (\mathbf{C}_j + \mathbf{C}_k)^{-1} \\
 &\stackrel{(A.42)}{=} (\mathbf{C}_j + \mathbf{C}_k)^{-1} \cdot [(\underline{\mu}_j - \underline{\mu}_k) \cdot (\underline{\mu}_j - \underline{\mu}_k)^\top - (\mathbf{C}_j + \mathbf{C}_k)] \cdot (\mathbf{C}_j + \mathbf{C}_k)^{-1}
 \end{aligned} \tag{A.44}$$

Damit ergibt sich die Herleitung nach Williams:

$$\begin{aligned}
 \int_{\mathbb{R}^N} f(\underline{x}, \underline{\eta}) \cdot \frac{\partial f(\underline{x}, \underline{\eta})}{\partial \mathbf{C}_k} \, d\underline{x} &\stackrel{(A.44)(A.38)}{=} w_k^2 \cdot \sum_{j=1}^m w_j^2 \cdot \mathcal{N}(\underline{\mu}_j, \underline{\mu}_k, \mathbf{C}_j + \mathbf{C}_k) \\
 &\quad \cdot (\mathbf{C}_j + \mathbf{C}_k)^{-1} \cdot [(\underline{\mu}_j - \underline{\mu}_k) \cdot (\underline{\mu}_j - \underline{\mu}_k)^\top - (\mathbf{C}_j + \mathbf{C}_k)] \cdot (\mathbf{C}_j + \mathbf{C}_k)^{-1} \cdot \mathbf{P}_k
 \end{aligned} \tag{A.45}$$

Partielle Ableitung nach der Kovarianz:

$$\begin{aligned}
 \nabla G(\mathbf{C}_k) &\stackrel{(A.30)(A.38)}{=} \text{Vec} \left(w_k^2 \cdot \mathbf{C}_k^{-1} \cdot \left[\sum_{j=1}^m w_j^2 \cdot \mathcal{N}(\underline{\mu}_j, \underline{\mu}_k, \mathbf{C}_j + \mathbf{C}_k) \right. \right. \\
 &\quad \cdot [\mathbf{M}_2^t + \underline{M}_1^t (\underline{M}_1^t)^\top - \underline{M}_1^t \underline{\mu}_k^\top - \underline{\mu}_k (\underline{M}_1^t)^\top + \underline{\mu}_k \underline{\mu}_k^\top - \mathbf{C}_k] \\
 &\quad \left. \left. - \sum_{i=1}^n \tilde{w}_i \cdot \mathcal{N}(\tilde{\underline{\mu}}_i, \underline{\mu}_k, \tilde{\mathbf{C}}_i + \mathbf{C}_k) \right. \right. \\
 &\quad \left. \left. \cdot [\mathbf{M}_2^s + \underline{M}_1^s (\underline{M}_1^s)^\top - \underline{M}_1^s \underline{\mu}_k^\top - \underline{\mu}_k (\underline{M}_1^s)^\top + \underline{\mu}_k \underline{\mu}_k^\top - \mathbf{C}_k] \right) \cdot \mathbf{C}_k^{-1} \cdot \mathbf{P}_k
 \end{aligned} \tag{A.46}$$

Diese ist identisch mit der Herleitung von Williams:

$$\begin{aligned}
\nabla G(\mathbf{C}_k) &\stackrel{(A.37)(A.45)}{=} w_k^2 \cdot \left[\sum_{j=1}^m w_j^2 \cdot \mathcal{N}(\underline{\mu}_j, \underline{\mu}_k, \mathbf{C}_j + \mathbf{C}_k) \right. \\
&\quad \cdot (\mathbf{C}_j + \mathbf{C}_k)^{-1} \cdot [(\underline{\mu}_j - \underline{\mu}_k) \cdot (\underline{\mu}_j - \underline{\mu}_k)^T - (\mathbf{C}_j + \mathbf{C}_k)] \cdot (\mathbf{C}_j + \mathbf{C}_k)^{-1} \cdot \mathbf{P}_k \quad (A.47) \\
&\quad - \sum_{i=1}^n \tilde{w}_i \cdot \mathcal{N}(\tilde{\underline{\mu}}_i, \underline{\mu}_k, \tilde{\mathbf{C}}_i + \mathbf{C}_k) \\
&\quad \left. \cdot (\tilde{\mathbf{C}}_i + \mathbf{C}_k)^{-1} \cdot [(\tilde{\underline{\mu}}_i - \underline{\mu}_k) \cdot (\tilde{\underline{\mu}}_i - \underline{\mu}_k)^T - (\tilde{\mathbf{C}}_i + \mathbf{C}_k)] \cdot (\tilde{\mathbf{C}}_i + \mathbf{C}_k)^{-1} \cdot \mathbf{P}_k \right]
\end{aligned}$$

Literaturverzeichnis

- [1] D. Schieferdecker and M. Huber, “Gaussian mixture reduction via clustering,” in *12th International Conference on Information Fusion*, July 6-9 2009.
- [2] B. Leiner, *Einführung in die Statistik*. Oldenbourg Wissenschaftsverlag, 2004.
- [3] J. L. Williams and P. S. Maybeck, “Cost-function-based Gaussian mixture reduction for target tracking,” in *6th International Conference on Information Fusion, Cairns, Australia*, July 08-11 2003, pp. 1047–1054.
- [4] S. Kullback, *Statistics and Information Theory*. John Wiley, New York, 1959.
- [5] P. C. Mahalanobis, “One generalized distance in statistics,” in *Proceedings of the National Institute of Science in India*, 1936.
- [6] M. WEST, “Approximating posterior distributions by mixtures,” *Journal of the Royal Statistical Society. Series B. Methodological*, vol. 55, no. 2, pp. 409–422, 1993.
- [7] D. J. Salmond, “Mixture reduction algorithms for target tracking in clutter,” *Signal and data processing of small targets 199, Orlando, FL*, Apr 16-18 1990.
- [8] A. R. Runnalls, “Kullback-Leibler Approach to Gaussian Mixture Reduction,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 3, pp. 989–999, Jul 2007.
- [9] M. F. Huber and U. D. Hanebeck, “Progressive Gaussian mixture reduction,” in *11th International Conference on Information Fusion*, 2008, pp. 1–8.
- [10] P. Arabie, L. Hubert, and G. de Soete, *Clustering and Classification*. World Scientific, 1996.
- [11] Z. Huang, “A fast clustering algorithm to cluster very large categorical data sets in data mining,” *Research Issues on Data Mining and Knowledge Discovery*, pp. 1–8, 1997.
- [12] G. B. Coleman and H. C. Andrews, “Image segmentation by clustering,” *Proceedings of the IEEE*, vol. 67, no. 5, pp. 773–785, 1979.
- [13] S. Lloyd, “Least squares quantization in PCM,” *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982.

- [14] D. Arthur and S. Vassilvitskii, “k-means++: The advantages of careful seeding,” in *SO-DA: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2007, pp. 1027–1035.
- [15] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical recipes in C*. Cambridge University Press New York, 1988.
- [16] J. L. Williams, “Gaussian mixture reduction for tracking multiple maneuvering targets in clutter,” Master’s thesis, Air Force Institute of Technology, Wright-Patterson Air Force Base, OH, 2003.
- [17] D. J. Salmond, “Mixture reduction algorithms for uncertain tracking,” *Technical Report 88004*, Royal Aerospace Establishment, 1988.
- [18] M. B., “Matrix calculus,” <http://www.ee.ic.ac.uk/hp/staff/dmb/matrix/calculus.html>, 2009.
- [19] K. Petersen and M. Pedersen, “The Matrix Cookbook,” http://www.imm.dtu.dk/pubdb/views/edoc_download.php/3274/pdf/imm3274.pdf, 2006.