



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



Diplomarbeit

Multivariate Dichteapproximation mit Dirac Mixtures

Thomas Peter Krekeler

1. November 2007

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

Betreuer: Dipl.-Inform. Oliver Schrempf

Eidesstattliche Erklärung

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

Karlsruhe, 1. November 2007

Thomas Peter Krekeler

„Den hat ein Bekannter von mir selbst geschrieben.“
Einen Bildschirmschoner der Homecomputer-Ära.
Ich war fasziniert. . . danke dafür.

Norman Vattmann, gestorben mit 13 Jahren.

Im Gedenken an meine liebe Cousine.

Birgit Grabarczyk, 25.03.1968 – 29.07.2007.

Inhaltsverzeichnis

Abbildungsverzeichnis	V
Algorithmenverzeichnis	IX
Verzeichnis der Beispiele	XI
Notation	XIII
1 Einleitung	1
2 Grundlagen	5
2.1 Dirac-Funktion	5
2.2 Dichte und Verteilung	6
2.2.1 Gleichverteilung	7
2.2.2 Gauß-Verteilung	7
2.2.3 Dirac-Verteilung	8
2.2.4 Mischverteilungen	10
2.3 Reihendarstellung und Approximation	11
2.4 Approximation formuliert als Optimierungsproblem	13
3 Stand der Technik	15
3.1 Stetige Grundfunktionen	15
3.2 Dirac-Impulse als Grundfunktionen	15
3.2.1 Simulated Annealing	16
3.2.2 Zufällige Stichproben	17
3.2.3 Greedy-Verfahren	17

3.2.4	Progressives Bayes	18
4	Problemformulierung	25
4.1	Ordnungsannahme	25
4.2	Differentialgleichungssystem	25
4.3	Parameteranzahl	26
4.4	Analytische Form des Gütemaßes	26
4.5	Bifurkationen	26
4.5.1	Separiertheit (lokal) optimaler Approximationen	27
4.6	Sackgassen	27
4.7	Zielsetzung	28
5	Progressives Bayes im Mehrdimensionalen	29
5.1	Notwendige Bedingung	29
5.2	Differentialgleichungssystem	32
5.3	Jakobi-Matrix	34
5.4	Irregularitäten	35
5.5	Aufwandsanalyse	38
6	Freie Parameter	41
6.1	Anforderungen	41
6.1.1	Startdichte	42
6.1.2	Progression	42
6.2	Spezielle Startdichten	43
6.2.1	Gleichverteilung	43
6.2.2	Quasi-ausgeartete Gauß-Mischdichten/Initiale Approximationen	46
6.3	Spezielle Progressionen	48
6.3.1	Variation der Parameter	48
6.3.2	Konstante Progression	49
6.3.3	Konvexe Kombination	49

7	Spezielle Kombinationen aus Startdichte und Progression	51
7.1	Progressive Verbesserung initialer Approximationen	51
7.1.1	Auswahl initialer Approximationen	52
7.1.2	Initiale Approximationen für Gleichverteilungen	53
7.1.3	Initiale Approximationen für Gauß-Verteilungen	60
7.2	Konvexe Kombination von Gleichverteilung und Zielverteilung	65
8	Suboptimale Approximation durch Clusterbildung	69
8.1	Approximation der Einzelcluster	69
8.2	Clusterbildung	73
8.3	Erreichen einer vorgegebenen Gesamtgüte	77
9	Vergleich zu anderen Verfahren	79
9.1	Beispiel Standardnormalverteilung	80
9.2	Beispiel Halbkreis	86
9.3	Beispiel ISAS Logo	91
10	Zusammenfassung und Ausblick	97
A	Mathematischer Anhang	101
A.1	Konventionen	101
A.2	Schreibweisen	102
A.3	Grundlegende Formeln	103
A.3.1	Verallgemeinerte Kettenregel	103
A.3.2	Partielle Ableitungen der max-Funktion	103
A.3.3	Leibniz-Regel für Parameterintegrale	105
A.3.4	Einbringen der Heaviside-Funktion in die Integrationsgrenzen	105
A.4	Approximation	105
A.4.1	Abkürzende Definitionen	106
A.4.2	Partielle Ableitungen von F	106
A.4.3	Das Integral von F	107
A.4.4	Das Integral von F über $d - 1$ Dimensionen	109

A.5	Progressive Verteilung	113
A.5.1	Allgemeiner Fall	113
A.5.2	Konvexe Kombination	116
A.5.3	Gleichverteilung	117
A.5.4	Achsenparallele Gauß-Mischverteilung	118
A.6	Partielle Ableitungen des Gütemaßes	120
A.6.1	Nach w_i	120
A.6.2	Nach x_{im}	121

Literaturverzeichnis		123
-----------------------------	--	------------

Abbildungsverzeichnis

2.1	Gleichverteilung.	8
2.2	Standardnormalverteilung.	9
2.3	Heaviside- und Dirac-Funktion.	10
2.4	Dirac-Mischverteilung.	11
3.1	Schematische Darstellung des Progressiven Bayes-Verfahrens.	19
5.1	Veranschaulichung der notwendigen Bedingung.	31
5.2	Achsenparallel ausgerichtete Dirac-Impulse.	37
6.1	Übertragung eindimensionaler Approximationen auf das Zweidimensionale.	45
6.2	Schematische Darstellung des Ansatzes zur progressiven Verbesserung initialer Approximationen.	46
7.1	Initiale Approximation einer Gleichverteilung ($\gamma = 0$).	54
7.2	Verbesserte Approximation einer Gleichverteilung ($\gamma \approx \frac{1}{6}$).	54
7.3	Verbesserte Approximation einer Gleichverteilung ($\gamma \approx \frac{1}{3}$).	55
7.4	Verbesserte Approximation einer Gleichverteilung ($\gamma = 1$).	55
7.5	Verschiedene initiale Approximationen für eine Gleichverteilung ($\gamma = 0$).	56
7.6	Verschiedene verbesserte Approximationen für eine Gleichverteilung ($\gamma = 1$).	57
7.7	Bifurkationen Teil 1: Fast identische initiale Approximationen ($\gamma = 0$).	58
7.8	Bifurkationen Teil 2: Sehr verschiedene verbesserte Approximationen ($\gamma = 1$).	58
7.9	Separiertheit (lokal) optimaler Approximationen einer Gleichverteilung.	59
7.10	Initiale und progressiv verbesserte Positionen für Gauß-Dichten.	61
7.11	Initiale und progressiv verbesserte Gewichte für Gauß-Dichten.	61
7.12	Initiale und progressiv verbesserte Approximationen für Gauß-Dichten.	62

7.13	Progression zu einer initialen Approximation einer Gauß-Dichte ($\gamma = 0$).	63
7.14	Progression zu einer initialen Approximation einer Gauß-Dichte ($\gamma \approx \frac{1}{3}$).	64
7.15	Progression zu einer initialen Approximation einer Gauß-Dichte ($\gamma \approx \frac{2}{3}$).	64
7.16	Progression zu einer initialen Approximation einer Gauß-Dichte ($\gamma = 1$).	65
7.17	Progressive Approximation mit einer Gleichverteilung als Startverteilung ($\gamma = 0$).	66
7.18	Progressive Approximation mit einer Gleichverteilung als Startverteilung ($\gamma \approx \frac{1}{3}$).	67
7.19	Progressive Approximation mit einer Gleichverteilung als Startverteilung ($\gamma \approx \frac{2}{3}$).	67
7.20	Progressive Approximation mit einer Gleichverteilung als Startverteilung ($\gamma = 1$).	68
8.1	Cluster gebildet durch das Cluster-Verfahren, Teil 1.	75
8.2	Cluster gebildet durch das Cluster-Verfahren, Teil 2.	76
8.3	Initiale Gesamtapproximation als Summe der Approximationen der Cluster.	77
8.4	Verbesserte Gesamtapproximation als Summe der Approximationen der Cluster.	77
9.1	Ergebnisse des Cluster-Verfahrens für die Standardnormalverteilung.	81
9.2	Gütevergleich für die Standardnormalverteilung, Teil 1.	81
9.3	Gütevergleich für die Standardnormalverteilung, Teil 2.	82
9.4	Streuung der Güte für Zufallsstichproben für die Standardnormalverteilung.	83
9.5	Visueller Vergleich der Approximationen für die Standardnormalverteilung, Teil 1.	84
9.6	Visueller Vergleich der Approximationen für die Standardnormalverteilung, Teil 2.	85
9.7	Zieldichte des Halbkreis Beispiels.	86
9.8	Ergebnisse des Cluster-Verfahrens für das Halbkreis Beispiel.	87
9.9	Gütevergleich für das Halbkreis Beispiel, Teil 1.	87
9.10	Gütevergleich für das Halbkreis Beispiel, Teil 2.	88
9.11	Streuung der Güte für Zufallsstichproben für das Halbkreis Beispiel.	88
9.12	Visueller Vergleich der Approximationen für das Halbkreis Beispiel, Teil 1.	89
9.13	Visueller Vergleich der Approximationen für das Halbkreis Beispiel, Teil 2.	90
9.14	Zieldichte für das ISAS Logo Beispiel.	91
9.15	Ergebnisse des Cluster-Verfahrens für das ISAS Logo Beispiel.	92
9.16	Gütevergleich für das ISAS Logo Beispiel.	92
9.17	Visueller Vergleich der Approximationen für das ISAS Logo Beispiel, Teil 1.	93

9.18	Visueller Vergleich der Approximationen für das ISAS Logo Beispiel, Teil 2. . . .	94
9.19	Streuung der Güte für Zufallsstichproben für das ISAS Logo Beispiel.	95

Algorithmenverzeichnis

1	Numerisches Lösungsverfahren für das Differentialgleichungssystem.	22
2	Approximation eines Clusters.	72
3	Approximation mit dem Cluster-Verfahren.	73
4	Erstellen eines Clusters.	74

Verzeichnis der Beispiele

Beispiel 1.1:	Robotersteuerung	1
Beispiel 1.1:	Robotersteuerung (Fortsetzung)	2
Beispiel 1.1:	Robotersteuerung (Fortsetzung)	2
Beispiel 5.1:	Notwendige Bedingung am Beispiel einer Gleichverteilung	31
Beispiel 5.1:	Notwendige Bedingung am Beispiel einer Gleichverteilung (Fortsetzung) . . .	32
Beispiel 5.2:	Achsenparallel ausgerichtete Dirac-Impulse	36
Beispiel 6.1:	Übertragung einer optimalen eindimensionalen Approximation	44
Beispiel 7.1:	Motivation des Positionsschemas für eine Gleichverteilung	53
Beispiel 7.2:	Verschiedene initiale Approximationen im Vergleich	56
Beispiel 7.3:	Bifurkationen bei achsenparallel ausgerichteten Dirac-Impulsen	57
Beispiel 7.4:	Separiertheit (lokal) optimaler Approximationen	59
Beispiel 7.5:	Positionsschema für Gauß-Dichten an Hand ihrer Konstanzbereiche	60
Beispiel 7.6:	Halbkreis als Zielverteilung	65
Beispiel 8.1:	Initiale Approximationen für achsenparallele Gauß-Mischdichten	70
Beispiel 8.2:	Abstandsmaß für achsenparallele Gauß-Dichten	73
Beispiel 8.3:	Cluster für das ISAS Logo	74

Notation

Konventionen

- x Skalar
- \underline{x} Spaltenvektor
- \underline{x}^T Transponierte des Vektors \underline{x}
- \mathbb{N} Natürliche Zahlen
- \mathbb{R} Reelle Zahlen
- \mathbb{R}^+ Nicht negative reelle Zahlen
- Ende eines Beispiels, Satzes oder Beweises
- Ende einer Definition

Funktionssymbole

- $\mathcal{N}(x, \mu, \sigma)$ Gauß-Dichtefunktion mit Mittelwert μ und Standardabweichung σ
- $\Phi(x, \mu, \sigma)$ Gauß-Verteilungsfunktion mit Mittelwert μ und Standardabweichung σ
- $\operatorname{erf}(x)$ Gauß'sche Fehlerfunktion
- $\delta(\underline{x})$ Dirac-Funktion
- $H(\underline{x})$ Heaviside-Funktion
- $1_M(\underline{x})$ Charakteristische Funktion der Menge M
- $\mathcal{U}(\underline{x}, \underline{\alpha}, \underline{\beta})$ Verteilungsfunktion der Gleichverteilung auf dem Intervall $[\underline{\alpha}, \underline{\beta}]$

- $\dot{g}(\underline{x})$ Gradient der Funktion g
- $\frac{\partial}{\partial x_i} g(\underline{x})$ Partielle Ableitung von g in Richtung von x_i

- γ Progressionsparameter
- $\underline{\eta}$ Parametervektor einer Approximationsverteilung

- $\tilde{f}(\underline{x}, \gamma)$ Progressive Dichte mit Progressionsparameter γ
- $\tilde{F}(\underline{x}, \gamma)$ Progressive Verteilungsfunktion mit Progressionsparameter γ

- $f(\underline{x}, \underline{\eta})$ Approximationsdichte mit Parametervektor $\underline{\eta}$
- $F(\underline{x}, \underline{\eta})$ Verteilungsfunktion einer Approximation mit Parametervektor $\underline{\eta}$

KAPITEL 1

Einleitung

Diese Arbeit befasst sich mit der Approximation von mehrdimensionalen Dichten durch Dirac-Mischdichten. Die Definition dieser Begriffe und eine Einführung in die Grundlagen ist im nachfolgenden Kapitel 2 zu finden. Eine Motivation für die Nutzung von Dirac-Mischdichten im Kontext der Systemtheorie erfolgt in Abschnitt 3.2. In diesem Abschnitt soll die Thematik umgangssprachlich an Hand eines Beispiels motiviert werden.

Viele technische Systeme sind durch einen internen Zustand charakterisiert. Aus Eingabe und Zustand des Systems zum aktuellen Zeitpunkt ergibt sich der Folgezustand zum nächsten Zeitpunkt. Mathematisch wird das Verhalten eines solchen Systems durch die Systemabbildung beschrieben. Sie ist eine Funktion, welche eine Eingabe zusammen mit einem Zustand wieder auf einen Zustand abbildet.

Beispiel 1.1: Robotersteuerung

Das System besteht aus einem Roboter in einem abgeschlossenen Raum. Der Roboter kann sich mit Hilfe seiner Räder bewegen. Im Raum sind Hindernisse verteilt.

Eingabe Stellgrößen für die Bewegung des Roboters, zum Beispiel Änderung der Fahrtrichtung und Geschwindigkeit.

Zustand Geschwindigkeit sowie Ausrichtung und Position des Roboters im Raum.

Der Roboter wird zu Beginn auf der Startposition platziert. Eine Zielposition wird vorgegeben und die Aufgabe des Roboters besteht nun darin, sich von der Start- zur Zielposition zu bewegen. Ein Sensor erlaubt dem Roboter, den Abstand zum nächsten Hindernis beziehungsweise zur nächsten Wand in Blickrichtung zu ermitteln.



In der realen Welt wird das System durch Störgrößen beeinflusst. Sie werden typischerweise als stochastisches Rauschen modelliert, dem so genannten Systemrauschen. Die Verteilungen der Störgrößen werden als bekannt angenommen, während ihre konkreten Ausprägungen normalerweise nicht bekannt sind.

Beispiel 1.1: Robotersteuerung (Fortsetzung)

Der Boden im Raum ist im Wesentlichen glatt, es gibt jedoch kleine Unebenheiten. Sie bilden eine Störgröße mit einem — wenn auch kleinen — Einfluss auf die Fahrtrichtung des Roboters. ■

Auch unter Kenntnis aller Eingaben ist der Zustand des Systems folglich im Verlauf der Zeit nicht mehr exakt bekannt und kann daher nur in Form einer Verteilung über alle möglichen Zustände angegeben werden: Einige Zustandsbereiche sind zum aktuellen Zeitpunkt wahrscheinlich, andere sind unwahrscheinlich.

Aus der Verteilung für den Zustand zum aktuellen Zeitpunkt, der aktuellen Eingabe sowie den stochastischen Kenngrößen des Systemrauschens ergibt sich über die Systemabbildung die Verteilung des Zustands zum nächsten Zeitpunkt. Die Berechnung dieser Verteilung zum nächsten Zeitpunkt wird Prädiktion genannt. Die Prädiktion kann wiederholt über mehrere Zeitschritte angewendet werden. In diesem Fall wird auf Grund des Systemrauschens immer unsicherer, wie der tatsächliche Zustand des Systems aussieht.

Um diese Unsicherheit zu vermindern, können Messungen am System vorgenommen werden. Gemessen wird der Systemzustand selbst oder Kenngrößen, welche mehr oder weniger stark mit dem Zustand korreliert sind. Normalerweise unterliegen diese Messungen ebenfalls einem Rauschen, dem so genannten Messrauschen. Mit Hilfe der Messung kann aus der prädierten Verteilung eine verbesserte Schätzung des Zustands berechnet werden. Dieser Berechnungsvorgang wird „Filtern“ genannt. Durch die Messung wird die Unsicherheit bezüglich des tatsächlichen Zustands geringer. Die verbesserte Schätzung hat wiederum die Form einer Verteilung über dem Zustandsraum.

Beispiel 1.1: Robotersteuerung (Fortsetzung)

Der Roboter führt mit seinem Sensor eine Messung für den Abstand zum nächsten Hindernis durch. Das Ergebnis ist „1 Meter“. Der Roboter besitzt eine — mehr oder weniger exakte — Karte des Raums samt Hindernissen, so dass für jede mögliche Position und Ausrichtung des Roboters im Raum der Abstand zum nächsten Hindernis — mehr oder weniger genau — bekannt ist. Auf Grund der Prädiktion vor der Messung waren zwei Zustandsbereiche etwa gleich wahrscheinlich, einer von ihnen mit einem Abstand von etwa 2 Metern zum nächsten Hindernis, der andere mit einem Abstand von etwa 1.1 Metern. Die Messung ist für den zweiten Zustandsbereich wesentlich plausibler als für den ersten, so dass ihm nach dem Filterschritt eine höhere Wahrscheinlichkeit zukommt als dem ersten. Da die Messung selbst mit einer Unsicherheit behaftet ist, sinkt die Wahrscheinlichkeit für den ersten Zustandsbereich allerdings nicht ganz auf Null. ■

Wird der Prädiktions- beziehungsweise Filterschritt mathematisch exakt durchgeführt, so wird dieser auch als *Bayes'scher* Prädiktions- beziehungsweise Filterschritt bezeichnet. Typischerweise ändert sich dabei der analytische Typ oder zumindest die Anzahl der Parameter der Verteilung. Um Prädiktion und Filterung rekursiv anwenden zu können, ist der Erhalt des Typs und der Anzahl der Parameter der Verteilung wünschenswert, in der Praxis sogar notwendig.

Dazu wird die Verteilung respektive deren Dichte approximiert. Die Approximation erfolgt normalerweise über Grundfunktionen, die durch gewichtete Summen (Reihenentwicklung) zur eigentlichen Approximationsfunktion zusammengesetzt werden. Diese Arbeit beschäftigt sich mit Dirac-Mischdichten als Approximationsfunktionen mit der Zielsetzung, eine möglichst gute Approximation zu erreichen. Was Dirac-Mischdichten sind und „möglichst gut“ bedeutet, wird im nachfolgenden Kapitel erläutert.

Grundlagen

2.1 Dirac-Funktion

Definition 2.1.1 Eine Distribution ist eine lineare sowie stetige Abbildung eines Testfunktionsraums auf die reellen Zahlen. Ein Testfunktionsraum ist ein Vektorraum, dessen Elemente Funktionen sind. \square

Distributionen sind verallgemeinerte Funktionen. Für eine mathematisch fundierte Einführung siehe zum Beispiel [1]. Definition 2.1.1 soll hier genügen, insbesondere da in dieser Arbeit nur eine Distribution betrachtet wird: die Dirac-Funktion δ . Sie ist auch unter dem Namen Dirac'sche Delta-Funktion oder — mathematisch korrekter — Delta-Distribution bekannt. Statt eine formale Definition zu geben, soll sie hier über ihre für diese Arbeit relevanten Eigenschaften eingeführt werden.

Der Funktionswert

$$\delta(\underline{x}) = \begin{cases} 0 & \underline{x} \neq 0 \\ \text{undefiniert} & \underline{x} = 0 \end{cases}$$

der Dirac-Funktion wird für jedes $\underline{x} \in \mathbb{R}^d$ als Null angesehen, außer in der Nullstelle. Dort ist sie nicht definiert. Oft wird ihr Funktionswert an dieser Stelle als unendlich betrachtet. Daher wird auch von Dirac-Impuls an der Stelle Null gesprochen. Wird der Dirac-Impuls durch eine lineare Transformation an eine andere Stelle $\underline{y} \in \mathbb{R}^d$ verschoben,

$$f(\underline{x}) := \delta(\underline{x} - \underline{y}) \ ,$$

so wird f auch als Dirac-Impuls an der Stelle \underline{y} bezeichnet.

Die Dirac-Funktion ist integrierbar und es gilt

$$\int_{\mathbb{R}^d} \delta(\underline{x}) d\underline{x} = 1 \ .$$

Eine der wertvollsten Eigenschaften der Dirac-Funktion ist die *Ausblendeigenschaft*. Für jedes $\underline{y} \in \mathbb{R}^d$ und jede Funktion f^t aus dem zu Grunde liegenden Testfunktionsraum gilt

$$\int_{\mathbb{R}^d} \delta(\underline{x} - \underline{y}) f^t(\underline{x}) d\underline{x} = f^t(\underline{y}) \ . \tag{2.1}$$

In den Abbildungen dieser Arbeit wird ein Dirac-Impuls durch einen Marker dargestellt. Abbildung 2.3 zeigt neben der Dirac-Funktion die eng mit ihr verbundene Heaviside-Funktion, die in Abschnitt 2.2.3 eingeführt wird.

2.2 Dichte und Verteilung

Definition 2.2.1 Eine Dichte $f(\underline{x})$ ist eine integrierbare, nicht negative Funktion. Gilt zusätzlich

$$\int_{\mathbb{R}^d} f(\underline{x}) d\underline{x} = 1 \quad ,$$

so heißt $f(\underline{x})$ auch Wahrscheinlichkeitsdichte. □

In dieser Arbeit wird keine Normierung für Dichte vorausgesetzt, im Unterschied zu einer Wahrscheinlichkeitsdichte ist das Integral über eine Dichte also nicht notwendigerweise gleich Eins. Weiterhin soll Definition 2.2.1 den Fall $f(\underline{x}) = \delta(\underline{x})$ der Dirac-Funktion umfassen, auch wenn $\delta(\underline{x})$ streng genommen keine Funktion sondern eine Distribution ist.

Definition 2.2.2 Für eine gegebene Dichte $f(\underline{x})$ heißt

$$F : \mathbb{R}^d \rightarrow \mathbb{R}, \quad F(\underline{x}) := \int_{-\infty}^{\underline{x}} f(\underline{t}) d\underline{t}$$

die Verteilungsfunktion von $f(\underline{x})$. □

Da eine Dichte immer nicht negativ und integrierbar ist, gilt

- $\lim_{\underline{x} \rightarrow -\infty} F(\underline{x}) = 0$,
- $\lim_{\underline{x} \rightarrow \infty} F(\underline{x}) = \int_{-\infty}^{\infty} f(\underline{x}) d\underline{x} < \infty$ und
- $\underline{x} \leq \underline{y} \Rightarrow F(\underline{x}) \leq F(\underline{y})$ für alle $\underline{x}, \underline{y} \in \mathbb{R}^d$.

Die Verteilungsfunktion einer Dichte ist also nicht negativ, beschränkt und monoton wachsend.

Ist $f(\underline{x})$ eine Wahrscheinlichkeitsdichte, dann ist das Bildmaß des Borel-Lebesgue-Maßes von f eine Wahrscheinlichkeitsverteilung. Das Bildmaß des Borel-Lebesgue-Maßes einer nicht notwendigerweise normierten Dichte wird in dieser Arbeit als Verteilung bezeichnet. Eine formale Definition dieser Begriffe würde zu sehr in die Wahrscheinlichkeitstheorie abschweifen, ohne für den Kern dieser Arbeit von Nutzen zu sein, denn: Verteilung meint in dieser Arbeit meist die Verteilungsfunktion einer Dichte. Eine fundierte Einführung in die Wahrscheinlichkeitstheorie bietet beispielsweise [2].

2.2.1 Gleichverteilung

Für die Definition der Gleichverteilung ist die Einführung der charakteristischen Funktion einer Menge hilfreich.

Definition 2.2.3 Die charakteristische Funktion

$$1_M(\underline{x}) := \begin{cases} 1 & \underline{x} \in M \\ 0 & \text{sonst} \end{cases}$$

einer Menge $M \subseteq \mathbb{R}^d$ ist gleich Eins genau für diejenigen Elemente, welche in der Menge liegen, und Null überall sonst. Insbesondere gilt für die einelementige Menge $M = \{0\}$

$$1_{\{0\}}(x) = \begin{cases} 1 & x = 0 \\ 0 & \text{sonst} \end{cases} .$$

□

Definition 2.2.4 Die Dichte $f(\underline{x})$ der Gleichverteilung $\mathcal{U}(\underline{\alpha}, \underline{\beta})$ auf dem Intervall $[\underline{\alpha}, \underline{\beta}]$, $\underline{\alpha}, \underline{\beta} \in \mathbb{R}^d$ mit $\underline{\alpha} \leq \underline{\beta}$, ist definiert als

$$f(\underline{x}) := \left\{ \prod_{p=1}^d \{\beta_p - \alpha_p\} \right\}^{-1} \cdot 1_{[\underline{\alpha}, \underline{\beta}]}(\underline{x}) .$$

Sie ist konstant auf dem Intervall $[\underline{\alpha}, \underline{\beta}]$ und Null überall sonst. Ihre Verteilungsfunktion ist

$$\mathcal{U}(\underline{x}, \underline{\alpha}, \underline{\beta}) = \prod_{p=1}^d \left\{ \frac{x_p - \alpha_p}{\beta_p - \alpha_p} 1_{[\alpha_p, \beta_p]}(x_p) \right\} .$$

□

Die Verteilungsfunktion einer mehrdimensionalen Gleichverteilung ist das Produkt der Verteilungsfunktionen eindimensionaler Gleichverteilungen,

$$\mathcal{U}(\underline{x}, \underline{\alpha}, \underline{\beta}) = \prod_{p=1}^d \mathcal{U}(x_p, \alpha_p, \beta_p)$$

und somit achsenparallel ausgerichtet. Sie ist linear in jeder einzelnen Dimension, als Ganzes jedoch nicht, wie in Abbildung 2.1 einer zweidimensionalen Gleichverteilung in der seitlichen Ansicht leicht zu erkennen ist.

2.2.2 Gauß-Verteilung

Definition 2.2.5 Für einen Mittelwert $\underline{\mu} \in \mathbb{R}^d$ und eine symmetrische positiv-semidefinite Kovarianzmatrix $\Sigma \in \mathbb{R}^{d \times d}$ ist die Dichte einer d -dimensionalen Gauß-Verteilung gegeben durch

$$\mathcal{N}(\underline{x}, \underline{\mu}, \Sigma) = \frac{1}{\{2\pi\}^{d/2} \cdot |\det(\Sigma)|^{1/2}} \cdot \exp\left(-\frac{1}{2} \cdot \{\underline{x} - \underline{\mu}\}^T \cdot \Sigma^{-1} \cdot \{\underline{x} - \underline{\mu}\}\right) .$$

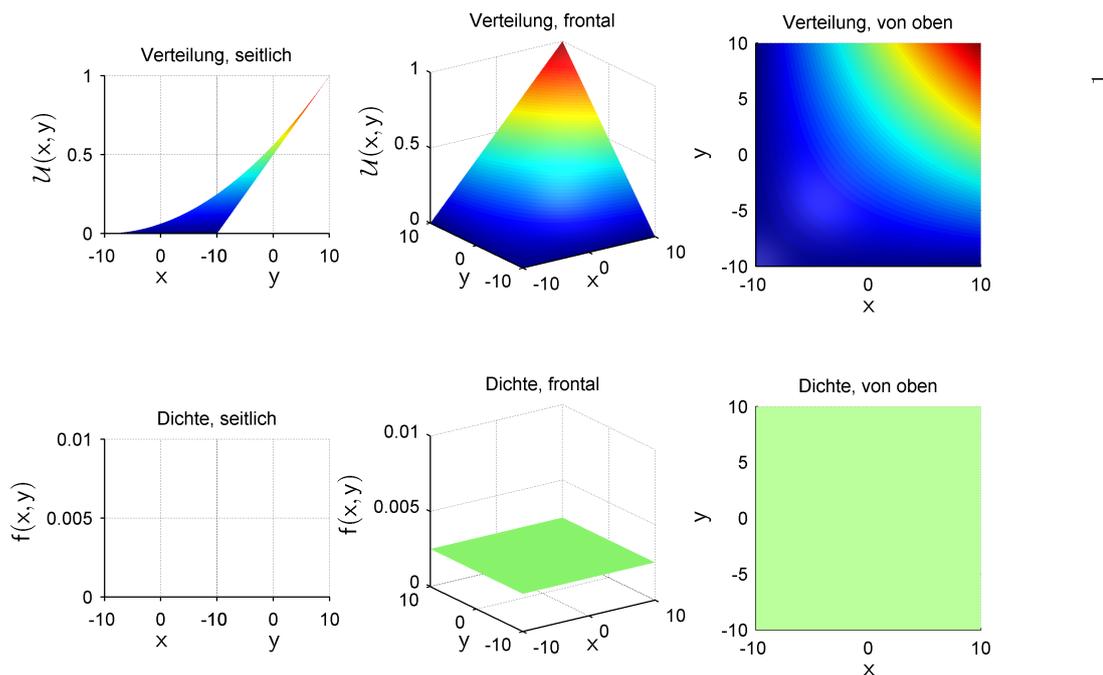


Abbildung 2.1: Zweidimensionale Gleichverteilung auf dem Intervall $[-10, 10]$ und ihre Dichte f . Die obere Zeile zeigt die Verteilungsfunktion aus verschiedenen Perspektiven. Darunter ist jeweils ihre Dichte aus der gleichen Perspektive zu sehen.

Ihre Verteilungsfunktion wird mit $\Phi(\underline{x}, \underline{\mu}, \Sigma)$ bezeichnet. Die Dichte

$$\mathcal{N}(\underline{x}, \underline{\mu}, \underline{\sigma}) := \mathcal{N}(\underline{x}, \underline{\mu}, \Sigma) = \prod_{q=1}^d \mathcal{N}(x_q, \mu_q, \sigma_q)$$

einer achsenparallelen Gauß-Verteilung mit Mittelwert $\underline{\mu}$ und Kovarianzmatrix $\Sigma = \text{diag}(\underline{\sigma})$, $\underline{\sigma} \in \mathbb{R}^d$ mit $\sigma_q > 0$, ist das Produkt eindimensionaler Gauß-Dichten. Ihre Verteilungsfunktion

$$\Phi(\underline{x}, \underline{\mu}, \underline{\sigma}) := \Phi(\underline{x}, \underline{\mu}, \Sigma) = \prod_{q=1}^d \Phi(x_q, \mu_q, \sigma_q)$$

ist wiederum das Produkt eindimensionaler Gauß-Verteilungsfunktionen. $\underline{\sigma}$ ist der Vektor der Standardabweichungen in Richtung der Hauptachsen. \square

Abbildung 2.2 zeigt eine zweidimensionale Standardnormalverteilung, d.h. eine achsenparallele Gauß-Verteilung mit $\underline{\mu} = \underline{0}$ und $\underline{\sigma} = \underline{1}$.

2.2.3 Dirac-Verteilung

Definition 2.2.6 Die Heaviside-Funktion ist im Eindimensionalen definiert als

$$H : \mathbb{R} \rightarrow \mathbb{R}, H(x) := \begin{cases} 0 & x < 0 \\ \frac{1}{2} & x = 0 \\ 1 & x > 0 \end{cases} .$$

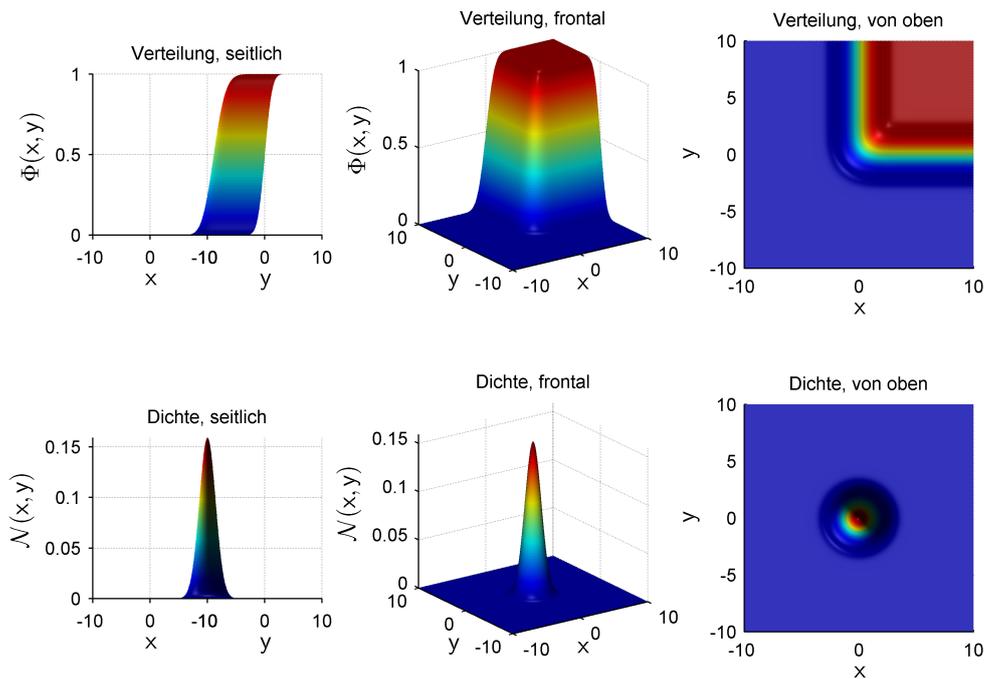


Abbildung 2.2: Zweidimensionale Standardnormalverteilung und ihre Dichte. Die obere Zeile zeigt die Verteilungsfunktion aus verschiedenen Perspektiven. Darunter ist jeweils ihre Dichte aus der gleichen Perspektive zu sehen.

Im Mehrdimensionalen ist die Heaviside-Funktion

$$H : \mathbb{R}^d \rightarrow \mathbb{R}, \quad H(\underline{x}) := \prod_{p=1}^d H(x_p)$$

das Produkt eindimensionaler Heaviside-Funktionen an den Stellen der Einzelkomponenten. \square

Wie in [12] gilt:

Lemma 2.2.7 Die Dirac-Funktion kann für jedes $x \in \mathbb{R}$ als Grenzwert

$$\delta(x) = \lim_{\sigma \rightarrow \infty} \mathcal{N}(x, 0, \sigma)$$

beziehungsweise

$$\delta(\underline{x}) = \lim_{\sigma \rightarrow \infty} \prod_{p=1}^d \mathcal{N}(x_p, 0, \sigma)$$

für jedes $\underline{x} \in \mathbb{R}^d$ mit Hilfe der Gauß-Dichtefunktion \mathcal{N} beschrieben werden. Die in Definition 2.2.6 verwendete Festlegung $H(0) = \frac{1}{2}$ erlaubt andererseits für jedes $x \in \mathbb{R}$ die Grenzwertdarstellung

$$H(x) = \lim_{\sigma \rightarrow \infty} \Phi(x, 0, \sigma) \quad (2.2)$$

der Heaviside-Funktion mit Hilfe der Gauß-Verteilungsfunktion Φ und im Mehrdimensionalen für jedes $\underline{x} \in \mathbb{R}^d$

$$H(\underline{x}) = \lim_{\sigma \rightarrow \infty} \prod_{p=1}^d \Phi(x_p, 0, \sigma) . \quad (2.3)$$

In diesem Sinne ist die Heaviside-Funktion die Verteilungsfunktion der Dirac-Funktion. \blacksquare

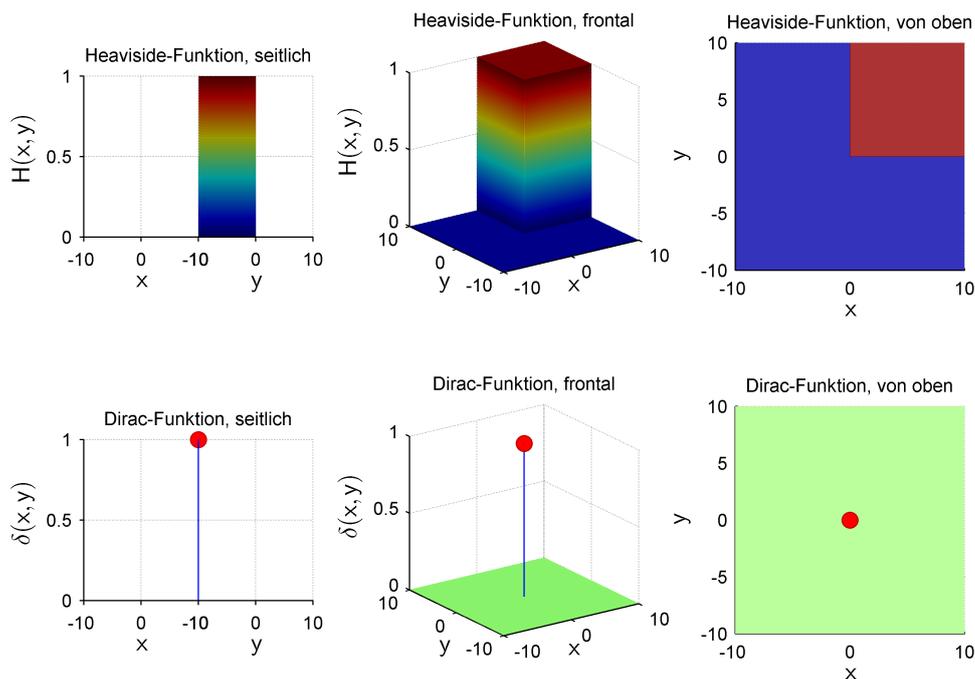


Abbildung 2.3: Zweidimensionale Heaviside- und Dirac-Funktion. Die erste Zeile zeigt die Heaviside-Funktion, die zweite die Dirac-Funktion in Form eines Markers, jeweils aus verschiedenen Perspektiven.

Abbildung 2.3 zeigt die zweidimensionale Heaviside-Funktion zusammen mit ihrer Dichte in Form der Dirac-Funktion.

2.2.4 Mischverteilungen

Definition 2.2.8 Sind $f_l(\underline{x})$, $l = 1, \dots, L$, Dichten mit Verteilungsfunktionen $F_l(\underline{x})$ und $w_l \in \mathbb{R}^+$ nicht negative Gewichte, dann ist

$$f(\underline{x}) := \sum_{l=1}^L w_l f_l(\underline{x})$$

wieder eine Dichte mit Verteilungsfunktion

$$F(\underline{x}) = \sum_{l=1}^L w_l F_l(\underline{x}) .$$

Die zugehörige Verteilung wird als Mischverteilung bezeichnet und $f(\underline{x})$ selbst als Mischdichte. Die Dichten $f_l(\underline{x})$ sind die Komponenten der Mischdichte $f(\underline{x})$. □

Eine spezielle Form der Mischverteilung ist die *Dirac-Mischverteilung*. Dirac-Mischdichten nehmen in dieser Arbeit eine zentrale Rolle als Approximationsdichten ein. Jede Komponente einer Dirac-Mischdichte

$$f(\underline{x}) := \sum_{l=1}^L w_l \delta(\underline{x} - \underline{x}_l)$$

ist ein Dirac-Impuls. Ihre Verteilungsfunktion

$$F(\underline{x}) = \sum_{l=1}^L w_l H(\underline{x} - \underline{x}_l)$$

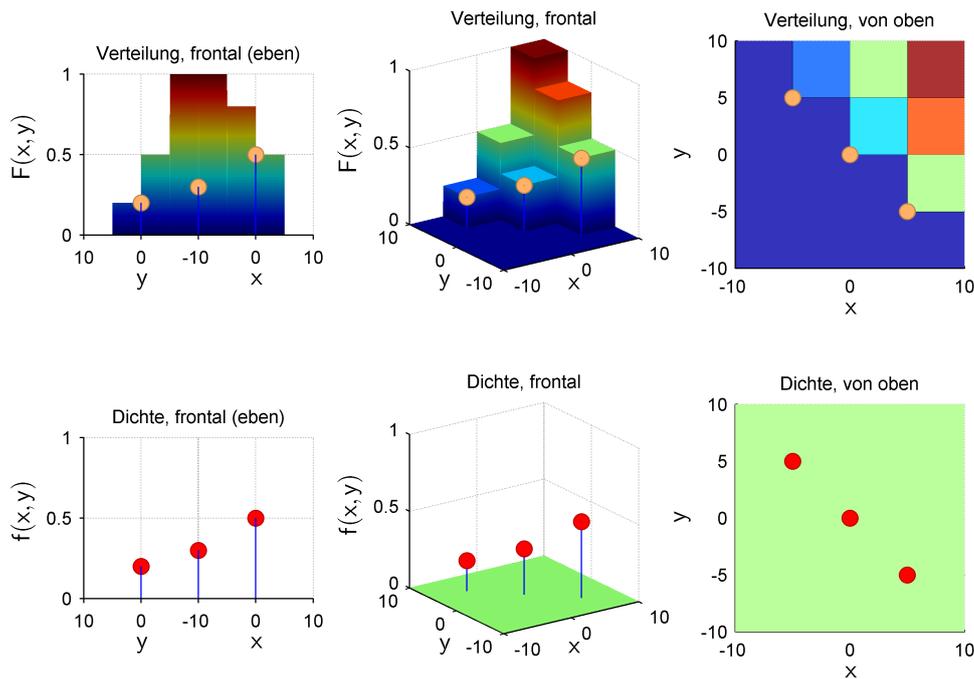


Abbildung 2.4: Dirac-Mischverteilung mit drei Dirac-Impulsen. Die obere Zeile zeigt die Verteilungsfunktion aus verschiedenen Perspektiven. Die Stellen, an denen sich eine Heaviside-Komponente befindet, sind markiert. Deutlich zu erkennen sind die Sprungstellen an den Überschneidungsbereichen der Heaviside-Funktionen (der grüne Bereich oben sowie der orangene und rote Bereich in der Ansicht von oben). Unterhalb der Verteilungsfunktion ist jeweils die Dichte aus der gleichen Perspektive zu sehen.

ist eine Heaviside-Mischfunktion aus mehreren linear verschobenen Heaviside-Komponenten. Die Gewichte $w_l \geq 0$ zusammen mit den Positionen \underline{x}_l der Dirac-Impulse bilden den Parametervektor der Dichte $f(\underline{x})$ beziehungsweise der Verteilungsfunktion $F(\underline{x})$. Abbildungen 2.4 zeigt eine Dirac-Mischverteilung mit drei Komponenten.

Neben der Dirac-Mischverteilung ist die *Gauß-Mischverteilung* von besonderem Interesse. Wie der Name schon andeutet, ist jede Komponente der Dichte einer Gauß-Mischverteilung die Dichte einer Gauß-Verteilung. *Achsenparallele Gauß-Mischdichten* finden auf Grund ihrer Einfachheit verbunden mit ihrer Ausdrucksmächtigkeit in vielen Bereichen der Systemtheorie Verwendung und sollen in dieser Arbeit als Originaldichten (siehe nächster Abschnitt) untersucht werden. In einer achsenparallelen Gauß-Mischdichte ist jede Komponente achsenparallel ausgerichtet entsprechend Definition 2.2.5.

2.3 Reihendarstellung und Approximation

Definition 2.3.1 Ein Funktionenraum \mathfrak{F} ist einen Vektorraum, dessen Elemente Funktionen sind. \mathfrak{F} ist vollständig, falls jede Linearkombination

$$\bar{f} = \sum_{l=1}^{\infty} w_l f_l$$

mit Gewichten w_l aus dem \mathfrak{F} zu Grunde liegenden Körper und Funktionen f_l aus \mathfrak{F} wieder in \mathfrak{F} liegt. \bar{f} heißt Reihenentwicklung von $\tilde{f} \in \mathfrak{F}$, falls

$$\bar{f} = \tilde{f}$$

gilt.

Sei $\mathcal{B} \subseteq \mathfrak{F}$ eine Menge von Grundfunktionen. Besitzt \mathcal{B} die Eigenschaft, dass jede Funktion $\tilde{f} \in \mathfrak{F}$ eine eindeutige Darstellung als Linearkombination von Funktionen aus \mathcal{B} besitzt, dann wird \mathcal{B} auch Menge von Basisfunktionen genannt. \square

Oft wird schon dann von einer Reihenentwicklung gesprochen, wenn \bar{f} bis auf einer Nullmenge mit \tilde{f} übereinstimmt. Im Folgenden wird vorausgesetzt, dass der Funktionenraum eine Menge von Dichten ist. Wird die Reihenentwicklung \bar{f} einer Dichte \tilde{f} entsprechend Definition 2.3.1 nach endlich vielen Gliedern $L \in \mathbb{N}$ abgebrochen,

$$f := \sum_{l=1}^L w_l f_l ,$$

so stimmen f und \bar{f} im Allgemeinen nicht mehr überein. f kann der unendlichen Reihenentwicklung dennoch nahe kommen,

$$f \approx \bar{f} = \tilde{f} ,$$

wenn hinreichend viele Glieder verwendet werden.

Allgemeiner, auch ohne die Voraussetzung eines Funktionenraums oder einer Darstellung als Reihenentwicklung, sei folgende Definition gegeben.

Definition 2.3.2 Ist \tilde{f} eine „gegebene“ Dichte aus einer Menge \mathcal{O} von Dichten und f eine „frei gewählte“ Dichte aus einer vorgegebenen Menge \mathcal{A} von Dichten mit $f \approx \tilde{f}$, dann heißt f eine Approximation von \tilde{f} . Die Verteilung von f wird als Approximationsverteilung der Verteilung von \tilde{f} bezeichnet. \tilde{f} und ihre Verteilung werden auch Originaldichte und Originalverteilung oder einfach kurz Original genannt. \square

Diese Definition lässt offen, wann $f \approx \tilde{f}$ gilt. Tatsächlich wird oft jede Dichte aus \mathcal{A} als Approximation von \tilde{f} betrachtet, wobei einige der Dichten aus \mathcal{A} das Original \tilde{f} besser approximieren als andere. „Besser“ kann mathematisch durch ein Gütemaß beschrieben werden:

Definition 2.3.3 Ein Gütemaß $G : \mathcal{O} \times \mathcal{A} \rightarrow \mathbb{R}^+$ ist eine nicht negative Abbildung und bestimmt den Abstand von f zu \tilde{f} . $G(\tilde{f}, f)$ sagt aus, wie gut die Approximation f das Original \tilde{f} approximiert. Je kleiner $G(\tilde{f}, f)$ ist, desto besser approximiert f die Originaldichte \tilde{f} . Der Funktionswert $G(\tilde{f}, f)$ des Gütemaßes wird auch Güte genannt. \square

G muss nicht notwendigerweise die Kriterien einer Metrik — wie etwa eine Dreiecksungleichung — erfüllen. In Abhängigkeit von der konkreten Anwendung können verschiedene Gütemaße gewählt werden. Der Abstand im quadratischen Mittel zwischen zwei Funktionen ist sehr anschaulich und wird oft benutzt, um einen Abstand zwischen zwei Funktionen zu messen. Eine

wichtige Anforderung in dieser Arbeit ist, dass das Gütemaß für Dirac-Mischdichten definiert ist. Während der Abstand im quadratischen Mittel zwischen den Dichten in diesem Fall Probleme bereitet, ist er auf den Verteilungsfunktionen wohldefiniert und berücksichtigt die Dichten indirekt. In dieser Arbeit wird

$$G(\tilde{f}, f) := \int_{\underline{c}_1}^{\underline{c}_2} \left\{ \tilde{F}(\underline{x}) - F(\underline{x}) \right\}^2 d\underline{x} \quad (2.4)$$

als Gütemaß verwendet für $\underline{c}_1, \underline{c}_2 \in \mathbb{R}^d$ mit $\underline{c}_1 \leq \underline{c}_2$. Die Begrenzung des Integrationsbereichs erweist sich als hilfreich in der mathematischen Analyse.

2.4 Approximation formuliert als Optimierungsproblem

Sei \mathcal{A} eine Menge von Approximationsdichten entsprechend Definition 2.3.2 und \tilde{f} eine zu approximierende Originaldichte. Eine bestmögliche Approximation f_{\min} aus \mathcal{A} für \tilde{f} entsprechend einem Gütemaß G zu bestimmen, entspricht der Berechnung von

$$f_{\min} = \operatorname{argmin}_{f \in \mathcal{A}} (G(\tilde{f}, f)) \quad (2.5)$$

Dabei kann es mehrere Approximationen f_{\min} aus \mathcal{A} geben, so dass $G(\tilde{f}, f_{\min})$ minimal wird.

In dieser Arbeit sollen Dirac-Mischdichten als Approximationsdichten betrachtet werden. Zu einer Dirac-Mischdichte

$$f(\underline{x}, \underline{\eta}) := \sum_{l=1}^L w_l \delta(\underline{x} - \underline{x}_l)$$

ist $\underline{\eta} := (w_1, \dots, w_L, \underline{x}_1^T, \dots, \underline{x}_L^T)^T$ der Parametervektor dieser Dichte, bestehend aus den Gewichten w_l und Positionen \underline{x}_l der einzelnen Dirac-Impulse. Der Parametervektor $\underline{\eta}$ legt die Dirac-Mischdichte eindeutig — wenn auch nicht eineindeutig — fest. Daher wird $\underline{\eta}$ mitunter auch mit der Dichte $f(\underline{x}, \underline{\eta})$ identifiziert: „die Approximation $\underline{\eta}$ “ meint dann die Dichte $f(\underline{x}, \underline{\eta})$ betrachtet als Approximationsdichte für eine Originaldichte \tilde{f} .

Ist \tilde{f} , beziehungsweise ihre Verteilung \tilde{F} , fest vorgegeben, so kann das Gütemaß aus (2.4) in Abhängigkeit vom Parametervektor $\underline{\eta}$ geschrieben werden als

$$G : \mathbb{R}^{+L} \times \mathbb{R}^{L \cdot d} \rightarrow \mathbb{R}^+, \quad G(\underline{\eta}) := \int_{\underline{c}_1}^{\underline{c}_2} \left\{ \tilde{F}(\underline{x}) - F(\underline{x}, \underline{\eta}) \right\}^2 d\underline{x} \quad (2.6)$$

für $\underline{c}_1, \underline{c}_2 \in \mathbb{R}^d$ mit $\underline{c}_1 \leq \underline{c}_2$. (2.5) lautet dann

$$\underline{\eta}_{\min} = \operatorname{argmin}_{\underline{\eta} \in \mathbb{R}^{+L} \times \mathbb{R}^{L \cdot d}} (G(\underline{\eta})) \quad .$$

Umgangssprachlich formuliert besteht die Aufgabe, eine bestmögliche Approximation für \tilde{f} zu bestimmen darin, eine Stelle des globalen Minimums von G zu finden — also in einem Optimierungsproblem für die Funktion G .

Definition 2.4.1 $\underline{\eta} \in \mathbb{R}^{+L} \times \mathbb{R}^{L \cdot d}$ ist die Stelle eines lokalen Minimums von G , falls es eine Umgebung U von $\underline{\eta}$ gibt, so dass für jedes $\underline{\xi} \in U$ $G(\underline{\xi}) \geq G(\underline{\eta})$ gilt. Die zu $\underline{\eta}$ gehörige Approximationsdichte $f(\underline{x}, \underline{\eta})$ wird in diesem Fall als lokal optimale Approximation bezeichnet. Gilt

sogar $U = \mathbb{R}^{+L} \times \mathbb{R}^{L \cdot d}$, so ist $\underline{\eta}$ eine globale Minimalstelle und $f(\underline{x}, \underline{\eta})$ wird auch global optimale Approximation genannt.

Eine Approximation $f(\underline{x}, \underline{\eta})$ ist suboptimal, falls die Güte $G(\underline{\eta})$ zwar „nicht weit“ vom globalen Minimum von G entfernt, $\underline{\eta}$ aber nicht notwendigerweise eine lokale oder gar globale Minimalstelle ist. \square

„Nicht weit“ wird dabei nicht näher festgelegt und ergibt sich gegebenenfalls aus dem Kontext. Mit lokal, global und sub-optimaler Approximation werden auch die Verteilung von $f(\underline{x}, \underline{\eta})$ und ihre Verteilungsfunktion $F(\underline{x}, \underline{\eta})$ bezeichnet. „Optimale Approximation“ ohne einen Zusatz von lokal, global oder sub(-optimal) meint in der Regel eine global optimale Approximation.

Stand der Technik

3.1 Stetige Grundfunktionen

Die Verwendung von Dirac-Mischdichten als Approximationsdichten im Kontext der Systemtheorie wird im nachfolgenden Abschnitt motiviert. Da auch viele andere Typen von Funktionen zur Approximation von Dichten in der Literatur genutzt werden, sollen hier einige wichtige Reihendarstellungen kurz genannt werden, die wie in Abschnitt 2.3 beschrieben als Approximationsfunktionen eingesetzt werden können.

Die *Fourierreihenentwicklung* nutzt Sinus- und Kosinus-Funktionen als Grundfunktionen. Über die Fouriertransformation steht sie im engen Zusammenhang mit der Dirac-Funktion. Fast alle praktisch relevanten Funktionen besitzen eine Fourierreihenentwicklung (siehe zum Beispiel [9]). Zu den Standardwerkzeugen der Analysis zählt die *Taylor-Entwicklung*, die als Grundfunktionen Polynome und für die Koeffizienten die Ableitungen der zu approximierenden Funktion im Entwicklungspunkt benutzt ([8]). Ein Vergleich der orthogonalen *Edgeworth-Entwicklung* mit *Gauß-Mischdichten* sowie den allgemeineren *Exponentialdichten* hinsichtlich ihrer Universalität, der Eigenschaft einer Dichte sowie der Berechenbarkeit von Parametern aus Momenten und umgekehrt ist in [12] zu finden.

Für eine ausführliche Einführung sowie die spezifischen Vor- und Nachteile der angeführten Reihendarstellungen sei auf die angeführten Literaturhinweise verwiesen.

3.2 Dirac-Impulse als Grundfunktionen

Dieser Abschnitt begründet die Verwendung von Dirac-Mischdichten in der Systemtheorie im Kontext rekursiver Filter. Für eine ausführlichere Darstellung siehe [12].

Ist f_k^e die Zustandsdichte für ein System zum Zeitpunkt k , so hat der Bayes'sche Prädiktions-schritt für ein (im Allgemeinen) nicht lineares, zeitdiskretes und wertekontinuierliches System die Form

$$f_{k+1}^p(\underline{x}_{k+1}) = \int_{\mathbb{R}^d} f_k^e(\underline{x}_k) \cdot f_{k+1}^a(\underline{x}_{k+1} | \underline{x}_k, \underline{u}_k) d\underline{x}_k . \quad (3.1)$$

Dabei ist f_{k+1}^p die Dichte des prädizierten Zustands und f_{k+1}^a die bedingte Dichte für den Zustand zum Zeitpunkt $k+1$ bei gegebenem Zustand \underline{x}_k und Eingabe \underline{u}_k zum Zeitpunkt k . f_{k+1}^a ergibt sich aus der Systemabbildung und der Verteilung des Systemrauschens. Wird f_k^e durch eine Dirac-Mischdichte

$$f(\underline{x}, \underline{\eta}) = \sum_{l=1}^L w_l \delta(\underline{x} - \underline{y}_l)$$

mit $\underline{\eta} := (w_1, \dots, w_L, \underline{y}_1^T, \dots, \underline{y}_L^T)^T$ approximiert, dann gilt mit Hilfe des Transformationssatzes und unter Ausnutzung von $\int g d\mu = \sum_{n=1}^{\infty} \int g d\mu_n$ für eine integrierbare Funktion g und Maße μ_n mit $\mu := \sum_{n=1}^{\infty} \mu_n$ (siehe beispielsweise [7])

$$f_{k+1}^p(\underline{x}_{k+1}) \approx \int_{\mathbb{R}^d} f(\underline{x}_k, \underline{\eta}) \cdot f_{k+1}^a(\underline{x}_{k+1} | \underline{x}_k, \underline{u}_k) d\underline{x}_k = \sum_{l=1}^L w_l f_{k+1}^a(\underline{x}_{k+1} | \underline{y}_l, \underline{u}_k) . \quad (3.2)$$

Für das Gleichheitszeichen wird dabei die Ausblendeigenschaft (2.1) der Dirac-Funktion ausgenutzt. Eine komplizierte analytische oder sehr rechenaufwendige numerische Berechnung des Integrals in (3.1) ist nicht erforderlich. Der Prädiktionsschritt nimmt somit eine besonders einfache Form an.

Genauer betrachtet ist die Darstellung von f_{k+1}^p in (3.2) eine Linearkombination der Dichten

$$g_{kl}(\underline{x}_{k+1}) := f_{k+1}^a(\underline{x}_{k+1} | \underline{y}_l, \underline{u}_k)$$

zu den vorliegenden Eingaben \underline{u}_k und den Positionen \underline{y}_l der Dirac-Impulse der Approximationsdichte. Diese spezielle Form der Darstellung der prädizierten Dichte f_{k+1}^p kann dazu genutzt werden, den Typ der Zustandsdichte während der rekursiven Filterung zu erhalten: Entspricht das Systemrauschen einer Gauß-Mischverteilung, so hat g_{kl} die Form einer Gauß-Mischdichte. Aus (3.2) folgt, dass der prädizierte Zustand zum Zeitpunkt $k+1$ ebenfalls annähernd einer Gauß-Mischverteilung entspricht. Ist das Messrauschen linear und Gauß-mischverteilt, so setzt sich dies fort auf den geschätzten Zustand zum Zeitpunkt $k+1$, f_{k+1}^e ist also ebenfalls eine Gauß-Mischdichte. Wird für den initialen Zustand zum Zeitpunkt $k=1$ eine Gauß-Mischdichte vorausgesetzt, so bleibt diese analytische Form der Schätzdichte während der Rekursion folglich erhalten. Zudem genügt für die Approximation ein Verfahren, das eine Gauß-Mischdichte durch eine Dirac-Mischdichte approximiert.

Die nachfolgenden Unterabschnitte beschreiben verschiedene Verfahren, um Dichten durch Dirac-Mischdichten zu approximieren. Spezifische Vor- und Nachteile sollen dabei zur Sprache kommen.

3.2.1 Simulated Annealing

In Abschnitt 2.4 wurde die Approximation einer Dichte als Optimierungsproblem beschrieben, bei dem eine globale Minimalstelle $\underline{\eta}_{\min}$ der Gütefunktion $G(\underline{\eta})$ zu finden ist. Eine Standardtechnik zum Auffinden von Minimalstellen einer Funktion ist der Simulated Annealing Algorithmus, den es in verschiedenen Varianten gibt. Er gehört zur Klasse der Monte-Carlo-Algorithmen, die Zufallsmomente benutzen, um ein Problem zu lösen. Für einen mathematischen Zugang siehe beispielsweise [4].

Für eine Gauß-Mischverteilung ist das Integral in Definition (2.6) von G im Mehrdimensionalen nicht analytisch lösbar, selbst wenn sie achsenparallel ausgerichtet ist. Die numerische Berechnung ist sehr kostenintensiv. Da der Simulated Annealing Algorithmus typischerweise sehr viele Auswertungen der zu minimierenden Funktion voraussetzt, ist dieser Ansatz für das Mehrdimensionale weniger gut geeignet. Zudem garantiert der Simulated Annealing Algorithmus zwar unter gewissen Voraussetzungen das Auffinden einer globalen Minimalstelle, dazu muss das Abkühlen der Temperatur aber so langsam geschehen, dass diese Aussage mehr einen theoretischen Wert besitzt.

3.2.2 Zufällige Stichproben

Ein weiterer Monte-Carlo-Algorithmus zum Erstellen von Approximationen in Form von Dirac-Mischdichten für eine Originaldichte besteht in der Verwendung zufälliger Stichproben. Dazu werden L Einzelstichproben $\underline{x}_1, \dots, \underline{x}_L$ der Originaldichte erzeugt, welche in der Approximationsdichte dann den Positionen der Dirac-Impulse entsprechen. Die Gewichte für die Dirac-Impulse werden alle gleich groß gewählt, $w_l := \frac{1}{L}$. Je mehr Stichproben benutzt werden, desto besser ist die Approximation. Dahinter verbirgt sich der Satz von Glivenko–Cantelli über die Konvergenz der empirischen Verteilungsfunktion (siehe beispielsweise [7]).

Diese Methode überzeugt insbesondere durch ihre Einfachheit und wird in dieser Hinsicht eigentlich nur noch durch das äquidistante Abtasten der Zustandsdichte übertroffen. Eine Methode zur Erzeugung von Stichproben für eine eindimensionale Verteilung ist die Inversionsmethode. Im Mehrdimensionalen kann sie bei achsenparallelen Verteilungen für jede Achse separat genutzt werden. Für die Normalverteilung werden Korrelationen durch die Multiplikation mit der Cholesky-Zerlegung der Kovarianzmatrix berücksichtigt.

Von Nachteil ist, dass nach dem Satz von Glivenko–Cantelli mit wachsendem L zwar eine Konvergenz der Approximationsverteilung gegen die Originalverteilung erfolgt, diese Konvergenz im Allgemeinen aber relativ langsam ist. Zudem ist die Wahrscheinlichkeit, mit der dieser Ansatz eine globale Minimalstelle von G für ein festgehaltenes L findet, gleich 0, sofern die Menge der globalen Minimalstellen von G eine Nullmenge — also insbesondere eine endliche oder abzählbar unendliche Menge — ist. Gleiches gilt für lokale Minimalstellen von G .

3.2.3 Greedy-Verfahren

In [6] wird ein Greedy-Verfahren zum Erzeugen von Approximationen für Originaldichten in Form von Dirac-Mischdichten im eindimensionalen Fall vorgestellt. Es basiert auf einer optimalen Platzierung eines Dirac-Impulses in einem vorgegebenen Intervall $[c_1, c_2]$ der Originaldichte. Jedem Dirac-Impuls wird das Intervall von seinem linken Nachbarn bis zu seinem rechten Nachbarn zugeordnet. Gibt es keinen Nachbarn, so bildet das initial vorgegebene Intervall $[c_1, c_2]$ die entsprechende Grenze. Das Verfahren ist iterativ und führt nach dem initialen Setzen eines Dirac-Impulses in jeder Iteration im Wesentlichen zwei Schritte durch:

1. Wähle einen Dirac-Impuls D .

Kriterium: Größter Fehler zur Originaldichte auf dem zum Dirac-Impuls gehörigen Intervall I bezüglich des Gütemaßes.

2. Bisektion:

(a) Unterteile I an der Position von D in zwei Teilintervalle I_1 und I_2 mit $I = I_1 \cup I_2$ und $I_1 \cap I_2 = \emptyset$.

(b) Entferne D .

(c) Platziere in jedem der Bereiche I_1 und I_2 je einen neuen Dirac-Impuls.

Kriterium: Optimalität bezüglich des Gütemaßes.

Das Verfahren hat einen geringen Laufzeitaufwand und kann Approximationen mit einer großen Anzahl von Dirac-Impulsen erzeugen. Zudem liegt in jedem Iterationsschritt eine Approximation vor.

Zur Übertragung in das Mehrdimensionale werden die verschiedenen Achsen separat betrachtet. In Schritt 2 wird dann zusätzlich eine Achse ausgewählt, entlang derer die Bisektion statt findet. Da hierbei alle anderen Achsen außer Acht gelassen werden, handelt es sich um einen heuristischen Ansatz.

Die mit diesem Verfahren erzeugten Approximationen sind schon im Eindimensionalen nur suboptimal. Um so mehr gilt dies in der mehrdimensionalen Variante. Wie der Vergleich verschiedener Verfahren in Kapitel 9 zeigt, beginnt die Güte der erzeugten Approximationen ab einer gewissen Anzahl von eingesetzten Dirac-Impulsen für komplizierter geformte Originaldichten zu stagnieren. Auch mit einer höheren Anzahl von Dirac-Impulsen wird also keine wesentlich bessere Approximation erreicht.

3.2.4 Progressives Bayes

Damit G an der Stelle $\underline{\eta}_{\min}$ ein globales Minimum annimmt, also

$$\underline{\eta}_{\min} = \operatorname{argmin}_{\underline{\eta}}(G(\underline{\eta}))$$

gilt, muss der Gradient $\frac{\partial}{\partial \underline{\eta}} G(\underline{\eta})$ von G an der Stelle $\underline{\eta}_{\min}$ gleich Null sein,

$$\frac{\partial}{\partial \underline{\eta}} G(\underline{\eta}_{\min}) \stackrel{!}{=} 0 \quad . \quad (3.3)$$

Die Differenzierbarkeit von G nach $\underline{\eta}$ wird hier vorausgesetzt. Diese notwendige, jedoch nicht hinreichende Bedingung wird auch von lokalen Minimalestellen erfüllt, von denen es im Allgemeinen leider viele gibt. In der Regel kann $\underline{\eta}_{\min}$ nicht analytisch aus (3.3) bestimmt werden.

In [5] wird für beliebige Approximationsdichten im Allgemeinen und für Gauß-Mischdichten als Approximationsdichten im Konkreten dargelegt, wie $\underline{\eta}_{\min}$ progressiv bestimmt werden kann.

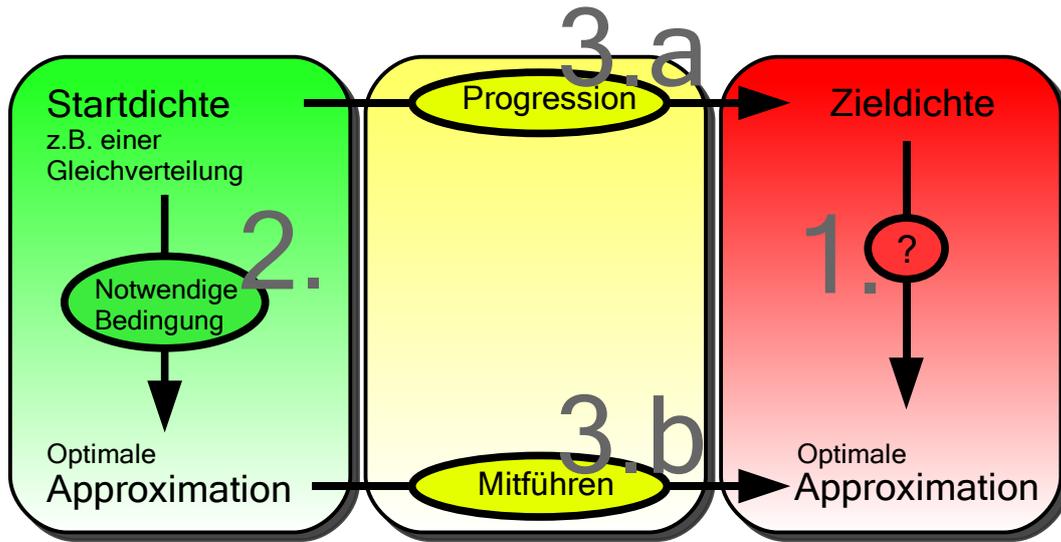


Abbildung 3.1: Schematische Darstellung des Progressiven Bayes-Verfahrens. Gesucht wird eine optimale Approximation für eine gegebene Zieldichte, die allerdings nicht direkt bestimmt werden kann (1.). Begonnen wird daher mit einer einfachen Startdichte, für die eine optimale Approximation über die notwendige Bedingung bestimmt wird (2.). Anschließend wird die Startdichte in die Zieldichte überführt (3.a) unter gleichzeitiger Mitführung und Anpassung der optimalen Approximation der Startdichte (3.b).

In [10] und [11] wird dieses Verfahren für den Fall von Dirac-Mischdichten als Approximationsdichten präsentiert. Die drei Arbeiten beschränken sich dabei auf den eindimensionalen Fall.

Die Idee des Progressiven Bayes-Verfahrens ist (siehe Abbildung 3.1), mit einer einfachen Startdichte \tilde{f}_0 zu beginnen, für die eine optimale Approximation in Form einer Stelle η_{\min}^0 des globalen Minimums von G bekannt ist. Ist die Startdichte einfach genug, so kann η_{\min}^0 aus der notwendigen Bedingung (3.3) bestimmt werden. Im Fall von Dirac-Mischdichten als Approximationsdichten gilt dies im Eindimensionalen beispielsweise für die Familie der Gleichverteilungen.

Die Startdichte wird anschließend progressiv in die Zieldichte = Originaldichte \tilde{f} überführt. Während dieser Überführung wird die optimale Approximation der Startdichte mitgeführt und den Änderungen der progressiven Dichte angepasst, so dass diese optimal bleibt. Am Ende der progressiven Überführung der Start- in die Zieldichte liegt dann eine optimale Approximation für die Zieldichte vor.

Formal wird dazu ein Progressionsparameter $\gamma \in [0, 1]$ eingeführt und eine progressive Dichte definiert:

Definition 3.2.1 Gegeben sei eine zu approximierende Originaldichte $\tilde{f} : \mathbb{R}^d \rightarrow \mathbb{R}^+$ sowie eine Startdichte \tilde{f}_0 . Dann heißt die Funktion $\tilde{f} : \mathbb{R}^d \times [0, 1] \rightarrow \mathbb{R}$ progressive Dichte mit Startdichte \tilde{f}_0 und Zieldichte \tilde{f} , falls die folgenden Bedingungen erfüllt sind:

1. Für jedes $\gamma \in [0, 1]$ ist $\tilde{f}(\underline{x}, \gamma)$ — betrachtet als Funktion von \underline{x} — eine Dichte.

2. Es gilt $\tilde{f}(\underline{x}, 0) = \tilde{f}_0(\underline{x})$ für jedes $\underline{x} \in \mathbb{R}^d$.

3. Es gilt $\tilde{f}(\underline{x}, 1) = \tilde{f}(\underline{x})$ für jedes $\underline{x} \in \mathbb{R}^d$.

Die progressive Dichte wird auch als progressive Version der Originaldichte \tilde{f} betrachtet. Der Vorgang, die Startdichte mit Hilfe der progressiven Dichte in die Zieldichte zu überführen, wird Progression genannt. Progression und progressive Dichte werden mitunter synonym benutzt.

Auf Seiten der Verteilungsfunktion wird

$$\tilde{F} : \mathbb{R}^d \times [0, 1] \rightarrow \mathbb{R}^+, \quad \tilde{F}(\underline{x}, \gamma) := \int_{-\infty}^{\underline{x}} \tilde{f}(\underline{t}, \gamma) d\underline{t}$$

als progressive Verteilungsfunktion mit Startverteilungsfunktion $\tilde{F}_0(\underline{x})$ und Zielverteilungsfunktion $\tilde{F}(\underline{x})$ bezeichnet, analog wird von der progressiven Verteilung bezüglich der beteiligten Verteilungen gesprochen.

Mit der progressiven Dichte zum Progressionsschritt/Progressionszeitpunkt γ ist die Dichte $\tilde{f}(\underline{x}, \gamma)$ — betrachtet als Funktion von \underline{x} — gemeint, wie zuvor werden analoge Sprechweisen für die Verteilung und die Verteilungsfunktion verwendet.

□

Die erste Forderung für die progressive Dichte kann gegebenenfalls abgeschwächt werden, so dass beispielsweise auch Funktionen mit negativen Funktionswerten während der Progression erlaubt sind. Dann kann allerdings nicht ohne Weiteres von der progressiven Verteilung gesprochen werden.

Neben der Startdichte zusammen mit ihrer Approximation in Form von $\underline{\eta}_{\min}^0$ ist die Wahl der Progression ein freier Parameter des Progressiven Bayes-Verfahrens zur Approximation einer vorgegebenen Zieldichte. D.h. Startdichte und ihre Approximation $\underline{\eta}_{\min}^0$ sowie die progressive Dichte können beliebig gewählt werden, solange die progressive Dichte tatsächlich die Startdichte in die vorgegebene Zieldichte überführt und $\underline{\eta}_{\min}^0$ (3.3) für die Startdichte erfüllt. Die Approximation einer Zieldichte mit Hilfe des Progressiven Bayes-Verfahrens wird auch als *progressive Approximation der Zieldichte* bezeichnet.

Um die optimale Startapproximation in Form der globalen Minimalstelle $\underline{\eta}_{\min}^0$ von G während der Progression mitführen zu können, wird eine progressive Version des Gütemaßes benötigt.

Definition 3.2.2 Für $\underline{c}_1, \underline{c}_2 \in \mathbb{R}^d$ mit $\underline{c}_1 \leq \underline{c}_2$ wird das (2.6) entsprechende progressive Gütemaß durch

$$G(\underline{\eta}, \gamma) := \int_{\underline{c}_1}^{\underline{c}_2} \left\{ \tilde{F}(\underline{x}, \gamma) - F(\underline{x}, \underline{\eta}) \right\}^2 d\underline{x} \quad (3.4)$$

definiert.

□

Bezeichnet $\underline{\eta}_{\min}(\gamma)$ für ein festgehaltenes γ die Stelle des globalen Minimums von $G(\underline{\eta}, \gamma)$ — betrachtet als Funktion von $\underline{\eta}$ —, dann lautet (3.3)

$$\frac{\partial}{\partial \underline{\eta}} G(\underline{\eta}_{\min}(\gamma), \gamma) \stackrel{!}{=} 0 . \quad (3.5)$$

Die Funktion $\underline{\eta}_{\min}(\gamma)$ ist durch diese Gleichung implizit als eine Lösung von

$$\frac{\partial}{\partial \underline{\eta}} G(\underline{\eta}, \gamma) \stackrel{!}{=} 0 \quad (3.6)$$

für die von γ abhängige Funktion $\underline{\eta} = \underline{\eta}(\gamma)$ gegeben. Der Satz über implizit definierte Funktionen liefert dann unter geeigneten Voraussetzungen an $\tilde{F}(\underline{x}, \gamma)$ die Existenz einer Lösung $\underline{\eta}(\gamma)$ von (3.6) mit $\underline{\eta}(0) = \underline{\eta}_{\min}^0$ als Startwert. Die Lösung ist nicht unbedingt eindeutig, was anschaulich schon deshalb klar ist, da (3.5) nur eine notwendige, aber keine hinreichende Bedingung für ein globales Minimum für ein festgehaltenes γ darstellt.

Bestimmen einer Lösung

Wie kann nun eine Lösung von (3.6) bestimmt werden? Als erstes sei festgehalten, dass (3.6) auch unter infinitesimal kleinen Änderungen von γ erhalten bleibt,

$$\frac{\partial}{\partial \gamma} \left\{ \frac{\partial}{\partial \underline{\eta}} G(\underline{\eta}, \gamma) \right\} \stackrel{!}{=} 0 . \quad (3.7)$$

Das ergibt sich sofort aus der Grenzwertdarstellung der Ableitung nach γ , da (3.6) für jedes γ aus $[0, 1]$ gefordert wird. Aus (3.7) wird dann ein äquivalentes, nicht lineares gewöhnliches Differentialgleichungssystem

$$\mathbf{P} \cdot \underline{\dot{\eta}} \stackrel{!}{=} \underline{b} \quad (3.8)$$

erster Ordnung aufgestellt. Dabei ist \mathbf{P} eine quadratische Matrix, deren Zeilen- und Spaltenanzahl der Länge von $\underline{\eta}$ entspricht. \mathbf{P} hängt von $\underline{\eta}$ und γ , jedoch nicht von $\underline{\dot{\eta}}$ oder höheren Ableitungen von $\underline{\eta}$ ab. Das Differentialgleichungssystem ist also explizit. Als Startwert ist $\underline{\eta}(0) = \underline{\eta}_{\min}^0$ vorgegeben.

Das Differentialgleichungssystem ist zwar nicht analytisch, dafür aber numerisch lösbar. Algorithmus 1 zeigt den in [11] verwendeten Algorithmus. Er geht progressiv vor und ist unterteilt in einen Prädiktor- sowie einen Korrektorschritt mit adaptiver Schrittweitenkontrolle für den Prädiktorschritt.

Prädiktorschritt Der Prädiktorschritt setzt den Funktionswert $\underline{\eta}(\gamma_{\text{alt}})$ als bekannt für ein gegebenes γ_{alt} voraus und erstellt daraus eine „Prädiktion“ $\underline{\eta}_{\text{tmp}}$ für den Funktionswert der Funktion $\underline{\eta}$ zum Progressionsschritt $\gamma_{\text{neu}} := \gamma_{\text{alt}} + \Delta\gamma$. Dabei ist $\Delta\gamma$ die Schrittweite. Konkret wird die Prädiktion über den Gradienten der Funktion $\underline{\eta}$ an der Stelle γ_{alt} berechnet,

$$\underline{\eta}_{\text{tmp}} := \underline{\eta}(\gamma_{\text{alt}}) + \Delta\gamma \cdot \underline{\dot{\eta}}(\gamma_{\text{alt}}) \approx \underline{\eta}(\gamma_{\text{neu}}) .$$

$\underline{\dot{\eta}}(\gamma_{\text{alt}})$ wird dazu aus (3.8) ermittelt. Dies ist möglich, da die Werte γ_{alt} und $\underline{\eta}(\gamma_{\text{alt}})$ nach Voraussetzung bekannt sind, und \mathbf{P} sowie \underline{b} in (3.8) daher konkrete Zahlenwerte enthalten. (3.8) stellt in diesem Fall somit ein *Lineares Gleichungssystem* mit den Vektor $\underline{y} := \underline{\dot{\eta}}$ der unbekanntenen Größen dar und kann nach \underline{y} aufgelöst werden, woraus sich dann $\underline{\dot{\eta}}(\gamma_{\text{alt}}) := \underline{y}$ ergibt.

Algorithmus 1 Predictor–corrector method with step–size control for parameter tracking.

```

1:  $\gamma := 0$ 
2:  $\underline{\eta}_0 := \underline{\eta}(\gamma = 0)$  //  $G(\underline{\eta}_0, \gamma = 0) = \underline{0}$ 
3:  $\Delta\gamma := \gamma_{\text{step\_min}}$ 
4: repeat
5:    $\gamma := \gamma + \Delta\gamma$ 
6:    $\underline{\eta}_{\text{tmp}} := \text{Predictor}(\underline{\eta}, \gamma)$ 
7:    $\left[ \underline{\eta}_{\text{tmp}}, \text{success} \right] := \text{Corrector}(\underline{\eta}_{\text{tmp}})$ 
8:   if success then
9:      $\underline{\eta} := \underline{\eta}_{\text{tmp}}$ 
10:    Increase( $\Delta\gamma, \gamma_{\text{step\_max}}$ )
11:   else
12:      $\gamma := \gamma - \Delta\gamma$ 
13:     Decrease( $\Delta\gamma, \gamma_{\text{step\_min}}$ )
14:   end if
15: until  $\gamma = 1$ 

```

Korrektorschritt Ziel des Korrektorschritts ist die Verbesserung der Prädiktion $\underline{\eta}_{\text{tmp}}$ aus dem vorangegangenen Prädiktorschritt. Sei

$$g : \mathbb{R}^{+L} \times \mathbb{R}^{L \cdot d} \rightarrow \mathbb{R}^+$$

$$g(\underline{\xi}) := \frac{\partial}{\partial \underline{\xi}} G(\underline{\xi}, \gamma_{\text{neu}}) = \left(\frac{\partial}{\partial \xi_1} G(\underline{\xi}, \gamma_{\text{neu}}), \dots, \frac{\partial}{\partial \xi_{L \cdot \{d+1\}}} G(\underline{\xi}, \gamma_{\text{neu}}) \right) .$$

Da $\underline{\eta}$ eine Lösung von (3.6) sein soll, gilt für die Funktion g an der Stelle $\underline{\eta}(\gamma_{\text{neu}})$

$$g(\underline{\eta}(\gamma_{\text{neu}})) \stackrel{!}{=} 0$$

Weil $\underline{\eta}_{\text{tmp}}$ ungefähr gleich $\underline{\eta}(\gamma_{\text{neu}})$ gilt, muss g in der Nähe von $\underline{\eta}_{\text{tmp}}$ eine Nullstelle besitzen, nämlich $\underline{\eta}(\gamma_{\text{neu}})$ selbst. Eine Nullstelle in der Nähe von $\underline{\eta}_{\text{tmp}}$ kann mit Hilfe des Newtonverfahrens bestimmt werden.

Dazu wird im Fall einer eindimensionalen Funktion g_1 die Zahlenfolge

$$\xi_{k+1} := \xi_k - \frac{g_1(\xi_k)}{\frac{\partial}{\partial \xi_k} g_1(\xi_k)} \quad (3.9)$$

definiert. Unter geeigneten Voraussetzungen konvergiert die Zahlenfolge (ξ_k) gegen eine Nullstelle, siehe zum Beispiel [8]. Zu (3.9) ist

$$\frac{\partial}{\partial \xi_k} g_1(\xi_k) \cdot \{\xi_{k+1} - \xi_k\} = -g_1(\xi_k)$$

äquivalent, sofern $\frac{\partial}{\partial \xi_k} g_1(\xi_k) \neq 0$ vorausgesetzt wird. In dieser Form ist die Formel auch im mehrdimensionalen Fall verwendbar,

$$\dot{g}(\underline{\xi}_k) \cdot \left\{ \underline{\xi}_{k+1} - \underline{\xi}_k \right\} = -g(\underline{\xi}_k) .$$

Da g ein Vektorfeld ist, ist ihre Ableitung definiert über die Jakobi-Matrix

$$\dot{g}(\underline{\xi}) = \mathbf{J}(\underline{\xi}) = \begin{pmatrix} \frac{\partial}{\partial \xi_1} \frac{\partial}{\partial \xi_1} G(\underline{\xi}, \gamma_{\text{neu}}) & \cdots & \frac{\partial}{\partial \xi_{L-\{d+1\}}} \frac{\partial}{\partial \xi_1} G(\underline{\xi}, \gamma_{\text{neu}}) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial \xi_1} \frac{\partial}{\partial \xi_{L-\{d+1\}}} G(\underline{\xi}, \gamma_{\text{neu}}) & \cdots & \frac{\partial}{\partial \xi_{L-\{d+1\}}} \frac{\partial}{\partial \xi_{L-\{d+1\}}} G(\underline{\xi}, \gamma_{\text{neu}}) \end{pmatrix}. \quad (3.10)$$

Beginnend mit $k = 0$ und $\underline{\xi}_0 := \underline{\eta}_{\text{tmp}}$ ergibt sich $\underline{\xi}_{k+1}$ nun durch Lösen des Linearen Gleichungssystems

$$\dot{g}(\underline{\xi}_k) \cdot \underline{\xi} = -g(\underline{\xi}_k) \quad (3.11)$$

für $\underline{\xi}$. Anschließend ist

$$\underline{\xi}_{k+1} := \underline{\xi} - \underline{\xi}_k$$

zu setzen.

Konvergiert $g(\underline{\xi}_k)$ schnell genug gegen $\underline{0}$, so liefert der Korrektorschritt ein neues $\underline{\eta}_{\text{tmp}} := \underline{\xi}_k$ und $\text{success} := \text{true}$ zurück. Andernfalls ist das Ergebnis das alte $\underline{\eta}_{\text{tmp}}$ und $\text{success} := \text{false}$.

Schrittweitenkontrolle Ist die Schrittweite zu groß gewählt, so ist die Prädiktion zu weit von einer Nullstelle entfernt und das Newtonverfahren im Korrektorschritt konvergiert nicht oder zumindest nicht schnell genug. Andererseits besteht der Wunsch, die Schrittweite möglichst groß zu wählen, um wenige Iterationen ausführen zu müssen. Dadurch wird der Rechenaufwand gering gehalten. Wie groß die Schrittweite gewählt werden kann — bei gleichzeitiger Konvergenz des Newtonverfahrens — hängt von der lokalen Gestalt von G in einem Bereich um $\underline{\eta}_{\text{tmp}}$ und γ_{neu} herum ab. Daher wird die Schrittweite in Abhängigkeit vom Erfolg einer Iteration — bestehend aus Prädiktor- und Korrektorschritt — angepasst.

Problemformulierung

[5], [10] und [11] behandeln das Progressive Bayes-Verfahren im Fall eindimensionaler Dichten. Das Progressive Bayes-Verfahren soll in dieser Arbeit mit Dirac-Mischdichten als Approximationsdichten im mehrdimensionalen Fall angewendet werden. Der letzte Abschnitt in diesem Kapitel enthält eine ausführliche Beschreibung der Zielsetzung. Die folgenden Abschnitte betrachten zuvor verschiedene Aspekte, die im mehrdimensionalen Fall zu berücksichtigen sind.

4.1 Ordnungsannahme

In [11] findet bei der Herleitung des Differentialgleichungssystems (3.8) aus Gleichung (3.7) eine Vorauswertung der auftretenden Heaviside-Funktionen statt. Diese Vorauswertung ist möglich, da eine Ordnung der Positionen der Dirac-Impulse der Approximationsdichte $f(x, \underline{\eta}) = \sum_{l=1}^L w_l \delta(x - x_l)$ vorausgesetzt wird, $x_l < x_{l+1}$ für alle $1 \leq l < L$. Sie führt im Eindimensionalen im Allgemeinen zu keiner Beschränkung der möglichen Approximationen.

Für $d > 1$ besitzt \mathbb{R}^d zwar eine partielle, jedoch keine totale Ordnung. Eine Ordnungsannahme für die Positionen der Dirac-Impulse in der mehrdimensionale Approximationsdichte $f(\underline{x}, \underline{\eta}) = \sum_{l=1}^L w_l \delta(\underline{x} - \underline{x}_l)$ führt daher zu einer starken Beschränkung der Lösungsmenge des Differentialgleichungssystems (3.8). Schon deshalb soll für diese Arbeit keine Ordnungsannahme getroffen werden. Die fehlende Ordnungsannahme zieht mehrere Schwierigkeiten nach sich.

4.2 Differentialgleichungssystem

Das Aufstellen des Differentialgleichungssystems (3.8) gehört zum Kern des Progressiven Bayes-Verfahrens. Ohne eine Ordnungsannahme wird das Aufstellen einer (3.8) entsprechenden Differentialgleichung aufwendiger, ebenso die Berechnung der Jakobi-Matrix \mathbf{J} für den Korrektorschritt im numerischen Lösungsverfahren. Wie sich später zeigen wird, enthalten \mathbf{P} und \mathbf{J} Irregularitäten in Form von Dirac-Impulsen.

4.3 Parameteranzahl

Mit steigender Anzahl der Parameter in $\underline{\eta}$ wird die analytische Gestalt der Funktion G immer komplizierter, es entstehen also immer mehr lokale Extremstellen.

Die Anzahl der notwendigen Parameter zur Festlegung der Position eines einzelnen Dirac-Impulses entspricht gerade der Dimension d selbst. Zudem wird bei wachsender Dimension eine immer höhere Anzahl von Dirac-Impulsen benötigt, um eine vorgegebene Güte durch eine optimalen Approximation zu erreichen. Insgesamt steigt die Anzahl der notwendigen Parameter mit steigender Dimension stark an.

4.4 Analytische Form des Gütemaßes

Nicht nur durch die höhere Parameteranzahl, wie im vorherigen Abschnitt beschrieben, wird die analytische Form des Gütemaßes im Mehrdimensionalen komplizierter.

Durch die Ordnungsannahme im Eindimensionalen ist die partielle Ableitung $\frac{\partial}{\partial x_i} G(\underline{\eta}, \gamma)$ von G in Richtung einer Positionskomponente x_i von keiner anderen Positionskomponente x_j , $j \neq i$, abhängig. Natürlich besteht durch die Ordnungsannahme selbst ein Zusammenhang zwischen den Positionen der Dirac-Impulse. Innerhalb der durch den linken und rechten Nachbarn gesetzten Grenze hat die Änderung der Position eines Dirac-Impulses allerdings nur einen lokalen Einfluss auf die Gestalt der Verteilungsfunktion F . Selbst wenn die Änderung innerhalb der vorgegebenen Grenzen groß ist, führt dies nur zu einer verhältnismäßig kleinen Änderung der Güte.

Im Mehrdimensionalen gilt dies nicht mehr. In Abbildung 2.4 ist eine zweidimensionale Dirac-Mischdichte und ihre Verteilungsfunktion zu sehen. Deutlich zu erkennen sind die Überschneidungsbereiche in der Verteilungsfunktion, die sich durch Summierung der einzelnen Heaviside-Funktionen ergeben. Eine Änderung der Position eines Dirac-Impulses wirkt sich auf diese Überschneidungsbereiche aus und führt zu einer relativ großen Änderung an der Verteilungsfunktion als Ganzer. Damit ist auch die Änderung der Güte verhältnismäßig groß, so dass das Progressive Bayes-Verfahren im Mehrdimensionalen wesentlich sensibler auf Positionsänderungen der Dirac-Impulse reagiert als im Eindimensionalen.

4.5 Bifurkationen

Die höhere Anzahl der Parameter in $\underline{\eta}$ zusammen mit der komplizierteren Form des Gütemaßes bei gleicher Parameteranzahl im Vergleich zum Eindimensionalen verschärft das Problem von Bifurkationen im Mehrdimensionalen. Bifurkationen sind Verzweigungen, die im Verlauf der Progression auftreten können.

Definition 4.5.1 Seien $\underline{\eta}_i$, $i = 1, 2$, zwei verschiedene Lösungen von (3.6) die bis zu einem $\gamma \in (0, 1)$ übereinstimmen, also $\forall \vartheta \in [0, \gamma) : \underline{\eta}_1(\vartheta) = \underline{\eta}_2(\vartheta)$, und in γ stetig sind. Unterscheiden sich $\underline{\eta}_1$ und $\underline{\eta}_2$ nun in jeder noch so kleinen Umgebung $[\gamma, \gamma + \epsilon)$ mit $\epsilon > 0$, dann tritt eine Bifurkation an der Stelle γ auf. Anders gesagt: Eine stetige Fortsetzung der Teillösung $\underline{\eta} : [0, \gamma) \rightarrow \mathbb{R}^{+L} \times \mathbb{R}^{L \cdot d}$, $\underline{\eta}(\vartheta) := \underline{\eta}_1(\vartheta)$, von (3.6) ist an der Stelle γ nicht eindeutig möglich. \square

Besitzt (3.6) mindestens eine Lösung, dann ist eine Fortsetzung einer Teillösung immer möglich. Diese kann jedoch unstetig sein.

Bifurkationen hängen ganz wesentlich von der gewählten Progression sowie der Startdichte und ihrer Approximation ab. Ziel im Progressiven Bayes-Verfahren ist, die Startapproximation optimal zu wählen und diese Optimalität während der Progression zu erhalten. Bifurkationen stellen in diesem Zusammenhang eine Gegenkraft dar und führen gegebenenfalls zu lediglich lokal optimalen Approximationen im Verlauf der Progression. Somit besteht der Wunsch durch eine geeignete Wahl von Startdichte samt Approximation und Progression, Bifurkationen möglichst zu vermeiden.

4.5.1 Separiertheit (lokal) optimaler Approximationen

Bifurkationen entstehen insbesondere dann, wenn im Verlauf der Progression (lokal) optimale Approximationen auftreten, die nicht, oder nicht hinreichend gut, separiert sind von anderen (lokal) optimalen Approximationen.

Definition 4.5.2 Für ein fest gewähltes $\gamma \in [0, 1]$ heißt $\underline{\eta} \in \mathbb{R}^{+L} \times \mathbb{R}^{L \cdot d}$ separiert, falls es eine Umgebung U von $\underline{\eta}$ gibt, so dass für alle $\underline{\xi} \in U$ mit $\underline{\xi} \neq \underline{\eta}$ gilt: $G(\underline{\xi}, \gamma) > G(\underline{\eta}, \gamma)$. \square

Insbesondere ist $\underline{\eta}$ in diesem Fall eine (lokale) Minimalstelle von $G(\underline{\eta}, \gamma)$ betrachtet als Funktion von $\underline{\eta}$. Definition 4.5.2 sagt, dass $\underline{\eta}$ darüber hinaus auch noch die einzige (lokale) Minimalstelle in einer ganzen Umgebung von $\underline{\eta}$ ist.

4.6 Sackgassen

Das größte Problem im Zusammenhang mit Bifurkationen ist, dass einige Fortsetzungen von Teillösungen von (3.6) nicht nur zu lediglich lokal — und nicht global — optimalen Approximationen führen, sondern sogar in Sackgassen enden können.

Definition 4.6.1 Eine Sackgasse ist eine Teillösung $\underline{\eta} : [0, \gamma) \rightarrow \mathbb{R}^{+L} \times \mathbb{R}^{L \cdot d}$, $\gamma \in (0, 1)$, von (3.6) die nicht stetig fortgesetzt werden kann. \square

Anders gesagt geht durch eine Sackgasse nicht nur ein Teil der Optimalität der Approximation während der Progression verloren. Statt dessen kann die Approximation gar nicht mehr bis zum Ende der Progression mitgeführt werden. In diesem Fall wird im numerischen Lösungsverfahren

aus Abschnitt 3.2.4 das eigentliche Ziel, eine Approximation für die Zieldichte zu bestimmen, nicht erreicht.

Etwas weniger strikt wird im Kontext des numerischen Lösungsverfahrens von einer Sackgasse gesprochen, wenn das Newtonverfahren im Korrektorschritt selbst dann nicht konvergiert, wenn die Schrittweite für den Prädiktorschritt sehr klein gewählt wurde.

4.7 Zielsetzung

Ziel dieser Arbeit ist, das Progressive Bayes-Verfahren im Mehrdimensionalen zur Approximation von Dichten anzuwenden. Dabei sollen Dirac-Mischdichten als Approximationsdichten benutzt werden. Die Form der zu approximierenden Dichten wird soweit wie möglich als beliebig angenommen. Mit Hilfe einer Implementierung für den konkreten Fall zweidimensionaler, achsenparalleler Gauß-Mischdichten sollen die Ergebnisse mit anderen Verfahren verglichen werden.

Bis auf die Vorgabe, innerhalb des vorgegebenen Intervalls $[\underline{c}_1, \underline{c}_2]$ zu liegen, sollen die Positionen der Dirac-Impulse der Approximationsdichte keinen weiteren Einschränkungen unterliegen. Insbesondere wird also keine Ordnungsannahme getroffen, welche die Menge der möglichen Approximationen stark einschränken würde. Ebenso unterliegen die Gewichte lediglich der Vorgabe, positiv zu sein. Insbesondere wird keine Gleichgewichtung vorausgesetzt.

Ein (3.8) entsprechendes explizites Differentialgleichungssystem soll aufgestellt werden, so dass das numerische Lösungsverfahren aus [11] eingesetzt werden kann. Dazu ist insbesondere die Jakobi-Matrix \mathbf{J} für den Korrektorschritt zu berechnen.

Für die Anwendung des Verfahrens sind geeignete Progressionen sowie Startdichten samt Startapproximationen zu untersuchen beziehungsweise zu finden.

Progressives Bayes im Mehrdimensionalen

5.1 Notwendige Bedingung

Die notwendige Bedingung (3.6) für eine Stelle $\underline{\eta}_{\min}$ des globalen Minimums von G für ein gegebenes γ lautet

$$\frac{\partial}{\partial \underline{\eta}} G(\underline{\eta}_{\min}, \gamma) \stackrel{!}{=} 0 .$$

Äquivalent dazu ist die Forderung, dass die partiellen Ableitungen

$$\frac{\partial}{\partial w_i} G(\underline{\eta}_{\min}, \gamma) \stackrel{!}{=} 0 \quad (5.1)$$

und

$$\frac{\partial}{\partial x_{im}} G(\underline{\eta}_{\min}, \gamma) \stackrel{!}{=} 0 \quad (5.2)$$

des progressiven Gütemaßes in Richtung der Gewichte sowie der einzelnen Positionskomponenten verschwinden, $1 \leq i \leq L$ und $1 \leq m \leq d$.

Diese partiellen Ableitung werden in Abschnitt A.6 berechnet. Das Ergebnis lautet

$$\frac{\partial}{\partial w_i} G(\underline{\eta}, \gamma) = 2 \left\{ \int_{\underline{x}_i}^{c_2} F(\underline{x}, \underline{\eta}) d\underline{x} - \int_{\underline{x}_i}^{c_2} \tilde{F}(\underline{x}, \gamma) d\underline{x} \right\} \quad (5.3)$$

und

$$\frac{\partial}{\partial x_{im}} G(\underline{\eta}, \gamma) = 2w_i \left\{ \int_{x_{iq}}^{c_{2q}} \tilde{F}(\underline{x}|_{x_m \equiv x_{im}}, \underline{\eta}) dx_q - \int_{x_{iq}}^{c_{2q}} F(\underline{x}|_{x_m \equiv x_{im}}, \underline{\eta}) dx_q \right\} \quad (5.4)$$

Die Schreibweise $\underline{x}|_{x_m \equiv x_{im}}$ bedeutet, dass an der m .ten Stelle des Vektors \underline{x} der Term x_{im} eingesetzt wird. $\int_{x_{iq}}^{c_{2q}} \dots dx_q$ meint ein Mehrfachintegral, in welchem nach x_q , $q = 1, \dots, m -$
 $q=1, \dots, d; q \neq m$

$1, m + 1, \dots, d$ integriert wird. Für eine ausführlichere Einführung dieser Schreibweisen siehe Abschnitt A.2.

Die Gleichungen (5.3) und (5.4) setzen sich aus den Termen

$$R(i, \underline{\eta}) := \int_{\underline{x}_i}^{c_2} F(\underline{x}, \underline{\eta}) d\underline{x}$$

sowie

$$R(i, m, \underline{\eta}) := \int_{x_{i_q}}^{c_{2q}} F(\underline{x}|_{x_m \equiv x_{im}}, \underline{\eta}) dx_q$$

$q=1, \dots, d; q \neq m$

für die Verteilungsfunktion $F(\underline{x}, \underline{\eta})$ der Approximation und

$$\tilde{R}(i, \underline{\eta}, \gamma) := \int_{\underline{x}_i}^{c_2} \tilde{F}(\underline{x}, \gamma) d\underline{x}$$

sowie

$$\tilde{R}(i, m, \underline{\eta}, \gamma) := \int_{x_{i_q}}^{c_{2q}} \tilde{F}(\underline{x}|_{x_m \equiv x_{im}}, \underline{\eta}) dx_q$$

$q=1, \dots, d; q \neq m$

für die progressive Verteilungsfunktion $\tilde{F}(\underline{x}, \gamma)$ zusammen. Die anschauliche Bedeutung dieser Terme soll hier im Zusammenhang mit den notwendigen Bedingungen (5.1) und (5.2) an Hand eines Beispiels verdeutlicht werden. Dazu sei als erstes (5.1) mit Hilfe von (5.3) als

$$\frac{\partial}{\partial w_i} G(\underline{\eta}_{\min}, \gamma) = 2 \left\{ R(i, \underline{\eta}_{\min}) - \tilde{R}(i, \underline{\eta}_{\min}, \gamma) \right\} \stackrel{!}{=} 0 \quad (5.5)$$

sowie (5.1) mit Hilfe von (5.4) als

$$\frac{\partial}{\partial x_{im}} G(\underline{\eta}_{\min}, \gamma) = 2w_i \left\{ \tilde{R}(i, m, \underline{\eta}_{\min}, \gamma) - R(i, m, \underline{\eta}_{\min}) \right\} \stackrel{!}{=} 0 \quad (5.6)$$

formuliert.

$R(i, \underline{\eta})$ ist definiert als das Integral der Verteilungsfunktion $F(\underline{x}, \underline{\eta})$ der Approximation über dem Intervall $[\underline{x}_i, c_2]$, analog ist $\tilde{R}(i, \underline{\eta}, \gamma)$ das Integral der progressiven Verteilungsfunktion $\tilde{F}(\underline{x}, \gamma)$ über $[\underline{x}_i, c_2]$. Im Allgemeinen unterscheidet sich die Verteilungsfunktion $F(\underline{x}, \underline{\eta})$ der Approximation von der progressiven Verteilungsfunktion $\tilde{F}(\underline{x}, \gamma)$ für ein gegebenes γ selbst dann, wenn $\underline{\eta}$ tatsächlich optimal gewählt ist. Aus (5.5) folgt jedoch, dass für eine optimale Approximation zumindest die Integrale über dem Intervall $[\underline{x}_i, c_2]$ übereinstimmen müssen,

$$R(i, \underline{\eta}_{\min}) \stackrel{!}{=} \tilde{R}(i, \underline{\eta}_{\min}, \gamma) . \quad (5.7)$$

Genauer gesagt ist (5.7) eine notwendige Bedingung für die optimale Wahl des Gewichts w_i .

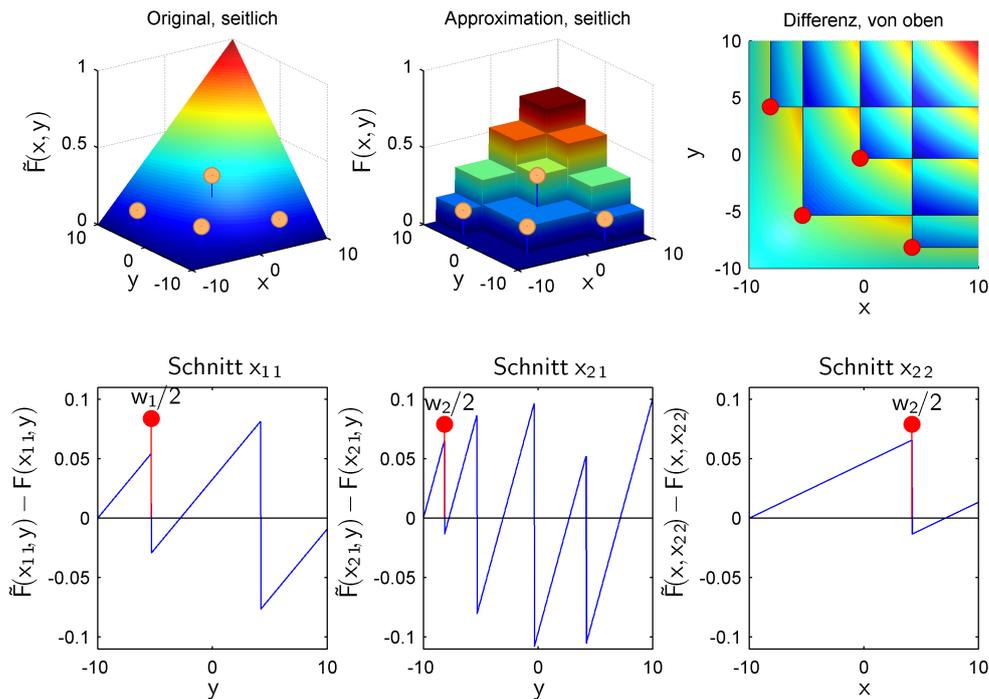


Abbildung 5.1: Veranschaulichung der notwendigen Bedingung für eine (lokal) optimale Approximation. Dargestellt ist eine zweidimensionale Gleichverteilungsfunktion (links oben) zusammen mit der Verteilungsfunktion einer lokal optimalen Approximation (oben Mitte). Rechts oben ist die Differenz der beiden Verteilungsfunktionen zu sehen. In der unteren Zeile sind verschiedene Schnitte durch die Differenz dargestellt. \underline{x}_1 ist die Position des Dirac-Impulses links unten, \underline{x}_2 die des Dirac-Impulses rechts unten in der Ansicht von oben.

Beispiel 5.1: Notwendige Bedingung am Beispiel einer Gleichverteilung

Abbildung 5.1 zeigt eine Gleichverteilung mit einer lokal optimalen Approximation. (5.7) besagt, dass das Integral der Differenzen in der Teilabbildung rechts oben gleich Null sein muss, jeweils über dem Intervall ausgehend von der Position eines Dirac-Impulses bis zur Integrationsobergrenze $\underline{c}_2 = 10$. ■

$R(i, m, \underline{\eta})$ ist das Integral über die Verteilungsfunktion $F(\underline{x}, \underline{\eta})$ der Approximation mit auf x_{im} festgehaltener m -ter Komponente, analog gilt dies für $\tilde{R}(i, m, \underline{\eta}, \gamma)$ und die progressive Verteilungsfunktion $\tilde{F}(\underline{x}, \gamma)$. Entsprechend (5.6) müssen auch diese Integrale für jede Positionskomponente x_{im} , $i = 1, \dots, L$ und $m = 1, \dots, d$ in einer optimalen Approximation übereinstimmen,

$$R(i, m, \underline{\eta}_{\min}) \stackrel{!}{=} \tilde{R}(i, m, \underline{\eta}_{\min}, \gamma) , \tag{5.8}$$

damit diese optimal platziert sind. Für die Äquivalenz von (5.6) und (5.8) werden positive Gewichte w_i vorausgesetzt. Es ist offensichtlich, dass eine optimale Approximation keine Gewichte enthalten kann, die gleich Null sind, da der zugehörige Dirac-Impuls schlicht überflüssig wäre.

Beispiel 5.1: Notwendige Bedingung am Beispiel einer Gleichverteilung (Fortsetzung)

In der unteren Zeile von Abbildung 5.1 sind verschiedene Schnitte der Differenz aus progressiver Verteilungsfunktion und Verteilungsfunktion der Approximation zu sehen. Jeder Schnitt erfolgt durch den über x_{im} in der Überschrift angegebenen i -ten Dirac-Impuls entlang der zu m komplementären Achse. Dabei entspricht $m = 1$ der x -Achse und $m = 2$ der y -Achse. \underline{x}_1 ist die Position des Dirac-Impulses in der Nähe von $\begin{pmatrix} -5 \\ -5 \end{pmatrix}$, \underline{x}_2 die des Dirac-Impulses in der Nähe von $\begin{pmatrix} 5 \\ -8 \end{pmatrix}$. Eingezeichnet ist jeweils ein Marker an der Position x_{im} mit einer Höhe, die der Hälfte des Gewichts w_i entspricht. Die Höhe des Markers entspricht damit gerade der Differenz des links- beziehungsweise rechtsseitigen Grenzwertes an der Position des Markers in der jeweiligen Unterabbildung. Dafür ist zu beachten, dass der Funktionswert $H(0)$ der Heaviside-Funktion an der Stelle 0 gerade $\frac{1}{2}$ ist.

(5.8) besagt, dass für jeden solchen Schnitt das Integral ab der markierten Position x_{im} gleich Null ist, beziehungsweise sein muss, da die Approximation lokal optimal ist. ■

5.2 Differentialgleichungssystem

In der Wiederholung des Progressiven Bayes-Verfahrens in Abschnitt 3.2.4 wurde schon erwähnt, dass das Differentialgleichungssystem

$$\mathbf{P} \cdot \underline{\dot{\eta}} \stackrel{!}{=} \underline{b}$$

im Eindimensionalen aus der notwendigen Bedingung

$$\frac{\partial}{\partial \underline{\eta}} G(\underline{\eta}_{\min}, \gamma) \stackrel{!}{=} 0 \tag{5.9}$$

hergeleitet wird, und zwar mit dem Ansatz, dass diese Bedingung für eine optimale Approximation während der Progression erhalten bleibt, also

$$\frac{\partial}{\partial \gamma} \left\{ \frac{\partial}{\partial \underline{\eta}} G(\underline{\eta}_{\min}, \gamma) \right\} \stackrel{!}{=} 0 \tag{5.10}$$

gilt. (5.10) ergibt sich dabei aus der Grenzwertdarstellung der Ableitung nach γ aus (5.9), da (5.9) für jedes γ aus $[0, 1]$ gefordert wird.

Im Mehrdimensionalen kann genauso verfahren werden. Ausgangspunkt dazu bilden die Forderungen (5.7) und (5.8), die ebenfalls unter infinitesimalen Änderungen von γ erhalten bleiben. Es folgt die notwendige Bedingung

$$\frac{\partial}{\partial \gamma} R(i, \underline{\eta}) \stackrel{!}{=} \frac{\partial}{\partial \gamma} \tilde{R}(i, \underline{\eta}, \gamma) \tag{5.11}$$

und analog

$$\frac{\partial}{\partial \gamma} R(i, m, \underline{\eta}) \stackrel{!}{=} \frac{\partial}{\partial \gamma} \tilde{R}(i, m, \underline{\eta}, \gamma) \tag{5.12}$$

Zu beachten ist, dass $\underline{\eta}$ hier als Funktion von γ zu betrachten ist. Mit Hilfe der verallgemeinerten Kettenregel für Ableitungen (siehe auch Abschnitt A.3.1) folgt aus (5.11)

$$\begin{aligned} & R^b(i, \underline{\eta}, \gamma) \frac{\partial}{\partial \gamma} \gamma + \sum_{l=1}^L \left\{ \frac{\partial}{\partial w_l} R(i, \underline{\eta}) \right\} \dot{w}_l + \sum_{l=1}^L \sum_{p=1}^d \left\{ \frac{\partial}{\partial x_{lp}} R(i, \underline{\eta}) \right\} \dot{x}_{lp} \\ \stackrel{!}{=} & \tilde{R}^b(i, \underline{\eta}, \gamma) \frac{\partial}{\partial \gamma} \gamma + \sum_{l=1}^L \left\{ \frac{\partial}{\partial w_l} \tilde{R}(i, \underline{\eta}, \gamma) \right\} \dot{w}_l + \sum_{l=1}^L \sum_{p=1}^d \left\{ \frac{\partial}{\partial x_{lp}} \tilde{R}(i, \underline{\eta}, \gamma) \right\} \dot{x}_{lp} \end{aligned} \quad (5.13)$$

und aus (5.12)

$$\begin{aligned} & R^b(i, m, \underline{\eta}, \gamma) \frac{\partial}{\partial \gamma} \gamma + \sum_{l=1}^L \left\{ \frac{\partial}{\partial w_l} R(i, m, \underline{\eta}) \right\} \dot{w}_l + \sum_{l=1}^L \sum_{p=1}^d \left\{ \frac{\partial}{\partial x_{lp}} R(i, m, \underline{\eta}) \right\} \dot{x}_{lp} \\ \stackrel{!}{=} & \tilde{R}^b(i, m, \underline{\eta}, \gamma) \frac{\partial}{\partial \gamma} \gamma + \sum_{l=1}^L \left\{ \frac{\partial}{\partial w_l} \tilde{R}(i, m, \underline{\eta}, \gamma) \right\} \dot{w}_l + \sum_{l=1}^L \sum_{p=1}^d \left\{ \frac{\partial}{\partial x_{lp}} \tilde{R}(i, m, \underline{\eta}, \gamma) \right\} \dot{x}_{lp} . \end{aligned} \quad (5.14)$$

Dabei bezeichnet $R^b(i, \underline{\eta}, \gamma) := \frac{\partial}{\partial \gamma} R(i, \underline{\eta})$ die Ableitung von $R(i, \underline{\eta})$ für ein von γ *unabhängiges* $\underline{\eta}$. Da $R(i, \underline{\eta})$ nicht von γ abhängt — es sei denn, $\underline{\eta}$ wird als Funktion von γ betrachtet (!) —, gilt $R^b(i, \underline{\eta}, \gamma) = 0$ und analog $R^b(i, m, \underline{\eta}, \gamma) := \frac{\partial}{\partial \gamma} R(i, m, \underline{\eta}) = 0$. Genauso bezeichnet $\tilde{R}^b(i, \underline{\eta}, \gamma) := \frac{\partial}{\partial \gamma} \tilde{R}(i, \underline{\eta}, \gamma)$ die Ableitung von $\tilde{R}(i, \underline{\eta}, \gamma)$ nach γ , wobei $\underline{\eta}$ als von γ *unabhängig* zu betrachten ist und analog $\tilde{R}^b(i, m, \underline{\eta}, \gamma) := \frac{\partial}{\partial \gamma} \tilde{R}(i, m, \underline{\eta}, \gamma)$. Da $\tilde{R}(i, \underline{\eta}, \gamma)$ und $\tilde{R}(i, m, \underline{\eta}, \gamma)$ explizit von γ abhängen, sind ihre partiellen Ableitungen nach γ im Gegensatz zu denen von $R(i, \underline{\eta})$ und $R(i, m, \underline{\eta})$ im Allgemeinen nicht gleich Null.

Unter Beachtung von $R^b(i, \underline{\eta}, \gamma) = 0$, $R^b(i, m, \underline{\eta}, \gamma) = 0$ und $\frac{\partial}{\partial \gamma} \gamma = 1$ können die Terme in (5.13) zu

$$\begin{aligned} \underbrace{\tilde{R}^b(i, \underline{\eta}, \gamma)}_{=: b_{\text{cod}(i)}} & \stackrel{!}{=} \underbrace{\sum_{l=1}^L \left\{ \frac{\partial}{\partial w_l} R(i, \underline{\eta}) - \frac{\partial}{\partial w_l} \tilde{R}(i, \underline{\eta}, \gamma) \right\} \dot{w}_l}_{=: \mathbf{P}(\text{cod}(i), \text{cod}(l))} \\ & + \underbrace{\sum_{l=1}^L \sum_{p=1}^d \left\{ \frac{\partial}{\partial x_{lp}} R(i, \underline{\eta}) - \frac{\partial}{\partial x_{lp}} \tilde{R}(i, \underline{\eta}, \gamma) \right\} \dot{x}_{lp}}_{=: \mathbf{P}(\text{cod}(i), \text{cod}(l, p))} \end{aligned} \quad (5.15)$$

und die in (5.14) zu

$$\underbrace{\tilde{R}^b(i, m, \underline{\eta}, \gamma)}_{=: b_{\text{cod}(i,m)}} \stackrel{!}{=} \underbrace{\sum_{l=1}^L \left\{ \frac{\partial}{\partial w_l} R(i, m, \underline{\eta}) - \frac{\partial}{\partial w_l} \tilde{R}(i, m, \underline{\eta}, \gamma) \right\}}_{=: \mathbf{P}(\text{cod}(i,m), \text{cod}(l))} \dot{w}_l$$

$$+ \underbrace{\sum_{l=1}^L \sum_{p=1}^d \left\{ \frac{\partial}{\partial x_{lp}} R(i, m, \underline{\eta}) - \frac{\partial}{\partial x_{lp}} \tilde{R}(i, m, \underline{\eta}, \gamma) \right\}}_{=: \mathbf{P}(\text{cod}(i,m), \text{cod}(l,p))} \dot{x}_{lp}$$
(5.16)

zusammengefasst werden. Für $i = 1, \dots, L$ und $m = 1, \dots, d$ bilden (5.15) und (5.16) das gesuchte explizite Differentialgleichungssystem. Die Einträge $\mathbf{P}(\alpha, \beta)$ einer Zeile α und einer Spalte β in der Matrix \mathbf{P} sowie die Einträge b_α des Vektors \underline{b} können aus ihnen direkt abgelesen werden. Die Funktion

$$\text{cod} : \{1, \dots, L\} \cup \{(l, m) | 1 \leq l \leq L, 1 \leq m \leq d\} \rightarrow \{1, \dots, L \cdot \{d + 1\}\}$$

$$\text{cod}(i) := i$$

$$\text{cod}(i, m) := L + d \cdot \{i - 1\} + m$$

dient dabei zur Kodierung eines Zeilen– beziehungsweise Spaltenindex α beziehungsweise β .

Das Differentialgleichungssystem ist jetzt mit Hilfe der R , \tilde{R} und \tilde{R}^b Terme ausgedrückt. Die Berechnung der R Terme und ihrer partiellen Ableitungen erfolgt in Abschnitt A.4 des mathematischen Anhangs. Einige dieser Terme enthalten Irregularitäten in Form von Dirac-Impulsen, siehe dazu Abschnitt 5.4. Die Berechnung der \tilde{R} Terme sowie ihrer partiellen Ableitungen samt \tilde{R}^b erfolgt für eine beliebige progressive Verteilungsfunktion sowie für den Fall einer Gleichverteilung und den Fall einer achsenparallelen Gauß-Mischverteilung in jedem Progressionsschritt γ in Abschnitt A.5.

5.3 Jakobi-Matrix

Die Jakobi-Matrix \mathbf{J} aus (3.10) enthält die zweiten partiellen Ableitung des Gütemaßes in Richtung der Gewichte und Positionskomponenten für $\gamma = \gamma_{\text{neu}}$. Die zweiten partiellen Ableitungen können leicht auf Basis der Gleichung (5.3) und (5.4) entsprechenden Darstellungen

$$\frac{\partial}{\partial w_i} G(\underline{\eta}, \gamma) = 2 \left\{ R(i, \underline{\eta}) - \tilde{R}(i, \underline{\eta}, \gamma) \right\}$$

sowie

$$\frac{\partial}{\partial x_{im}} G(\underline{\eta}, \gamma) = 2w_i \left\{ \tilde{R}(i, m, \underline{\eta}, \gamma) - R(i, m, \underline{\eta}) \right\}$$

der ersten partiellen Ableitungen berechnet werden. Mit Hilfe der Kodierungsfunktion cod aus dem vorhergehenden Abschnitt gilt

$$\begin{aligned}
 \mathbf{J}(\text{cod}(i), \text{cod}(l)) &= \frac{\partial}{\partial w_l} \frac{\partial}{\partial w_i} G(\underline{\eta}, \gamma) = 2 \frac{\partial}{\partial w_l} R(i, \underline{\eta}) \\
 \mathbf{J}(\text{cod}(i), \text{cod}(l, p)) &= \frac{\partial}{\partial x_{lp}} \frac{\partial}{\partial w_i} G(\underline{\eta}, \gamma) = 2 \left\{ \frac{\partial}{\partial x_{lp}} R(i, \underline{\eta}) - \frac{\partial}{\partial x_{lp}} \tilde{R}(i, \underline{\eta}, \gamma) \right\} \\
 \mathbf{J}(\text{cod}(i, m), \text{cod}(l)) &= \frac{\partial}{\partial w_l} \frac{\partial}{\partial x_{im}} G(\underline{\eta}, \gamma) \\
 &= \begin{cases} 2 \left\{ \tilde{R}(i, m, \underline{\eta}, \gamma) - R(i, m, \underline{\eta}) - w_i \frac{\partial}{\partial w_i} R(i, m, \underline{\eta}) \right\} & i = l \\ -2w_i \frac{\partial}{\partial w_l} R(i, m, \underline{\eta}) & i \neq l \end{cases} \\
 \mathbf{J}(\text{cod}(i, m), \text{cod}(l, p)) &= \frac{\partial}{\partial x_{lp}} \frac{\partial}{\partial x_{im}} G(\underline{\eta}, \gamma) = 2w_i \left\{ \frac{\partial}{\partial x_{lp}} \tilde{R}(i, m, \underline{\eta}, \gamma) - \frac{\partial}{\partial x_{lp}} R(i, m, \underline{\eta}) \right\}
 \end{aligned}$$

für jedes $i, l = 1, \dots, L$ und $m, p = 1, \dots, d$. Dazu wird ausgenutzt, dass $\tilde{R}(i, \underline{\eta}, \gamma)$ und $\tilde{R}(i, m, \underline{\eta}, \gamma)$ das Gewicht w_l nicht enthalten. Wie für den vorhergehenden Abschnitt sind die Berechnungen der verbleibenden Terme in den entsprechenden Unterabschnitten von Anhang A zu finden.

5.4 Irregularitäten

Das Gütemaß basiert auf den Verteilungsfunktionen F und \tilde{F} . F besteht aus einer Linearkombination von Heaviside-Funktionen. Die Heaviside-Funktion ist nicht stetig im Nullpunkt und folglich an dieser Stelle nicht glatt. Das Gütemaß ist es somit ebenso wenig. Da es das Integral über einen Term mit Heaviside-Funktionen bildet, enthalten einige der partiellen Ableitungen des Gütemaßes wiederum Heaviside-Funktionen, insbesondere sind diese partiellen Ableitungen nicht überall stetig. Einige der zweiten partiellen Ableitungen enthalten Dirac-Funktionen als Ableitungen der Heaviside-Funktionen: im vorhergehenden Abschnitt wurde gezeigt, dass die zweiten partiellen Ableitungen des Gütemaßes mit Hilfe der R Terme ausgedrückt werden können. Konkret beinhaltet die partielle Ableitung $\frac{\partial}{\partial x_{lm}} R(i, m, \underline{\eta})$, $i, l = 1, \dots, L$, $m = 1, \dots, d$, Dirac-Funktionen, siehe (A.31) und (A.33). Neben \mathbf{J} enthält somit auch \mathbf{P} in den Einträgen $\mathbf{P}(\text{cod}(i, m), \text{cod}(l, m))$ Irregularitäten in Form von Dirac-Funktionen, siehe (5.16).

Im Eindimensionalen hilft die Ordnungsannahme für die Positionen und die damit verbundene Vorauswertung der Heaviside-Funktionen, diese Irregularitäten zu vermeiden. Da für den mehrdimensionalen Fall keine Ordnungsannahme getroffen werden soll, muss in irgendeiner anderen Form mit ihnen umgegangen werden. Schließlich kann mit Dirac-Funktionen zwar analytisch gerechnet werden, für die numerische Lösung des Differentialgleichungssystems sind sie jedoch „unhandlich“.

Glücklicherweise ist das Problem nicht so groß, wie es im ersten Moment scheinen mag. Für das Folgende wird vorausgesetzt, dass \tilde{F} hinreichend glatt ist und keine weiteren Irregularitäten erzeugt. Für den Rest dieses Abschnitts wird γ als fest und $G(\underline{\eta}, \gamma)$ als Funktion von $\underline{\eta}$ betrachtet.

Gesucht wird eine Stelle des globalen Minimums von G . Als erstes sei festgehalten, dass die partiellen Ableitungen von G fast überall stetig sind. Unstetig ist lediglich die partielle Ableitung $\frac{\partial}{\partial x_{im}} G(\underline{\eta}, \gamma)$ für genau diejenigen $\underline{\eta} = (w_1, \dots, w_L, \underline{x}_1^T, \dots, \underline{x}_L^T)^T$ für die es ein von i verschiedenes $l \in \{1, \dots, L\}$ gibt, so dass $x_{im} = x_{lm}$ gilt. Siehe dazu Gleichung (A.50) und (A.29) sowie Definition (A.16). Anschaulich besitzt die Approximationsdichte $f(\underline{x}, \underline{\eta})$ in einem solchen Fall also mindestens zwei Dirac-Impulse i und l , die in der m .ten Achse parallel ausgerichtet sind.

Es ist zu erwarten, dass G in den Stellen des globalen Minimums eine glatte Funktion ist und ihre partiellen Ableitungen an diesen Stellen stetig sind. Tatsächlich veranschaulicht das nachfolgende Beispiel 5.2, dass Approximationen mit achsenparallel ausgerichteten Dirac-Impulsen die Überschneidungsbereiche der Heaviside-Funktionen nicht optimal ausnutzen. Im Umkehrschluss gilt: In den Stellen des globalen Minimums ist G *in der Regel* eine glatte Funktion, an denen keine Irregularitäten auftreten.

In Abhängigkeit von der konkret vorgelegten Dichte \tilde{f} kann eine optimale Approximation zwar sehr wohl zwei achsenparallel ausgerichtete Dirac-Impulse enthalten. Für hinreichend glatte Dichten \tilde{f} ist dies jedoch untypisch. Für die Praxis gilt zudem:

1. Mit dem numerischen Lösungsverfahren eine Stelle $\underline{\eta}$ zu treffen, in der (mindestens) zwei Dirac-Impulse *genau* achsenparallel ausgerichtet sind, ist unwahrscheinlich.
2. Die Rechengenauigkeit ist begrenzt. Sollte der Fall $x_{im} = x_{lm}$ eintreten, dann werden x_{im} und x_{lm} als infinitesimal verschieden betrachtet.

Für die Implementierung wurde daher $\delta(0) := \lim_{x \rightarrow 0} \delta(x) = 0$ verwendet. Erfreulicherweise zeigt sich, dass das Verfahren selbst bei einem Start aus einem achsenparallelen Grid für die Positionen der Dirac-Impulse sinnvolle Ergebnisse liefert, oder doch zumindest liefern kann. Der Unterabschnitt „Bifurkationen“ von Abschnitt 7.1.2 zeigt allerdings auch, dass an Stellen $\underline{\eta}$ mit achsenparallel ausgerichteten Dirac-Impulsen vermehrt Bifurkationen auftreten können. Dabei handelt es sich jedoch nicht um eine Folge aus der Festsetzung $\delta(0) := 0$. Es bestätigt lediglich die intuitive Erwartung, dass G in der Umgebung von nicht glatten Stellen starken Änderungen unterworfen sein und somit in dieser Umgebung viele lokale Extremstellen besitzen kann.

Beispiel 5.2: Achsenparallel ausgerichtete Dirac-Impulse

Abbildung 5.2 zeigt links oben eine Approximation der Standardnormalverteilung, die mit dem Übertragungsansatz aus Abschnitt 6.2.1 aus einer eindimensionalen Approximation gewonnen wurde. Sie ist als Ü-16 bezeichnet. Die mit Hilfe dieses Ansatzes gewonnenen Approximationen zeichnen sich durch achsenparallel ausgerichtete Dirac-Impulse aus.

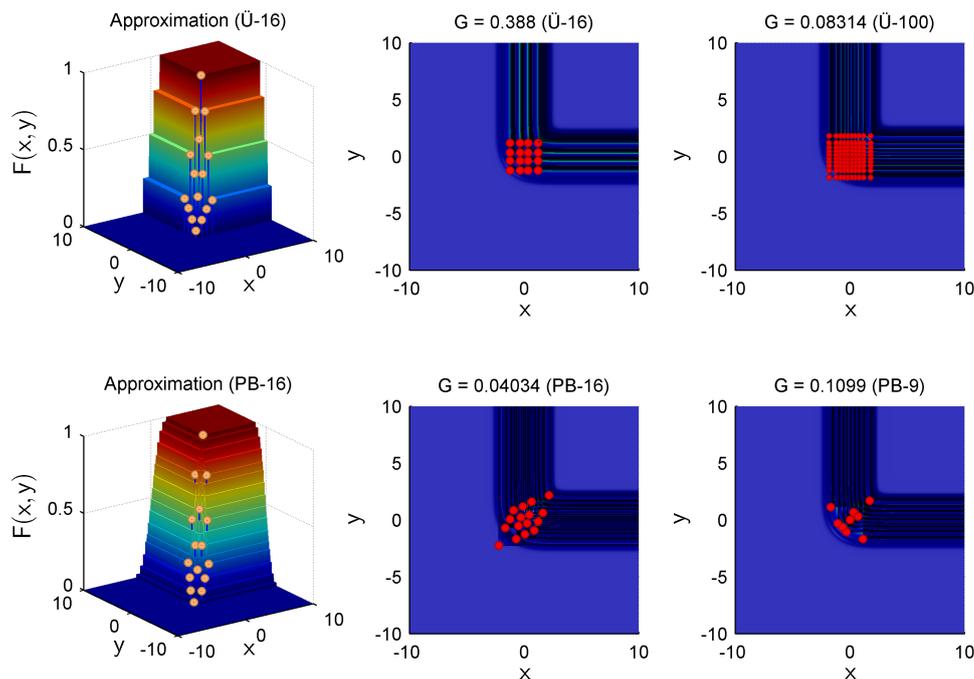


Abbildung 5.2: Approximationen mit achsenparallel ausgerichteten Dirac-Impulsen im Vergleich zu lokal optimalen Approximationen der Standardnormalverteilung. Die beiden linken Teilabbildungen zeigen die Verteilungsfunktionen verschiedener Approximationen der Standardnormalverteilung. Rechts daneben ist jeweils die quadratische Differenz zur Standardnormalverteilung abgebildet. Rechts außen sind die quadratischen Differenzen für zwei weitere Approximationen zu sehen.

Links unten in der Abbildung ist eine Approximation der Standardnormalverteilung zu sehen, die mit Hilfe des Progressiven Bayes-Verfahrens erstellt wurde, bezeichnet als PB-16. Obwohl beide Approximationen die gleiche Anzahl von 16 Dirac-Impulsen nutzen, ist diese Approximation wesentlich „glatter“ als die Approximation aus dem Übertragungsansatz. Die Ursache dafür sind die gut genutzten Überschneidungsbereiche der Heaviside-Funktionen.

Jeweils rechts neben den beiden Approximationen ist ihre quadratische Differenz zur Standardnormalverteilung abgebildet. Die Stellen, an denen sich Dirac-Impulse befinden, sind durch Marker gekennzeichnet. Die Gridstruktur der Approximation aus dem Übertragungsansatz ist deutlich zu erkennen.

Oben rechts in der Abbildung ist die Quadratische Differenz einer Approximation \ddot{U} -100 mit 100 Dirac-Impulsen zur Standardnormalverteilung zu sehen, die ebenfalls mit dem Übertragungsansatz gewonnen wurde. Darunter ist die Quadratische Differenz zur Standardnormalverteilung einer mit dem Progressiven Bayes-Verfahren gewonnenen Approximation PB-9 abgebildet. Sie besitzt lediglich

9 Dirac-Impulse.

In den Überschriften der Teilabbildungen, welche quadratische Differenzen zeigen, ist jeweils die Güte G für die zugehörigen Approximationen vermerkt. Die Güte von PB-9 ist mit 0.1099 zwar etwas schlechter als die von Ü-100 mit 0.08314, kann aber beinahe als gleich gut betrachtet werden und dass, obwohl PB-9 weniger als ein Zehntel der Dirac-Impulse zur Approximation nutzt als Ü-100. Ü-16 und PB-16 nutzen gleich viele Dirac-Impulse, die Approximation PB-16 aus dem Progressiven Bayes-Verfahren erreicht aber mit 0.04034 eine fast um eine ganze Größenordnung bessere Güte als Ü-16 mit 0.388. ■

5.5 Aufwandsanalyse

Für die Aufwandsanalyse des Progressiven Bayes-Verfahrens sei L die Anzahl der Dirac-Impulse in der Approximation sowie d die Dimension des zu Grunde liegenden \mathbb{R}^d . Abkürzend sei $n := L \cdot \{d + 1\}$.

Vorberechnete Terme Die ersten und zweiten partiellen Ableitungen des progressiven Gütemaßes in Richtung der Gewichte und Positionen kann entsprechend ihren Darstellungen (A.49) und (A.50) sowie denen aus Abschnitt 5.3 auf Basis der Terme $R(i, \underline{\eta})$, $R(i, m, \underline{\eta})$, $\tilde{R}(i, \underline{\eta}, \gamma)$ und $\tilde{R}(i, m, \underline{\eta}, \gamma)$ und deren partiellen Ableitungen berechnet werden. Für die \tilde{R} Terme werden dazu $O(\{L \cdot d\}^2)$ Operationen benötigt, sofern sie auf Basis der den Definitionen (A.13) bis (A.21) entsprechenden Terme berechnet werden.

Der Aufwand für die \tilde{R} Terme hängt natürlich wesentlich von der verwendeten progressiven Dichte ab. Für achsenparallele Gauß-Mischdichten in jedem Progressionsschritt γ mit K_P Komponenten müssen $O(L \cdot d \cdot K_P)$ Auswertungen der Gauß-Dichte- und Verteilungsfunktion zur Berechnung der Terme $T(i, q, k)$ und $S(i, k, q)$ aus Abschnitt A.5.4 erfolgen. Zusätzlich werden $O(L \cdot d^2 \cdot K_P)$ Additionen und Multiplikationen benötigt, um die \tilde{R} Terme darauf aufbauend auszuwerten.

Prädiktorschritt Der wesentliche Aufwand im Prädiktorschritt liegt zum einen in der Berechnung der Matrix \mathbf{P} und des Vektors \underline{b} sowie dem anschließenden Lösen des Linearen Gleichungssystems $\mathbf{P} \cdot \underline{\eta} \stackrel{!}{=} \underline{b}$, siehe Abschnitt „Prädiktorschritt“ in 3.2.4. Das Gauß'sche Eliminationsverfahren kann zum Lösen des Linearen Gleichungssystems eingesetzt werden. Es hat einen Aufwand von $O(n^3)$ für die $n \times n$ Matrix \mathbf{P} . Die Berechnung der Matrix \mathbf{P} und des Vektors \underline{b} kann auf Basis der R und \tilde{R} Terme geschehen und verursacht entsprechend der Darstellung von \mathbf{P} und \underline{b} in Abschnitt 5.2 nur einen zusätzlichen Aufwand von $O(n \cdot n + n)$.

Korrektorschritt Der Korrektorschritt führt mehrere Iterationen aus, um eine Nullstelle mit Hilfe des Newtonverfahrens zu bestimmen (siehe Abschnitt „Korrektorschritt“ in 3.2.4).

Sei I_N die Anzahl dieser Iterationen. Für I_N wird typischerweise eine relativ kleine Obergrenze, zum Beispiel 10, gewählt, ab welcher der Korrektorschritt erfolglos abbricht. Für jede Iteration sind die $n \times n$ Jakobi-Matrix $\mathbf{J} = \mathbf{J}(\xi) = \dot{g}(\xi)$ sowie $\frac{\partial}{\partial \xi} G(\xi, \gamma_{\text{neu}}) = g(\xi)$ aufzustellen. Diese Berechnungen können entsprechend der Darstellung von \mathbf{J} in Abschnitt 5.3 und den Darstellungen der partiellen Ableitungen von G aus (A.49) und (A.50) auf Basis der vorberechneten R und \tilde{R} Terme in $O(n \cdot n + n)$ erfolgen. Anschließend ist das Lineare Gleichungssystem (3.11) zu lösen, was mit dem Gauß'schen Eliminationsverfahren einen Aufwand von $O(n^3)$ erzeugt.

Iterationen Die Anzahl I der Iterationen bestehend aus Prädiktor- und Korrektorschritten variiert auf Grund der adaptiven Schrittweitenkontrolle. Eine Obergrenze wird durch die minimale Schrittweite induziert. Da diese Obergrenze in der Regel sehr weit von der tatsächlich benötigten Anzahl von Iterationen entfernt ist, wird I hier als gegeben vorausgesetzt.

Im späteren Abschnitt 6.2.2 werden initiale Approximationen als Startapproximationen eingeführt, die dann in Abschnitt 7.1 mit Hilfe des Progressiven Bayes-Verfahrens zu Approximationen der Zieldichte verbessert werden. Ist eine Gauß-Mischdichte mit K Komponenten als Zieldichte gegeben, so ist eine entsprechend Abschnitt 7.1 gewählte progressive Dichte zur progressiven Verbesserung initialer Approximationen in jedem Progressionsschritt eine Gauß-Mischdichte mit $K_P = L + K$ Komponenten. Das Progressive Bayes-Verfahren benötigt in diesem Fall

$$O(I \cdot I_N \cdot \{L \cdot \{L + K\} \cdot d\})$$

Auswertungen der Gauß-Dichte- und Verteilungsfunktion. Zusätzlich werden

$$O(I \cdot I_N \cdot \{L \cdot \{L + K\} \cdot d^2\})$$

mathematische Grundoperationen in Form von Additionen und Multiplikationen benötigt, um die Linearen Gleichungssysteme aufzustellen und weitere

$$O(I \cdot I_N \cdot \{L \cdot d\}^3)$$

Grundoperationen, um diese mit dem Gauß'schen Eliminationsverfahren zu lösen.

Freie Parameter

Das Progressive Bayes-Verfahren ist ein Verfahren zur Bestimmung einer Approximationsdichte für eine gegebene Originaldichte $\tilde{f}(\underline{x})$. Der Grundansatz, diese Approximation durch eine Progression zu bestimmen, ist recht allgemein. So gibt das Verfahren selber den Typ der Approximationsdichten nicht vor. Das Aufstellen des Differentialgleichungssystems ist stark von der Auswahl des Typs abhängig. In dieser Arbeit werden ausschließlich Dirac-Mischdichten als Approximationsdichten verwendet. Nach Auswahl dieses Dichtetyps zur Approximation sind immer noch folgende Parameter frei wählbar:

- Die Startdichte $\tilde{f}_0(\underline{x})$.
- Die Anzahl L der Dirac-Impulse in der Approximation.
- Die Startapproximation $\underline{\eta}_0$ der Startdichte in Form des Parametervektors $\underline{\eta}_0$ aus Gewichten und Positionen der Dirac-Impulse.
- Die progressive Dichte $\tilde{f}(\underline{x}, \gamma)$ zur Startdichte $\tilde{f}_0(\underline{x})$ mit Zieldichte $\tilde{f}(\underline{x})$.

Dieses Kapitel betrachtet die Anforderungen an die Startdichte, ihre Approximation sowie die Progression. Einige ausgewählte Spezialfälle werden dargestellt und bewertet.

6.1 Anforderungen

Ziel beim Bestimmen einer Approximation ist es, eine global optimale Approximation zu finden. Mit dem Progressiven Bayes-Verfahren kann eine optimale Approximation für die Zieldichte bestimmt werden, indem mit einer optimalen Approximation für die Startdichte begonnen und die Optimalität der Approximation während der Progression erhalten wird. Daraus leiten sich verschiedene Anforderungen an Startdichte und Progression ab.

6.1.1 Startdichte

Die Anzahl L der Dirac-Impulse zur Approximation wird typischerweise in Abhängigkeit von gewünschter Güte und verfügbarer Rechenzeit gewählt. Die Kenntnis einer optimalen Approximation der Startdichte bezieht sich genauer gesagt immer auf ein vorgegebenes L . Möglicherweise ist eine optimale Approximation nur für bestimmte L bekannt, zum Beispiel nur für gerade Anzahlen. Offensichtlich ist es wünschenswert, eine optimale Approximation für möglichst viele, am besten alle $L \geq 1$ zu kennen, um so die Auswahl von L hinsichtlich der gewünschten Güte und zur Verfügung stehenden Rechenzeit nicht zusätzlich beschränken zu müssen.

Die Startdichte geht Hand in Hand mit der Progression, da sie durch die Progression möglichst glatt in die Zieldichte überführt werden soll. Andernfalls kann es zu einem harten Bruch zwischen der progressiven Dichte für $\gamma = 0$ — also der Startdichte — und den für γ nahe bei Null liegenden progressiven Dichten zum Progressionszeitpunkt γ kommen. Die Kenntnis einer optimalen Approximation für die Startdichte wäre mithin wertlos, da sie nicht progressiv in Richtung der Zieldichte angepasst werden kann: das Verfahren endet sofort in einer Sackgasse. Selbst wenn die Progression die Startdichte glatt in die Zieldichte überführt, so hat die Startdichte einen großen Einfluss auf die im Verlauf der Progression auftretenden Bifurkationen und Sackgassen.

Die Startapproximation hängt unmittelbar mit der Startdichte zusammen. Die Idealvorstellung besteht in einer global optimalen Approximation der Startdichte. Die globale Optimalität ist jedoch nicht zwingend erforderlich, lokale Optimalität genügt schon, um als Startwert zur Lösung des Differentialgleichungssystems eingesetzt werden zu können. Für das numerische Lösungsverfahren ist es sogar ausreichend, wenn die Approximation $\underline{\eta}_0$ nahe genug bei einer lokalen Minimalstelle von $G(\underline{\eta}) = G(\underline{\eta}, 0)$ liegt, so dass das Newtonverfahren im Korrektorschritt nach der ersten Prädiktion noch (schnell genug) konvergiert.

6.1.2 Progression

Die Progression in Form der progressiven Dichte hat einen wesentlichen Einfluss auf das Laufzeitverhalten des Progressiven Bayes-Verfahrens. Neben den schon im vorhergehenden Abschnitt genannten Anforderungen im Zusammenhang mit der Startdichte sollte die Berechnung der Matrizen \mathbf{P} und \mathbf{J} mit möglichst wenig Rechenaufwand erfolgen können. Insbesondere ist eine rein numerische Berechnung der Integrale in den \tilde{R} Termen und ihren partiellen Ableitungen aus Abschnitt 5.1 beziehungsweise A.5.1 in der Regel viel zu langsam. Die analytische Lösbarkeit dieser Integrale ist praktisch zwingend erforderlich und hängt wesentlich von der Form der progressiven Dichte ab.

Damit die Approximation bei Änderung von γ mitgeführt werden kann, muss $\frac{\partial}{\partial \eta} G(\underline{\eta}, \gamma)$ in Richtung von γ ableitbar sein. Dies setzt eine gewisse Glattheit der Progression voraus, die zur Vermeidung von Bifurkationen und Sackgassen möglichst groß sein sollte. Zusätzlich muss die Progression glatt sein, um große Schrittweiten im Prädiktorschritt zu ermöglichen, was über

die Vermeidung von Bifurkationen und Sackgassen aus einer rein mathematischen Perspektive hinaus geht.

6.2 Spezielle Startdichten

6.2.1 Gleichverteilung

Für diesen Abschnitt, genauso wie für die Berechnung der Terme in Abschnitt A.5.3, wird davon ausgegangen, dass das Intervall der betrachteten Gleichverteilung mit dem Integrationsintervall $[\underline{c}_1, \underline{c}_2]$ im Gütemaß übereinstimmt.

Konkrete Vorteile der Gleichverteilung als Startverteilung sind ihre mathematisch leichte Handhabung und ein geringer Rechenaufwand in der Progression. Zu erwarten ist, dass die Gleichverteilung nur eine optimale Approximation besitzt, gegebenenfalls in an den Achsen gespiegelten Versionen in Abhängigkeit von der Dimension d . Die Dirac-Impulse in einer optimalen Approximation sollten relativ gleichmäßig über das gesamte Intervall $[\underline{c}_1, \underline{c}_2]$ verteilt und insbesondere dadurch für viele Zielverteilungen geeignet sein. Zudem ist zu erwarten, dass die optimale Approximation von anderen lokal optimalen Approximationen separiert ist.

Die Dichte einer Gleichverteilung ist konstant auf dem ganzen betrachteten Intervall. Für den eindimensionalen Fall ergibt sich daraus die intuitiv richtige Vorstellung, dass die Dirac-Impulse einer optimalen Approximation äquidistant über dem ganzen Intervall verteilt sein und gleiche Gewichte haben sollten. Im Mehrdimensionalen trifft dies jedoch nicht zu, siehe dazu den nachfolgenden Unterabschnitt „Übertragung aus dem Eindimensionalen“.

Das Finden einer optimalen Approximation für die Gleichverteilung im Mehrdimensionalen erweist sich als nicht ganz einfach. In den drei folgenden Unterabschnitten werden drei Ansätze zum Finden von (lokal) optimalen Approximationen für die Gleichverteilung beschrieben.

Analytischer Ansatz

Die notwendigen Bedingungen (5.7)

$$R(i, \underline{\eta}_{\min}) \stackrel{!}{=} \tilde{R}(i, \underline{\eta}_{\min}, \gamma)$$

sowie (5.8)

$$R(i, m, \underline{\eta}_{\min}) \stackrel{!}{=} \tilde{R}(i, m, \underline{\eta}_{\min}, \gamma)$$

für ein globales Minimum des Gütemaßes an der Stelle $\underline{\eta}_{\min} = (w_1, \dots, w_L, \underline{x}_1^T, \dots, \underline{x}_L^T)^T$ lauten für die Gleichverteilung konkret

$$\begin{aligned} & \sum_{l=1}^L w_l \prod_{q=1}^d \{c_{2q} - \max(x_{iq}, x_{lq})\} \\ & \stackrel{!}{=} \left\{ \prod_{q=1}^d \{c_{2q} - c_{1q}\} \right\}^{-1} \cdot \prod_{q=1}^d \left\{ \frac{1}{2} \{c_{2q}^2 - x_{iq}^2\} + c_{2q} \{c_{2q} - x_{iq}\} \right\} \end{aligned}$$

und

$$\begin{aligned} & \sum_{l=1}^L w_l \prod_{q=1, q \neq m}^d \{c_{2q} - \max(x_{iq}, x_{lq})\} H(x_{im} - x_{lm}) \\ & \stackrel{!}{=} \left\{ \prod_{q=1}^d \{c_{2q} - c_{1q}\} \right\}^{-1} \cdot \{x_{im} + c_{2q}\} \prod_{q=1, q \neq m}^d \left\{ \frac{1}{2} \{c_{2q}^2 - x_{iq}^2\} + c_{2q} \{c_{2q} - x_{iq}\} \right\} . \end{aligned}$$

Siehe Darstellung (A.24) beziehungsweise (A.29) von $R(i, \underline{\eta})$ beziehungsweise $R(i, m, \underline{\eta})$ und Abschnitt A.5.3 für $\tilde{R}(i, \underline{\eta}, \gamma)$ und $\tilde{R}(i, m, \underline{\eta}, \gamma)$.

Dieses Gleichungssystem kann leider nicht ohne Weiteres analytisch nach w_i und x_{im} für $i = 1, \dots, L$ und $m = 1, \dots, d$ aufgelöst werden. Zudem beschränkt sich die Lösungsmenge nicht auf eine globale Minimalstelle des Gütemaßes und ihrer an den Achsen gespiegelten Versionen, sondern beinhaltet auch lokale Minimalstellen. Ein Blick auf das Gleichungssystem lässt erahnen, dass es von diesen einige gibt.

Wird eine — natürlich von vorne herein die Lösungsmenge beschränkende — Ordnungsannahme für die Positionen getroffen, so können die max- und Heaviside-Funktionen ausgewertet werden. Übrig bleibt dann ein nicht lineares polynomielles Gleichungssystem in den freien Variablen w_i und x_{im} . Das Finden einzelner Lösungen mit numerischen Lösungsverfahren ist dann gegebenenfalls möglich, ohne allerdings eine Aussage hinsichtlich der lokalen oder globalen Optimalität zu liefern.

Übertragung aus dem Eindimensionalen

Sind w_i und x_i , $i = 1, \dots, L$, die Gewichte und Positionen einer eindimensionalen Approximation einer eindimensionalen Gleichverteilung, so kann daraus eine Approximation einer zweidimensionalen Gleichverteilung konstruiert werden. Dazu werden x_1 bis x_L jeweils auf der x - und y -Achse aufgetragen. Jeder Schnittpunkt im entstehenden Raster bildet eine Stelle, an der ein Dirac-Impuls der zweidimensionalen Approximation sitzt. Das Gewicht an der Stelle $\begin{pmatrix} x_i \\ x_j \end{pmatrix}$, $i, j = 1, \dots, L$, ist gleich dem Produkt $w_i \cdot w_j$.

Beispiel 6.1: Übertragung einer optimalen eindimensionalen Approximation

Abbildung 6.1 zeigt die Übertragung einer optimalen eindimensionalen Approximation auf eine zweidimensionale Gleichverteilung. Dabei wurde keine Normierung der Gewichte auf Eins vorausgesetzt. Die entstehende zweidimensionale Approximation ist lokal optimal. ■

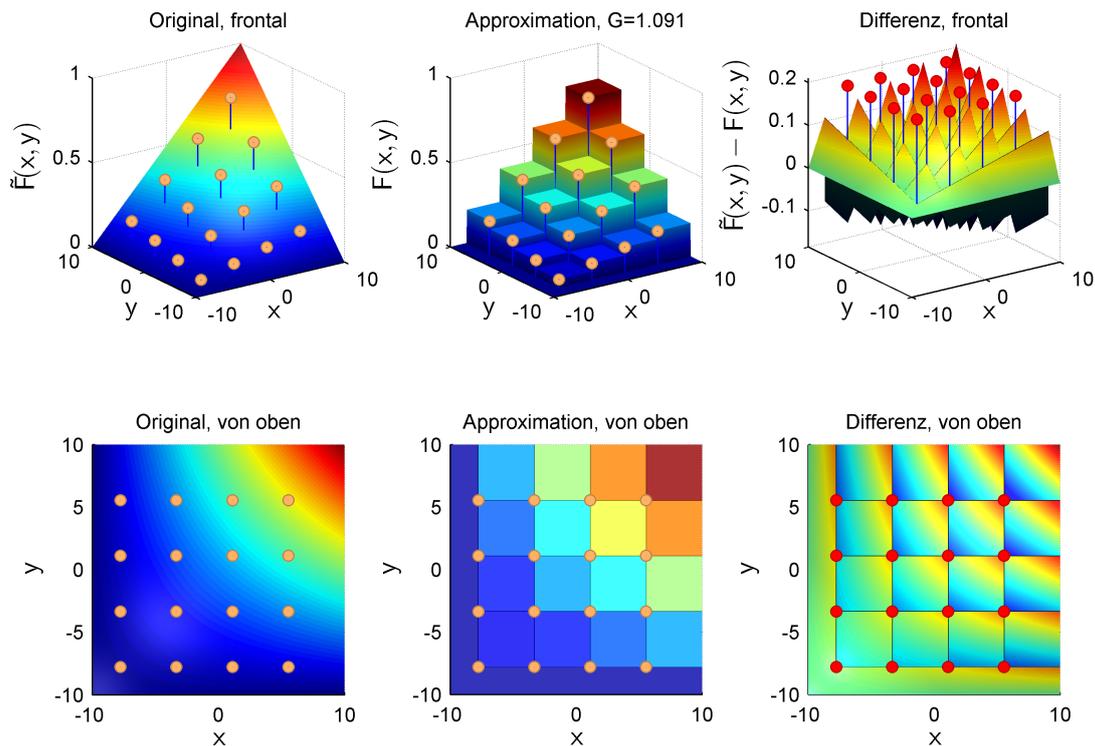


Abbildung 6.1: Zweidimensionale Gleichverteilung approximiert mit 16 Dirac-Impulsen basierend auf einer optimalen Approximation einer eindimensionalen Gleichverteilung. Dargestellt ist die Verteilungsfunktion der Gleichverteilung in den beiden linken Teilabbildungen, sowie die Verteilungsfunktion der Approximation in der Mitte. Markiert sind jeweils die Positionen, an denen sich Heaviside-Funktionen in der Approximation befinden. In der Darstellung der Differenz (rechts) sind analog die Positionen der Dirac-Impulse markiert. Die Höhe der Marker ist auf den Maximalwert der Differenz normiert und spiegelt wider, dass alle Dirac-Impulse das gleiche Gewicht haben.

Viele Variationen einer Gridstruktur der Positionen der Dirac-Impulse in einer Approximation sind möglich, auch ohne auf diese Übertragungstechnik zurück zugreifen. Sie sind alle schon deshalb nicht global optimal, weil sie die Überschneidungsbereiche der Heaviside-Funktionen nicht nutzen.

Der Übertragungsansatz kann auch auf andere achsenparallel ausgerichtete Verteilungen angewendet werden, zum Beispiel auf die Standardnormalverteilung (siehe auch Beispiel 5.2).

Progressive Approximation

Eine weitere Möglichkeit, Approximationen für die Gleichverteilung zu bestimmen, ist das Progressive Bayes-Verfahren selbst zu benutzen. Offensichtlich kann die Gleichverteilung in diesem Fall nicht selbst die Startverteilung sein. Schließlich wird ja deshalb eine optimale Approximation für die Gleichverteilung gesucht, um sie überhaupt als Startverteilung nutzen zu können.

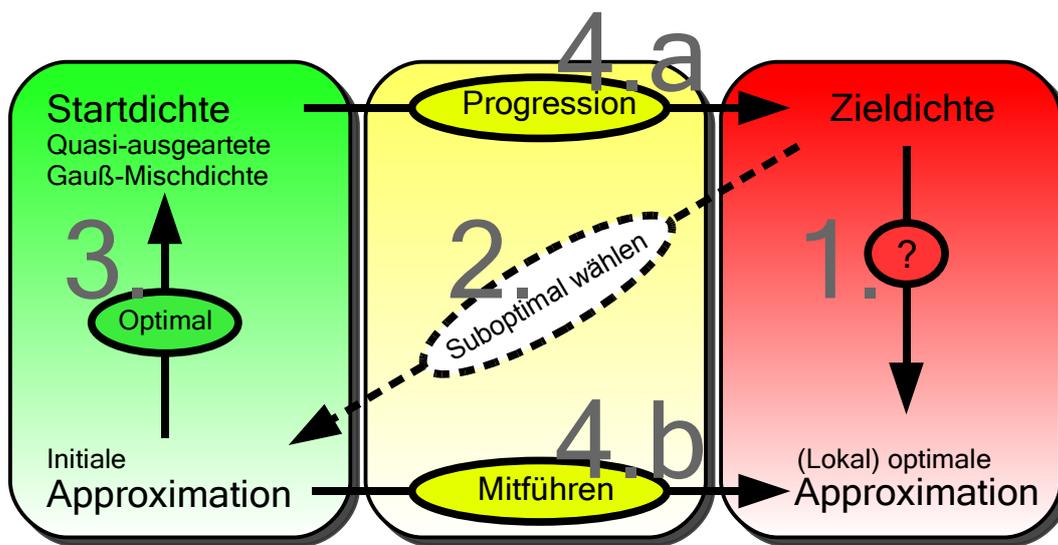


Abbildung 6.2: Schematische Darstellung zur progressiven Verbesserung initialer Approximationen (siehe Abschnitt 7.1), basierend auf dem Progressiven Bayes-Verfahren — vergleiche auch mit Abbildung 3.1. Gesucht wird eine (lokal) optimale Approximation für eine gegebene Zieldichte, die allerdings nicht direkt bestimmt werden kann (1.). Da im Mehrdimensionalen aus den notwendigen Bedingungen (5.7) und (5.8) keine optimale Approximation für die Gleichverteilung bestimmt werden kann, wird statt einer Startdichte eine initiale Approximation vorgegeben. Diese wird an Hand der Zieldichte gewählt (2.) und eine dazu passende Startdichte wird konstruiert (3.). Anschließend wird die Startdichte in die Zieldichte überführt (4.a) unter gleichzeitiger Mitführung und Anpassung der initiale Approximation (4.b).

In Abschnitt 7.1.2 werden lokal optimale Approximationen der Gleichverteilung mit Hilfe initialer Approximationen als Startapproximationen bestimmt. Initiale Approximationen werden im folgenden Abschnitt eingeführt.

6.2.2 Quasi-ausgeartete Gauß-Mischdichten/Initiale Approximationen

Das Finden einer optimalen Approximation für die Gleichverteilung ist ein offenes Problem. Da die Gleichverteilung als eine der einfachsten Verteilungen überhaupt betrachtet werden kann, stellt sich die Frage, ob es überhaupt eine einfache Verteilung gibt, für welche eine optimale Approximation leicht ersichtlich ist. Für eine d -dimensionale Hyperebene als Verteilungsfunktion, welche linear ist und somit ebenfalls als etwas Einfaches betrachtet werden kann, ist das Auffinden einer optimalen Approximation beispielsweise nicht wirklich leichter als für eine d -dimensionale Gleichverteilung: das Gleichungssystem aus den notwendigen Bedingungen (5.7) und (5.8) enthält weiterhin max- und Heaviside-Funktionen sowie quadratische Terme in den freien Variablen, siehe auch Abschnitt „Analytischer Ansatz“ in 6.2.1.

Die Grundidee des Progressiven Bayes-Verfahren ist, mit einer einfachen Startdichte zu beginnen, für die eine optimale Approximation bekannt ist, diese dann progressiv in die Zieldichte zu überführen und die optimale Approximation dabei mitzuführen, siehe Abbildung 3.1.

Auch wenn eine *optimale* Approximation der Zieldichte nicht bekannt ist, besteht doch oft zumindest eine Vorstellung davon, wie eine *suboptimale* Approximation der Zieldichte aussehen könnte — und sei es auch einfach nur durch Ziehen einer zufälligen Stichprobe. Eine etwas andere Sichtweise auf die Grundidee des Progressiven Bayes-Verfahrens ist, mit einer suboptimalen Approximation der Zieldichte als Startapproximation zu beginnen und diese in eine bessere lokal optimale Approximation für die Zieldichte zu überführen.

Definition 6.2.1 Eine initiale Approximation ist eine (suboptimale) Approximation der Zieldichte, die als Startapproximation benutzt wird. \square

Damit kehrt sich das Problem, zu einer Startdichte eine optimale Approximation zu finden, um (siehe Abbildung 6.2): Gegeben ist eine initiale Approximation, die als Startapproximation genutzt werden soll. Gibt es eine Startdichte, welche die initiale Approximation als optimale Approximation besitzt?

Die Antwort darauf fällt positiv aus: Wähle als Startdichte die initiale Approximationsdichte selbst. Tatsächlich ist für die so gewählte Startdichte die vorgegebene initiale Approximationsdichte die einzige *optimale* Approximation, positive Gewichte und paarweise verschiedene Dirac-Impulse in der gegebenen initialen Approximationsdichte vorausgesetzt.

Allerdings führt beim Aufstellen der Matrizen \mathbf{P} und \mathbf{J} der Typ der Approximationsdichten in Form von Dirac-Mischdichten zu Irregularitäten, siehe Abschnitt 5.4. Um weitere Irregularitäten durch die Verwendung einer Dirac-Mischdichte als Startdichte zu vermeiden, kann diese durch eine quasi-ausgeartete Gauß-Mischdichte ersetzt werden.

Definition 6.2.2 Eine quasi-ausgeartete Gauß-Mischdichte ist eine Gauß-Mischdichte, für welche die Einträge in der Kovarianzmatrix jeder einzelnen Komponente sehr klein sind. \square

Nach Lemma 2.2.7 entspricht eine achsenparallele Gauß-Dichte mit sehr kleinen Varianzen fast einem Dirac-Impuls, falls der Mittelwert der Gauß-Dichte mit der Position des Dirac-Impulses übereinstimmt.

Definition 6.2.3 Gegeben sei eine initiale Approximation $f(\underline{x}, \underline{\eta}) = \sum_{l=1}^L w_l \delta(\underline{x} - \underline{x}_l)$. Dann wird jede achsenparallele quasi-ausgeartete Gauß-Mischdichte $\tilde{f}_0(\underline{x}) = \sum_{l=1}^L w_l \mathcal{N}(\underline{x}, \underline{x}_l, \underline{\sigma}_l)$ mit Standardabweichungen $\underline{\sigma}_l \approx \underline{0}$ als zu $f(\underline{x}, \underline{\eta})$ gehörig bezeichnet. \square

Sind die Standardabweichungen einer zu einer initialen Approximation $f(\underline{x}, \underline{\eta})$ gehörigen achsenparallelen quasi-ausgearteten Gauß-Mischdichte $\tilde{f}_0(\underline{x})$ klein genug, so ist $f(\underline{x}, \underline{\eta})$ fast eine optimale Approximation von $\tilde{f}_0(\underline{x})$ und $\tilde{f}_0(\underline{x})$ kann als Startdichte mit Startapproximation $f(\underline{x}, \underline{\eta})$ im numerischen Lösungsverfahren benutzt werden.

Wie eine geeignete progressive Dichte zu einer gegebenen initialen Approximation aufgestellt werden kann, zeigt Abschnitt 7.1. Dabei wird die Form der Startdichte als quasi-ausgeartete Gauß-Mischdichte gewinnbringend ausgenutzt.

Der größte Vorteil initialer Approximationen ist die freie Wahl der Gewichte und Positionen ihrer Dirac-Impulse. Ihre Auswahl kann — muss aber nicht notwendigerweise — in Abhängigkeit von der Zieldichte erfolgen, wodurch das Einbringen von Wissen über die Zieldichte möglich wird, was den Begriff „initiale Approximation“ motiviert.

Andererseits liegt in der Wahl von Gewichten und Positionen auch der größte Nachteil: eine geeignete Wahl ist nicht unbedingt offensichtlich. Zudem ist die Approximation am Ende der Progression — sprich die Approximation der Zieldichte — stark von der initialen Approximation abhängig. Die initiale Approximation bestimmt also wesentlich, wie nahe die lokal optimale Approximation am Ende der Progression einer global optimalen Approximation der Zieldichte kommen kann.

6.3 Spezielle Progressionen

6.3.1 Variation der Parameter

Eine sehr einfache Progression kann gegebenenfalls aufgestellt werden, falls die Zieldichte

$$\tilde{f}(\underline{x}) = \tilde{f}(\underline{x}, \underline{\kappa})$$

eine parametrische Form hat für einen bestimmten Parametervektor $\underline{\kappa} = \underline{\kappa}_1$. Der Parametervektor

$$\underline{\kappa} = \underline{\kappa}(\gamma)$$

kann dann vom Progressionsparameter γ abhängig gemacht werden, so dass

$$\tilde{f}(\underline{x}, \gamma) := \tilde{f}(\underline{x}, \underline{\kappa}(\gamma))$$

als progressive Dichte gewählt werden kann. Für $\gamma = 1$ muss $\underline{\kappa}(1) = \underline{\kappa}_1$ gelten, damit die progressive Dichte am Ende der Progression tatsächlich der Zieldichte entspricht. Eine Schwierigkeit bei der ansonsten freien Wahl für $\underline{\kappa}(\gamma)$ besteht in der Festlegung von $\underline{\kappa}(0)$ und damit der Startdichte der Progression, da eine geeignete Startapproximation gefunden werden muss, die zur Startdichte passt. Zudem muss die entstehende Progression die Anforderungen aus Abschnitt 6.1.2 in einem ausreichendem Maße erfüllen.

In Anlehnung an [11] kann die Progression für eine eindimensionale Gauß-Mischdichte beispielsweise aufgestellt werden, indem die Varianz σ_k der k -ten Gauß-Komponente für kleinere γ immer weiter ausgedehnt wird,

$$\sigma_k(\gamma) := \frac{1 + \epsilon}{\gamma + \epsilon} \sigma_k$$

für ein fest gewähltes kleines $\epsilon \in \mathbb{R}$. Der Parametervektor $\underline{\kappa}(\gamma) = \underline{\kappa}(\gamma, \epsilon)$ ist folglich vom vorgegebenen ϵ abhängig. Gewichte und Mittelwerte der Gauß-Komponenten bleiben während der Progression konstant. Dann entspricht die progressive Dichte für $\gamma = 1$ der Zieldichte,

$$\tilde{f}(x, \underline{\kappa}(1, \epsilon)) = \tilde{f}(x, \underline{\kappa}_1) = \tilde{f}(x)$$

während die Startverteilungsfunktion $\tilde{F}_\epsilon(x) := \tilde{F}(x, \underline{\kappa}(0, \epsilon))$ einer endlichen Gleichverteilungsfunktion $\mathcal{U}(x, \alpha(\epsilon), \beta(\epsilon))$ auf dem von ϵ abhängigen Intervall $[\alpha(\epsilon), \beta(\epsilon)]$ um so ähnlicher ist, je kleiner ϵ gewählt wird. Dabei sind für ein festes ϵ die Unter- und Obergrenzen $\alpha(\epsilon)$ und $\beta(\epsilon)$ so gewählt, dass der Abstand im quadratischen Mittel zur Startverteilungsfunktion $\tilde{F}_\epsilon(x)$ minimal ist,

$$(\alpha(\epsilon), \beta(\epsilon)) := \operatorname{argmin}_{(\alpha(\epsilon), \beta(\epsilon))} \int_{-\infty}^{\infty} \left\{ \tilde{F}_\epsilon(x) - \mathcal{U}(x, \alpha(\epsilon), \beta(\epsilon)) \right\}^2 dx .$$

Es folgt, dass es zu jedem $\Delta > 0$ ein $\epsilon > 0$ gibt, so dass für alle $x \in \mathbb{R}$ gleichzeitig

$$\left| \tilde{F}(x, \underline{\kappa}(0, \epsilon)) - \mathcal{U}(x, \alpha(\epsilon), \beta(\epsilon)) \right| < \Delta$$

gilt.

Vorteilhaft an dieser Vorgehensweise ist, dass die entstehende Startverteilung praktisch als Gleichverteilung betrachtet werden kann, für die im Eindimensionalen eine optimale Approximation leicht bestimmt werden kann. Da die Integrationsgrenzen \underline{c}_1 und \underline{c}_2 in der Definition des Gütemaßes fest sind und nicht mit ϵ geändert werden, wandert allerdings immer mehr Masse der Verteilung aus dem betrachteten Intervall $[\underline{c}_1, \underline{c}_2]$ ab, je kleiner ϵ gewählt wird. Das führt leider dazu, dass der Prädiktorschritt dazu neigt, die Dirac-Impulse aus dem Intervall $[\underline{c}_1, \underline{c}_2]$ herauszuschieben. Durch Fixieren eines Dirac-Impulses an der Position \underline{c}_1 und eines weiteren bei \underline{c}_2 sowie mit Hilfe der Ordnungsannahme kann dies im eindimensionalen Fall verhindert werden. Im Mehrdimensionalen scheitert dieser Ansatz auf Grund der fehlenden Ordnungsannahme.

6.3.2 Konstante Progression

Definition 6.3.1 Für eine gegebene Zieldichte $\tilde{f}(\underline{x})$ heißt

$$\tilde{f}(\underline{x}, \gamma) := \tilde{f}(\underline{x})$$

konstante Progressionsdichte zur Zieldichte $\tilde{f}(\underline{x})$. $\tilde{f}(\underline{x}, \gamma)$ hängt effektiv nicht von γ ab und entspricht in jedem Progressionsschritt γ der Zieldichte.

Für sich genommen macht diese Progressionsform im Rahmen des Progressiven Bayes-Verfahrens wenig Sinn, sie kann jedoch in Verbindung mit einer konvexen Kombination eingesetzt werden, die im nächsten Abschnitt eingeführt wird.

6.3.3 Konvexe Kombination

Definition 6.3.2 Sind $f^*(\underline{x}, \gamma)$ sowie $f^\circ(\underline{x}, \gamma)$ zwei progressive Dichten, dann heißt die progressive Dichte

$$\tilde{f}(\underline{x}, \gamma) := \{1 - \gamma\} f^*(\underline{x}, \gamma) + \gamma f^\circ(\underline{x}, \gamma)$$

konvexe Kombination von f^* und f° . Ihre Startdichte entspricht der Startdichte von f^* während ihre Zieldichte der Zieldichte von f° entspricht. \square

Die konvexe Kombination hat mehrere Vorteile. So dürfen f^* und $\overset{\circ}{f}$ beliebig sein. Sie müssen keiner parametrischen Form entsprechen, insbesondere müssen sie nicht ein und dieselbe parametrische Form haben.

Die für das Aufstellen der Matrizen \mathbf{P} und \mathbf{J} benötigten Terme können leicht auf entsprechende Terme für f^* und $\overset{\circ}{f}$ zurückgeführt werden, siehe Abschnitt A.5.2. Außerdem ist der zusätzlich induzierte Rechenaufwand sehr gering.

Als Spezialfall kann $f^*(\underline{x}, \gamma) := \tilde{f}_0(\underline{x})$ als konstant entsprechend einer vorgegebenen Start- und $\overset{\circ}{f}(\underline{x}, \gamma) := \tilde{f}(\underline{x})$ entsprechend einer vorgegebenen Zieldichte gewählt werden.

Die konvexe Kombination erfüllt die Anforderung der Glattheit nicht besser als dies f^* und $\overset{\circ}{f}$ selber tun, führt aber andererseits selbst zu keinen weiteren Irregularitäten.

Spezielle Kombinationen aus Startdichte und Progression

7.1 Progressive Verbesserung initialer Approximationen

In Abschnitt 6.2.2 wurden initiale Approximationen und quasi-ausgeartete Gauß-Mischdichten als Startdichten eingeführt. Dieser Abschnitt beschäftigt sich mit der Frage, wie eine solche Startdichte glatt in die Zieldichte überführt werden kann. Für die spezielle Wahl von Gleichverteilung und Gauß-Verteilungen als Zielverteilungen wird anschließend jeweils ein Schema initialer Approximationen für den zweidimensionalen Fall vorgestellt, auf die der im Folgenden vorgestellte Ansatz zur progressiven Verbesserung initialer Approximationen erfolgreich angewendet werden kann.

Definition 7.1.1 *Gegeben sei eine Originaldichte $\tilde{f}(\underline{x})$ mit initialer Approximation $f(\underline{x}, \underline{\eta}_0)$. Ist $\tilde{f}(\underline{x}, \gamma)$ eine progressive Dichte mit Zieldichte $\tilde{f}(\underline{x})$ und einer zu $f(\underline{x}, \underline{\eta}_0)$ passenden Startdichte $\tilde{f}_0(\underline{x})$, so wird jede aus der Startapproximation $f(\underline{x}, \underline{\eta}_0)$ mit Hilfe des Progressiven Bayes-Verfahrens gewonnene Approximation der Zieldichte $\tilde{f}(\underline{x})$ als progressiv verbesserte Version der initialen Approximation $f(\underline{x}, \underline{\eta}_0)$ betrachtet. \square*

Es bleibt zu zeigen, wie eine geeignete progressive Dichte aufgestellt werden kann.

Gegeben sei dazu eine initiale Approximation $f(\underline{x}, \underline{\eta}_0)$ mit positiven Gewichten w_l und paarweise verschiedenen Positionen \underline{x}_l der Dirac-Impulse, $l = 1, \dots, L$. Zu dieser initialen Approximation sei

$$\tilde{f}_0(\underline{x}) := \sum_{l=1}^L \omega_l \mathcal{N}(\underline{x}, \underline{\mu}_l, \underline{\sigma}_l)$$

eine der Definition 6.2.3 entsprechende zugehörige Startdichte in Form einer achsenparallelen quasi-ausgearteten Gauß-Mischdichte mit Gewichten $\omega_l = w_l$ und Standardabweichungen $\underline{\sigma}_l \approx 0$. Jede ihrer Komponenten fixiert einen Dirac-Impuls der initialen Approximation in ihrem Mittelwert $\underline{\mu}_l = \underline{x}_l$. Damit im Verlauf der Progression eine bessere Approximation der Zieldichte erreicht werden kann, muss diese Fixierung gelöst und den Dirac-Impulsen damit mehr Spielraum in ihrer Positionierung gelassen werden. Ein langsames Lösen der Fixierung in

der progressiven Version

$$\tilde{f}_0(\underline{x}, \gamma) := \sum_{l=1}^L \omega_l \mathcal{N}(\underline{x}, \underline{\mu}_l, \underline{\sigma}_l(\gamma))$$

von $\tilde{f}_0(\underline{x})$ kann durch ein langsames Aufweiten der Standardabweichungen $\underline{\sigma}_l(\gamma)$ im Verlauf der Progression erreicht werden. Für $\gamma = 0$ gilt dabei $\underline{\sigma}_l(0) = \underline{\sigma}_l$.

$f_0(\underline{x}, \gamma)$ ist eine progressive Dichte mit Startdichte $\tilde{f}_0(\underline{x})$ und Zieldichte $\tilde{f}_0(\underline{x}, 1)$. $\tilde{f}_0(\underline{x}, 1)$ ist im Allgemeinen ungleich der vorgegebenen Originaldichte $\tilde{f}(\underline{x})$. Mit Hilfe einer konvexen Kombination

$$\tilde{f}(\underline{x}, \gamma) := \{1 - \gamma\} \cdot \tilde{f}_0(\underline{x}, \gamma) + \gamma \cdot \tilde{f}(\underline{x})$$

aus $\tilde{f}_0(\underline{x}, \gamma)$ und der konstanten Progression zur Zieldichte $\tilde{f}(\underline{x})$ ergibt sich eine progressive Dichte $\tilde{f}(\underline{x}, \gamma)$, welche die Startdichte $\tilde{f}_0(\underline{x})$ in die Zieldichte $\tilde{f}(\underline{x})$ überführt.

Abgesehen von der Festlegung $\underline{\sigma}_l(0) = \underline{\sigma}_l$ kann die von γ abhängige Funktion $\underline{\sigma}_l(\gamma)$ beliebig gewählt werden. Tatsächlich können auch ω_l und $\underline{\mu}_l$ von γ abhängig gemacht werden. Die Berechnungen in Abschnitt A.5.4 für die zum Aufstellen der Matrizen \mathbf{P} und \mathbf{J} benötigten Terme gehen von diesem allgemeinen Fall aus. Die Kunst beim Aufstellen einer Progression besteht jedoch ganz wesentlich darin, nicht unnötig für Bifurkationen oder Sackgassen zu sorgen. Eine Faustregel ist daher, die progressive Dichte möglichst wenig in Abhängigkeit von γ zu ändern. Für die im Rahmen dieser Arbeit berechneten progressiven Approximationen wurden daher nur die Standardabweichungen variiert und zwar in Form einer Linearkombination $\underline{\sigma}_l(\gamma) := \{1 - \gamma\} \cdot \underline{\sigma}_0 + \gamma \cdot \underline{\sigma}_1$ aus für die von l unabhängigen Werte $\underline{\sigma}_0 \approx \underline{0}$ und $\underline{\sigma}_1$. $\underline{\sigma}_1$ bestimmt, wie schnell die Standardabweichungen aufgeweitet werden. Ein Wert von $\underline{\sigma}_1 = \underline{1}$ hat sich dabei als praktikabel erwiesen.

Die auf diese Weise erstellten progressiven Dichten überführen die zur initialen Approximation gehörige Startdichte glatt in die vorgegebene Zieldichte im Sinne der Ableitbarkeit von $\frac{\partial}{\partial \eta} G(\underline{\eta}, \gamma)$ nach γ — unter geeigneten Voraussetzungen an die Zieldichte $\tilde{f}(\underline{x})$. Abschnitt 6.1.2 fordert darüber hinaus Glattheit im Sinne von wenigen Bifurkationen oder gar Sackgassen sowie der Möglichkeit großer Schrittweiten im Prädiktorschritt des numerischen Lösungsverfahrens. Diese Punkte werden zumindest hinreichend gut erfüllt, wie die Anwendung dieser Progressionsform im Rahmen von Kapitel 8 zeigt. In den Abbildungen 7.13 bis 7.16 zu Beispiel 7.5 ist eine progressive Dichte und ihre Verteilungsfunktion zu verschiedenen Progressionszeitpunkten γ zu sehen, die der hier vorgestellten Progressionsform entspricht.

7.1.1 Auswahl initialer Approximationen

Eine initiale Approximation wird durch die Gewichte und Positionen ihrer Dirac-Impulse eindeutig bestimmt. Auch wenn sich die Idee und der Begriff „initiale Approximation“ an einer Wahl der Gewichte und Positionen in Abhängigkeit von der Zieldichte anlehnt, ist sie prinzipiell nicht weiter beschränkt, solange die Gewichte positiv und die Positionen paarweise verschieden sind.

Praktisch wird die Wahl dadurch eingeschränkt, dass die Progression nicht in eine Sackgasse laufen soll. Je mehr Dirac-Impulse in der initialen Approximation verwendet werden, desto schwieriger ist es im Allgemeinen, die Progression so zu wählen, dass dies nicht geschieht.

Im Speziellen führt die Wahl einer lokal optimalen Approximation der Zieldichte als initiale Approximation oft zu einer Sackgasse, da anschaulich betrachtet wenig Spielraum während der Progression besteht. Eine Strategie ist daher, alle Gewichte gleich groß zu wählen, so dass ihre Summe dem Integral über der Zieldichte entspricht. Somit bleibt der progressiven Approximation zumindest das Potential der Gewichtsoptimierung. Zudem ist die konkrete Festlegung geeigneter Gewichte in der Regel wesentlich komplizierter, als das Festlegen der Positionen.

7.1.2 Initiale Approximationen für Gleichverteilungen

In diesem Abschnitt wird ein Positionsschema für die Positionen der Dirac-Impulse einer initialen Approximation zur progressiven Approximation einer zweidimensionalen Gleichverteilung auf dem Intervall $[\underline{c}_1, \underline{c}_2]$, das mit dem Integrationsintervall im Gütemaß übereinstimmt, aufgestellt. Die Gewichte werden alle gleich groß gewählt, so dass ihre Summe Eins ergibt.

Beispiel 7.1: Motivation des Positionsschemas für eine Gleichverteilung

In Abschnitt 6.2.1 wird eine optimale Approximation für eine eindimensionale Gleichverteilung auf eine zweidimensionale Gleichverteilung übertragen. In der entstehenden Approximation befinden sich die Dirac-Impulse auf einem achsenparallel ausgerichteten Grid. Diese Approximation ist lokal, aber nicht global optimal. Wird diese Approximation als initiale Approximation gewählt, so macht sich das im vorherigen Abschnitt beschriebene Problem für lokal optimale initiale Approximationen bemerkbar: der Spielraum in der Progression ist zu klein und die progressive Verbesserung läuft sofort in eine Sackgasse.

In Abbildung 7.1 ist eine Variation der Approximation aus Abschnitt 6.2.1 betrachtet als initiale Approximation mit zugehöriger quasi-ausgearteter Gauß-Mischverteilung zu sehen. Die Summe der Gewichte ergibt nun Eins, zudem ist das Grid im betrachteten Integrationsbereich zentriert. Die progressive Verbesserung dieser initialen Approximation ist in den Abbildungen 7.2 bis 7.4 für $\gamma \approx 1/6$, $\gamma \approx 1/3$ sowie $\gamma = 1$ zu sehen. Die entstehende lokal optimale Approximation der Gleichverteilung hat eine Güte von 0.2317 und ist damit besser als die Güte der Approximation aus der Übertragung aus dem Eindimensionalen mit 1.091 (siehe Abbildung 6.1). ■

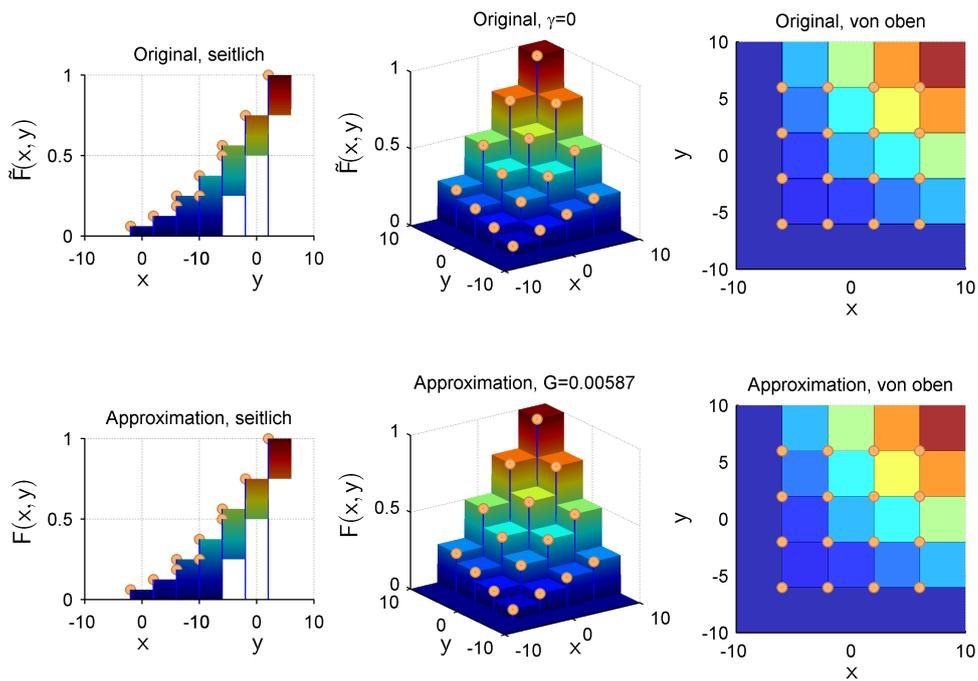


Abbildung 7.1: Initiale Approximation zur Approximation einer zweidimensionalen Gleichverteilung. Dargestellt ist in der oberen Zeile die progressive Verteilungsfunktion für $\gamma = 0$ und darunter jeweils die Verteilungsfunktion der initialen Approximation. Angegeben ist die Güte G der initialen Approximation für die dargestellte progressive Verteilung zum Progressionszeitpunkt $\gamma = 0$.

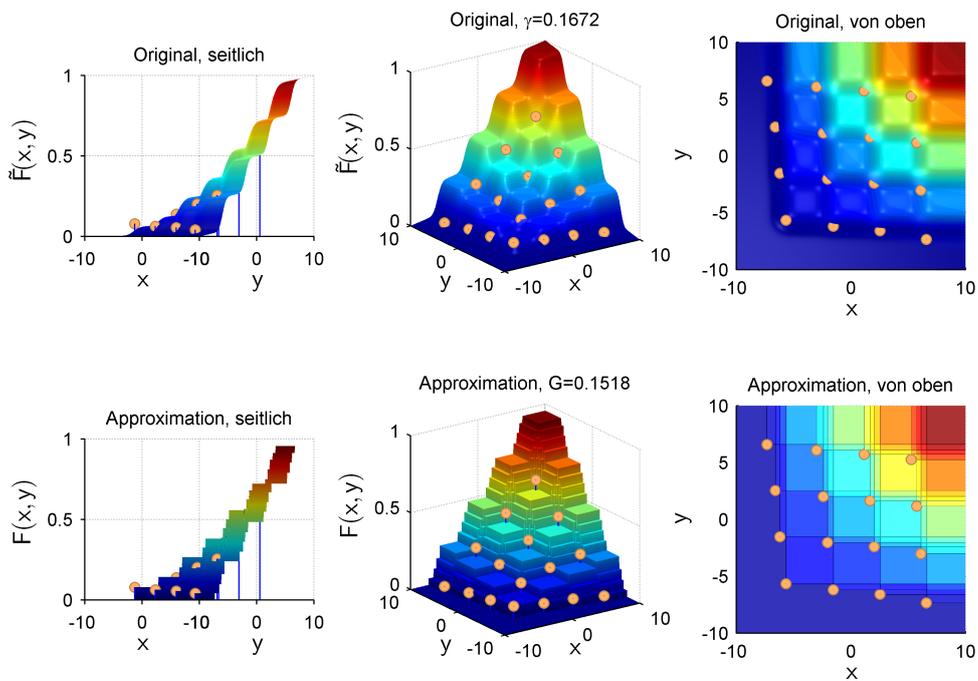


Abbildung 7.2: Progressive Verbesserung der initialen Approximation aus Abbildung 7.1, $\gamma \approx 1/6$.

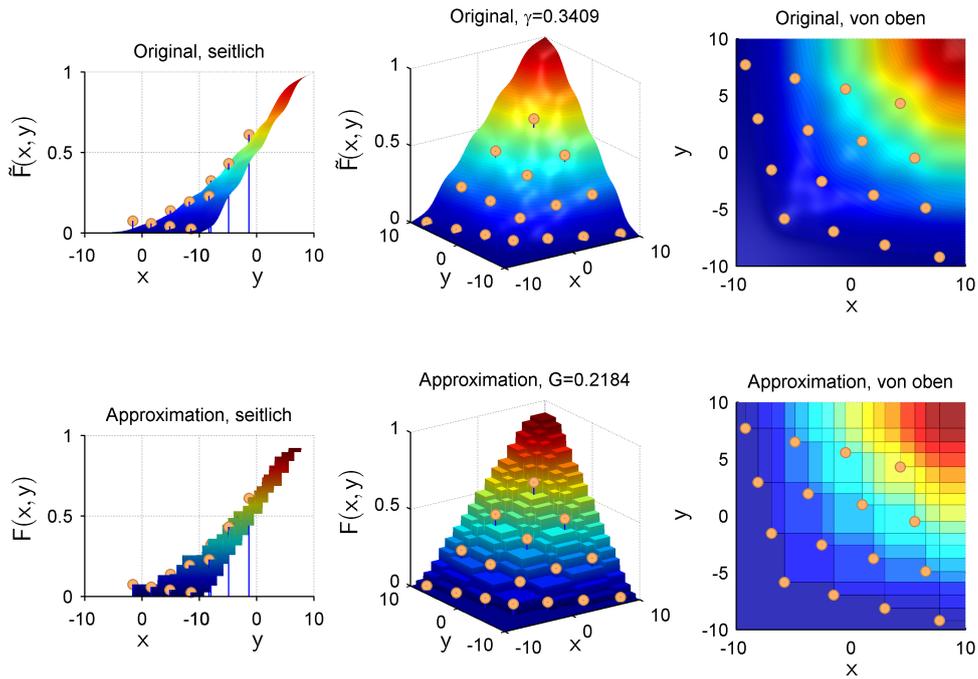


Abbildung 7.3: Progressive Verbesserung der initialen Approximation aus Abbildung 7.1, $\gamma \approx 1/3$.

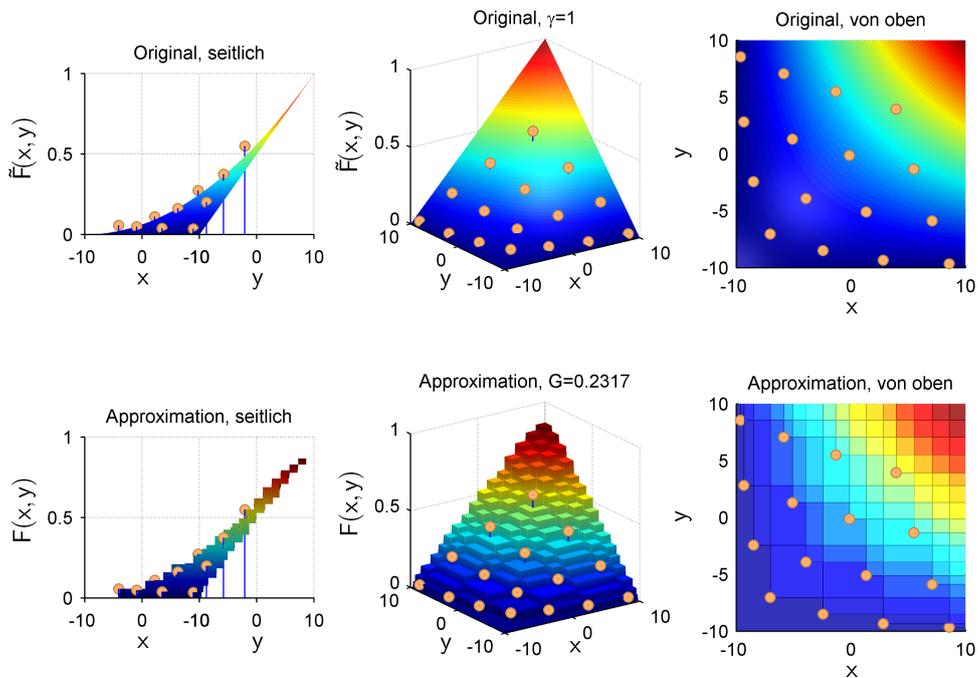


Abbildung 7.4: Progressiv verbesserte Approximation der Gleichverteilung ausgehend von der initialen Approximation aus Abbildung 7.1 ($\gamma = 1$).

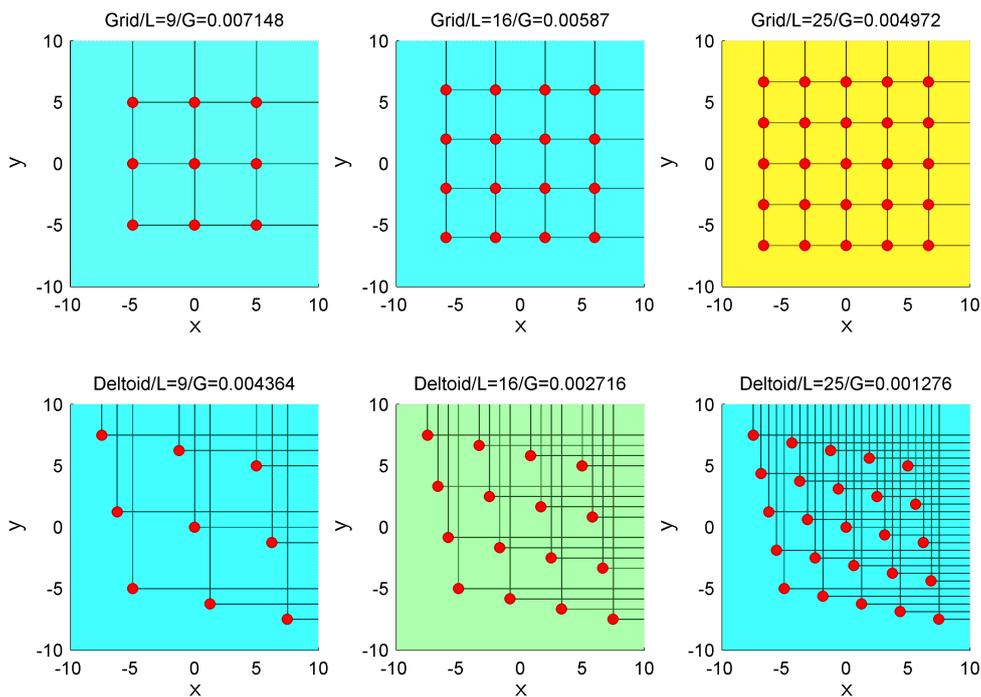


Abbildung 7.5: Differenzen aus Startverteilungsfunktion und initialer Approximation für 6 verschiedene initiale Approximationen in einer Ansicht von oben. Markiert sind die Positionen der Dirac-Impulse der jeweiligen initialen Approximation. Die Überschriften der Teilabbildung zeigen jeweils die Form, Anzahl L der Dirac-Impulse und die Güte G zum Progressionszeitpunkt $\gamma = 0$ an.

Die Form der Positionen der Approximation in Abbildung 7.4 zeigt Ähnlichkeiten zu einem Deltoiden. Das achsenparallel ausgerichtete Grid aus Abbildung 7.1 kann variiert werden, indem es als Deltoid betrachtet und anschließend in Richtung der Haupt- und Nebendiagonalen gestreckt oder gestaucht wird, wodurch sich ein Positionsschema und damit ein Schema initialer Approximationen für die zweidimensionale Gleichverteilung ergibt.

Beispiel 7.2: Verschiedene initiale Approximationen im Vergleich

Abbildung 7.5 zeigt in jeder Teilabbildung die Differenz $\tilde{F}(\underline{x}, \gamma) - F(\underline{x}, \underline{\eta})$ aus Startverteilungsfunktion und initialer Approximation für 6 verschiedene initiale Approximationen. Die oberen drei Bilder zeigen einen achsenparallel ausgerichteten Deltoiden (Grid) für verschiedene Anzahlen L von Dirac-Impulsen. Die mittlere initiale Approximation stimmt mit der aus Abbildungen 7.1 überein. Darunter ist für die jeweils gleiche Anzahl L ein Deltoid zu sehen, der in Richtung der Hauptdiagonale etwas gestaucht und in Richtung der Nebendiagonale etwas gestreckt wurde. Abbildung 7.6 zeigt die jeweils progressiv verbesserte Approximation. Für $L = 9$ und $L = 25$ ergeben sich aus dem verzerrten Deltoiden bessere Approximationen als aus der jeweiligen achsenparallelen Version darüber. Für $L = 16$ ergibt sich die gleiche Approximation. ■

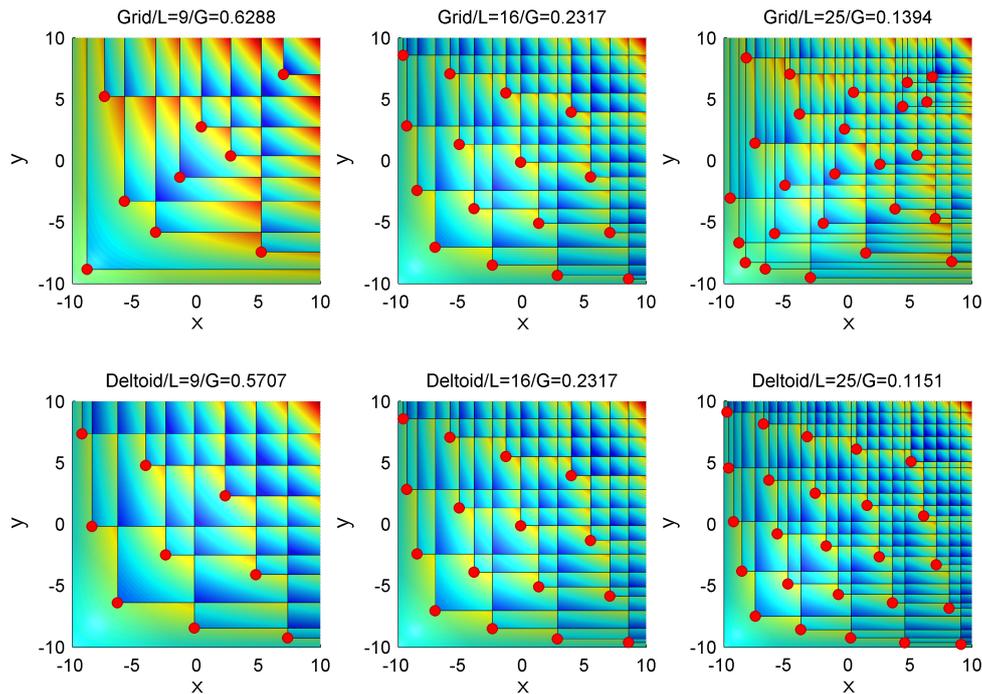


Abbildung 7.6: Differenzen zur Gleichverteilung der progressiv verbesserten Approximationen aus Abbildung 7.5 ($\gamma = 1$).

Bifurkationen

Es ist kein Zufall, dass die progressiv verbesserten Approximationen in Abbildung 7.6 aus den achsenparallel ausgerichteten Deltoiden in Abbildung 7.5 eine sehr andere Form haben als die initialen Approximationen selbst, während die verbesserten Approximationen aus dem verzerrten Deltoiden diesen immer noch verhältnismäßig ähnlich sehen. In Abschnitt 5.4 wurde schon darauf hingewiesen, dass für Approximationen η , welche Dirac-Impulse mit achsenparallel ausgerichteten Positionen enthalten, Irregularitäten in der Gütefunktion G bestehen. Beispiel 7.3 zeigt, dass in der Gegend dieser Irregularitäten Bifurkationen entstehen.

Beispiel 7.3: Bifurkationen bei achsenparallel ausgerichteten Dirac-Impulsen

Abbildung 7.7 zeigt im rechten Teil die Differenz vier verschiedener initialer Approximationen in Deltoid-Form zu ihrer jeweiligen quasi-ausgearteten Gauß-Mischverteilung als Startverteilung. Der Streckungsfaktor α in Richtung der Hauptdiagonale und β in Richtung der Nebendiagonale beträgt (etwa) $(\alpha, \beta) = (0.4, 0.4)$ mit einer additiven Variation von $\Delta = (-5 \cdot 10^{-15}, -5 \cdot 10^{-15})$, $\Delta = (-5 \cdot 10^{-15}, 0)$, $\Delta = (0, -5 \cdot 10^{-15})$ beziehungsweise $\Delta = (5 \cdot 10^{-15}, 5 \cdot 10^{-15})$ für die verschiedenen Deltoide. Der Unterschied in den Streckungsfaktoren ist also minimal. Im linken Teil der Abbildung sind alle Dirac-Mischdichten der initialen Approximationen gleichzeitig eingezeichnet.

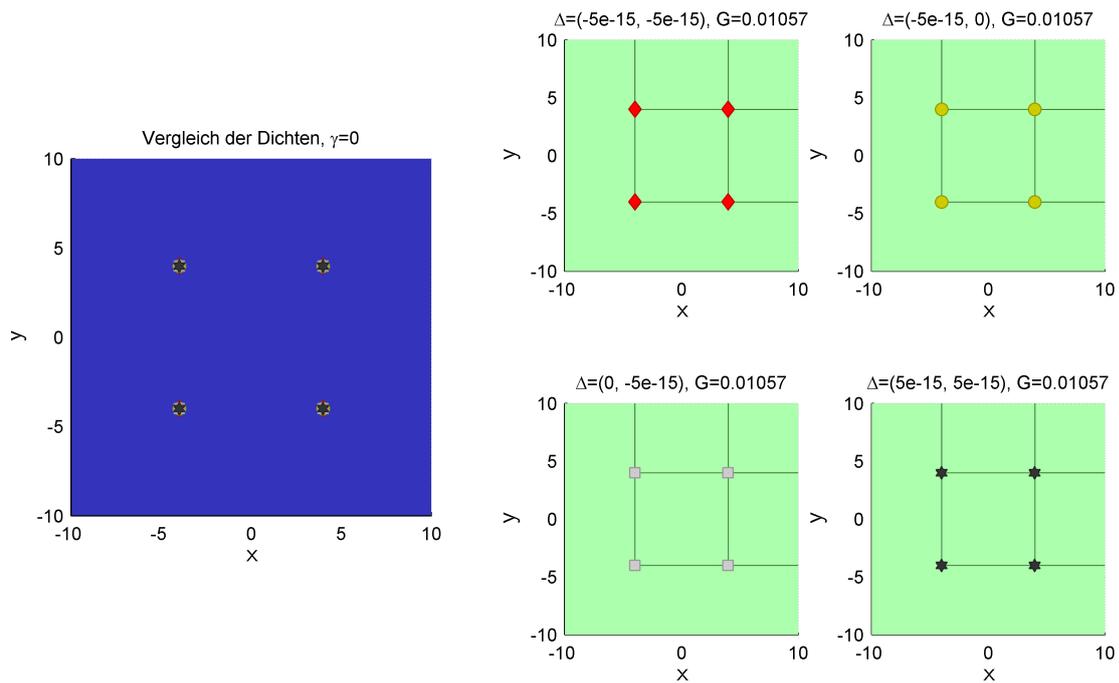


Abbildung 7.7: Beispiel zu Bifurkationen an Hand von vier initialen Approximationen. Von oben ist im rechten Teil jeweils die Differenz der Startverteilung zur initialen Approximation zu sehen. Zusätzlich enthält der Titel jeder der rechten Teilabbildungen die Güte G zum aktuellen Progressionsschritt $\gamma = 0$. Das linke Teilbild zeigt alle Dirac-Mischdichten der Approximationen gleichzeitig.

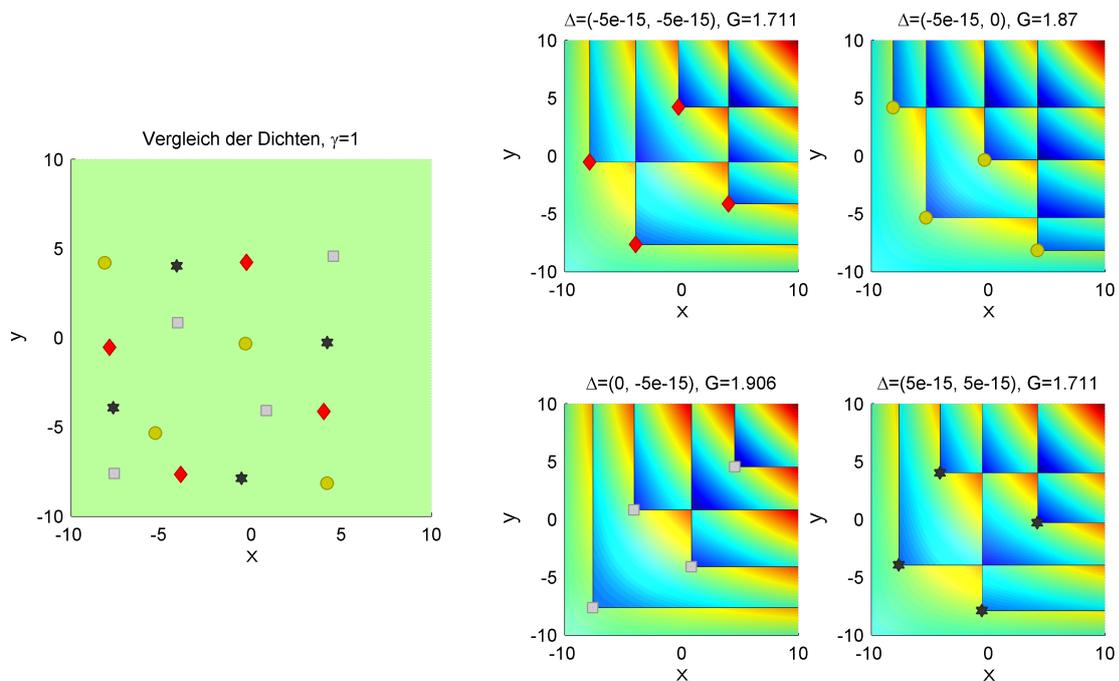


Abbildung 7.8: Progressiv verbesserte Approximationen aus Abbildung (7.7) ($\gamma = 1$).

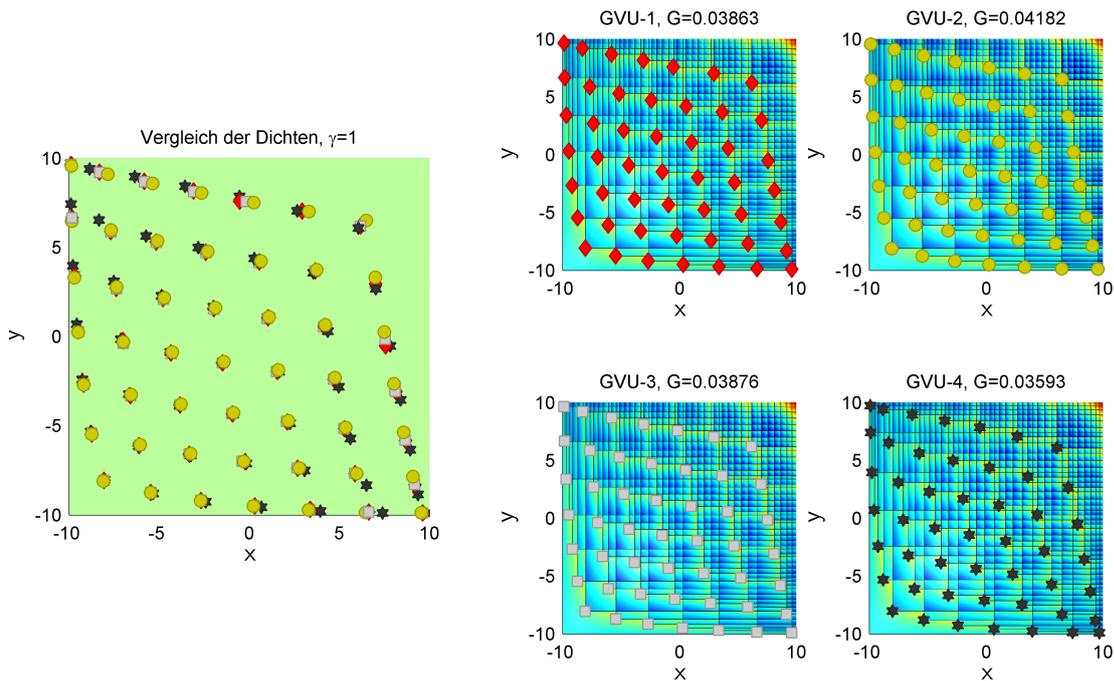


Abbildung 7.9: Separiertheit (lokal) optimaler Approximationen der Gleichverteilung. Dargestellt ist im rechten Teil der Abbildung die Differenz der Gleichverteilung zu jeweils einer lokal optimalen Approximation. Auf der linken Seite sind alle Dirac-Mischdichten der Approximationen in ein Teilbild eingezeichnet.

Abbildung 7.8 zeigt die Differenz der Gleichverteilung zu den progressiv verbesserten Approximationen. Entstanden sind vier sehr verschiedene Approximationen der Gleichverteilung, obwohl die initialen Approximationen sich sehr ähnlich sind. Tatsächlich gilt für $i, j \in \{-5, \dots, 5\}$ und die Streckungsfaktoren $(\alpha', \beta') = (\alpha, \beta) + (i, j) \cdot 10^{-15}$:

- Ist $\alpha' < \beta'$, so entspricht die progressiv verbesserte Approximation der initialen Approximation zum Deltoiden mit den Streckungsfaktoren (α', β') jeweils der Approximation rechts oben im rechten Teil von Abbildung 7.8.
- Ist $\alpha' > \beta'$ so ergibt sich die Approximation links unten im rechten Teil der Abbildung 7.8.
- Für $\alpha' = \beta'$ ergeben sich unterschiedliche der dargestellten Approximationen in Abbildung 7.8.

Anders gesagt: Für $\alpha' = \beta'$ induziert die zugehörige initiale Approximation eine Bifurkation gleich zu Beginn der Progression. Ursache sind die in diesem Fall achsenparallel ausgerichteten Dirac-Impulse. ■

Beispiel 7.4: Separiertheit (lokal) optimaler Approximationen

Abschließend vermittelt Abbildung 7.9 noch einen Eindruck davon, dass die (lokal) optimalen Approximationen der Gleichverteilung für größere Anzahlen von Dirac-Impulsen nicht sonderlich gut separiert sind. Dargestellt ist im rechten Teil der Abbildung die Differenz der Gleichverteilung zu progressiv

verbesserten Approximationen ausgehend von initialen Approximationen basierend auf dem Deltoid-Ansatz. Der linke Teil zeigt wiederum alle Dirac-Mischdichten der verbesserten Approximationen in einem Bild. ■

7.1.3 Initiale Approximationen für Gauß-Verteilungen

Dem hier vorgestellten Schema von initialen Approximationen für eine zweidimensionale Gauß-Dichte liegt die intuitive Vorstellung zu Grunde, dass die Positionen der Dirac-Impulse einer optimalen Approximation auf konzentrischen Ellipsoiden um den Mittelwert liegen könnten — entsprechend der Konstanzbereiche der Gauß-Dichte in Form der Kovarianzellipsen. Die Hauptachsen der Ellipsoide ergeben sich aus den Eigenvektoren der Kovarianzmatrix der gegebenen Gauß-Dichte. Für eine achsenparallel ausgerichtete Gauß-Dichte liegen die Hauptachsen des Ellipsoids parallel zu den Achsen des Koordinatensystems. Die Ausdehnungen der Ellipsoide ergeben sich aus den Wurzeln der Eigenwerte der Kovarianzmatrix, im achsenparallelen Fall somit aus den Standardabweichungen in Richtung der jeweiligen Achse.

Bei Vorgabe der Anzahl L_E der Ellipsoide wird der äußerste Ellipsoid so platziert, dass die Ausdehnung in Richtung seiner Hauptachsen gerade zweimal der Wurzel des zugehörigen Eigenwerts entspricht. Die inneren Ellipsoide folgen in äquidistantem Abstand. Auf jedem Ellipsoid werden ebenfalls äquidistant eine vorgegebene Anzahl L_D von Dirac-Impulsen platziert. Die Gesamtanzahl der Dirac-Impulse L ist somit gleich $L_E \cdot L_D$. Um die Dirac-Impulse besser über die Gesamtfläche zu verteilen, werden sie auf jedem Ellipsoid beginnend mit einem leichten Versatz zum vorherigen Ellipsoiden platziert, so dass sich ein Spiralmuster ergibt. Alle Gewichte der Dirac-Impulse werden gleich groß gewählt, so dass ihre Summe der Gewichtung der Gauß-Dichte entspricht, beziehungsweise gleich Eins ist, falls keine (explizite) Gewichtung vorgenommen wird.

Beispiel 7.5: Positionsschema für Gauß-Dichten an Hand ihrer Konstanzbereiche

Abbildung 7.10 zeigt in jeder Spalte jeweils von oben die Dichte einer gewichteten, achsenparallelen Gauß-Verteilung $G-k$, $k = 1, \dots, 3$, mit Mittelwert $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ und Standardabweichungen $\underline{\sigma}_k$ von $\begin{pmatrix} 3 \\ 3 \end{pmatrix}$, $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ beziehungsweise $\begin{pmatrix} 0.2 \\ 2 \end{pmatrix}$ in Richtung der Hauptachsen. Die Gewichte ω_k sind 0.8, 1, und 0.5. In der oberen Zeile sind die Positionen der Dirac-Impulse in der initialen Approximation markiert. Die Parameter (L_E, L_D) sind $(6, 14)$, $(3, 14)$ beziehungsweise $(3, 8)$. In der Zeile darunter sind zum Vergleich die Positionen der jeweils progressiv verbesserten Approximation zu sehen. Die Struktur der Positionen der Dirac-Impulse hat sich nur relativ wenig geändert, während die Gewichte in der progressiv verbesserten Approximation der Form der jeweiligen Zieldichte besser angepasst sind: Abbildung 7.11 zeigt das Gleiche wie Abbildung 7.10 jedoch aus einer seitlichen Perspektive. Die Höhe der eingezeichneten Dirac-Impulse wurde zur besseren Darstellung auf die Maximalhöhe der jeweiligen Zieldichte normiert.

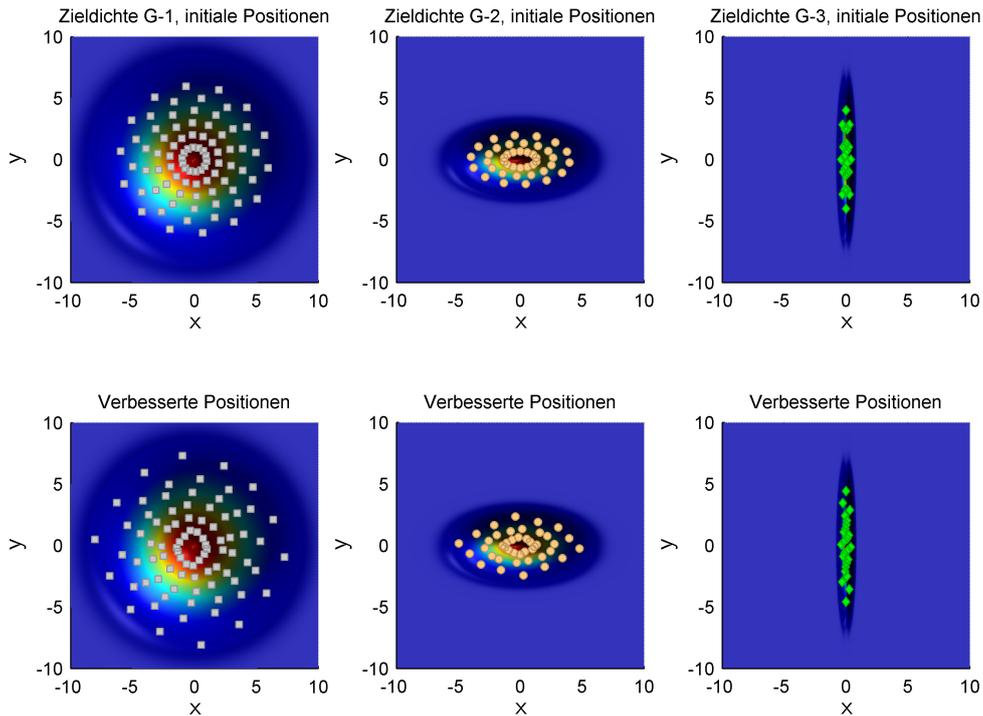


Abbildung 7.10: Initiale Positionen im Vergleich zu den progressiv verbesserten Positionen. Zu sehen ist in jeder Spalte die Dichte einer anderen (gewichteten) Gauß-Verteilung. In der oberen Zeile sind die Positionen der Dirac-Impulse der initialen Approximation markiert, in der unteren die der progressiv verbesserten Approximation.

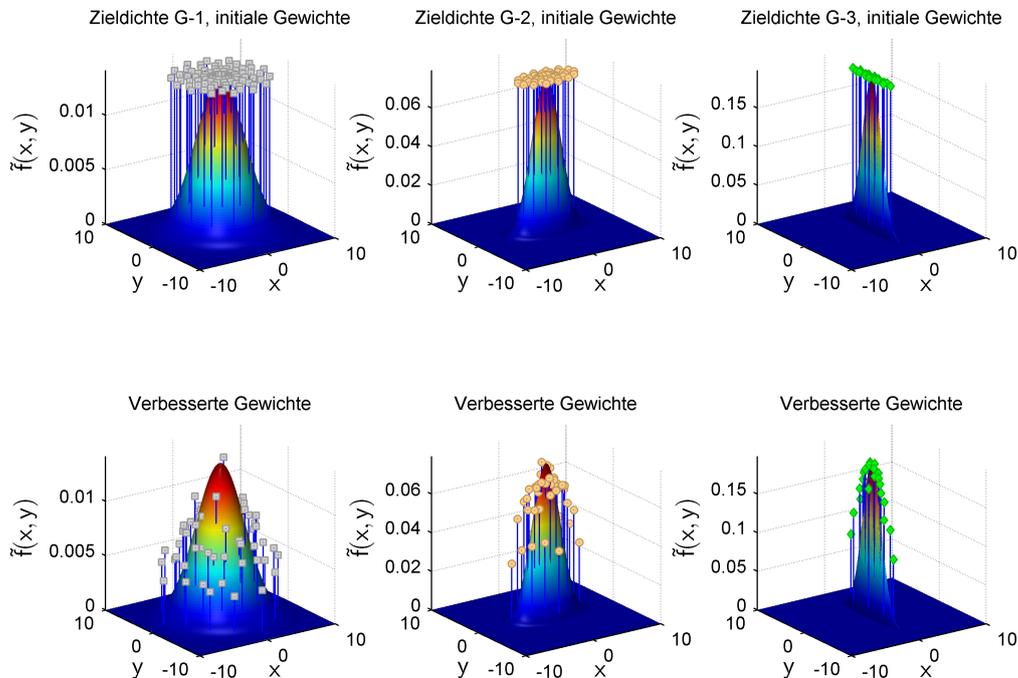


Abbildung 7.11: Initiale Gewichte im Vergleich zu den progressiv verbesserten Gewichten. Diese Abbildung zeigt das Gleiche wie Abbildung 7.10 nur aus einer anderen Perspektive. Die Höhen der eingezeichneten Dirac-Impulse sind zur besseren Darstellung auf die Maximalhöhe der jeweiligen Zieldichte normiert.

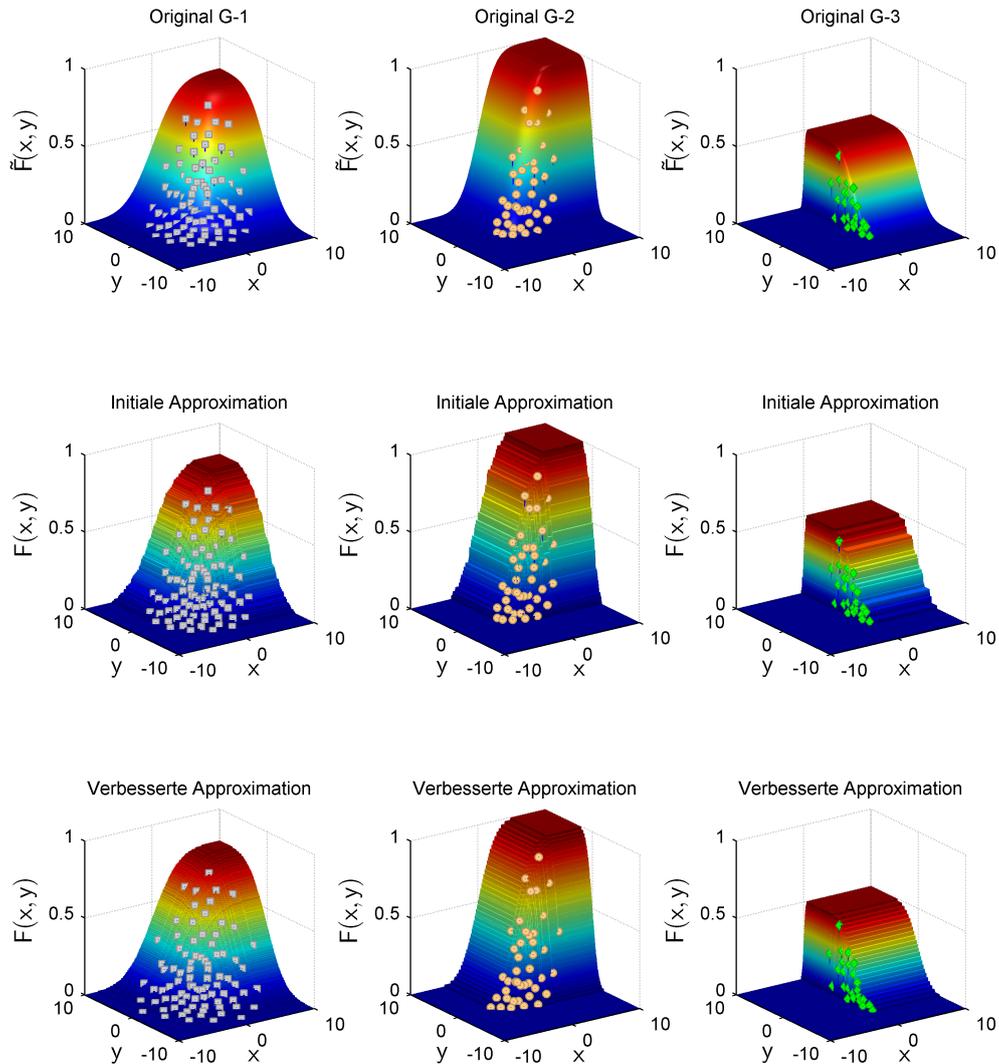


Abbildung 7.12: Initiale und progressiv verbesserte Approximationen im Vergleich. Dargestellt ist jeweils die Verteilungsfunktion der Zielverteilung (erste Zeile), der initialen Approximation (mittlere Zeile) sowie der progressiv verbesserten Approximation (untere Zeile) zu den Gauß-Dichten G-1, G-2 und G-3 aus Beispiel 7.5. In den Zielverteilungen sind die Positionen der Dirac-Impulse der initialen Approximationen markiert.

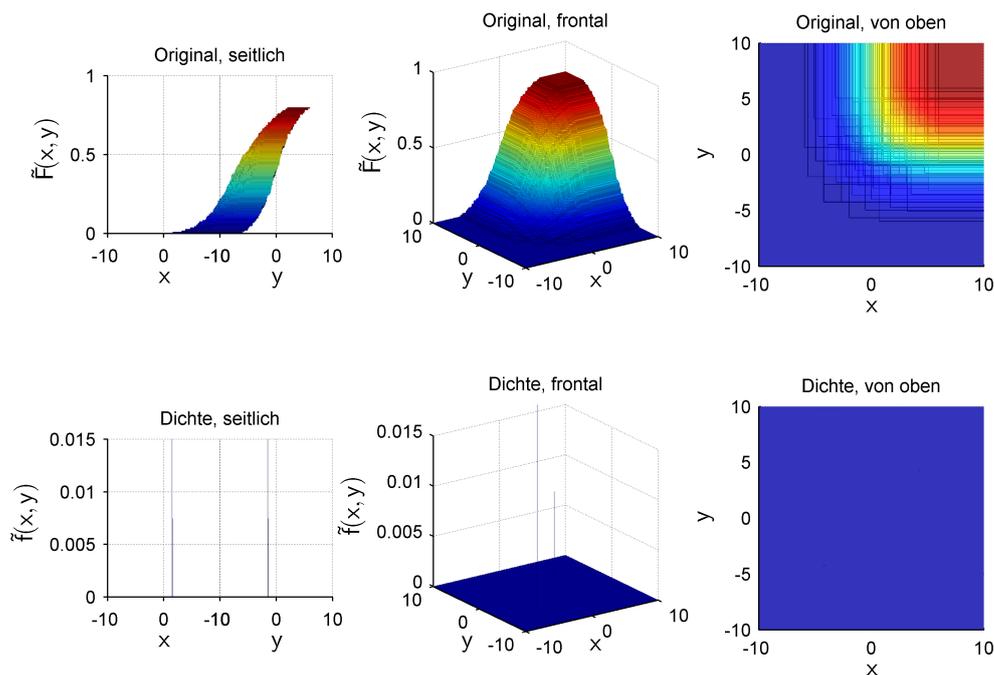


Abbildung 7.13: Zur ersten Zielverteilung G-1 aus Beispiel 7.5 ist in der oberen Zeile die Verteilungsfunktion der progressiven Verteilung für $\gamma = 0$ aus verschiedenen Perspektiven zu sehen, darunter jeweils die Dichte aus der gleichen Perspektive. Da die Varianzen der Gauß-Mischdichte, die hier die Startdichte bildet, sehr klein sind, ist von ihr allerdings wenig zu sehen.

Abbildung 7.12 zeigt einen Vergleich auf der Seite der Verteilungsfunktionen. Die Ähnlichkeit der initialen Approximation mit der jeweiligen Zielverteilung ist deutlich zu erkennen. Die progressive Verbesserung führt zu einer besseren Anpassung in der Feinstruktur.

Die Ähnlichkeit der initialen Approximation ergibt sich aus der Ähnlichkeit der Verteilungsfunktionen der Startverteilung in Form einer quasi-ausgearteten Gauß-Mischverteilung zur Zielverteilung. Abbildung 7.13 bis 7.16 zeigen die progressive Verteilungsfunktion und Dichte zu den Progressionszeitpunkten $\gamma = 0$, $\gamma \approx 1/3$, $\gamma \approx 2/3$ und $\gamma = 1$ für G-1.

■

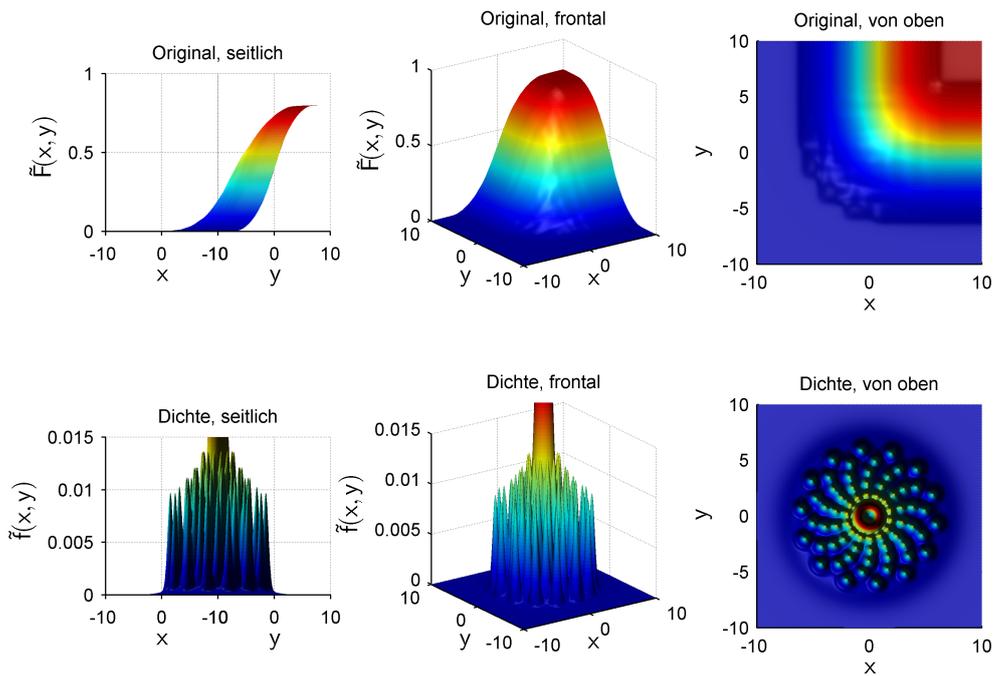


Abbildung 7.14: Progressive Verteilungsfunktion und Dichte für G-1 aus Beispiel 7.5 ($\gamma \approx 1/3$).

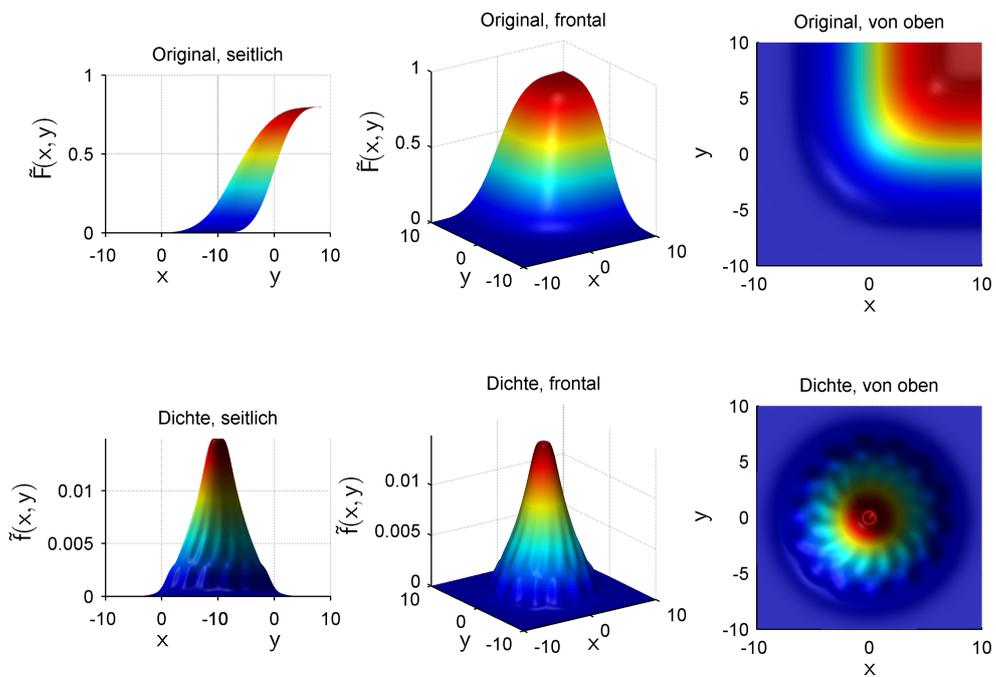


Abbildung 7.15: Progressive Verteilungsfunktion und Dichte für G-1 aus Beispiel 7.5 ($\gamma \approx 2/3$).

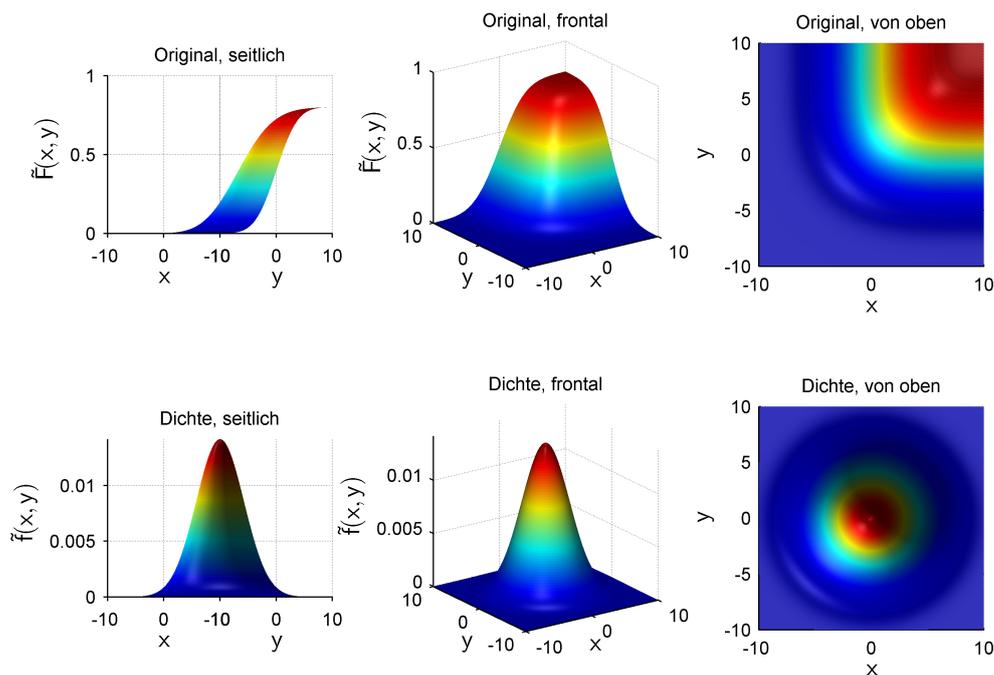


Abbildung 7.16: Progressive Verteilungsfunktion und Dichte für G-1 aus Beispiel 7.5 ($\gamma = 1$).

7.2 Konvexe Kombination von Gleichverteilung und Zielverteilung

Aus dem Ansatz zur progressiven Verbesserung initialer Approximationen liegen lokal optimale Approximationen der Gleichverteilung vor. Bei Verwendung der Gleichverteilung als Startverteilung in Kombination mit diesen nur lokal und nicht global optimalen Approximationen sind vermehrt Sackgassen im Verlauf der Progression sowie eine nur lokal optimale Approximation der Zielverteilung zu erwarten. Im Rahmen dieser Arbeit hat sich gezeigt, dass mit zunehmender Anzahl von Dirac-Impulsen dabei das Problem der Sackgassen tatsächlich immer größer wird. Ob dies auch bei Einsatz einer global optimalen Startapproximation gilt, steht nicht fest. Jedoch lässt die kompliziertere Form der Gütefunktion mit zunehmender Dimension sowie die abnehmende Separiertheit der (lokal) optimalen Approximationen der Gleichverteilung bei steigender Anzahl von Dirac-Impulsen (siehe Abbildung 7.9) dies vermuten.

Das folgende Beispiel nutzt als Progressionsform eine konvexe Kombination der Gleichverteilung mit der vorgegebenen Zielverteilung und demonstriert, dass diese sehr einfache Progressionsform zumindest prinzipiell funktioniert, wenn auch nur für eine begrenzte Menge von Zielverteilungen auf Grund der nur lokal — und nicht global — optimalen Startapproximationen.

Beispiel 7.6: Halbkreis als Zielverteilung

Dieses Beispiel zeigt eine Approximation der Zielverteilung aus Abschnitt 9.2 ausgehend von einer Gleichverteilung. Die lokal optimale Startapproximation umfasst 25 Dirac-Impulse und wurde mit dem Ansatz zur progressiven Verbesserung initialer Approximationen gewonnen (siehe Abschnitt

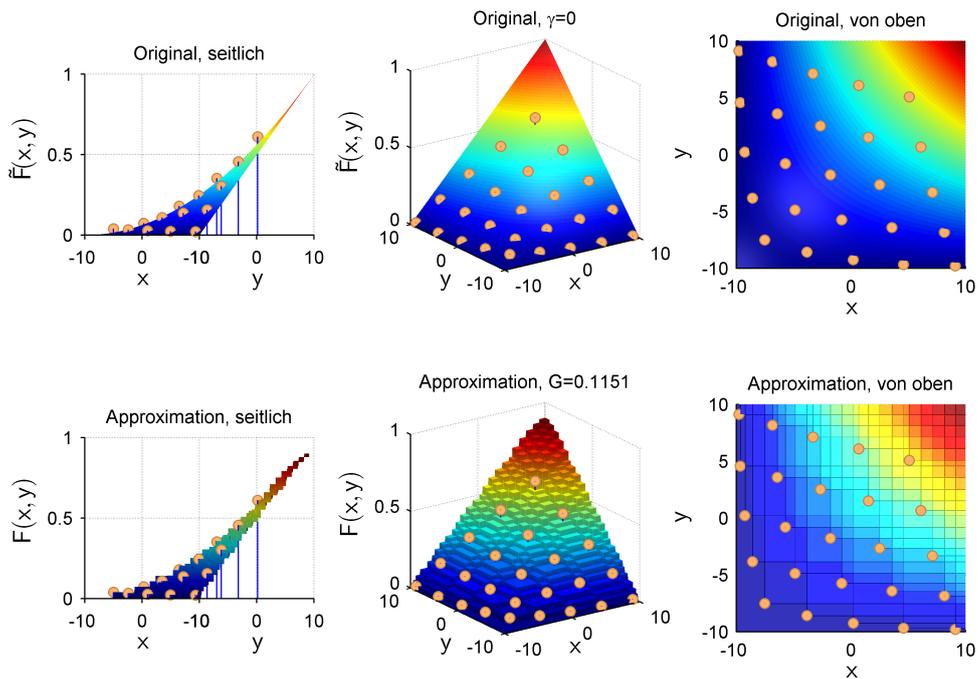


Abbildung 7.17: Startverteilungsfunktion in Form der Gleichverteilung (obere Zeile) mit einer lokal optimalen Startapproximation (untere Zeile) ($\gamma = 0$).

7.1.2). Die Zieldichte ist eine achsenparallele Gauß-Mischdichte mit 16 Komponenten, die zwei separate Moden und einen Halbkreis ausbilden (siehe Abbildung 9.7). Abbildung 7.17 bis 7.20 zeigen in der oberen Zeile die progressive Verteilung aus verschiedenen Perspektiven und darunter jeweils die zugehörige Approximation für $\gamma = 0$, $\gamma \approx 1/3$, $\gamma \approx 2/3$ und $\gamma = 1$. ■

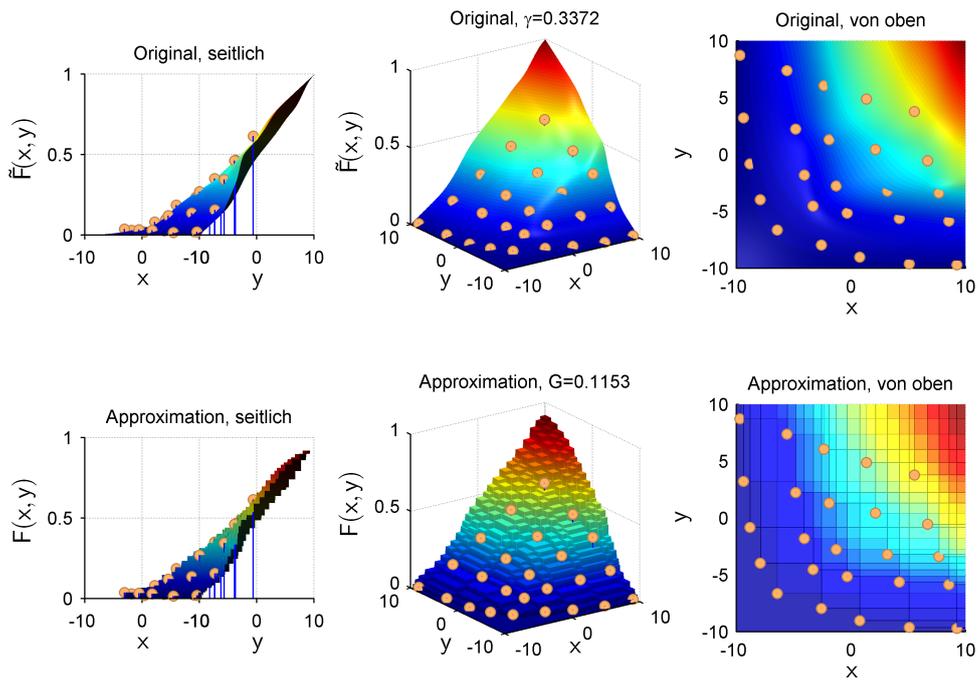


Abbildung 7.18: Progressive Verteilungsfunktion (obere Zeile) mit zugehöriger Approximation (untere Zeile) für $\gamma \approx 1/3$.

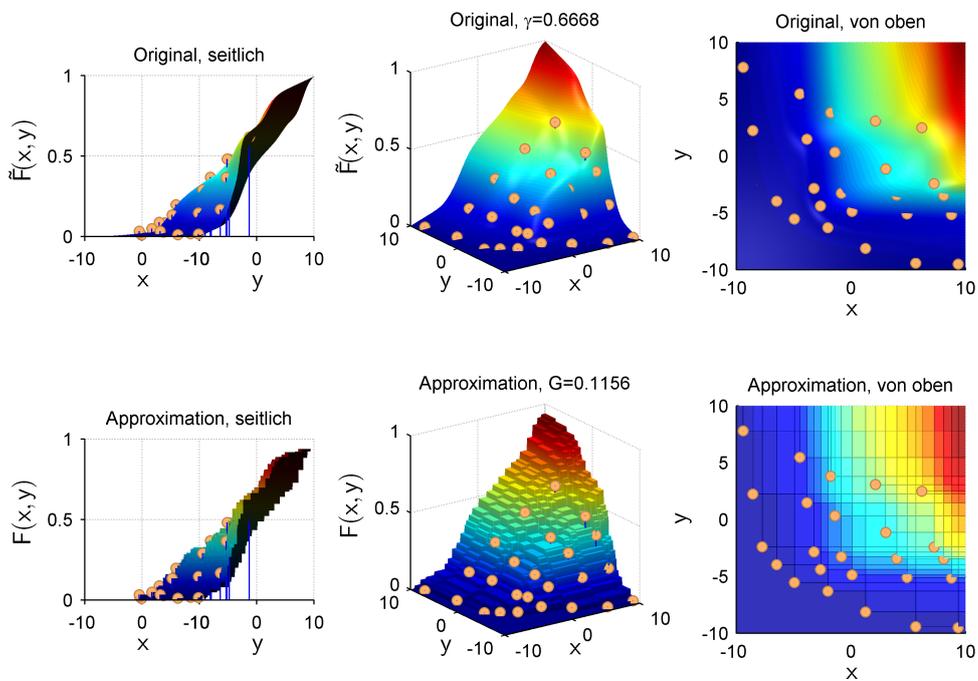


Abbildung 7.19: Progressive Verteilungsfunktion (obere Zeile) mit zugehöriger Approximation (untere Zeile) für $\gamma \approx 2/3$.

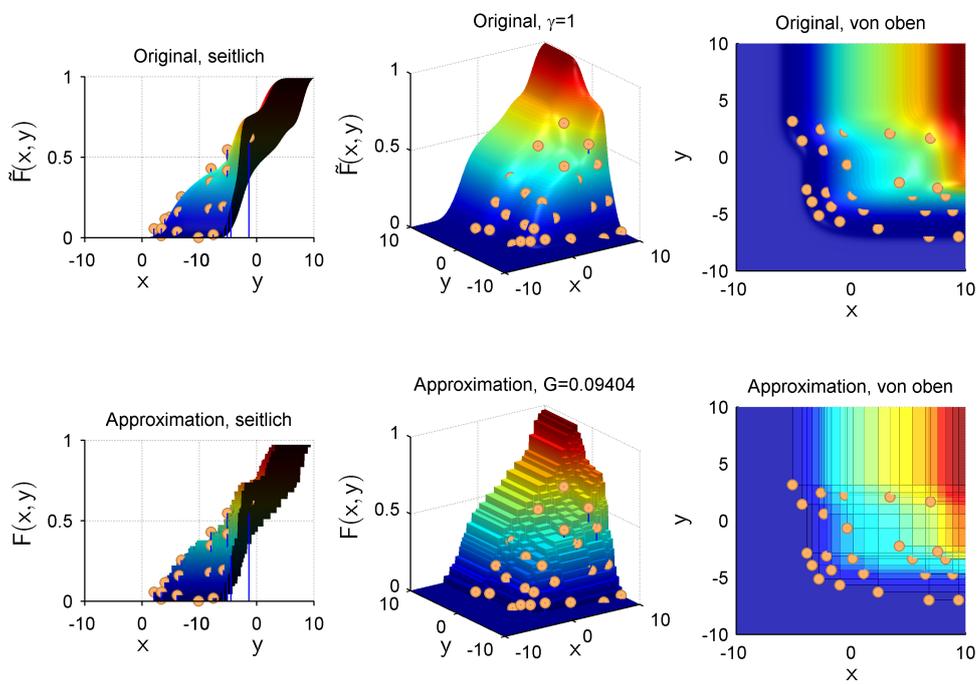


Abbildung 7.20: Zielverteilungsfunktion (obere Zeile) mit zugehöriger Approximation (untere Zeile) ($\gamma = 1$).

Suboptimale Approximation durch Clusterbildung

Mit steigender Anzahl von Dirac-Impulsen in der Approximation steigt der Laufzeitaufwand für das Progressive Bayes-Verfahren erheblich an, siehe Abschnitt 5.5. Zudem werden Sackgassen immer wahrscheinlicher. Kompliziert geformte Zieldichten benötigen jedoch eine angemessen hohe Anzahl von Dirac-Impulsen zur Approximation, um eine ausreichend gute Güte zu erreichen. Dieses Kapitel beschreibt ein Approximationsverfahren basierend auf Clusterbildung — im Folgenden Cluster-Verfahren genannt —, das mit diesen Problemen für den Fall von Mischdichten als Zieldichten umgehen kann.

Definition 8.0.1 Sei $\tilde{f}(\underline{x}) = \sum_{k=1}^K \omega_k \tilde{f}_k(\underline{x})$ eine Mischdichte. Ein Cluster C ist eine Teilmenge von $\{1, \dots, K\}$ und wird mit der Mischdichte $\tilde{f}_C(\underline{x}) := \sum_{k \in C} \omega_k \tilde{f}_k(\underline{x})$ identifiziert. \square

Die Grundidee ist, dass eine Gauß-Mischdichte, deren einzelnen Komponenten in etwa gleiche Mittelwerte sowie Kovarianzmatrizen besitzen, einer einzelnen Gauß-Dichte recht ähnlich sieht. Das in Abschnitt 7.1.3 präsentierte Schema initialer Approximationen für Gauß-Dichten sollte daher auch bei so geformten Gauß-Mischdichten in Verbindung mit dem Ansatz zur progressiven Verbesserung initialer Approximationen aus Abschnitt 7.1 gute Ergebnisse liefern. Eine beliebige Gauß-Mischdichte wird dann in mehrere Cluster zerlegt, die separat approximiert werden. Die Gesamtapproximation ist die Summe aller Einzelapproximationen.

Da die einzelnen Cluster separat approximiert werden, ist die Gesamtapproximation als Summe der Einzelapproximationen nicht mehr (lokal) optimal bezüglich des Gütemaßes. Das Verfahren erzeugt also prinzipiell nur suboptimale Approximationen.

8.1 Approximation der Einzelcluster

Normalerweise soll für die Gesamtapproximation der Zieldichte eine bestimmte Güte erreicht werden. Mithin muss mit der Approximation für jeden Cluster ebenfalls eine bestimmte Güte G_{\min} erreicht werden. G_{\min} wird hier als Eingabeparameter zur Approximation eines Einzelclusters betrachtet. Je kleiner G_{\min} gewählt ist, desto mehr Dirac-Impulse müssen in der

Approximation im Allgemeinen zum Einsatz kommen. Andererseits bestimmt die Anzahl L der Dirac-Impulse, die zur Approximation der Zieldichte zum Einsatz kommen, auch ganz wesentlich den Rechenaufwand des Progressiven Bayes-Verfahrens. Daher wird im Folgenden von einer faktisch gegebenen Obergrenze L_{\max} für L ausgegangen.

Eine Möglichkeit, eine Approximation für den Cluster zu bestimmen, ist, eine initiale Approximation durch Ziehen einer zufälligen Stichprobe $\{\underline{x}_1, \dots, \underline{x}_L\}$ der Zieldichte wie in Abschnitt 3.2.2 beschrieben zu konstruieren und diese dann progressiv zu verbessern. Läuft die progressive Verbesserung in eine Sackgasse, so wird eine neue Stichprobe gezogen und erneut begonnen. Ist die progressive Verbesserung dagegen erfolgreich und erreicht die Güte der progressiv verbesserten Approximation mindestens G_{\min} , so ist die Approximation erfolgreich abgeschlossen. Andernfalls ist die Anzahl der verwendeten Dirac-Impulse möglicherweise zu gering, so dass der Stichprobenumfang L für den nächsten Versuch erhöht werden muss. Ist L schon gleich L_{\max} , dann gibt es eventuell keine Approximation mit höchstens L_{\max} Dirac-Impulsen, die eine Güte von mindestens G_{\min} erreicht und die Approximation des Clusters scheitert.

Gegebenenfalls kann eine bessere Kandidatenmenge von initialen Approximationen aus der Form des Clusters oder dem zu Grunde liegenden Dichtetyp abgeleitet werden, als einfach auf zufällige Stichproben zurück zugreifen.

Beispiel 8.1: Initiale Approximationen für achsenparallele Gauß-Mischdichten

Sei $\tilde{f}_C(\underline{x}) = \sum_{k \in C} \omega_k \mathcal{N}(\underline{\mu}_k, \underline{\sigma}_k)$ ein Cluster bestehend aus d -dimensionalen achsenparallelen Gauß-Dichten. Der achsenparallele Ellipsoid mit Mittelpunkt $\underline{\mu}_k$ und Ausdehnungen $2 \cdot \underline{\sigma}_k$ in Richtung der Hauptachsen ist ein Konstanzbereich der Dichte $\mathcal{N}(\underline{\mu}_k, \underline{\sigma}_k)$, $k \in C$. Sei

$$\underline{\mu} := \sum_{k=1}^K \omega_k \underline{\mu}_k$$

der gewichtete Mittelwert der Mittelwerte sowie

$$\sigma_q := \sigma_{kq} + |\mu_q - \mu_{kq}|,$$

$q = 1, \dots, d$. In diesem Sinne fasst $\tilde{f}_\Sigma(\underline{x}) := \{\sum_{k \in C} \omega_k\} \mathcal{N}(\underline{\mu}, \underline{\sigma})$ alle Komponenten von $\tilde{f}_C(\underline{x})$ in einer gewichteten achsenparallelen Gauß-Dichte zusammen. Auf diese kann nun das Schema für initiale Approximationen aus Abschnitt 7.1.3 angewendet werden. Sind sich alle Komponenten im Cluster C sehr ähnlich, dann gilt

$$\tilde{f}_\Sigma(\underline{x}) \approx \tilde{f}_C(\underline{x})$$

und eine dem Abschnitt 7.1.3 entsprechende initiale Approximation stellt immer noch eine recht gute suboptimale Approximation für $\tilde{f}_\Sigma(\underline{x})$ dar.

Anzumerken bleibt, dass die Ähnlichkeit der Komponenten in $\tilde{f}_C(\underline{x})$ keine zwingende Voraussetzung ist, sondern nur die Auswahl einer initialen Approximation entsprechend Abschnitt 7.1.3 motiviert. Für den Vergleich zu anderen Verfahren in Kapitel 9 werden die Kandidatenmengen initialer Approximationen wie gerade beschrieben konstruiert. ■

Sei im Folgenden eine Kandidatenmenge \mathfrak{J} initialer Approximationen gegeben. Ist \mathfrak{J} endlich, dann kann unter Berücksichtigung der folgenden Punkte nach einer geeigneten initialen Approximation $f(\underline{x}, \underline{\eta}_0)$ in \mathfrak{J} gesucht werden.

1. Die progressive Verbesserung von $f(\underline{x}, \underline{\eta}_0)$ läuft in keine Sackgasse.
2. Die vorgegebene Güte G_{\min} wird erreicht.
3. Die Anzahl L der Dirac-Impulse in $f(\underline{x}, \underline{\eta}_0)$ ist möglichst klein.

Der dritte Punkt verhindert eine „Überapproximation“ im Sinne von einer durch die progressiv verbesserte Approximation erreichten Güte, die wesentlich besser ist als die vorgegebene Güte G_{\min} . Auch wenn im ersten Augenblick gegen eine wesentlich bessere Güte als die vorgegebene nichts einzuwenden scheint, macht es letztlich wenig Sinn, wenn in einigen der Einzelcluster eine wesentlich bessere Güte erreicht wird als in anderen, für welche G_{\min} „gerade so“ erreicht wird: Die Güten der Cluster mit schlechteren Approximationen dominieren im Allgemeinen die Gesamtgüte. Eine wesentlich bessere Approximation in einem einzelnen Cluster würde die Gesamtgüte mitunter nicht wesentlich verbessern, könnte dafür aber das Verhältnis aus Güte zur benutzten Anzahl von Dirac-Impulsen deutlich verschlechtern.

Alle möglichen initialen Approximationen aus \mathfrak{J} auf die drei Kriterien hin zu untersuchen, ist recht aufwendig. Unter der heuristischen Annahme, dass die Güte einer progressiv verbesserten Approximation mit steigender Anzahl von Dirac-Impulsen immer besser wird, kann mit einem Halbierungsverfahren wie in Algorithmus 2 nach einer geeigneten initialen Approximation aus \mathfrak{J} gesucht werden.

Dazu wird \mathfrak{J} als Liste von Parametervektoren initialer Approximationen betrachtet und nach der Anzahl von Dirac-Impulsen der enthaltenen initialen Approximationen sortiert. Die Vergleichsoperation des Suchverfahrens berechnet für das gerade betrachtete Element $\underline{\eta}_0$ der Liste die progressiv verbesserte Approximation $\underline{\eta}$. Ist ihre Güte G nicht schlechter als G_{\min} , so wird im linken Teil weiter gesucht. Ist sie schlechter wird im rechten Teil weiter gesucht. Ein Spezialfall tritt ein, falls die progressive Verbesserung in eine Sackgasse läuft. In diesem Fall wird das Element ignoriert und weiter fortgefahren und zwar so lange, bis

1. ein Element gefunden wird, für das das Progressive Bayes-Verfahren erfolgreich angewendet werden kann. In diesem Fall wird wie zuvor anhand der erreichten Güte entschieden, ob im linken oder rechten Teil der Liste fortgefahren wird. Oder
2. eine vorgegebene Höchstzahl I_{\max} von zu ignorierenden Elementen überschritten wird. In diesem Fall wird angenommen, dass alle nachfolgenden Elemente ebenfalls zu Sackgassen in der progressiven Verbesserung führen. Weiter gesucht wird im linken Teil der Liste.

Von den betrachteten progressiv verbesserten Approximationen, die die Mindestgüte G_{\min} erreichen, wird diejenige mit der geringsten Anzahl von Dirac-Impulsen ausgegeben.

Algorithmus 2 Approximation eines Clusters.

```

1: function ( $\underline{\eta}_{\text{result}}$ , found, qualityReached) := ApproximateCluster( $\tilde{f}$ ,  $C$ ,  $G_{\text{min}}$ )
2:
3: ( $\underline{\eta}_{\text{result}}$ , found, qualityReached) := ( $\underline{0}$ , false, false)
4: ( $G_{\text{result}}$ ,  $L_{\text{result}}$ ,  $I$ ) := ( $\infty$ ,  $\infty$ , 0)
5:
6:  $\mathcal{J}$  := CreateCandidatesOfInitialApproximations( $\tilde{f}$ ,  $C$ ,  $G_{\text{min}}$ ,  $L_{\text{max}}$ )
7: Sort  $\mathcal{J}$  by Count of Dirac Impulses
8: while ( $\mathcal{J} \neq \emptyset$ ) do
9:      $n := \text{LengthOf}(\mathcal{J}) / 2$ 
10:     $\underline{\eta}_0 := \mathcal{J}[n]$ 
11:    ( $\underline{\eta}$ , success) := ImproveProgressive( $\underline{\eta}_0$ ,  $\tilde{f}$ ,  $C$ )
12:    if (success) then
13:         $G := \text{CalcQuality}(\underline{\eta})$ 
14:         $L := \text{GetCountOfDiracImpulses}(\underline{\eta})$ 
15:        if ( $L < L_{\text{result}} \ \&\& \ G \leq G_{\text{min}}$ 
16:            ||  $L == L_{\text{result}} \ \&\& \ G < G_{\text{result}}$ 
17:            ||  $L > L_{\text{result}} \ \&\& \ G < G_{\text{result}} \ \&\& \ G_{\text{result}} > G_{\text{min}}$ ) then
18:            ( $\underline{\eta}_{\text{result}}$ ,  $G_{\text{result}}$ ,  $L_{\text{result}}$ ) := ( $\underline{\eta}$ ,  $G$ ,  $L$ )
19:        end if
20:        if ( $G \leq G_{\text{min}}$ ) then
21:             $\mathcal{J} := \text{GetSubset}(\mathcal{J}, 0, n)$  // Startindex ist 0, letzter enthaltener Index ist  $n - 1$ 
22:        else
23:             $\mathcal{J} := \text{GetSubset}(\mathcal{J}, n + 1, \text{LengthOf}(\mathcal{J}))$ 
24:        end if
25:         $I := 0$ 
26:    else
27:        if ( $I \leq I_{\text{max}}$ ) then
28:             $\mathcal{J} := \text{RemoveElement}(\mathcal{J}, n)$ 
29:             $I := I + 1$ 
30:        else
31:             $\mathcal{J} := \text{GetSubset}(\mathcal{J}, 0, n)$ 
32:             $I := 0$ 
33:        end if
34:    end if
35: end while
36: (found, qualityReached) := ( $G_{\text{result}} < \infty$ ,  $G_{\text{result}} \leq G_{\text{min}}$ )

```

Algorithmus 3 Approximation mit dem Cluster-Verfahren. Dabei ist $\tilde{f} = \sum_{k=1}^K \tilde{f}_k$ und ein Index $k \in \{1, \dots, K\}$ wird mit der Komponente \tilde{f}_k identifiziert.

```

1: function ( $\underline{\eta}_{\text{result}}$ , found, qualityReached) := Approximate( $\tilde{f}$ ,  $G_{\text{min}}$ )
2:
3: ( $\underline{\eta}_{\text{result}}$ , found, qualityReached) := ( $\underline{0}$ , false, true)
4:  $K := \text{GetComponentCount}(\tilde{f})$ 
5:  $V := (1, \dots, K)$ 
6: while ( $V \neq \emptyset$ ) do
7:   first :=  $V[0]$ 
8:   Sort  $V$  by Distance to  $\tilde{f}_{\text{first}}$ 
9:   ( $C$ ,  $\underline{\eta}$ , found, reached) := CreateCluster( $\tilde{f}$ ,  $V$ , 0, LengthOf( $V$ ),  $G_{\text{min}}$ )
10:  if (!found) then
11:    break
12:  end if
13:  qualityReached := qualityReached && reached
14:   $V := V \setminus C$ 
15:   $\underline{\eta}_{\text{result}} := \text{Join}(\underline{\eta}_{\text{result}}, \underline{\eta})$ 
16: end while

```

8.2 Clusterbildung

Für die Clusterbildung wird ein Abstandsmaß zwischen den Komponenten der Zieldichte als gegeben vorausgesetzt.

Beispiel 8.2: Abstandsmaß für achsenparallele Gauß-Dichten

Ein einfaches Abstandsmaß zwischen zwei achsenparallelen Gauß-Dichten mit Mittelwerten $\underline{\mu}_{k_i}$ und Standardabweichungen $\underline{\sigma}_{k_i}$, $i = 1, 2$ ist $\left\| \underline{\mu}_{k_1} - \underline{\mu}_{k_2} \right\|_2 + \left\| \underline{\sigma}_{k_1} - \underline{\sigma}_{k_2} \right\|_2$ als Summe der Euklidnormen der Differenzen aus Mittelwerten und Standardabweichungen. Dieses Abstandsmaß wird für den Vergleich zu anderen Verfahren in Kapitel 9 verwendet. ■

In der Clusterbildung muss sichergestellt werden, dass jeder einzelne Cluster erfolgreich approximiert werden kann. Da dies einem beliebig vorgelegten Cluster nicht von vorne herein anzusehen ist, wird die Approximation der Einzelcluster mit der Clusterbildung selbst verschränkt.

Dazu wird eine List V der noch verbliebenen Komponenten der Zieldichte erstellt, siehe Algorithmus 3. Die Liste V wird sortiert nach dem Abstand zur ersten Komponente aus V . Mit Hilfe eines Halbierungsverfahrens wird der größte Cluster C , bestehend aus den ersten n Elementen aus V , gewählt, so dass die Approximation des Clusters erfolgreich verläuft und gleichzeitig die Mindestgüte G_{min} pro Cluster erreicht wird, siehe Algorithmus 4.

Algorithmus 4 Erstellen eines Clusters.

```

1: function ( $C, \underline{\eta}$ , found, qualityReached) := CreateCluster( $\tilde{f}, V, n_{\min}, n_{\max}, G_{\min}$ )
2:
3:  $n := (n_{\min} + n_{\max})/2$ 
4:  $C := \text{GetSubset}(V, 0, n + 1)$ 
5: ( $\underline{\eta}$ , found, qualityReached) := ApproximateCluster( $\tilde{f}, C, G_{\min}$ )
6: if ( $n_{\min} < n_{\max} - 1$ ) then
7:   if (!(found && qualityReached)) then
8:     ( $C, \underline{\eta}$ , found, qualityReached) := CreateCluster( $\tilde{f}, V, n_{\min}, n, G_{\min}$ )
9:   else
10:    ( $C2, \underline{\eta}_2$ , found2, reached2) := CreateCluster( $\tilde{f}, V, n + 1, n_{\max}, G_{\min}$ )
11:    if (found2 && reached2) then
12:      ( $C, \underline{\eta}$ ) := ( $C2, \underline{\eta}_2$ )
13:    end if
14:  end if
15: end if
    
```

Falls ein neuer Cluster erfolgreich gebildet wurde, wird V erneut sortiert nach dem Abstand zur ersten Gauß-Komponente aus V und das Verfahren läuft iterativ fort bis V leer und somit eine Approximation der Zieldichte in Form der Summe der Approximationen der Einzelcluster gefunden ist.

Die Cluster werden so groß wie möglich gewählt, um möglichst viele der Überschneidungsbereiche der Komponenten der Zieldichte in einem Cluster nutzen zu können. Je mehr Komponenten ein Cluster enthält, desto komplizierter wird andererseits seine Form und das verwendete Schema initialer Approximationen passt gegebenenfalls nicht mehr besonders gut. Die Vorgabe einer relativ kleinen Mindestgüte G_{\min} stellt eine hohe Anforderung an die Güte der Approximation der Einzelcluster und wirkt als Gegenkraft zum Zusammenfassen vieler Komponenten in einem Cluster. Für relativ große Mindestgüten G_{\min} bleibt jedoch Potential für eine geschicktere Clusterbildung über diese Arbeit hinaus bestehen.

Beispiel 8.3: Cluster für das ISAS Logo

Abbildungen 8.1 und 8.2 zeigen die durch das Cluster-Verfahren gebildeten Cluster für die Zieldichte aus Abschnitt 9.3 mit $G_{\min} = 10^{-5}$. Die Zieldichte bildet das Logo des Lehrstuhls für Intelligente Sensor-Aktor-Systeme (ISAS) der Universität Karlsruhe mit 1605 Gauß-Komponenten nach, siehe Abbildung 9.14. Gebildet werden 21 verschiedene Cluster C_n , $n = 1, \dots, 21$. Eingezeichnet sind jeweils die Dirac-Impulse der initialen Approximation, deren progressiv verbesserte Version die Approximation für den jeweiligen Cluster darstellt. Die initiale Gesamtapproximation der Zieldichte als Summe der initialen Approximationen der Einzelcluster ist in Abbildung 8.3 zu sehen. Abbildung 8.4 zeigt die Summe der progressiv verbesserten Approximationen der Einzelcluster und damit die Approximation der Zieldichte. ■

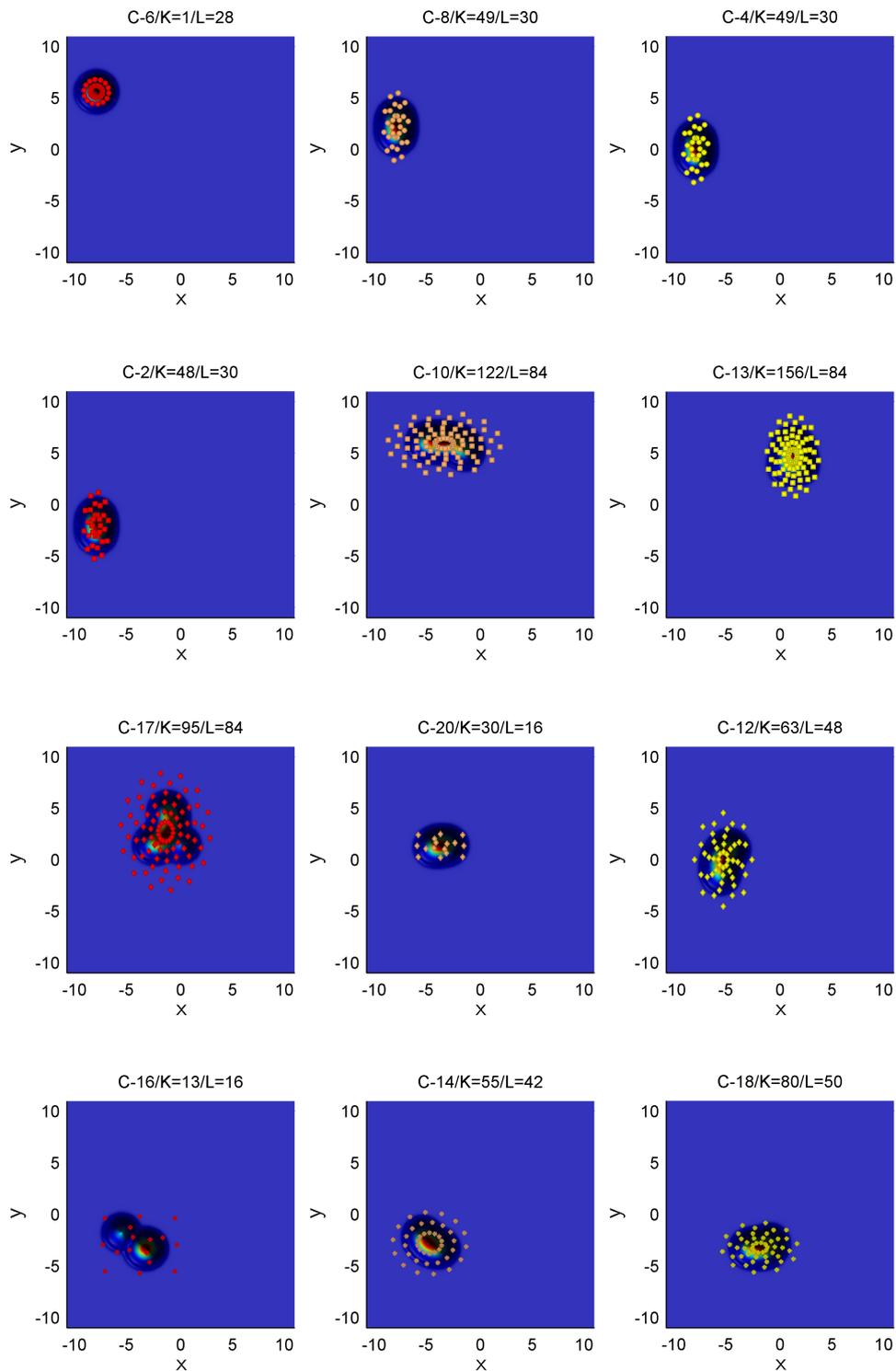


Abbildung 8.1: Cluster gebildet durch das Cluster-Verfahren zur Zieldichte aus Abschnitt 9.3. Dargestellt sind die Dichten der Cluster. In der Überschrift zu jeder Teilabbildung ist mit C- n die Nummer n des Clusters vermerkt. K ist die Anzahl der Komponenten im Cluster, L die Anzahl der Komponenten in der zugehörigen initialen Approximation. Die Positionen der Dirac-Impulse der initialen Approximation sind jeweils markiert.

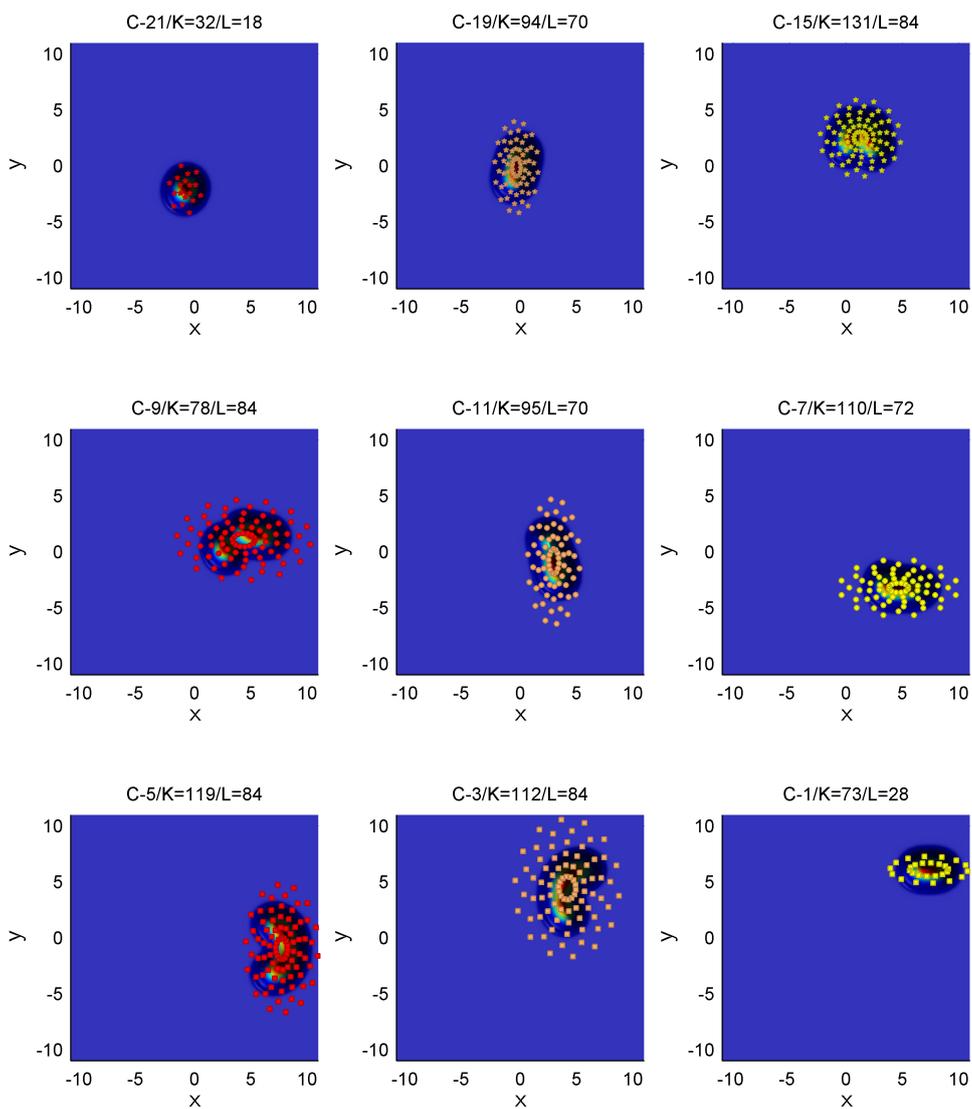


Abbildung 8.2: Fortsetzung von Abbildung 8.1.

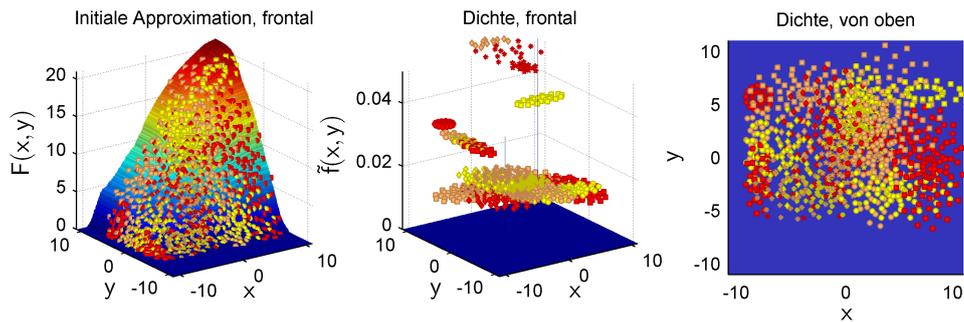


Abbildung 8.3: Darstellung der initialen Gesamtapproximation als Summe der initialen Approximationen aus Abbildung 8.1 und 8.2. Links ist die Verteilungsfunktion der Gesamtapproximation zu sehen. In der Mitte sowie rechts ist die quasi-ausgeartete Dichte der Gesamtstartverteilung abgebildet, von der auf Grund der sehr kleinen Varianzen allerdings wenig zu sehen ist. Mit eingezeichnet ist die Dichte der Gesamtapproximation. Die Grundlinien der Marker wurden zur besseren Übersicht weggelassen.

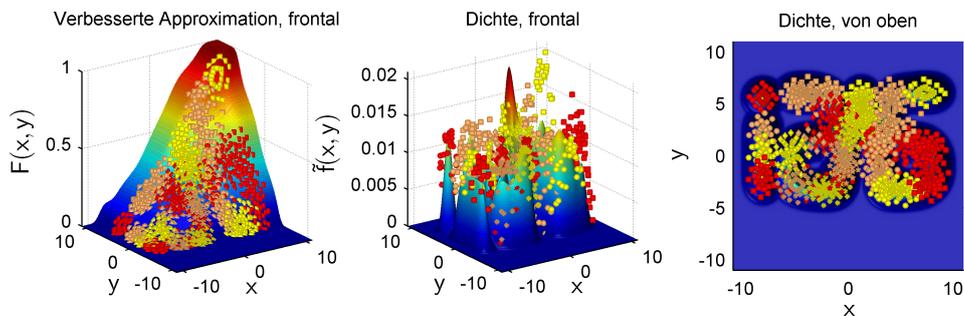


Abbildung 8.4: Progressiv verbesserte Version von Abbildung 8.3. Die Höhe der Dirac-Impulse in der mittleren Abbildung wurde auf die maximale Höhe der eingezeichneten Zieldichte normiert.

8.3 Erreichen einer vorgegebenen Gesamtgüte

Die Approximation der Einzelcluster hat als Parameter die mindestens zu erreichende Güte G_{\min} . Das Cluster-Verfahren besitzt G_{\min} als Eingabewert. Leider ist die Vorgabe der zu erreichenden Güte pro Cluster etwas unhandlich. Schöner wäre, eine zu erreichende Güte $G_{\min \text{ gesamt}}$ für die Gesamtapproximation vorgeben zu können.

Lemma 8.3.1 Sei K die Anzahl der Komponenten in der Zieldichte. Bezeichne $A(G_{\min})$ die Anzahl der Cluster, die das Cluster-Verfahren bei einer festgehaltenen Zieldichte für verschiedene Werte von G_{\min} bildet. Wird bei einer Anwendung des Cluster-Verfahrens für jeden Cluster die vorgegebene Güte G_{\min} erreicht, dann gilt für die Gesamtgüte G_{gesamt} der erstellten Gesamtapproximation die Abschätzung

$$G_{\text{gesamt}} \leq G_{\min} \cdot A(G_{\min}) .$$

■

Um eine vorgegebene Gesamtgüte $G_{\min\text{gesamt}}$ zu erreichen, kann die Anzahl $A \leq K$ der Cluster, die das Verfahren erstellen wird, geschätzt und

$$G_{\min} := \frac{G_{\min\text{gesamt}}}{A}$$

gewählt werden.

Allgemein macht für die Wahl von G_{\min} ein Wert aus dem Intervall $[\frac{G_{\min\text{gesamt}}}{K}, G_{\min\text{gesamt}}]$ Sinn:

- Ein kleiner Wert ist nicht sinnvoll, da das Verfahren mit $G_{\min} := \frac{G_{\min\text{gesamt}}}{K}$ die erforderliche Gesamtgüte schon erreicht, sofern G_{\min} in jedem Einzelcluster erreicht wird. Sollte dies nicht der Fall sein, dann ist es unwahrscheinlich — wenn auch nicht unmöglich —, dass dies für einen kleineren Wert von G_{\min} der Fall sein wird.
- Ein größerer Wert ist ebenfalls nicht sinnvoll, da mindestens ein Cluster gebildet wird und das Cluster-Verfahren für jeden Cluster versucht, diesen nicht mit einer wesentlich besseren Güte als der vorgegebenen zu approximieren, um so die Anzahl der Dirac-Impulse in der Approximation so gering wie möglich zu halten.

Da das Intervall begrenzt ist, kann ein Halbierungsverfahren eingesetzt werden, um ein geeignetes G_{\min} zu finden — unter der heuristischen Annahme, dass das Verfahren für kleinere G_{\min} immer eine bessere Gesamtgüte G_{gesamt} erzeugt. Die Statistiken in den Abbildungen 9.1, 9.8 und 9.15 aus den Anwendungen des Cluster-Verfahrens im Rahmen des Vergleichs in Kapitel 9 stützen diese heuristische Annahme. Im Hinblick auf den Laufzeitaufwand wurde für den Vergleich allerdings darauf verzichtet, G_{\min} automatisch aus einem vorgegebenen G_{gesamt} zu bestimmen.

Vergleich zu anderen Verfahren

In diesem Kapitel wird das Cluster-Verfahren aus Kapitel 8 mit anderen Verfahren verglichen. Dazu wurde im Rahmen dieser Arbeit eine Implementierung des Cluster-Verfahrens in Matlab erstellt. Sie beinhaltet eine Implementierung des Progressiven Bayes-Verfahrens für den zweidimensionalen Fall achsenparalleler Gauß-Mischverteilungen sowie Gleichverteilungen als Start-, Ziel- oder progressive Verteilung in jedem Progressionsschritt und ermöglicht über konvexe Kombinationen dieser progressiven Verteilungen eine Realisierung des Ansatzes zur progressiven Verbesserung initialer Approximationen.

Verglichen werden sollen die Ergebnisse dieser Implementierung bei Anwendung auf drei verschiedene Zieldichten mit den Ergebnissen einer am Lehrstuhl für Intelligente Sensor-Aktor-Systeme (ISAS) der Universität Karlsruhe entwickelten Implementierung des in Abschnitt 3.2.3 beschriebenen Greedy-Verfahrens sowie mit aus zufälligen Stichproben erstellten Approximationen wie in Abschnitt 3.2.2 beschrieben. Dazu wird in den drei folgenden Unterabschnitten jeweils eine der Zieldichten mit allen drei Verfahren approximiert. Die Güte wird hier — wie in den anderen Teilen dieser Arbeit auch — numerisch berechnet mit einer Rasterung von 800 mal 800 Punkten für das Integral in der Gütefunktion. Im Folgenden wird die Grundstruktur der Vergleiche in den Unterabschnitten dieses Kapitels beschrieben.

Für das Cluster-Verfahren werden verschiedene Mindestgüten pro Cluster vorgegeben. In der jeweils ersten Abbildung mit statistischen Auswertungen in den drei Unterabschnitten ist die erreichte Gesamtgüte der Approximation, die Anzahl ihrer Dirac-Impulse, die Anzahl der gebildeten Cluster sowie die benötigte Zeit in Abhängigkeit von der vorgegebenen Mindestgüte G_{\min} pro Cluster vermerkt. Gerechnet wurde auf einem Computer mit 3.2GHz Pentium 4 und einem Gigabyte Hauptspeicher. Es folgt jeweils ein Vergleich mit den beiden anderen Verfahren bezüglich der erreichten (Gesamt-)Güten bezogen auf die eingesetzte Anzahl von Dirac-Impulsen. Die Ergebnisse des Cluster-Verfahrens werden dabei mit „Clustern“ bezeichnet.

Das Greedy-Verfahren ist iterativ. In jeder Iteration liegt eine Approximation der Zieldichte vor, die Anzahl der Dirac-Impulse steigt dabei um Eins pro Iterationsschritt an. Neben der Zieldichte wird eine maximale Anzahl von Dirac-Impulsen vorgegeben und die Approximationen zu der angegebenen Anzahl von Dirac-Impulsen wird der entsprechenden Iteration entnommen. Im Vergleich werden die erreichten Güten mit „Greedy“ bezeichnet.

Für jede Zieldichte werden jeweils 10 zufällige Stichproben erstellt, jede mit einem Umfang von 8192 Einzelstichproben. Zu jeder i .ten dieser 10 Stichproben und einer vorgegebenen Anzahl L von Dirac-Impulsen werden die ersten L Einzelstichproben aus der Stichprobe als Positionen der Dirac-Impulse in einer Approximation $A_{i,L}$ aufgefasst. Alle Gewichte der Dirac-Impulse werden für die Approximation $A_{i,L}$ gleich groß gewählt, so dass ihre Summe Eins ergibt, was jeweils dem Integral über der Zieldichte entspricht. Zum Vergleich zu den anderen beiden Verfahren wird für jedes vorgegebene L der Mittelwert $G_L := \frac{1}{10} \sum_{i=1}^{10} G_{i,L}$ der Güte $G_{i,L}$ der Approximation $A_{i,L}$ gebildet. Diese Mittelwerte werden in den Statistiken mit „Samples“ bezeichnet. Die jeweils letzte Abbildung mit statistischen Auswertungen in den Unterabschnitten zeigt die Mittelwerte zusammen mit den empirischen Standardabweichungen, um einen Eindruck über die Streuung zu vermitteln. Schließlich kann eine durch eine zufällige Stichprobe erstellte Approximation prinzipiell beliebig schlecht als auch gut sein, auch wenn die Wahrscheinlichkeit dafür jeweils gegen Null tendiert.

Für alle y -Achsen der Abbildungen mit statistischen Auswertungen wird eine logarithmische Skala verwendet, da die abgetragenen Funktionswerte sehr große Bereiche abdecken. Konkrete Funktionswerte sind mit einem Kreuz für das Cluster-Verfahren oder einem Punkt für die beiden anderen Verfahren markiert. Sie werden linear interpoliert. Die angegebenen Verhältnisse von zwei Funktionswerten für zwei verschiedene Verfahren ergeben sich aus den Interpolationen. Einige der Funktionswerte sind textuell vermerkt, da sie nicht ganz so einfach an der logarithmischen Skala ablesbar sind.

9.1 Beispiel Standardnormalverteilung

Zielverteilung in diesem Abschnitt ist die Standardnormalverteilung auf dem Intervall $[-10, 10]$, wie sie in Abbildung 2.2 zu sehen ist. Sie soll einen Eindruck von der erreichbaren Güte durch die verschiedenen Verfahren für eine einfachst mögliche Gauß-Mischverteilung in Form der Standardnormalverteilung vermitteln.

In Abbildung 9.1 sind die Ergebnisse des Cluster-Verfahrens zu sehen. Abbildung 9.2 zeigt, dass das Cluster-Verfahren bezogen auf die eingesetzte Anzahl von Dirac-Impulsen die beste Güte erreicht. Gegenüber dem Greedy-Verfahren wird eine 6.6 bis 19fach bessere — sprich kleinere — Güte, gegenüber den zufälligen Stichproben sogar Faktor 15 bis 55, erreicht, jeweils bezogen auf die eingesetzte Anzahl von Dirac-Impulsen. Im Wesentlichen wächst der Unterschied mit zunehmender Anzahl von Dirac-Impulsen immer weiter an, erfährt jedoch gegenüber dem Greedy-Verfahren für die letzten der für das Cluster-Verfahren verfügbaren Werte eine Stagnation.

Hier zeigt sich deutlich ein Nachteil des Cluster-Verfahrens: Mit den voreingestellten Parametern für das Progressive Bayes-Verfahren und dem verwendeten Schema initialer Approximationen aus Beispiel 7.5 wird auch bei Vorgabe einer Mindestgüte von $G_{\min} = 0$ pro Cluster

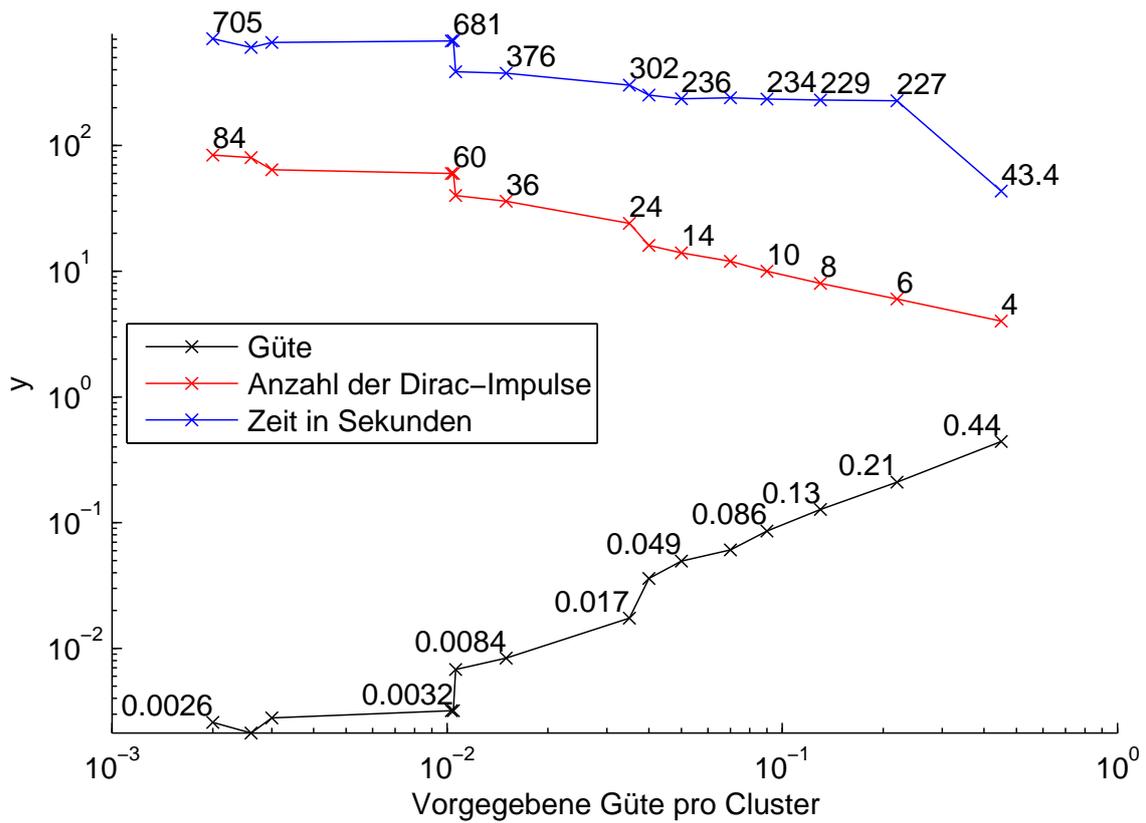


Abbildung 9.1: Ergebnisse des Cluster-Verfahrens für die Standardnormalverteilung.

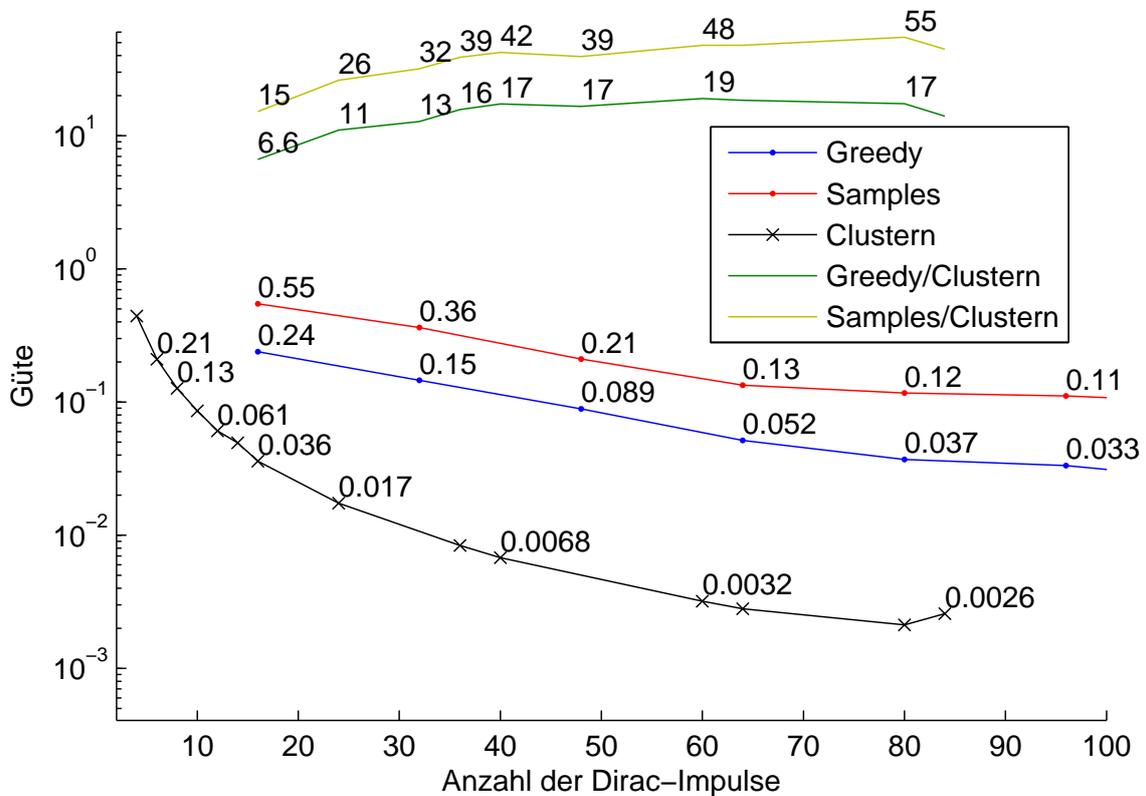


Abbildung 9.2: Vergleich der erreichten Güte der verschiedenen Verfahren für die Standardnormalverteilung.

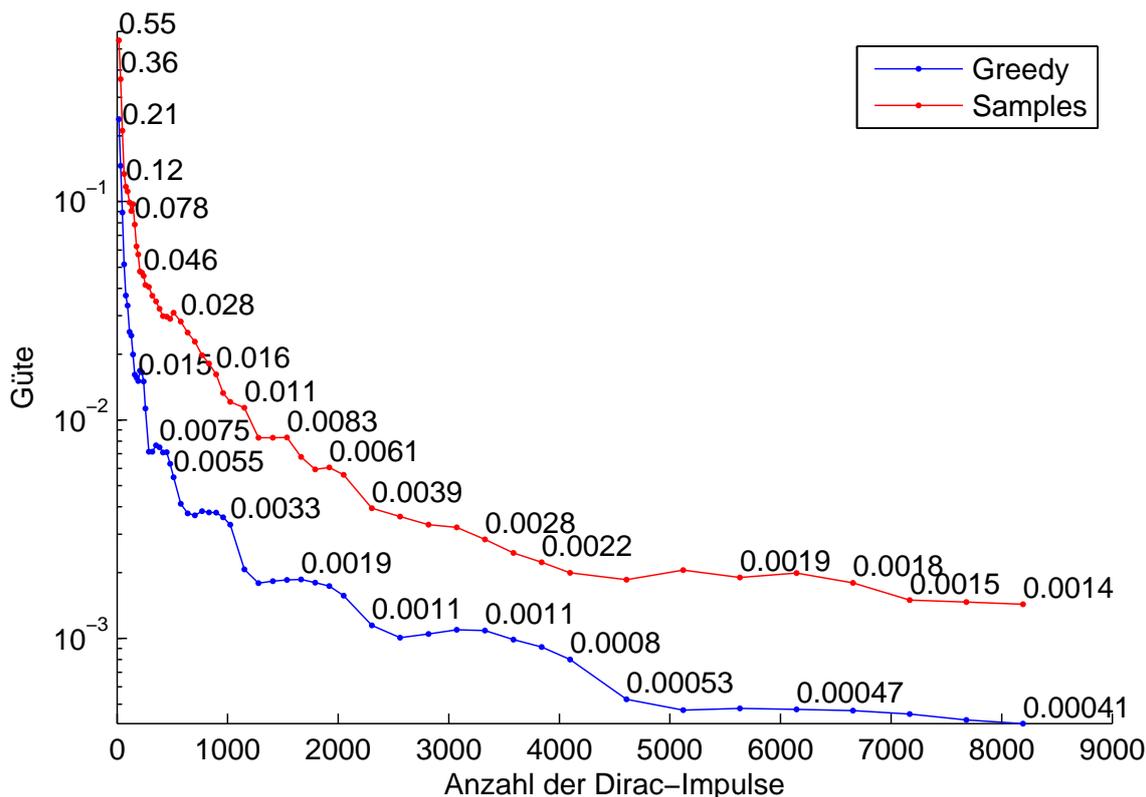


Abbildung 9.3: Erreichte Güte durch das Greedy-Verfahren und die zufälligen Stichproben für größere Anzahlen von Dirac-Impulsen für die Standardnormalverteilung.

nur eine Approximation erzeugt, die 84 Dirac-Impulse umfasst. Mit anderen Parametern oder initialen Approximationen kann eine bessere Güte erreicht werden. Um dem Cluster-Verfahren auf diese Weise jedoch keinen Vorteil zu verschaffen, wurden für alle drei Zieldichten in diesem Kapitel die gleichen Parameter und das gleiche Schema initialer Approximationen verwendet.

Sowohl mit dem Greedy-Verfahren als auch mit den zufälligen Stichproben kann die beste durch das Cluster-Verfahren erreichte Güte von etwa 0.0022 ohne Weiteres übertroffen werden — siehe Abbildung 9.3 —, und das mit einem sehr kleinen Bruchteil der durch das Cluster-Verfahren benötigten Laufzeit. Allerdings benötigt das Greedy-Verfahren dazu etwa 1000 Dirac-Impulse, während diese Güte als Mittelwert durch die zufälligen Stichproben erst bei etwa 4000 Dirac-Impulsen erreicht wird. Demgegenüber begnügt sich das Cluster-Verfahren mit 80 Dirac-Impulsen, um die Güte von 0.0022 zu erreichen. Abbildung 9.4 vermittelt einen Eindruck über die Streuung der Güte der mit den zufälligen Stichproben erstellten Approximationen.

Abbildung 9.5 und 9.6 vermitteln einen Eindruck davon, warum das Cluster-Verfahren bei gleicher Anzahl von Dirac-Impulsen eine deutlich bessere Güte erreicht als das Greedy-Verfahren. Dargestellt sind die durch das Greedy- sowie das Cluster-Verfahren erstellten Approximationen für verschiedene Anzahlen von Dirac-Impulsen. In jeder Zeile sind jeweils Approximationen mit etwa gleicher Anzahl von Dirac-Impulsen zu sehen. Da das Cluster-Verfahren schon auf Grund der vorgegebenen Beschränkung von $L_{\max} = 100$ (siehe Abschnitt 8) keine Approximation mit mehr als 100 Dirac-Impulsen für eine einzelne Gauß-Verteilung erzeugen kann, ist rechts in

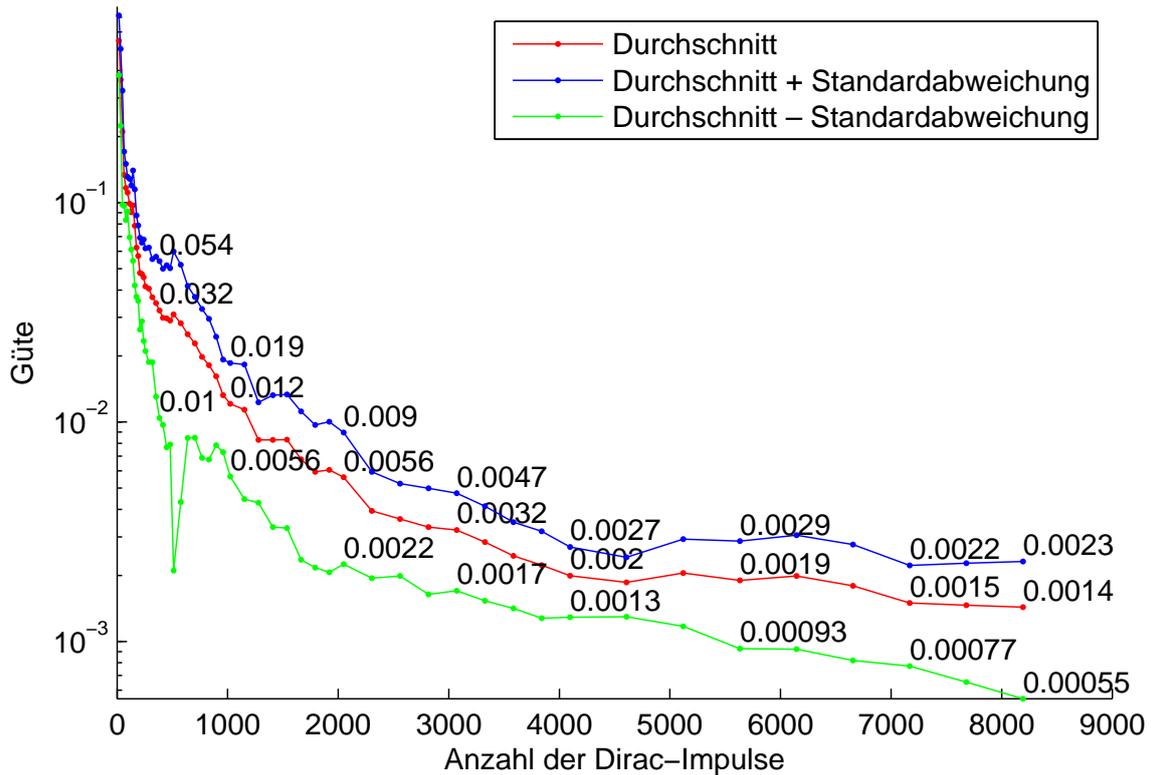


Abbildung 9.4: Streuung der Güte für die 10 verschiedenen Zufallsstichproben für die Standardnormalverteilung.

Abbildung 9.6 die Standardnormalverteilung im Vergleich zu ihrer Approximation mit 8192 Dirac-Impulsen durch das Greedy-Verfahren zu sehen.

Die dargestellten Approximationen durch das Greedy-Verfahren wirken gegenüber den Approximationen durch das Cluster-Verfahren „kantig“. Ein wesentlicher Grund dafür ist, dass das Greedy-Verfahren durch die Bisektionen viele achsenparallel ausgerichtete Dirac-Impulse erzeugt, wodurch die Überschneidungsbereiche der Heaviside-Funktionen nicht besonders gut genutzt werden. Das dem Cluster-Verfahren zu Grunde liegende Progressive Bayes-Verfahren nutzt diese Überschneidungsbereiche wesentlich besser.

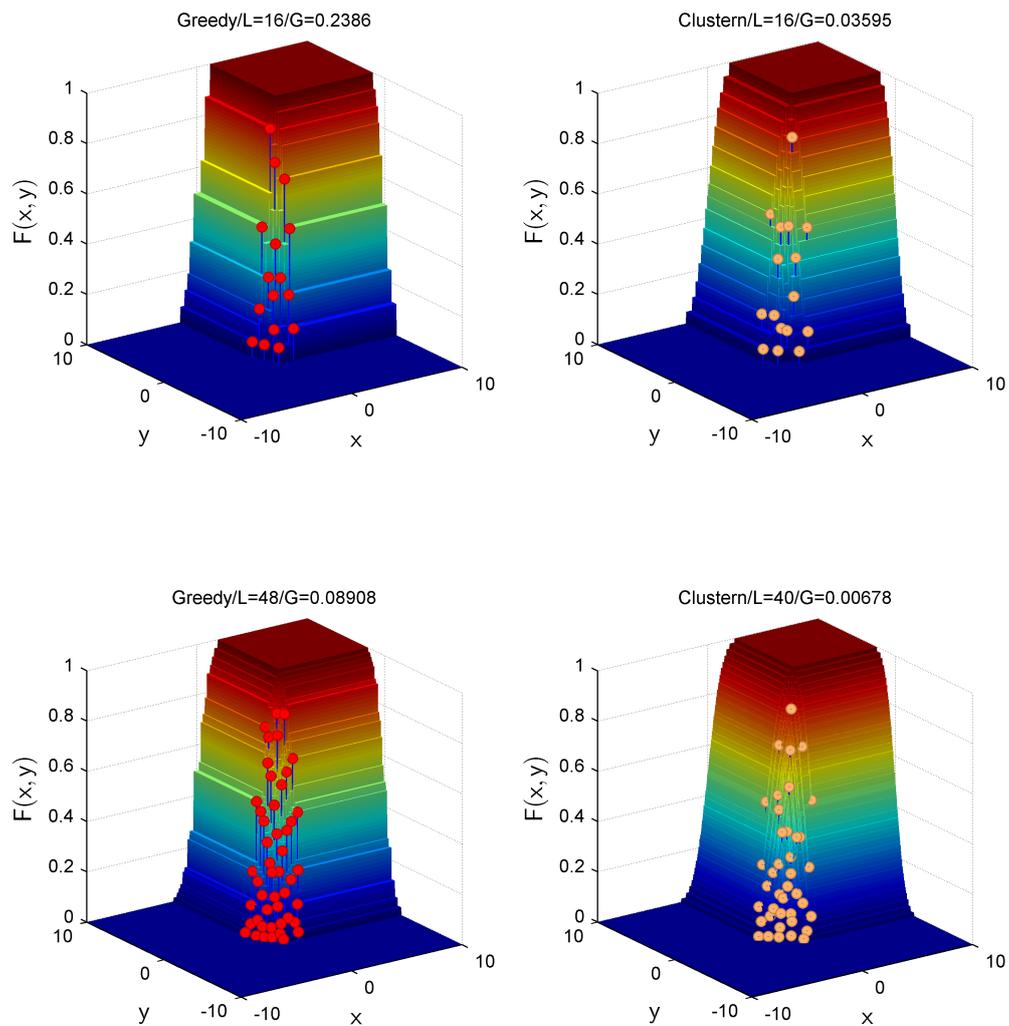


Abbildung 9.5: Vergleich der Approximationen aus dem Greedy-Verfahren zu denen aus dem Cluster-Verfahren bei etwa gleicher Anzahl von Dirac-Impulsen. Dargestellt sind die Verteilungsfunktionen der erzeugten Approximationen aus einer frontalen Perspektive. Der Titel nennt das Verfahren, die Anzahl der Dirac-Impulse sowie die Güte der Approximation.

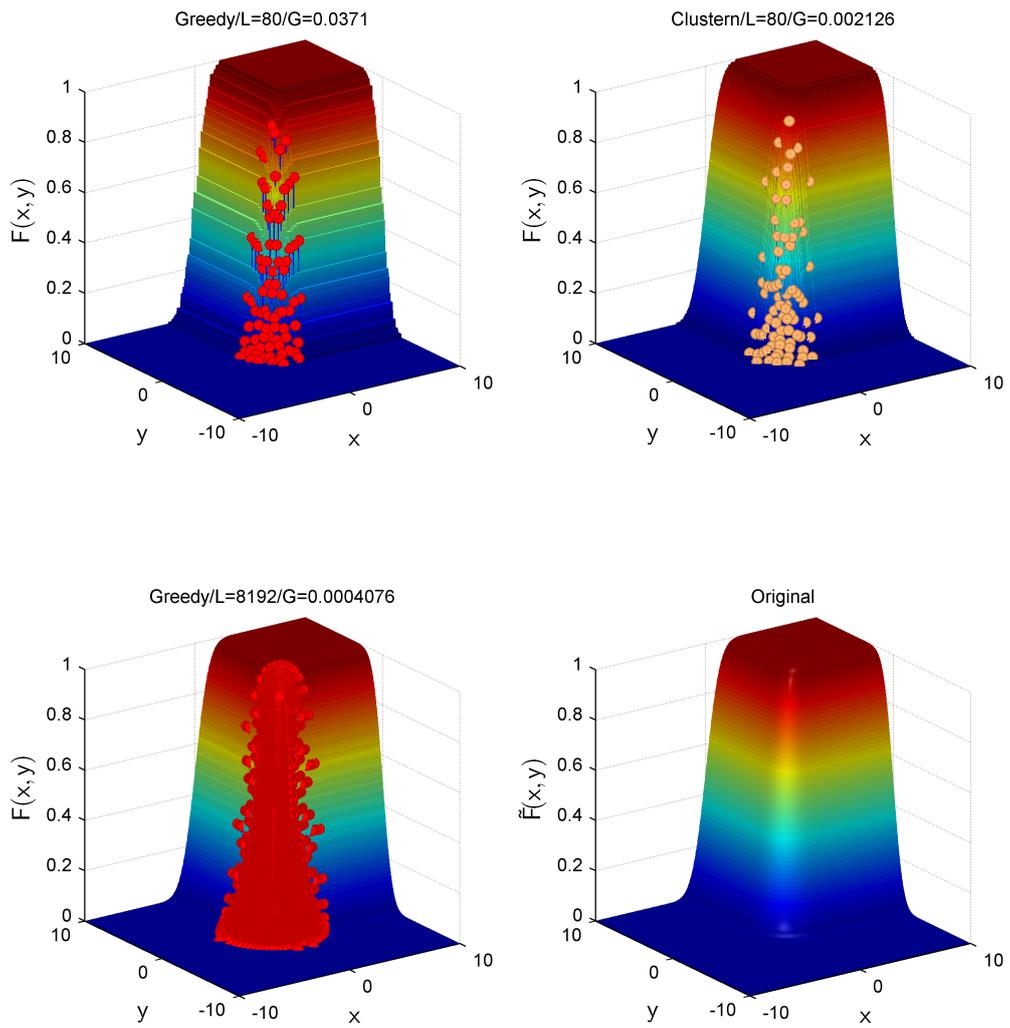


Abbildung 9.6: Fortsetzung von Abbildung 9.5.

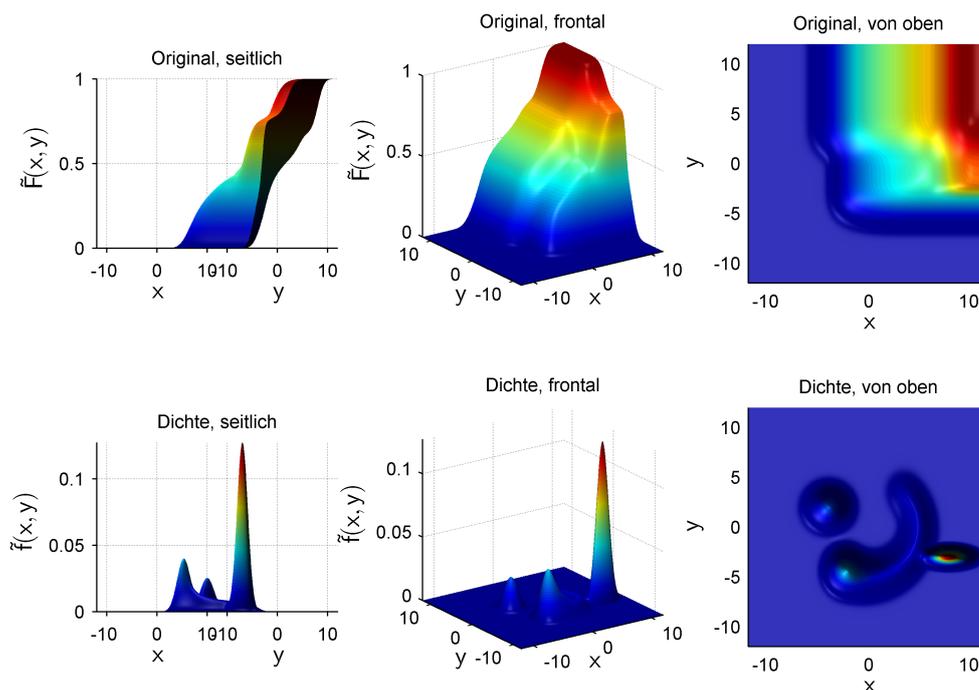


Abbildung 9.7: Zur Approximation verwendete Zieldichte für Abschnitt 9.2. Dargestellt ist die Verteilungsfunktion (obere Zeile) sowie die Dichte selbst (untere Zeile) aus verschiedenen Perspektiven.

9.2 Beispiel Halbkreis

Die in Abbildung 9.7 dargestellte Zieldichte auf dem Intervall $[-12, 12]$ ist eine Gauß-Mischdichte mit 16 Komponenten. Zwei dieser Komponenten bilden jeweils eine — mehr oder weniger — separierte Mode aus, während die anderen 14 einen Halbkreis bilden. Diese Zieldichte soll als Beispiel für Gauß-Mischdichten mit geringer bis mittlerer Komponentenanzahl dienen und zeigen, dass das Cluster-Verfahren auch mit separierten Moden umgehen kann.

Abbildung 9.8 zeigt die Ergebnisse des Cluster-Verfahrens. Qualitativ ähnelt der Verlauf der Funktionswerte der erreichten Güten durch die verschiedenen Verfahren in Abbildung 9.9 denen aus Abbildung 9.2 für die Standardnormalverteilung. Das Cluster-Verfahren erreicht wiederum die beste Güte bei einer festgehaltenen Anzahl von Dirac-Impulsen und der Abstand nimmt mit zunehmender Anzahl immer weiter zu. Gegenüber dem Greedy-Verfahren wird eine 3.6 bis 38fach bessere Güte bezogen auf die eingesetzte Anzahl von Dirac-Impulsen erreicht, gegenüber den zufälligen Stichproben beträgt der Faktor 25 bis über 200. Im Vergleich zur Standardnormalverteilung hat sich damit der Vorsprung des Cluster-Verfahrens zu den beiden anderen Verfahren vergrößert.

Selbst mit 8192 Dirac-Impulsen erreicht der Durchschnitt der zufälligen Stichproben nur eine Güte von 0.0057, siehe Abbildung 9.10. Diese Güte erreicht das Cluster-Verfahren schon mit 128 Dirac-Impulsen, siehe Abbildung 9.8. Das Greedy-Verfahren beginnt etwa ab 2000 Dirac-Impulsen zu stagnieren und kommt über eine Güte von 0.015 nicht hinaus.

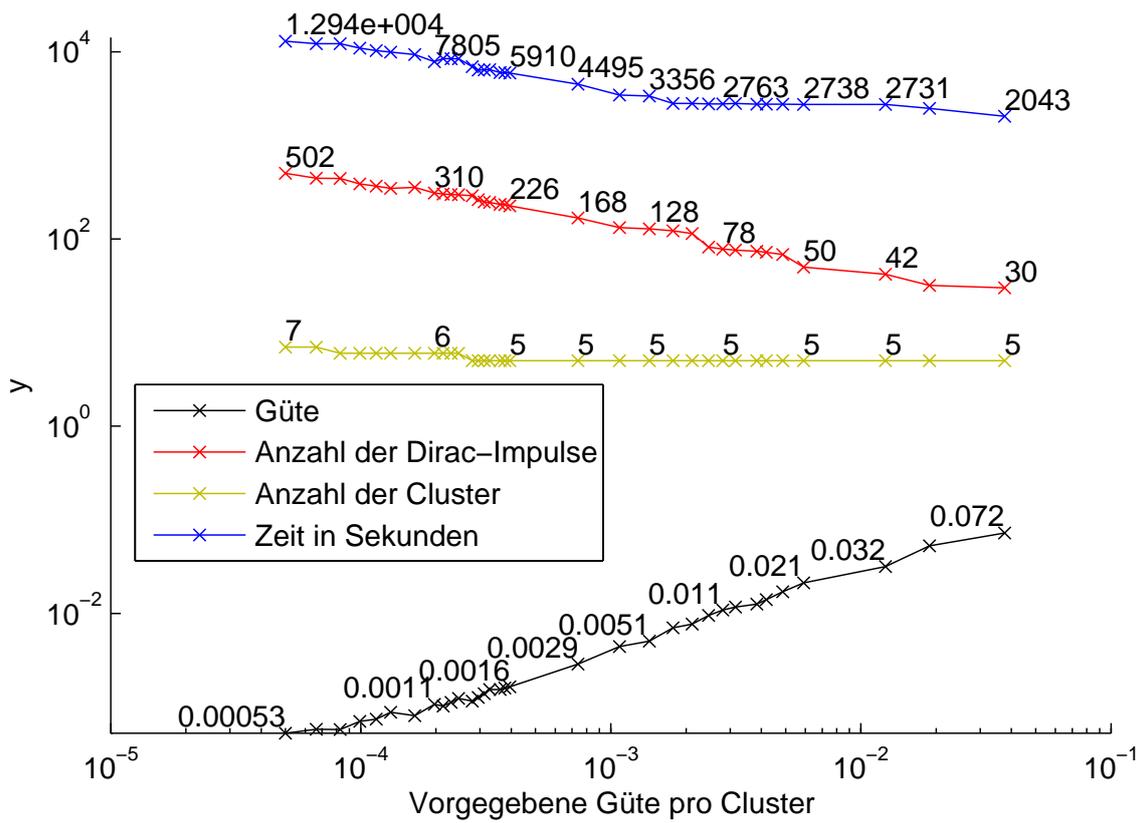


Abbildung 9.8: Ergebnisse des Cluster-Verfahrens für das Halbkreis Beispiel.

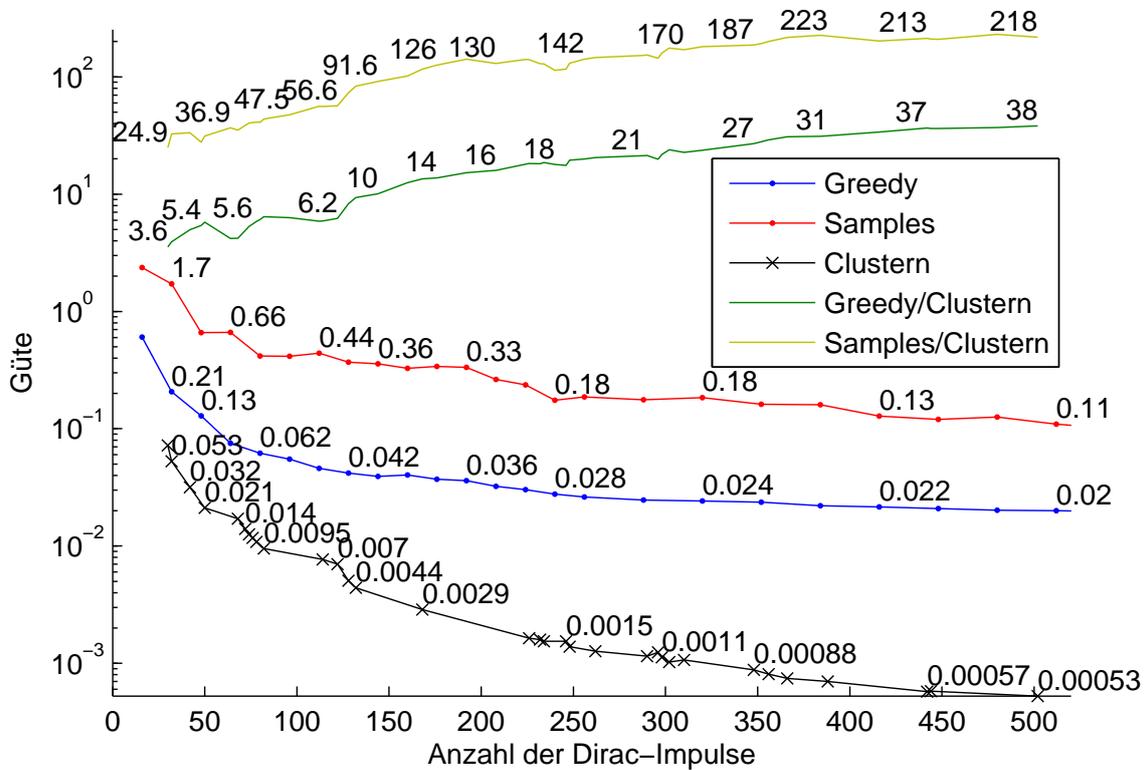


Abbildung 9.9: Vergleich der erreichten Güte der verschiedenen Verfahren für das Halbkreis Beispiel.

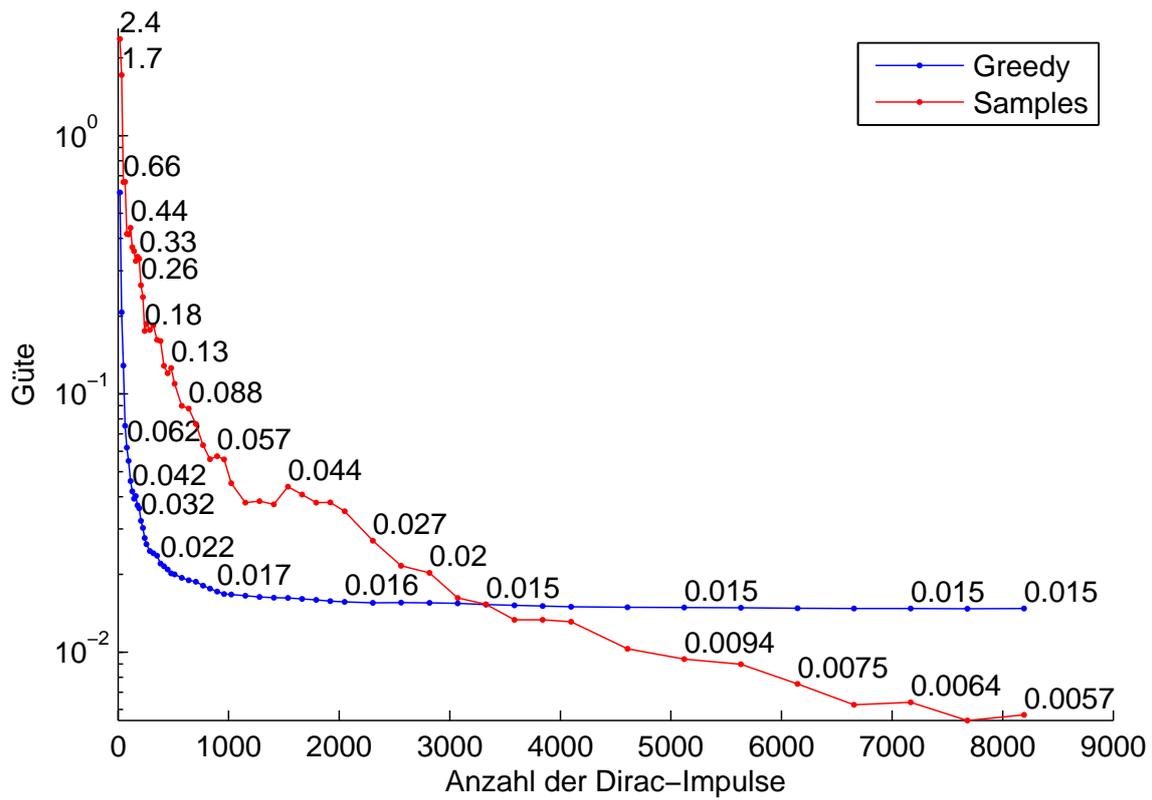


Abbildung 9.10: Erreichte Güte durch das Greedy-Verfahren und die zufälligen Stichproben für größere Anzahlen von Dirac-Impulsen für das Halbkreis Beispiel.

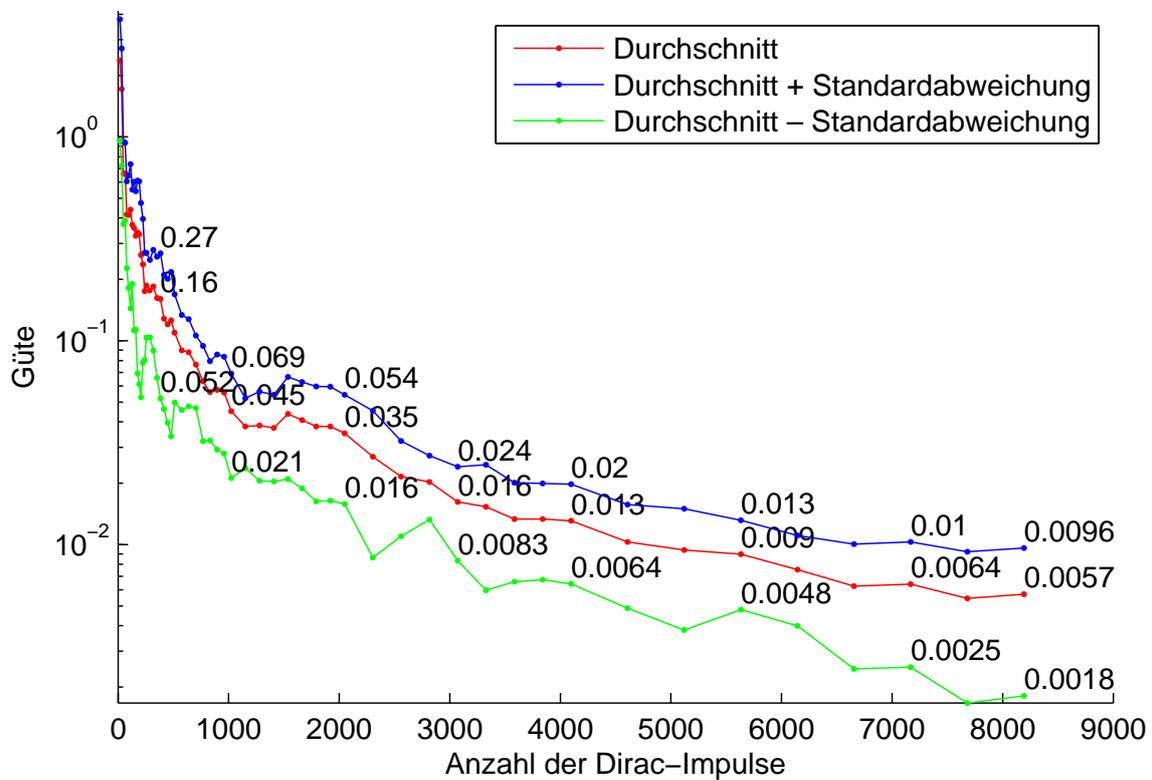


Abbildung 9.11: Streuung der Güte für die 10 verschiedenen Zufallsstichproben für das Halbkreis Beispiel.

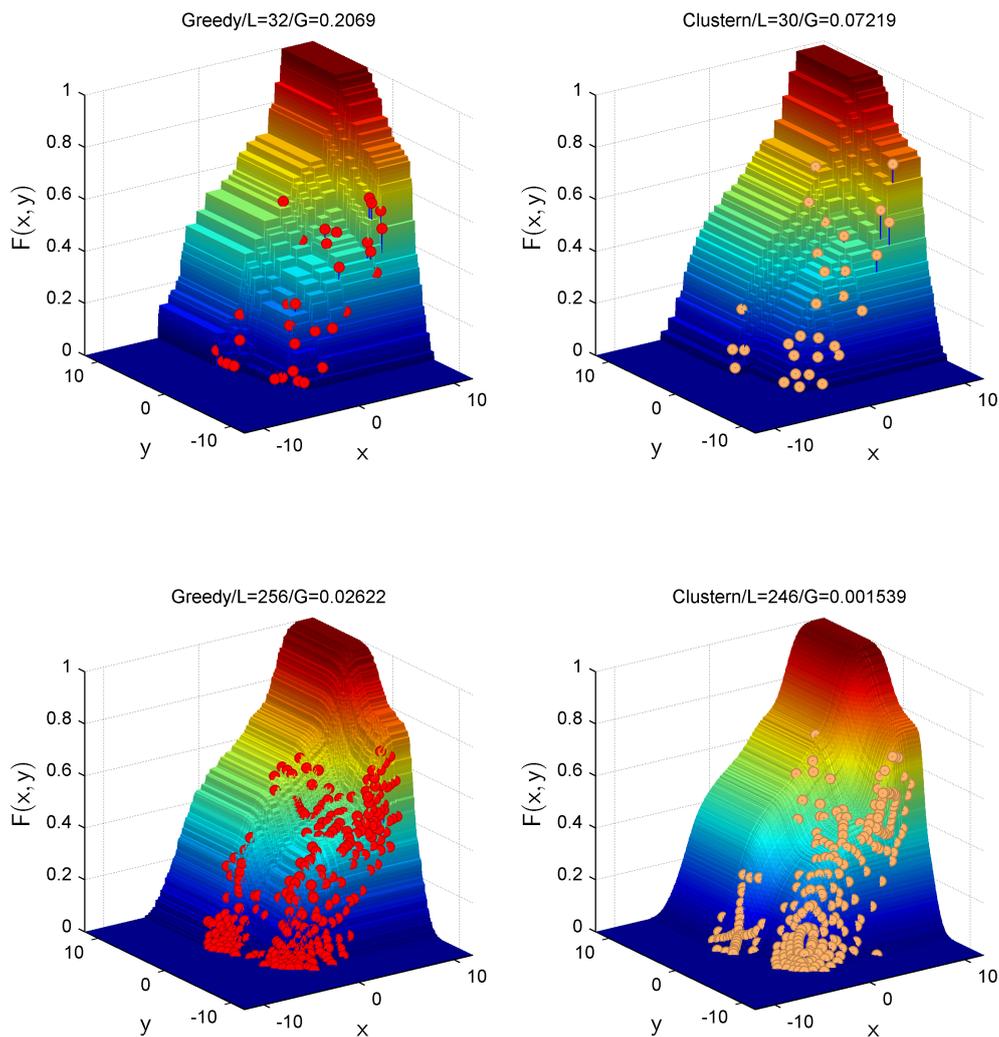


Abbildung 9.12: Vergleich der Approximationen aus dem Greedy-Verfahren zu denen aus dem Cluster-Verfahren bei etwa gleicher Anzahl von Dirac-Impulsen. Dargestellt sind die Verteilungsfunktionen der erzeugten Approximationen aus einer frontalen Perspektive. Der Titel nennt das Verfahren, die Anzahl L der Dirac-Impulse sowie die Güte G der Approximation.

Die Abbildungen 9.12 und 9.13 der Approximationen des Greedy- und des Cluster-Verfahrens für einige ausgewählte Anzahlen von Dirac-Impulsen sollen einen Eindruck der erreichten Qualität vermitteln. Das Greedy-Verfahren wurde hier zum Vergleich gewählt, da es für die Anzahlen von Dirac-Impulsen, zu denen Approximationen aus dem Cluster-Verfahren vorliegen, gegenüber den zufälligen Stichproben die deutlich bessere Güte liefert und erst ab etwa 3000 Dirac-Impulsen von den zufälligen Stichproben hinsichtlich der erreichten Güte geschlagen wird.

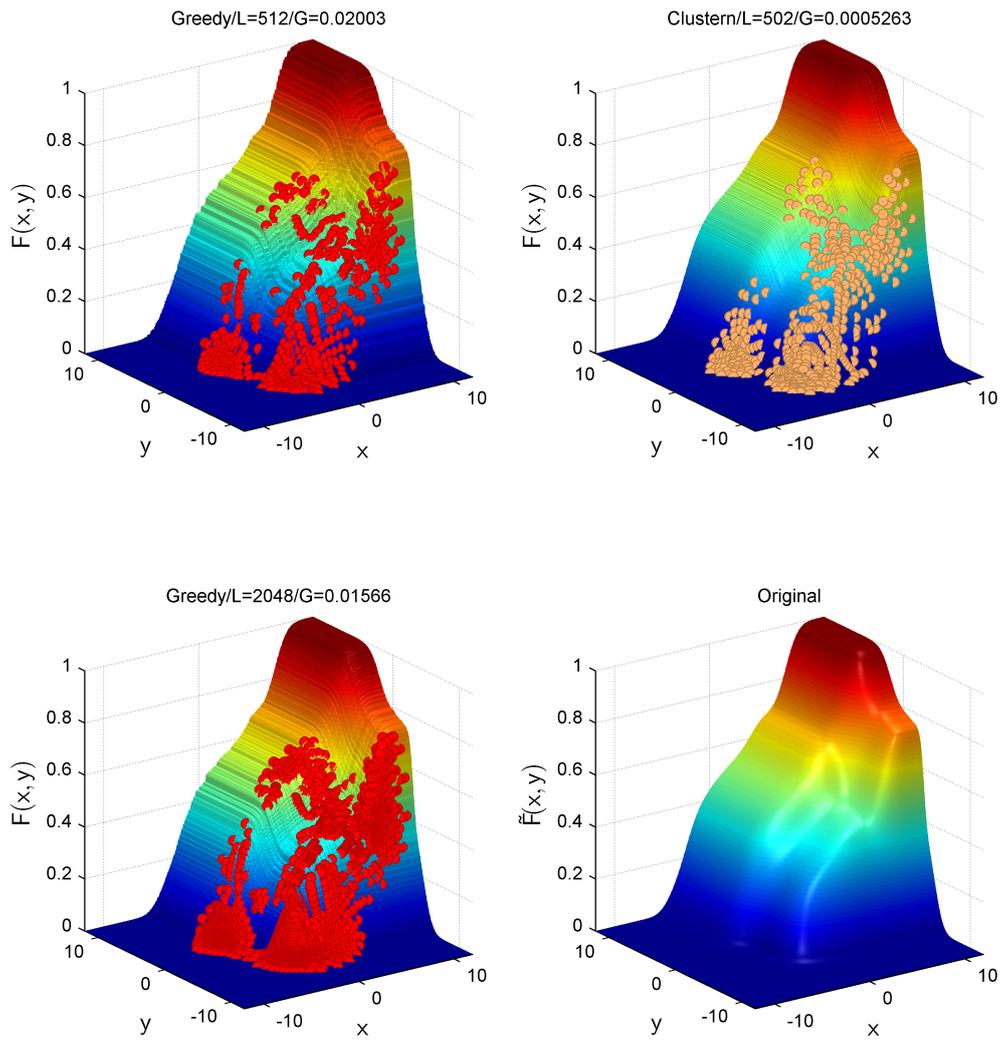


Abbildung 9.13: Fortsetzung von Abbildung 9.12.

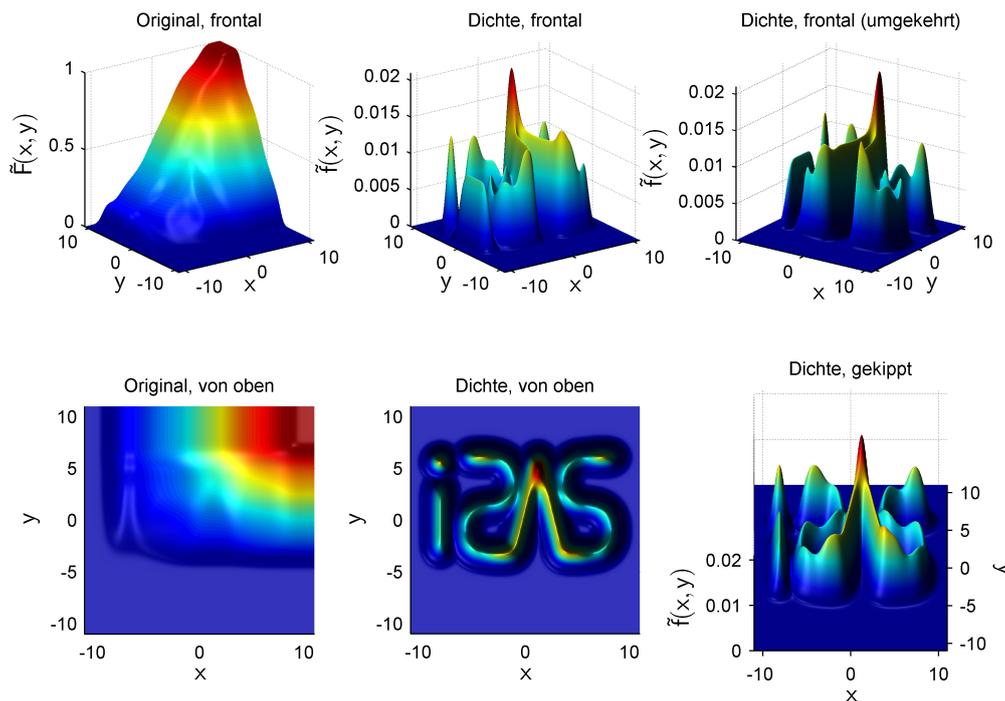


Abbildung 9.14: Zur Approximation verwendete Zieldichte für Abschnitt 9.3. Dargestellt ist die Verteilungsfunktion (links) sowie die Dichte selbst aus verschiedenen Perspektiven.

9.3 Beispiel ISAS Logo

Die in Abbildung 9.14 dargestellte Zieldichte auf dem Intervall $[-11, 11]$ ist eine Gauß-Mischdichte mit 1605 Komponenten. Sie bildet die Gestalt des Logos des Lehrstuhls für Intelligente Sensor-Aktor-Systeme (ISAS) der Universität Karlsruhe nach. Diese Zieldichte soll als Beispiel für kompliziert geformte Gauß-Mischdichten mit großer Komponentenanzahl und starker Überschneidung der Komponenten dienen.

Abbildung 9.15 zeigt die Ergebnisse des Cluster-Verfahrens. Die Zeiten zu den beiden kleinsten vorgegebenen Güten pro Cluster sind circa Angaben. Qualitativ ähnelt der Verlauf der Funktionswerte der erreichten Güten durch die verschiedenen Verfahren in Abbildung 9.16 denen aus Abbildung 9.2 für die Standardnormalverteilung und 9.9 für das Halbkreis Beispiel, wobei das Greedy-Verfahren allerdings zurückfällt. Das Cluster-Verfahren erreicht wiederum die beste Güte bei einer festgehaltenen Anzahl von Dirac-Impulsen und der Abstand nimmt mit zunehmender Anzahl immer weiter zu. Gegenüber den zufälligen Stichproben wird eine 26 bis 600fach bessere Güte bezogen auf die eingesetzte Anzahl von Dirac-Impulsen erreicht. Damit hat sich der Vorsprung des Cluster-Verfahrens ein weiteres mal vergrößert.

Selbst mit 8192 Dirac-Impulsen erreicht der Durchschnitt der zufälligen Stichproben „nur“ eine Güte von 0.0078, siehe Abbildung 9.16. Diese Güte erreicht das Cluster-Verfahren schon mit 360 Dirac-Impulsen. Das Greedy-Verfahren beginnt etwa ab 500 Dirac-Impulsen zu stagnieren und kommt über eine Güte von 0.28 nicht hinaus.

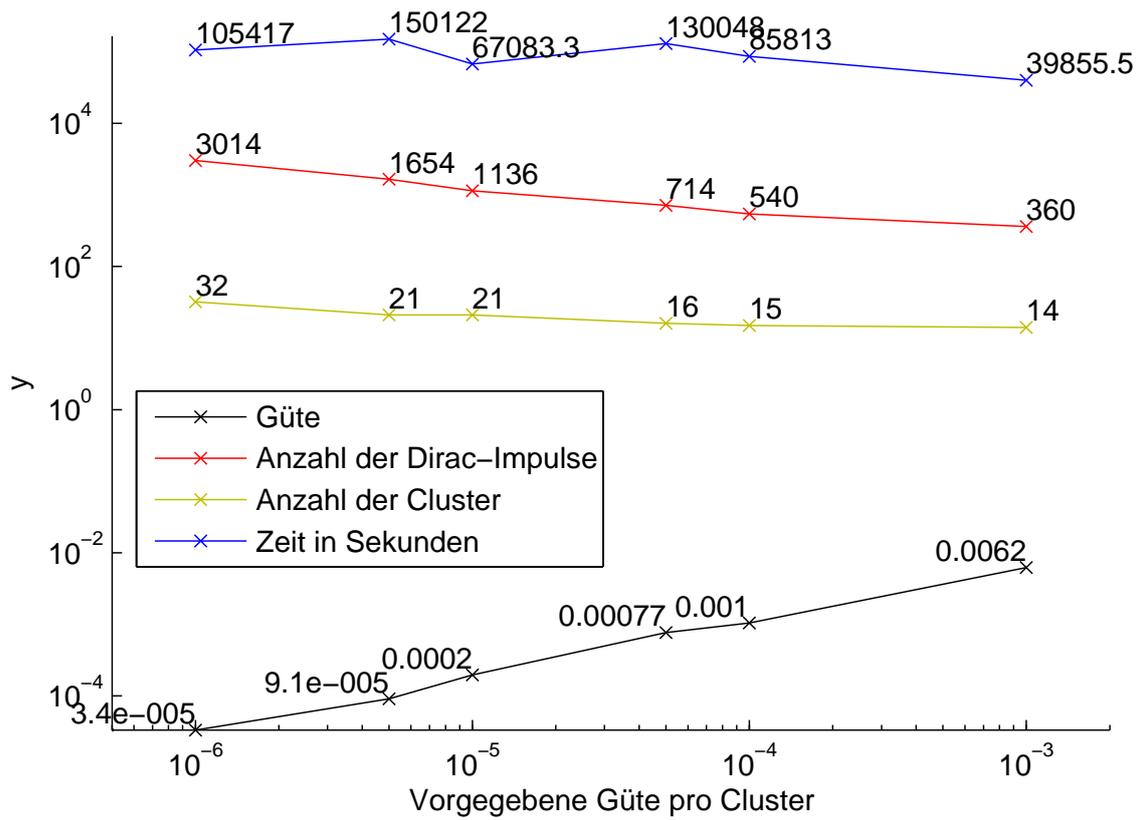


Abbildung 9.15: Ergebnisse des Cluster-Verfahrens für das ISAS Logo Beispiel.

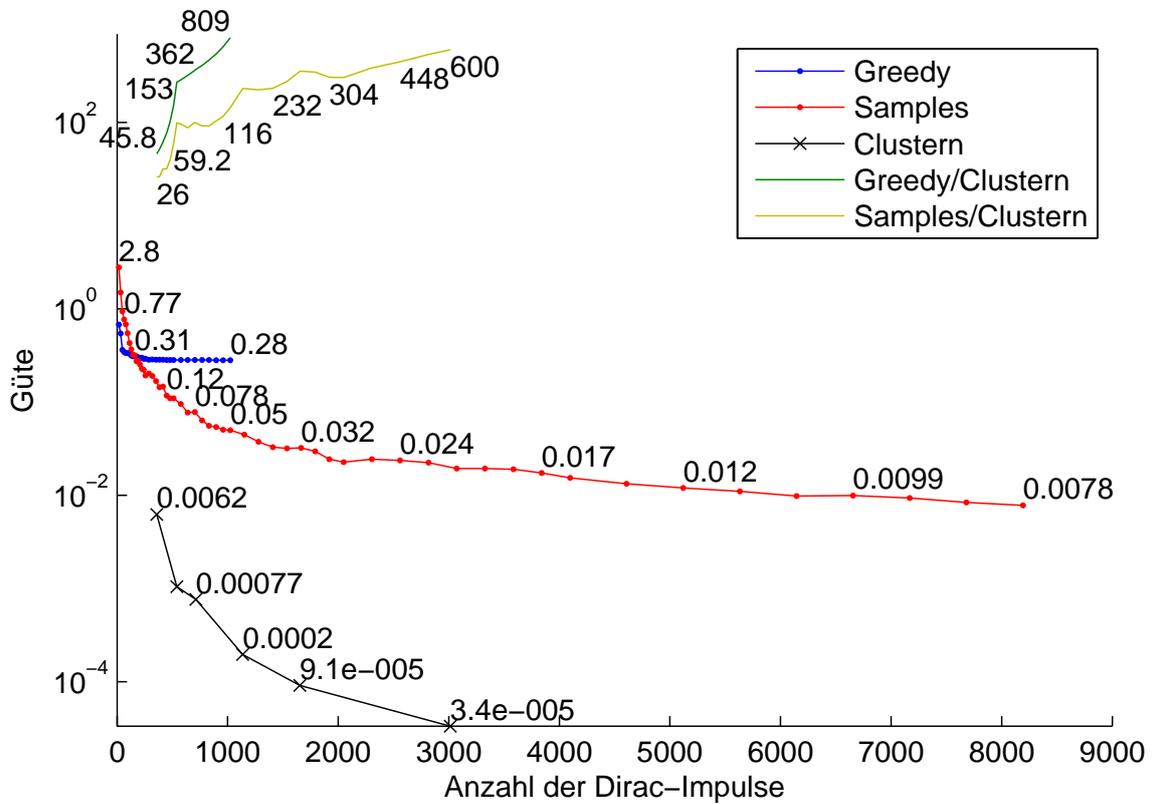


Abbildung 9.16: Vergleich der erreichten Güte der verschiedenen Verfahren für das ISAS Logo Beispiel.

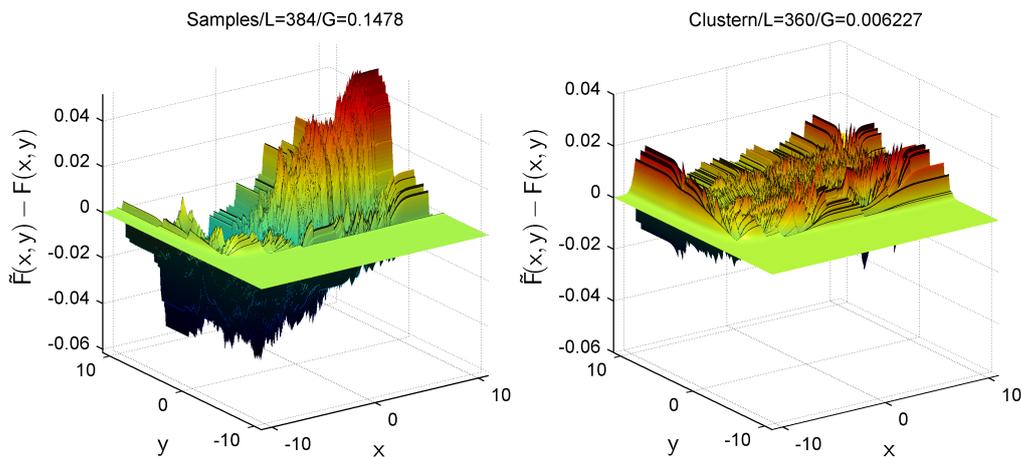


Abbildung 9.17: Vergleich der aus den zufälligen Stichproben konstruierten Approximationen zu denen aus dem Cluster-Verfahren bei etwa gleicher Anzahl von Dirac-Impulsen. Dargestellt ist jeweils die Differenz der Zielverteilungsfunktion zur Verteilungsfunktion der Approximation aus einer frontalen Perspektive. Der Titel nennt das Verfahren, die Anzahl L der Dirac-Impulse sowie die Güte G der Approximation.

Wie schon für die beiden anderen Beispiele sollen die Abbildungen 9.17 und 9.18 für einige ausgewählte Anzahlen von Dirac-Impulsen einen Eindruck von der Qualität der Approximationen vermitteln. Ist das Greedy-Verfahren für das Halbkreis Beispiel noch bis 3000 Dirac-Impulsen das Verfahren mit der besseren Güte gegenüber den zufälligen Stichproben, so gilt dies hier nur noch bis zu einer Anzahl von etwa 100 Dirac-Impulsen. Daher wurden für den visuellen Vergleich eine der zufälligen Stichproben ausgewählt die relativ nahe an den Mittelwerten aus den zufälligen Stichproben liegt, siehe auch Abbildung 9.19.

Da bei den hier vorliegenden hohen Anzahlen von Dirac-Impulsen ein visueller Vergleich der Verteilungsfunktionen relativ schwierig ist, wird in den Abbildungen 9.17 und 9.18 die Differenz der Zielverteilungsfunktion zu der Verteilungsfunktion der jeweiligen Approximation dargestellt. Der Funktionswertebereich ist in allen Unterabbildungen gleich groß, so dass der Vergleich leichter fällt.

Während sich der Fehler in Form der Differenzen für das Cluster-Verfahren recht gleichmäßig über den gesamten Bereich verteilt, gibt es in den Differenzen für die zufälligen Stichproben große Schwankungen: einige Bereiche werden gut, andere schlecht approximiert. Diese Aussage trifft im Übrigen auch für die Standardnormalverteilung und das Halbkreis Beispiel zu.

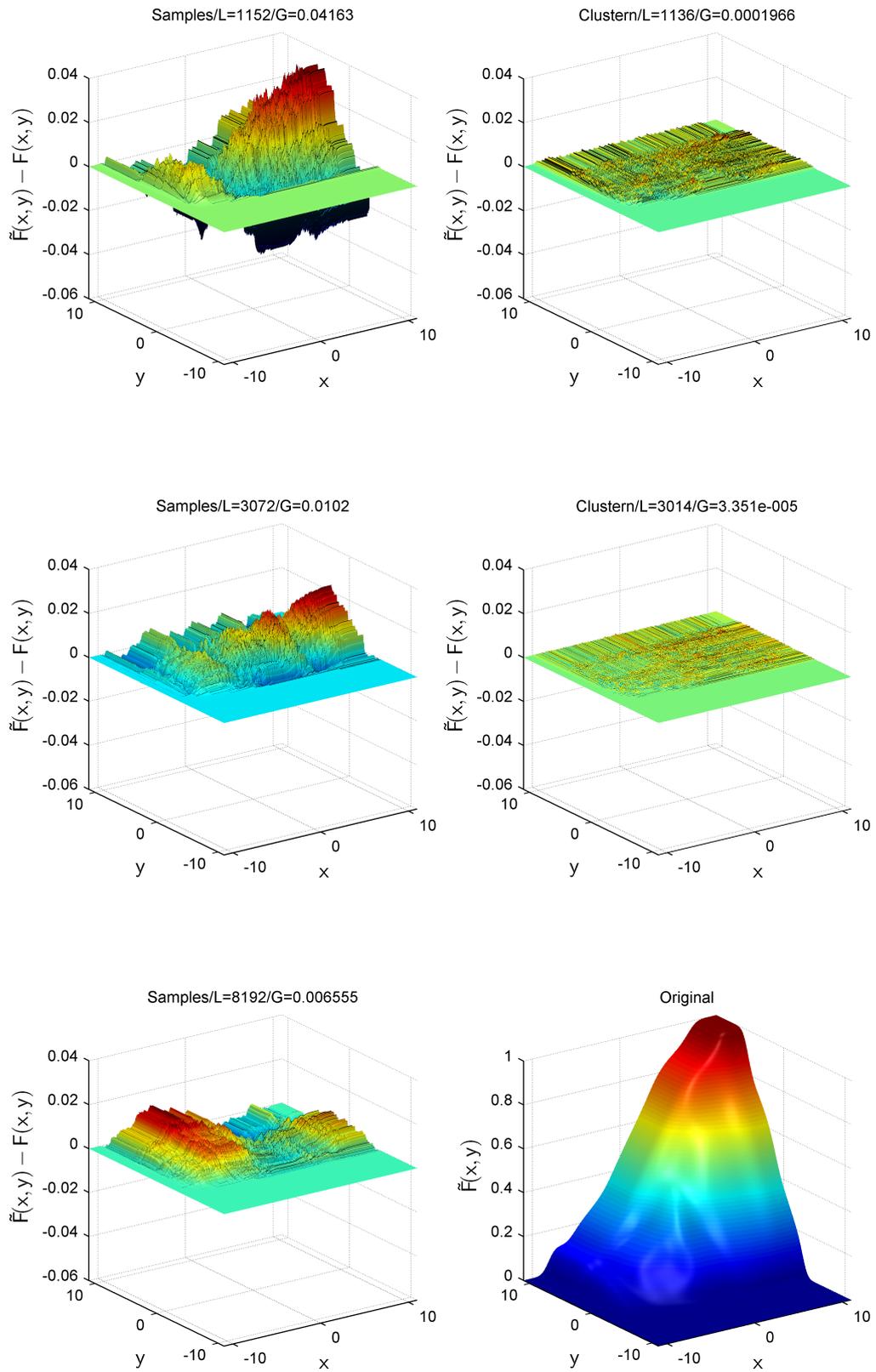


Abbildung 9.18: Fortsetzung von Abbildung 9.17.

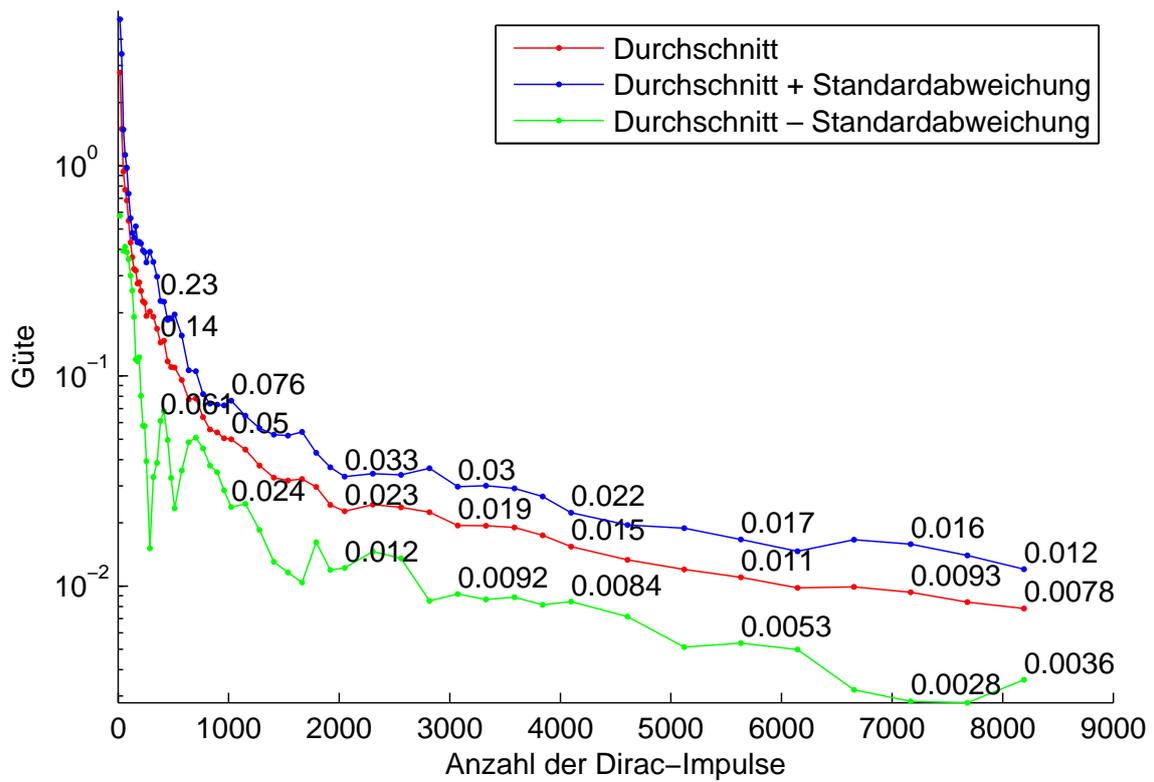


Abbildung 9.19: Streuung der Güte für die 10 verschiedenen Zufallsstichproben für das ISAS Logo Beispiel.

Zusammenfassung und Ausblick

Einleitend zu dieser Arbeit wurde die Verwendung von Dirac-Mischdichten als Grundfunktionen zur Approximation von Zustandsdichten motiviert (siehe Kapitel 1 sowie Abschnitt 3.2). Verschiedene Verfahren zur Approximation wurden betrachtet, insbesondere für den Fall mehrdimensionaler Zustandsdichten (siehe Abschnitt 3). Mit dem Ziel, Approximationen zu erstellen, die einer optimalen Approximation möglichst nahe kommen, wurde das Progressive Bayes-Verfahren zur Approximation von Dichten im Mehrdimensionalen mit Dirac-Mischdichten als Approximationsfunktionen näher untersucht. Zu den frei wählbaren Parametern des Progressiven Bayes-Verfahrens gehören Startdichte und Progression. Die Progression überführt die Startdichte in die gegebene zu approximierende Dichte (Zieldichte). Dabei wird eine optimale Approximation der Startdichte mitgeführt und angepasst, so dass am Ende der Progression eine (lokal) optimale Approximation der Zieldichte vorliegt. Dazu wurde ein explizites Differentialgleichungssystem aufgestellt (Abschnitt 5.2), das sich aus Integraltermen für die Verteilungsfunktion der Approximation sowie für die progressive Verteilungsfunktion und den partiellen Ableitungen dieser Terme aufbaut. Die Terme für die Approximation wurden in Abschnitt A.4 berechnet. Die Terme für die progressive Verteilung wurden ausgehend vom allgemeinen Fall (Abschnitt A.5.1) für Spezialfälle angegeben, in denen die progressive Verteilung in jedem Progressionsschritt γ eine Gleichverteilung (Abschnitt A.5.3), Gauß-Mischverteilung (Abschnitt A.5.4) oder konvexe Kombination (Abschnitt A.5.2) ist. Die Lösung des Differentialgleichungssystems erfolgt numerisch (siehe Abschnitt „Bestimmen einer Lösung“ in 3.2.4) und liefert eine (lokal) optimale Approximation der Zieldichte.

Anhand der Anforderungen an Startdichte und Progression (Abschnitt 6.1) wurde die Gleichverteilung als Startverteilung motiviert und untersucht (Abschnitt 6.2.1). Wie sich zeigt, kann eine optimale Approximation der Gleichverteilung im Mehrdimensionalen nicht analytisch aus der notwendigen Bedingung (3.6) bestimmt werden (Abschnitt „Analytischer Ansatz“ in 6.2.1). Die Übertragung einer optimalen Approximation aus dem Eindimensionalen in das Zweidimensionale erweist sich als nicht global optimal (Abschnitt „Übertragung aus dem Eindimensionalen“ in 6.2.1). Mit dem Progressiven Bayes-Verfahren wurden lokal optimale Approximationen der Gleichverteilung bestimmt (Abschnitt „Progressive Approximationen“ in 6.2.1 respektive Abschnitt 7.1.2) und in Form einer konvexen Kombination von Gleichverteilung als Startverteilung mit einer Zielverteilung angewendet (Abschnitt 7.2).

Das eingeführte Konzept initialer Approximationen (Abschnitt 6.2.2) geht von einer suboptimalen Approximation der Zieldichte aus und verbessert diese mit Hilfe des Progressiven Bayes-Verfahrens (Abschnitt 7.1). Ein großer Vorteil ist, dass die vorgegebene initiale Approximation nicht lokal optimal für die Zieldichte sein muss, tatsächlich kann sie prinzipiell beliebig gewählt werden. Allerdings hängt die progressiv verbesserte Approximation stark von der initialen Approximation ab, was zum einen den Begriff „initiale Approximation“ motiviert und das Einbringen von Wissen über eine gute Approximation der Zieldichte ermöglicht, andererseits aber auch den wesentlichen Nachteil dieses Ansatzes darstellt, da eine geeignete Wahl nicht notwendigerweise offensichtlich ist (siehe Abschnitt 7.1.2 und 7.1.3).

Auf Basis des Ansatzes zur progressiven Verbesserung initialer Approximationen wurde ein Cluster-Verfahren entwickelt, um eine möglichst große Klasse von Mischdichten im Allgemeinen und zweidimensionalen achsenparallelen Gauß-Mischdichten im Konkreten approximieren zu können (Kapitel 8). Das Cluster-Verfahren wurde mit dem Greedy-Verfahren aus [6] sowie aus zufälligen Stichproben konstruierten Approximationen entsprechend Abschnitt 3.2.2 verglichen und zeigt bei den verwendeten Beispielen eine um ein bis drei Größenordnungen bessere Güte bezogen auf die eingesetzte Anzahl der Dirac-Impulse in der Approximation (Kapitel 9). Erkauft wird dieser Vorteil mit einer hohen Laufzeit des zu Grunde liegenden Progressiven Bayes-Verfahrens, siehe Abschnitt 5.5.

Die Wahl der verschiedenen Parameter — beispielsweise die Geschwindigkeit mit der die Varianzen bei der progressiven Verbesserung aufgeweitet werden — ist recht sensibel und bietet noch einiges Potential, um die Approximation einzelner Gauß-Dichten zu verbessern. Das Cluster-Verfahren erreicht bei der Approximation einer Gauß-Mischdichte mit vielen überlappenden Komponenten bessere Güten als für eine einzelne Gauß-Dichte. Ein Aufspalten einer einzelnen Gauß-Dichte in mehrere gewichtete Gauß-Komponenten ist ein möglicher Ansatz, um für einzelne Gauß-Dichten — oder Gauß-Komponenten einer Gauß-Mischdichte — eine bessere Approximation mit Hilfe des Cluster-Verfahrens zu erreichen.

Je größer der Integrationsbereich $[\underline{c}_1, \underline{c}_2]$ im Gütemaß ist, desto schlechter wird das Konvergenzverhalten des Newton Verfahrens im Korrektorschritt. Das Cluster-Verfahren kann so angepasst werden, dass für jeden Cluster nur ein Integrationsbereich betrachtet wird, der die Hauptmasse der im Cluster liegenden Gauß-Dichten umfasst, während die Hauptmasse der Zieldichte gegebenenfalls deutlich über diesen Bereich hinaus geht.

Liegt zuviel Masse der Zielverteilung außerhalb des Integrationsbereichs $[\underline{c}_1, \underline{c}_2]$, so schiebt der Korrektorschritt die Dirac-Impulse aus den Bereich $[\underline{c}_1, \underline{c}_2]$ heraus (siehe auch Abschnitt 6.3.1). Eine mögliche Handhabung ist, den Integrationsbereich künstlich größer zu machen, und die resultierende Approximation anschließend wieder auf den eigentlichen Integrationsbereich zu beschränken. Eine andere Möglichkeit besteht darin, das Gütemaß selbst zu modifizieren, so dass der Gradient für Positionen außerhalb des Integrationsbereichs wieder in Richtung des Integrationsbereichs zeigt und die Positionen automatisch wieder in den Integrationsbereich geschoben werden.

Durch ein analytisch auswertbares Gütemaß könnten die hohen Kosten für die numerische Berechnung des Integrals innerhalb des Gütemaßes, das in dieser Arbeit benutzt wird, eingespart werden. Ebenfalls zu Gunsten einer geringeren Laufzeit kann auf einen Teil der Güte — bezogen auf eine gegebene Anzahl von Dirac-Impulsen — verzichtet werden. Die hier genutzte Clusterbildung selbst ist recht einfach gehalten. Mit mehr „Intelligenz“ können die Überschneidungsbereiche der Gauß-Komponenten mitunter noch besser genutzt und andererseits unnötige Versuche, Komponenten zusammen zufassen, die sich kaum überschneiden, vermieden werden.

Den möglichen Maßnahmen zur Laufzeitreduktion für das Cluster-Verfahren zum Trotz bleibt der relativ hohe O -Kalkül Aufwand aus dem zu Grunde liegenden Progressiven Bayes-Verfahren bestehen, siehe Abschnitt 5.5. Für Anwendungen, in denen hohe Anzahlen von Dirac-Impulsen in der Approximation hohe Kosten nach sich ziehen, kann sich der Einsatz dennoch lohnen. Zudem kann das Verfahren als Referenz für die durch andere Verfahren erreichbare Güte dienen.

Mathematischer Anhang

A.1 Konventionen

Zur Erinnerung sei hier eine kurze Wiederholung einiger Definitionen angeführt. Die Approximationsdichte ist eine Dirac-Mischdichte

$$f(\underline{x}, \underline{\eta}) := \sum_{l=1}^L w_l \delta(\underline{x} - \underline{x}_l)$$

mit Parametervektor $\underline{\eta} = (w_1, \dots, w_L, \underline{x}_1^T, \dots, \underline{x}_L^T)^T$ und Verteilungsfunktion

$$F(\underline{x}, \underline{\eta}) = \sum_{l=1}^L w_l H(\underline{x} - \underline{x}_l) \ .$$

$\underline{\eta} = \underline{\eta}(\gamma)$ und damit die Gewichte $w_i = w_i(\gamma)$ und Positionen $\underline{x}_i = \underline{x}_i(\gamma)$ hängen im Allgemeinen vom Progressionsparameter $\gamma \in [0, 1]$ ab. Mit

$$\frac{\partial}{\partial w_i} F(\underline{x}, \underline{\eta})$$

ist die partielle Ableitung der Verteilungsfunktion F in Richtung des i .ten Gewichts gemeint. w_i dient hierbei lediglich zur Identifikation der Richtung, in welche F abgeleitet werden soll.

Die Progressionsdichte wird mit

$$\tilde{f}(\underline{x}, \gamma)$$

und ihre Verteilung mit

$$\tilde{F}(\underline{x}, \gamma)$$

bezeichnet. Für jedes $\gamma \in [0, 1]$ soll $\tilde{f}(\underline{x}, \gamma)$ durch $f(\underline{x}, \underline{\eta}) = f(\underline{x}, \underline{\eta}(\gamma))$ beziehungsweise $\tilde{F}(\underline{x}, \gamma)$ durch $F(\underline{x}, \underline{\eta}) = F(\underline{x}, \underline{\eta}(\gamma))$ approximiert werden. Das verwendete progressive Gütemaß

$$G(\underline{\eta}, \gamma) := \int_{\underline{c}_1}^{\underline{c}_2} \left\{ \tilde{F}(\underline{x}, \gamma) - F(\underline{x}, \underline{\eta}) \right\}^2 d\underline{x}$$

ist der Abstand im quadratischen Mittel zwischen den Verteilungsfunktionen auf dem Intervall $[\underline{c}_1, \underline{c}_2]$. Als Nebenbedingung wird vorausgesetzt, dass die Positionen \underline{x}_i innerhalb des Integrationsintervalls liegen,

$$\underline{c}_1 \leq \underline{x}_i \leq \underline{c}_2$$

und dass die Gewichte nicht negativ sind,

$$w_i \geq 0$$

für jedes $1 \leq i \leq L$.

Folgende Konventionen werden in diesem Kapitel verwendet.

d	Dimension des betrachteten Raums
m, n, p, q, r	Variablen für die Dimension, $1 \leq m, n, p, q, r \leq d$
L	Anzahl der Dirac-Impulse in der Approximationsdichte
i, j, l	Variable für einen Dirac-Impuls, $1 \leq i, j, l \leq L$
w_i	Gewicht des i .ten Dirac-Impulses
\underline{x}_i	Position des i .ten Dirac-Impulses
x_{im}	m .te Komponente der Position \underline{x}_i
\underline{c}_2	Obergrenze des Integrationsintervalls für das Gütemaß
c_{2q}	q .te Komponente von \underline{c}_2
\underline{c}_1	Untergrenze des Integrationsintervalls für das Gütemaß
c_{1q}	q .te Komponente von \underline{c}_1

A.2 Schreibweisen

Um in längeren Herleitungen für mehr Übersicht zu sorgen, sind Teile der Formeln fett gesetzt. Die Kommentare am rechten Rand beziehen sich in der Regel auf diese fett gesetzten Formelteile.

In Anlehnung an die bekannte Schreibweise für Summen, $\sum_{p=1}^d \{ \dots \}$, wird

$$\int_{\substack{\alpha^{(p)} \\ p=1, \dots, d}}^{\beta^{(p)}} g(x_1, \dots, x_d) dx_p := \int_{\alpha^{(1)}}^{\beta^{(1)}} \cdots \int_{\alpha^{(d)}}^{\beta^{(d)}} g(x_1, \dots, x_d) dx_1 \cdots dx_d$$

als Schreibweise für ein Mehrfachintegral über die d dimensionale Funktion g benutzt, $\alpha^{(p)}, \beta^{(p)} \in \mathbb{R}$, $1 \leq p \leq d$. Für $1 \leq q \leq d$ und $y \in \mathbb{R}$ ist

$$\int_{\substack{\alpha^{(p)} \\ p=1, \dots, d, p \neq q}}^{\beta^{(p)}} g(x_1, \dots, x_{q-1}, y, x_{q+1}, \dots, x_d) dx_p \tag{A.1}$$

zu verstehen als

$$\int_{\alpha^{(1)}}^{\beta^{(1)}} \cdots \int_{\alpha^{(q-1)}}^{\beta^{(q-1)}} \int_{\alpha^{(q+1)}}^{\beta^{(q+1)}} \cdots \int_{\alpha^{(d)}}^{\beta^{(d)}} g(x_1, \dots, x_{q-1}, y, x_{q+1}, \dots, x_d) dx_1 \cdots dx_{q-1} dx_{q+1} \cdots dx_d .$$

Für den Vektor $\underline{x} = (x_1, \dots, x_d)^T$ meint

$$\underline{x}|_{x_q \equiv \tau} := (x_1, \dots, x_{q-1}, \tau, x_{q+1}, \dots, x_d)^T$$

die *symbolische* Ersetzung der q .ten Komponente in \underline{x} durch τ . (A.1) lautet dann vektoriell geschrieben

$$\int_{\alpha^{(p)}}^{\beta^{(p)}} g(\underline{x}|_{x_q \equiv y}) dx_p \quad (\text{A.2})$$

$p=1, \dots, d, p \neq q$

Die Betonung der symbolischen Ersetzung ist insofern wichtig, als dass $y = y(\gamma)$ mitunter eine Funktion von γ statt einer festen Zahl ist. In diesem Fall ist (A.2) eine Funktion von γ und

$$\frac{\partial}{\partial \gamma} \left\{ \int_{\alpha^{(p)}}^{\beta^{(p)}} g(\underline{x}|_{x_i \equiv y(\gamma)}) dx_p \right\}$$

$p=1, \dots, J, j \neq i$

meint die Ableitung dieser Funktion nach γ .

Für eine Teilmenge $M \subseteq \mathbb{R}^d$ bezeichnet $1_M(\underline{x})$ die charakteristische Funktion der Menge M , siehe Definition 2.2.3. Insbesondere gilt für $d = 1$ und $M = \{0\}$

$$1_{\{0\}}(x) = \begin{cases} 1 & x = 0 \\ 0 & \text{sonst} \end{cases}$$

A.3 Grundlegende Formeln

A.3.1 Verallgemeinerte Kettenregel

Sind $g_p(x)$, $1 \leq p \leq d$, Funktionen von x und ist g eine d dimensionale Funktion, dann ist die Ableitung der Verkettung von g nach g_1, \dots, g_d gegeben durch

$$\frac{\partial}{\partial x} g(g_1(x), \dots, g_d(x)) = \sum_{p=1}^d \left\{ \frac{\partial}{\partial g_p} g(g_1, \dots, g_d) \right\} \cdot \frac{\partial}{\partial x} g_p(x) \quad (\text{A.3})$$

Siehe beispielsweise [3]. Dabei meint $\frac{\partial}{\partial g_p} g(g_1, \dots, g_d)$ die partielle Ableitung von g in Richtung der p .ten Komponente.

A.3.2 Partielle Ableitungen der max-Funktion

Es gilt (siehe unten)

$$\frac{\partial}{\partial x} \max(x, y) = H(x - y) \quad (\text{A.4})$$

und analog $\frac{\partial}{\partial y} \max(x, y) = H(y - x)$. Mit der verallgemeinerten Kettenregel ergibt sich daraus für differenzierbare Funktionen $g(x)$ und $h(x)$

$$\boxed{\frac{\partial}{\partial x} \max(g(x), h(x)) = H(g(x) - h(x))\dot{g}(x) + H(h(x) - g(x))\dot{h}(x)} . \quad (\text{A.5})$$

Insbesondere ist

$$\boxed{\frac{\partial}{\partial x} \max(x, x) = 2H(x - x)} . \quad (\text{A.6})$$

Zur Herleitung von (A.4) sei

$$g(x) := \max(x, 0) \quad (\text{A.7})$$

Ihre Ableitung ist

$$\dot{g}(x) = \begin{cases} 0 & x < 0 \\ 1 & x > 0 \end{cases} .$$

Für $x = 0$ ist g streng genommen nicht differenzierbar. Aus der Darstellung

$$g(x) = \int_{-\infty}^x H(t) dt$$

von g mit Hilfe der Heaviside-Funktion ergibt sich die Rechtfertigung für die in dieser Arbeit passende Definition

$$\dot{g}(0) := H(0) = \frac{1}{2}$$

so dass insgesamt

$$\dot{g}(x) = H(x) \quad (\text{A.8})$$

gilt. Es folgt

$$\frac{\partial}{\partial x} \max(x, y) \quad (\text{Eigenschaft von max})$$

$$= \frac{\partial}{\partial x} \{\mathbf{max}(x - y, 0) + y\} \quad (\text{mit (A.7)})$$

$$= \frac{\partial}{\partial \mathbf{x}} \{g(x - y) + y\} \quad (\text{Linearität, } y \text{ unabhängig von } x, \text{ Kettenregel})$$

$$= \dot{g}(x - y) \quad (\text{mit (A.8)})$$

$$= H(x - y) .$$

A.3.3 Leibniz-Regel für Parameterintegrale

Für stetig differenzierbare Funktionen $\alpha(x)$, $\beta(x)$ und $g(x, y)$ von x und y gilt für die Ableitung des Parameterintegrals von g in den Grenzen von $\alpha(x)$ bis $\beta(x)$ (siehe [3])

$$\frac{\partial}{\partial x} \int_{\alpha(x)}^{\beta(x)} g(x, y) dy = \int_{\alpha(x)}^{\beta(x)} \left[\frac{\partial}{\partial x} g(x, y) \right] dy + g(x, \beta(x)) \dot{\beta}(x) - g(x, \alpha(x)) \dot{\alpha}(x) \quad . \quad (\text{A.9})$$

Ist $\beta(x) = c$ konstant, $\alpha(x) = x$ und hängt $g = g(y)$ nicht von x ab, so folgt

$$\frac{\partial}{\partial x} \int_x^c g(y) dy = -g(x) \quad . \quad (\text{A.10})$$

A.3.4 Einbringen der Heaviside-Funktion in die Integrationsgrenzen

Da $\{0\}$ eine Nullmenge ist, gilt

$$\int_{\alpha}^{\beta} H(x) dx = \int_{\alpha}^{\beta} 1_{[0, \infty)}(x) dx = \int_{\max(0, \alpha)}^{\beta} 1 dx$$

für $\alpha \leq \beta \in \mathbb{R}$ und allgemeiner

$$\int_{\alpha}^{\beta} H(x - y) g(x) dx = \int_{\max(y, \alpha)}^{\beta} g(x) dx$$

für jede integrierbare Funktion g . Dabei darf $\alpha = -\infty$ und/oder $\beta = \infty$ sein.

A.4 Approximation

In diesem Abschnitt werden die Integrale

$$R(i, \underline{\eta}) := \int_{\underline{x}_i}^{c_2} F(\underline{x}, \underline{\eta}) d\underline{x} \quad (\text{A.11})$$

und

$$R(i, m, \underline{\eta}) := \int_{\substack{x_{i_q} \\ q=1, \dots, d; q \neq m}}^{c_2} F(\underline{x}|_{x_m = x_{im}}, \underline{\eta}) dx_q \quad (\text{A.12})$$

$i = 1, \dots, L$, $m = 1, \dots, d$, analytisch ausgewertet und ihre partiellen Ableitungen berechnet. Sie sind von zentraler Bedeutung da die ersten und zweiten partiellen Ableitungen des Gütemaßes mit ihrer Hilfe dargestellt werden können, siehe dazu Abschnitt 5.1, 5.2 und 5.3.

Die Herleitung ist nicht ohne Weiteres mit einem Computeralgebrasystem möglich und wird hier vollständig durchgeführt. Die Ergebnisse enthalten Teilweise Dirac-Impulse. Zur Handhabung dieser Irregularitäten siehe Abschnitt 5.4.

A.4.1 Abkürzende Definitionen

Zur besseren Lesbarkeit werden nachfolgende Abkürzungen benutzt. Die Abhängigkeit von der Integrationsobergrenze \underline{c}_2 und dem Parametervektor $\underline{\eta} = (w_1, \dots, w_L, \underline{x}_1^T, \dots, \underline{x}_L^T)^T$ ist dabei nicht explizit aufgeführt.

$$\overset{\Pi}{\max}(i, l) := \prod_{q=1}^d \{c_{2q} - \max(x_{iq}, x_{lq})\} \quad (\text{A.13})$$

$$\overset{\Pi}{\max}(i, l, p) := \prod_{q=1, q \neq p}^d \{c_{2q} - \max(x_{iq}, x_{lq})\} \quad (\text{A.14})$$

$$\overset{\Pi}{\max}(i, l, r, p) := \begin{cases} 1 & d = 2 \\ \prod_{q=1, q \neq r, q \neq p}^d \{c_{2q} - \max(x_{iq}, x_{lq})\} & d > 2 \end{cases} \quad (\text{A.15})$$

$$\overset{\Pi H}{\max}(i, l, p) := \overset{\Pi}{\max}(i, l, p) H(x_{ip} - x_{lp}) \quad (\text{A.16})$$

$$\overset{\Pi H}{\max}(i, l, r, p) := \overset{\Pi}{\max}(i, l, r, p) H(x_{ip} - x_{lp}) \quad (\text{A.17})$$

$$\overset{\widehat{\Pi H}}{\max}(i, l, p) := \overset{\Pi}{\max}(i, l, p) H(x_{lp} - x_{ip}) \quad (\text{A.18})$$

$$\overset{\widehat{\Pi H}}{\max}(i, l, r, p) := \overset{\Pi}{\max}(i, l, r, p) H(x_{lp} - x_{ip}) \quad (\text{A.19})$$

$$\overset{\Pi H^2}{\max}(i, l, r, p) := \overset{\Pi H}{\max}(i, l, r, p) H(x_{ir} - x_{lr}) \quad (\text{A.20})$$

$$\overset{\widehat{\Pi H}^2}{\max}(i, l, r, p) := \overset{\widehat{\Pi H}}{\max}(i, l, r, p) H(x_{ir} - x_{lr}) \quad (\text{A.21})$$

A.4.2 Partielle Ableitungen von F

Nach w_i Die Linearität des Ableitungsoperators liefert

$$\boxed{\frac{\partial}{\partial w_i} F(\underline{x}, \underline{\eta}) = H(\underline{x} - \underline{x}_i)} \quad (\text{A.22})$$

Nach x_{im}

$$\frac{\partial}{\partial x_{im}} \mathbf{F}(\underline{\mathbf{x}}, \underline{\boldsymbol{\eta}}) \quad (\text{Def.})$$

$$\frac{\partial}{\partial x_{im}} \left\{ \sum_{l=1}^L w_l \mathbf{H}(\underline{\mathbf{x}} - \underline{\mathbf{x}}_l) \right\} \quad (\text{Def.})$$

$$= \frac{\partial}{\partial \mathbf{x}_{im}} \left\{ \sum_{l=1}^L w_l \prod_{p=1}^d H(x_p - x_{lp}) \right\} \quad (\text{Linearität})$$

$$= \sum_{l=1}^L w_l \left\{ \prod_{p=1, p \neq m}^d H(x_p - x_{lp}) \right\} \frac{\partial}{\partial x_{im}} \mathbf{H}(\mathbf{x}_m - \mathbf{x}_{lm}) \quad (= 0 \text{ für } l \neq i)$$

und somit

$$\boxed{\frac{\partial}{\partial x_{im}} F(\underline{\mathbf{x}}, \underline{\boldsymbol{\eta}}) = -w_i \left\{ \prod_{p=1, p \neq m}^d H(x_p - x_{ip}) \right\} \delta(x_m - x_{im})} \quad (\text{A.23})$$

A.4.3 Das Integral von F

$$\mathbf{R}(i, \underline{\boldsymbol{\eta}}) \quad (\text{Def.})$$

$$= \int_{\underline{\mathbf{x}}_i}^{\underline{\mathbf{c}}_2} \mathbf{F}(\underline{\mathbf{x}}, \underline{\boldsymbol{\eta}}) d\underline{\mathbf{x}} \quad (\text{Def.})$$

$$= \int_{\underline{\mathbf{x}}_i}^{\underline{\mathbf{c}}_2} \sum_{l=1}^L w_l \mathbf{H}(\underline{\mathbf{x}} - \underline{\mathbf{x}}_l) d\underline{\mathbf{x}} \quad (\text{Linearität})$$

$$= \sum_{l=1}^L w_l \int_{\underline{\mathbf{x}}_i}^{\underline{\mathbf{c}}_2} \mathbf{H}(\underline{\mathbf{x}} - \underline{\mathbf{x}}_l) d\underline{\mathbf{x}} \quad (\text{Def., Fubini, Unabhängigkeit der Komponenten})$$

$$= \sum_{l=1}^L w_l \prod_{q=1}^d \int_{\mathbf{x}_{iq}}^{\mathbf{c}_{2q}} \mathbf{H}(\mathbf{x}_q - \mathbf{x}_{lq}) d\mathbf{x}_q \quad (\text{Einbringen in die Integrationsgrenzen})$$

$$= \sum_{l=1}^L w_l \prod_{q=1}^d \int_{\max(\mathbf{x}_{iq}, \mathbf{x}_{lq})}^{\mathbf{c}_{2q}} \mathbf{1} d\mathbf{x}_q \quad (= \text{Länge des Integrationsintervalls})$$

$$= \sum_{l=1}^L w_l \prod_{q=1}^d \{ \mathbf{c}_{2q} - \max(\mathbf{x}_{iq}, \mathbf{x}_{lq}) \}$$

und schließlich mit Definition (A.13)

$$\boxed{R(i, \underline{\eta}) = \sum_{l=1}^L w_l \max^{\Pi}(i, l)} . \quad (\text{A.24})$$

Partielle Ableitungen

Nach x_{in}

$$\frac{\partial}{\partial x_{in}} R(i, \underline{\eta}) \quad (\text{Darstellung aus (A.24)})$$

$$= \frac{\partial}{\partial x_{in}} \left\{ \sum_{l=1}^L w_l \max^{\Pi}(i, l) \right\} \quad (\text{Def.})$$

$$= \frac{\partial}{\partial x_{in}} \left\{ \sum_{l=1}^L w_l \underbrace{\max^{\Pi}(i, l, n)}_{\text{frei von } x_{in}} \{c_{2n} - \max(x_{in}, x_{ln})\} \right\} \quad (\text{Linearität})$$

$$= \sum_{l=1}^L w_l \max^{\Pi}(i, l, n) \frac{\partial}{\partial x_{in}} \{c_{2n} - \max(x_{in}, x_{ln})\} \quad (\text{mit (A.4) für } l \neq i \text{ und (A.6) sonst})$$

$$= - \sum_{l=1}^L \{1 + 1_{\{0\}}(i-l)\} w_l \max^{\Pi}(i, l, n) H(x_{in} - x_{ln})$$

also mit Definition (A.16)

$$\boxed{\frac{\partial}{\partial x_{in}} R(i, \underline{\eta}) = - \sum_{l=1}^L \{1 + 1_{\{0\}}(i-l)\} w_l \max^{\Pi H}(i, l, n)} . \quad (\text{A.25})$$

Nach x_{jn} , $j \neq i$

$$\frac{\partial}{\partial x_{jn}} R(i, \underline{\eta}) \quad (\text{Darstellung aus (A.24)})$$

$$= \frac{\partial}{\partial x_{jn}} \left\{ \sum_{l=1}^L w_l \mathop{\text{m\AA}}^{\Pi} (i, l) \right\} \quad (\text{Def.})$$

$$= \frac{\partial}{\partial x_{jn}} \left\{ \sum_{l=1}^L w_l \underbrace{\mathop{\text{m\AA}}^{\Pi} (i, l, n)}_{\text{frei von } x_{jn}} \{c_{2n} - \max(x_{in}, x_{ln})\} \right\} \quad (\text{Linearit\AA})$$

$$= \sum_{l=1}^L w_l \mathop{\text{m\AA}}^{\Pi} (i, l, n) \frac{\partial}{\partial x_{jn}} \{c_{2n} - \max(x_{in}, x_{ln})\} \quad (= 0 \text{ f\AA } l \neq j, \text{ mit (A.4) f\AA } l = j)$$

$$= -w_j \mathop{\text{m\AA}}^{\Pi} (i, j, n) H(x_{jn} - x_{in})$$

also mit Definition (A.18)

$$\boxed{\frac{\partial}{\partial x_{jn}} R(i, \underline{\eta}) = -w_j \mathop{\text{m\AA}}^{\widehat{\Pi H}} (i, j, n)} \quad (\text{A.26})$$

Nach w_j Da $\mathop{\text{m\AA}}^{\Pi} (i, l)$ das Gewicht w_j nicht enth\AA, liefert die Linearit\AA des Ableitungsoperators ausgehend von (A.24)

$$\boxed{\frac{\partial}{\partial w_j} R(i, \underline{\eta}) = \mathop{\text{m\AA}}^{\Pi} (i, j)} \quad (\text{A.27})$$

Nach γ Werden w_l und x_{lp} als Funktionen von γ betrachtet dann folgt aus (A.24) unter Anwendung der verallgemeinerten Kettenregel (A.3)

$$\boxed{\frac{\partial}{\partial \gamma} R(i, \underline{\eta}) = \sum_{l=1}^L \left\{ \frac{\partial}{\partial w_l} R(i, \underline{\eta}) \right\} \dot{w}_l + \sum_{l=1}^L \sum_{p=1}^d \left\{ \frac{\partial}{\partial x_{lp}} R(i, \underline{\eta}) \right\} \dot{x}_{lp}} \quad (\text{A.28})$$

A.4.4 Das Integral von F \AAber $d - 1$ Dimensionen

Ziel ist die Berechnung von $R(i, m, \underline{\eta})$ aus Def. (A.12).

$$R(i, m, \underline{\eta}) \quad (\text{Def.})$$

$$= \int_{\substack{x_{iq} \\ q=1, \dots, d; q \neq m}}^{c_{2q}} [\mathbf{F}(\underline{\mathbf{x}}|_{\mathbf{x}_m \equiv \mathbf{x}_{im}}, \underline{\eta})] dx_q \quad (\text{Def.})$$

$$= \int_{\substack{x_{iq} \\ q=1, \dots, d; q \neq m}}^{c_{2q}} [\sum_{l=1}^L w_l H(\underline{x}|_{\mathbf{x}_m \equiv \mathbf{x}_{im}} - \underline{x}_l)] d\mathbf{x}_q \quad (\text{Linearität})$$

$$= \sum_{l=1}^L w_l \int_{\substack{x_{iq} \\ q=1, \dots, d; q \neq m}}^{c_{2q}} \mathbf{H}(\underline{x}|_{\mathbf{x}_m \equiv \mathbf{x}_{im}} - \underline{x}_l) dx_q \quad (\text{Def.})$$

$$= \sum_{l=1}^L w_l \int_{\substack{x_{iq} \\ q=1, \dots, d; q \neq m}}^{c_{2q}} \left[\mathbf{H}(\mathbf{x}_{im} - \mathbf{x}_{lm}) \left\{ \prod_{p=1, p \neq m}^d \mathbf{H}(\mathbf{x}_p - \mathbf{x}_{lp}) \right\} \right] d\mathbf{x}_q$$

(Linearität, Unabhängigkeit der Komponenten)

$$= \sum_{l=1}^L w_l H(\mathbf{x}_{im} - \mathbf{x}_{lm}) \prod_{q=1, q \neq m}^d \int_{x_{iq}}^{c_{2q}} \mathbf{H}(\mathbf{x}_q - \mathbf{x}_{lq}) dx_q$$

(Einbringen in die Integrationsgrenzen)

$$= \sum_{l=1}^L w_l H(\mathbf{x}_{im} - \mathbf{x}_{lm}) \prod_{q=1, q \neq m}^d \int_{\max(x_{iq}, x_{lq})}^{c_{2q}} \mathbf{1} d\mathbf{x}_q \quad (= \text{Länge des Integrationsintervalls})$$

$$= \sum_{l=1}^L w_l \left\{ \prod_{q=1, q \neq m}^d \{c_{2q} - \max(x_{iq}, x_{lq})\} \right\} H(\mathbf{x}_{im} - \mathbf{x}_{lm})$$

also mit Definition (A.16)

$$\boxed{R(i, m, \underline{\eta}) = \sum_{l=1}^L w_l \max_{\Pi H} (i, l, m)} \quad (\text{A.29})$$

Partielle Ableitungen

Nach w_j Die Linearität des Ableitungsoperators liefert aus (A.29) sofort

$$\boxed{\frac{\partial}{\partial w_j} R(i, m, \underline{\eta}) = \max_{\Pi H} (i, j, m)} \quad (\text{A.30})$$

Nach x_{im} Es gilt

$$\begin{aligned} & \frac{\partial}{\partial x_{im}} R(i, m, \underline{\eta}) && \text{(Darstellung aus (A.29))} \\ &= \frac{\partial}{\partial x_{im}} \left\{ \sum_{l=1}^L w_l \underbrace{\max(i, l, m)}_{\text{frei von } x_{im}} H(x_{im} - x_{lm}) \right\} && \text{(Linearität)} \\ &= \sum_{l=1}^L w_l \max(i, l, m) \frac{\partial}{\partial x_{im}} H(x_{im} - x_{lm}) && (= 0 \text{ f\"ur } i = l) \end{aligned}$$

also

$$\boxed{\frac{\partial}{\partial x_{im}} R(i, m, \underline{\eta}) = \sum_{l=1, l \neq i}^L w_l \max(i, l, m) \delta(x_{im} - x_{lm})} . \quad (\text{A.31})$$

Nach x_{in} , $n \neq m$ Es gilt

$$\begin{aligned} & \frac{\partial}{\partial x_{in}} R(i, m, \underline{\eta}) && \text{(Darstellung aus (A.29))} \\ &= \frac{\partial}{\partial x_{in}} \left\{ \sum_{l=1}^L w_l \max(i, l, m) H(x_{im} - x_{lm}) \right\} && \text{(Def.)} \\ &= \frac{\partial}{\partial x_{in}} \left\{ \sum_{l=1}^L w_l \underbrace{\max(i, l, m, n)}_{\text{frei von } x_{in}} \{c_{2n} - \max(x_{in}, x_{ln})\} H(x_{im} - x_{lm}) \right\} && \text{(Linearität)} \\ &= \sum_{l=1}^L w_l \max(i, l, m, n) H(x_{im} - x_{lm}) \frac{\partial}{\partial x_{in}} \{c_{2n} - \max(x_{in}, x_{ln})\} && \text{(mit (A.4) f\"ur } l \neq i \text{ und (A.6) sonst)} \\ &= - \sum_{l=1}^L \{1 + 1_{\{0\}}(i-l)\} w_l \max(i, l, m, n) H(x_{im} - x_{lm}) H(x_{in} - x_{ln}) \end{aligned}$$

also mit Definition (A.20)

$$\boxed{\frac{\partial}{\partial x_{in}} R(i, m, \underline{\eta}) = - \sum_{l=1}^L \{1 + 1_{\{0\}}(i-l)\} w_l \max(i, l, m, n)} . \quad (\text{A.32})$$

Nach x_{jm} , $j \neq i$ Es gilt

$$\frac{\partial}{\partial x_{jm}} R(i, m, \underline{\eta}) \quad (\text{Darstellung aus (A.29)})$$

$$= \frac{\partial}{\partial x_{jm}} \left\{ \sum_{l=1}^L w_l \underbrace{\max_{\Pi}(i, l, m)}_{\text{frei von } x_{jm}} H(x_{im} - x_{lm}) \right\} \quad (\text{Linearität})$$

$$= \sum_{l=1}^L w_l \max_{\Pi}(i, l, m) \frac{\partial}{\partial x_{jm}} H(x_{im} - x_{lm}) \quad (= 0 \text{ für } l \neq j)$$

also

$$\boxed{\frac{\partial}{\partial x_{jm}} R(i, m, \underline{\eta}) = -w_j \max_{\Pi}(i, j, m) \delta(x_{im} - x_{jm})} \quad (\text{A.33})$$

Nach x_{jn} , $j \neq i$, $n \neq m$ Es gilt

$$\frac{\partial}{\partial x_{jn}} R(i, m, \underline{\eta}) \quad (\text{Darstellung aus (A.29)})$$

$$\frac{\partial}{\partial x_{jn}} \left\{ \sum_{l=1}^L w_l \max_{\Pi H}(i, l, m) \right\} \quad (\text{Def.})$$

$$= \frac{\partial}{\partial x_{jn}} \left\{ \sum_{l=1}^L w_l \underbrace{\max_{\Pi}(i, l, m, n)}_{\text{frei von } x_{jn}} \{c_{2n} - \max(x_{in}, x_{ln})\} H(x_{im} - x_{lm}) \right\} \quad (\text{Linearität})$$

$$= \sum_{l=1}^L w_l \max_{\Pi}(i, l, m, n) H(x_{im} - x_{lm}) \frac{\partial}{\partial x_{jn}} \{c_{2n} - \max(x_{in}, x_{ln})\} \quad (= 0 \text{ für } l \neq j)$$

$$= w_j \max_{\Pi}(i, j, m, n) H(x_{im} - x_{jm}) \frac{\partial}{\partial x_{jn}} \{c_{2n} - \max(x_{in}, x_{jn})\} \quad (\text{mit (A.4)})$$

$$= -w_j \max_{\Pi}(i, j, m, n) H(x_{im} - x_{jm}) H(x_{jn} - x_{in})$$

also mit Definition (A.21)

$$\boxed{\frac{\partial}{\partial x_{jn}} R(i, m, \underline{\eta}) = -w_j \frac{\Pi \hat{H}^2}{\max(i, j, m, n)}} \quad . \quad (\text{A.34})$$

Nach γ Werden w_l und x_{lp} als Funktionen von γ betrachtet dann folgt aus (A.29) unter Anwendung der verallgemeinerten Kettenregel (A.3)

$$\boxed{\frac{\partial}{\partial \gamma} R(i, m, \underline{\eta}) = \sum_{l=1}^L \left\{ \frac{\partial}{\partial w_l} R(i, m, \underline{\eta}) \right\} \dot{w}_l + \sum_{l=1}^L \sum_{p=1}^d \left\{ \frac{\partial}{\partial x_{lp}} R(i, m, \underline{\eta}) \right\} \dot{x}_{lp}} \quad . \quad (\text{A.35})$$

A.5 Progressive Verteilung

A.5.1 Allgemeiner Fall

Die Integrale

$$\tilde{R}(i, \underline{\eta}, \gamma) := \int_{\underline{x}_i}^{c_2} \tilde{F}(\underline{x}, \gamma) d\underline{x} \quad (\text{A.36})$$

$$\tilde{R}(i, m, \underline{\eta}, \gamma) := \int_{\substack{x_{iq} \\ q=1, \dots, d; q \neq m}}^{c_{2q}} \tilde{F}(\underline{x}|_{x_m \equiv x_{im}}, \gamma) dx_q \quad (\text{A.37})$$

$$\tilde{R}(i, m, n, \underline{\eta}, \gamma) := \begin{cases} \tilde{F}(\underline{x}|_{x_m \equiv x_{im}, x_n \equiv x_{in}}, \gamma) & d = 2 \\ \int_{\substack{x_{iq} \\ q=1, \dots, d; q \neq m, q \neq n}}^{c_{2q}} \tilde{F}(\underline{x}|_{x_m \equiv x_{im}, x_n \equiv x_{in}}, \gamma) dx_q & \text{sonst} \end{cases} \quad (\text{A.38})$$

$$\tilde{R}^b(i, \underline{\eta}, \gamma) := \int_{\underline{x}_i}^{c_2} \left[\frac{\partial}{\partial \gamma} \tilde{F}(\underline{x}, \gamma) \right] d\underline{x} \quad (\text{A.39})$$

und

$$\tilde{R}^b(i, m, \underline{\eta}, \gamma) := \int_{\substack{x_{iq} \\ q=1, \dots, d; q \neq m}}^{c_{2q}} \left[\left[\frac{\partial}{\partial \gamma} \tilde{F}(\underline{x}, \gamma) \right]_{x_m \equiv x_{im}} \right] dx_q \quad (\text{A.40})$$

für $i = 1, \dots, L$ und $m, n = 1, \dots, d$ sind von zentraler Bedeutung, da die ersten und zweiten partiellen Ableitungen des Gütemaßes mit ihrer Hilfe dargestellt werden können, siehe Abschnitt 5.1, 5.2 und 5.3. Dazu werden zusätzlich die partiellen Ableitungen von $\tilde{R}(i, \underline{\eta}, \gamma)$ und $\tilde{R}(i, m, \underline{\eta}, \gamma)$ benötigt.

Einige dieser partiellen Ableitungen können schon im allgemeinen Fall einer beliebigen Progressionsverteilung \tilde{F} ausgewertet oder auf eines der anderen Integrale zurückgeführt werden. Dies geschieht in den folgenden Unterabschnitten.

Nicht wesentlich vereinfacht werden können

- $\tilde{R}(i, \underline{\eta}, \gamma)$, $\tilde{R}(i, m, \underline{\eta}, \gamma)$, und $\tilde{R}(i, m, n, \underline{\eta}, \gamma)$,
- $\tilde{R}^b(i, \underline{\eta}, \gamma)$ und $\tilde{R}^b(i, m, \underline{\eta}, \gamma)$
- sowie die partielle Ableitung $\frac{\partial}{\partial x_{im}} \tilde{R}(i, m, \underline{\eta}, \gamma)$.

Sie müssen für eine konkret vorgelegte Progressionsverteilung \tilde{F} separat berechnet werden, siehe Abschnitte A.5.2, A.5.3 und A.5.4.

Abschließend sei noch der Hinweis gegeben, dass $\tilde{R}^b(i, \underline{\eta}, \gamma)$ der partiellen Ableitung von $\tilde{R}(i, \underline{\eta}, \gamma)$ nach γ entspricht,

$$\tilde{R}^b(i, \underline{\eta}, \gamma) = \frac{\partial}{\partial \gamma} \tilde{R}(i, \underline{\eta}, \gamma)$$

sofern $\underline{\eta}$ als *unabhängig* von γ betrachtet wird. Ebenso gilt

$$\tilde{R}^b(i, m, \underline{\eta}, \gamma) = \frac{\partial}{\partial \gamma} \tilde{R}(i, m, \underline{\eta}, \gamma)$$

unter der gleichen Voraussetzung.

Partielle Ableitungen des Integrals von \tilde{F}

Nach w_j Da w_j in $\tilde{R}(i, \underline{\eta}, \gamma)$ nicht vorkommt, gilt

$$\boxed{\frac{\partial}{\partial w_j} \tilde{R}(i, \underline{\eta}, \gamma) = 0} \quad . \quad (\text{A.41})$$

Nach x_{in}

$$\frac{\partial}{\partial x_{in}} \tilde{R}(i, \underline{\eta}, \gamma) \quad (\text{Def.})$$

$$= \frac{\partial}{\partial x_{in}} \left\{ \int_{\mathbf{x}_{i_q}}^{c_{2_q}} \tilde{F}(\underline{\mathbf{x}}, \gamma) d\mathbf{x}_q \right\} \quad (\text{Fubini})$$

$$= \frac{\partial}{\partial \mathbf{x}_{in}} \left\{ \int_{x_{in}}^{c_{2_n}} \left[\int_{\mathbf{x}_{i_p}}^{c_{2_q}} \tilde{F}(\underline{\mathbf{x}}, \gamma) d\mathbf{x}_p \right] dx_n \right\} \quad (\text{Def.})$$

$$= \frac{\partial}{\partial \mathbf{x}_{in}} \left\{ \int_{x_{in}}^{c_{2_n}} \underbrace{\tilde{R}(i, n, \underline{\eta}, \gamma)}_{\text{frei von } x_{in}} dx_n \right\}$$

also mit (A.10)

$$\boxed{\frac{\partial}{\partial x_{in}} \tilde{R}(i, \underline{\eta}, \gamma) = -\tilde{R}(i, n, \underline{\eta}, \gamma)} \quad (\text{A.42})$$

Nach x_{jn} , $j \neq i$ Da x_{jn} in $\tilde{R}(i, \underline{\eta}, \gamma)$ nicht vorkommt, gilt

$$\boxed{\frac{\partial}{\partial x_{jn}} \tilde{R}(i, \underline{\eta}, \gamma) = 0} \quad (\text{A.43})$$

Nach γ Werden w_l und x_{lp} als Funktionen von γ betrachtet dann folgt unter Anwendung der verallgemeinerten Kettenregel (A.3) und Beachtung der Definition von $\tilde{R}^b(i, \underline{\eta}, \gamma)$

$$\boxed{\frac{\partial}{\partial \gamma} \tilde{R}(i, \underline{\eta}, \gamma) = \tilde{R}^b(i, \underline{\eta}, \gamma) + \sum_{l=1}^L \left\{ \frac{\partial}{\partial w_l} \tilde{R}(i, \underline{\eta}, \gamma) \right\} \dot{w}_l + \sum_{l=1}^L \sum_{p=1}^d \left\{ \frac{\partial}{\partial x_{lp}} \tilde{R}(i, \underline{\eta}, \gamma) \right\} \dot{x}_{lp}} \quad (\text{A.44})$$

Partielle Ableitungen des Integrals von \tilde{F} über $d - 1$ Dimensionen

Nach w_j Da w_j in $\tilde{R}(i, m, \underline{\eta}, \gamma)$ nicht vorkommt, gilt

$$\boxed{\frac{\partial}{\partial w_j} \tilde{R}(i, m, \underline{\eta}, \gamma) = 0} . \quad (\text{A.45})$$

Nach x_{in} , $n \neq m$ Analog zur Herleitung von $\frac{\partial}{\partial x_{in}} \tilde{R}(i, \underline{\eta}, \gamma)$ ergibt sich

$$\boxed{\frac{\partial}{\partial x_{in}} \tilde{R}(i, m, \underline{\eta}, \gamma) = -\tilde{R}(i, m, n, \underline{\eta}, \gamma)} . \quad (\text{A.46})$$

Nach x_{jn} , $j \neq i$ Da x_{jn} in $\tilde{R}(i, m, \underline{\eta}, \gamma)$ nicht vorkommt, gilt

$$\boxed{\frac{\partial}{\partial x_{jn}} \tilde{R}(i, m, \underline{\eta}, \gamma) = 0} . \quad (\text{A.47})$$

Nach γ Werden w_l und x_{lp} als Funktionen von γ betrachtet dann folgt unter Anwendung der verallgemeinerten Kettenregel (A.3) und Beachtung der Definition von $\tilde{R}^b(i, m, \underline{\eta}, \gamma)$

$$\boxed{\frac{\partial}{\partial \gamma} \tilde{R}(i, m, \underline{\eta}, \gamma) = \tilde{R}^b(i, m, \underline{\eta}, \gamma) + \sum_{l=1}^L \left\{ \frac{\partial}{\partial w_l} \tilde{R}(i, m, \underline{\eta}, \gamma) \right\} \dot{w}_l + \sum_{l=1}^L \sum_{p=1}^d \left\{ \frac{\partial}{\partial x_{lp}} \tilde{R}(i, m, \underline{\eta}, \gamma) \right\} \dot{x}_{lp}} . \quad (\text{A.48})$$

A.5.2 Konvexe Kombination

Für diesen Abschnitt sei die Verteilungsfunktion

$$\tilde{F}(\underline{x}, \gamma) := \{1 - \gamma\} \overset{*}{F}(\underline{x}, \gamma) + \gamma \overset{\circ}{F}(\underline{x}, \gamma)$$

der progressiven Verteilung in jedem Progressionsschritt γ eine konvexe Kombination einer Start- $\overset{*}{F}(\underline{x}, \gamma)$ mit einer Zielverteilungsfunktion $\overset{\circ}{F}(\underline{x}, \gamma)$. Die Funktionen $\overset{*}{R}$ beziehungsweise $\overset{\circ}{R}$ sind für $\overset{*}{F}$ beziehungsweise $\overset{\circ}{F}$ völlig analog zu den \tilde{R} Funktionen für \tilde{F} definiert und werden als bekannt vorausgesetzt. Siehe auch die beiden Folgeabschnitte für Gleich- und Gauß-Mischverteilungen.

Für $\tilde{R}^b(i, \underline{\eta}, \gamma)$ gilt

$$\tilde{R}^b(i, \underline{\eta}, \gamma) \quad (\text{Def.})$$

$$\int_{\underline{x}_i}^{\underline{c}_2} \left[\frac{\partial}{\partial \gamma} \tilde{F}(\underline{x}, \gamma) \right] d\underline{x} \quad (\text{Def.})$$

$$= \int_{\underline{x}_i}^{\underline{c}_2} \left[\frac{\partial}{\partial \gamma} \left\{ (1 - \gamma) \overset{*}{F}(\underline{x}, \gamma) + \gamma \overset{\circ}{F}(\underline{x}, \gamma) \right\} \right] d\underline{x} \quad (\text{Linearität und Produktregel})$$

$$= \int_{\underline{x}_i}^{\underline{c}_2} \left[-\overset{*}{F}(\underline{x}) - \gamma \frac{\partial}{\partial \gamma} \overset{*}{F}(\underline{x}, \gamma) + \overset{\circ}{F}(\underline{x}, \gamma) + \gamma \frac{\partial}{\partial \gamma} \overset{\circ}{F}(\underline{x}, \gamma) \right] d\underline{x} \quad (\text{Linearität und umordnen, Def.})$$

$$= \left\{ \overset{\circ}{R}(i, \underline{\eta}, \gamma) + \gamma \overset{\circ}{R}^b(i, \underline{\eta}, \gamma) \right\} - \left\{ \overset{*}{R}(i, \underline{\eta}, \gamma) + \gamma \overset{*}{R}^b(i, \underline{\eta}, \gamma) \right\} .$$

Für $\tilde{R}^b(i, m, \underline{\eta}, \gamma)$ ist die Rechnung analog. Insgesamt liefert die Linearität des Integrals folgende Resultate.

$$\tilde{R}(i, \underline{\eta}, \gamma) = (1 - \gamma) \overset{*}{R}(i, \underline{\eta}, \gamma) + \gamma \overset{\circ}{R}(i, \underline{\eta}, \gamma)$$

$$\tilde{R}(i, m, \underline{\eta}, \gamma) = (1 - \gamma) \overset{*}{R}(i, m, \underline{\eta}, \gamma) + \gamma \overset{\circ}{R}(i, m, \underline{\eta}, \gamma)$$

$$\tilde{R}(i, m, n, \underline{\eta}, \gamma) = (1 - \gamma) \overset{*}{R}(i, m, n, \underline{\eta}, \gamma) + \gamma \overset{\circ}{R}(i, m, n, \underline{\eta}, \gamma)$$

$$\tilde{R}^b(i, \underline{\eta}, \gamma) = \left\{ \overset{\circ}{R}(i, \underline{\eta}, \gamma) + \gamma \overset{\circ}{R}^b(i, \underline{\eta}, \gamma) \right\} - \left\{ \overset{*}{R}(i, \underline{\eta}, \gamma) + \gamma \overset{*}{R}^b(i, \underline{\eta}, \gamma) \right\}$$

$$\tilde{R}^b(i, m, \underline{\eta}, \gamma) = \left\{ \overset{\circ}{R}(i, m, \underline{\eta}, \gamma) + \gamma \overset{\circ}{R}^b(i, m, \underline{\eta}, \gamma) \right\} - \left\{ \overset{*}{R}(i, m, \underline{\eta}, \gamma) + \gamma \overset{*}{R}^b(i, m, \underline{\eta}, \gamma) \right\}$$

$$\frac{\partial}{\partial x_{im}} \tilde{R}(i, m, \underline{\eta}, \gamma) = (1 - \gamma) \frac{\partial}{\partial x_{im}} \overset{*}{R}(i, m, \underline{\eta}, \gamma) + \gamma \frac{\partial}{\partial x_{im}} \overset{\circ}{R}(i, m, \underline{\eta}, \gamma)$$

A.5.3 Gleichverteilung

Die Berechnung der benötigten Funktionen für die Gleichverteilung auf dem Intervall $[c_1, c_2]$ mit

$$\tilde{F}(\underline{x}, \gamma) := \prod_{q=1}^d \left\{ \frac{x_q - c_{1q}}{c_{2q} - c_{1q}} 1_{[c_{2q}, c_{1q}]}(x_q) \right\}$$

erfolgt ohne größere Mühe, insbesondere da die Unabhängigkeit von γ vorausgesetzt wird. Mit den Abkürzungen

$$U_2^\Pi(i) := \prod_{q=1}^d \left\{ \frac{1}{2} \{c_{2q}^2 - x_{iq}^2\} + c_{2q} \{c_{2q} - x_{iq}\} \right\}$$

$$U_2^\Pi(i, m) := \prod_{q=1, q \neq m}^d \left\{ \frac{1}{2} \{c_{2q}^2 - x_{iq}^2\} + c_{2q} \{c_{2q} - x_{iq}\} \right\}$$

$$U_2^\Pi(i, m, n) := \begin{cases} 1 & d = 2 \\ \prod_{q=1, q \neq m, q \neq n}^d \left\{ \frac{1}{2} \{c_{2q}^2 - x_{iq}^2\} + c_{2q} \{c_{2q} - x_{iq}\} \right\} & \text{sonst} \end{cases}$$

$$C := \prod_{q=1}^d \{c_{2q} - c_{1q}\}$$

ist das Resultat

$$\begin{aligned} \tilde{R}(i, \underline{\eta}, \gamma) &= \frac{1}{C} \cdot U_2^\Pi(i) \\ \tilde{R}(i, m, \underline{\eta}, \gamma) &= \frac{1}{C} \cdot \{x_{im} + c_{2m}\} U_2^\Pi(i, m) \\ \tilde{R}(i, m, n, \underline{\eta}, \gamma) &= \frac{1}{C} \cdot \{x_{im} + c_{2m}\} \{x_{in} + c_{2n}\} U_2^\Pi(i, m, n) \\ \tilde{R}^b(i, \underline{\eta}, \gamma) &= \frac{1}{C} \cdot U_2^\Pi(i) \\ \tilde{R}^b(i, m, \underline{\eta}, \gamma) &= \frac{1}{C} \cdot \{x_{im} + c_{2m}\} U_2^\Pi(i, m) \\ \frac{\partial}{\partial x_{im}} \tilde{R}(i, m, \underline{\eta}, \gamma) &= \frac{1}{C} \cdot U_2^\Pi(i, m) \end{aligned}$$

A.5.4 Achsenparallele Gauß-Mischverteilung

Die progressive Verteilung mit Verteilungsfunktion

$$\tilde{F}(x, \gamma) := \sum_{k=1}^K \omega_k(\gamma) \cdot \prod_{q=1}^d \Phi(x_q, \mu_{kq}(\gamma), \sigma_{kq}(\gamma))$$

sei zu jedem Progressionszeitpunkt γ eine achsenparallele Gauß-Mischverteilung mit $K \geq 1$ Komponenten. $\Phi(x, \mu, \sigma)$ bezeichnet dabei die Verteilungsfunktion einer eindimensionalen Gauß-Verteilung mit Mittelwert μ und Standardabweichung σ . Die Gewichte $\omega_k = \omega_k(\gamma)$, Mittelwerte $\mu_{kq} = \mu_{kq}(\gamma)$ und Standardabweichungen $\sigma_{kq} = \sigma_{kq}(\gamma)$ sollen dabei im Allgemeinen vom Progressionsparameter γ abhängig sein. Die Differenzierbarkeit von ω_k , μ_{kq} und σ_{kq} nach γ wird vorausgesetzt.

Die Berechnung der Terme kann mit einem Computeralgebrasystem erfolgen. Das Zwischenspeichern gemeinsamer Teilausdrücke reduziert den Laufzeitaufwand in der Implementierung erheblich. Dazu seien folgende Terme eingeführt:

$$\begin{aligned} T(i, q, k) &:= \int_{x_{iq}}^{c_{2q}} \left[\frac{\partial}{\partial \gamma} \Phi(x_q, \mu_{kq}, \sigma_{kq}) \right] dx_q \\ &= [\sigma_{kq} \dot{\sigma}_{kq} \mathcal{N}(x_q, \mu_{kq}, \sigma_{kq}) - \dot{\mu}_{kq} \Phi(x_q, \mu_{kq}, \sigma_{kq})]_{x_{iq}}^{c_{2q}} \end{aligned}$$

$$\begin{aligned} S(i, k, q) &:= \int_{x_{iq}}^{c_{2q}} \Phi(x_q, \mu_{kq}, \sigma_{kq}) dx_q \\ &= [\sigma_{kq}^2 \mathcal{N}(x_q, \mu_{kq}, \sigma_{kq}) + \{x_q - \mu_{kq}\} \Phi(x_q, \mu_{kq}, \sigma_{kq})]_{x_{iq}}^{c_{2q}} \end{aligned}$$

$$S^\Pi(i, k) := \prod_{q=1}^d S(i, k, q)$$

$$S^\Pi(i, k, m) := \prod_{q=1, q \neq m}^d S(i, k, q)$$

$$S^\Pi(i, k, m, n) := \begin{cases} 1 & d = 2 \\ \prod_{q=1, q \neq m, q \neq n}^d S(i, k, q) & \text{sonst} \end{cases}$$

Die Herleitung einer Darstellung der benötigten Funktionen mit Hilfe dieser Abkürzungen ist langwierig, das Resultat lautet:

$$\tilde{R}(i, \underline{\eta}, \gamma) = \sum_{k=1}^K \omega_k S^\Pi(i, k)$$

$$\tilde{R}(i, m, \underline{\eta}, \gamma) = \sum_{k=1}^K \omega_k \Phi(x_{im}, \mu_{km}, \sigma_{km}) S^\Pi(i, k, m)$$

$$\tilde{R}(i, m, n, \underline{\eta}, \gamma) = \sum_{k=1}^K \omega_k \Phi(x_{im}, \mu_{km}, \sigma_{km}) \Phi(x_{in}, \mu_{kn}, \sigma_{kn}) S^\Pi(i, k, m, n)$$

$$\tilde{R}^b(i, \underline{\eta}, \gamma) = \sum_{k=1}^K \dot{\omega}_k S^\Pi(i, k) + \sum_{k=1}^K \omega_k \left\{ \sum_{q=1}^d S^\Pi(i, k, q) T(i, q, k) \right\}$$

$$\begin{aligned} \tilde{R}^b(i, m, \underline{\eta}, \gamma) &= \sum_{k=1}^K \dot{\omega}_k \Phi(x_{im}, \mu_{km}, \sigma_{km}) S^\Pi(i, k, m) \\ &\quad + \sum_{k=1}^K \omega_k \left\{ \left\{ \sum_{q=1, q \neq m}^d S^\Pi(i, k, q, m) T(i, q, k) \right\} \Phi(x_{im}, \mu_{km}, \sigma_{km}) \right. \\ &\quad \left. + S^\Pi(i, k, m) \left[\frac{\partial}{\partial \gamma} \Phi(x, \mu_{km}, \sigma_{km}) \right]_{x \equiv x_{im}} \right\} \end{aligned}$$

$$\frac{\partial}{\partial x_{im}} \tilde{R}(i, m, \underline{\eta}, \gamma) = \sum_{k=1}^K \omega_k \mathcal{N}(x_{im}, \mu_{km}, \sigma_{km}) S^\Pi(i, k, m)$$

Für die Berechnung von $\tilde{R}^b(i, m, \underline{\eta}, \gamma)$ sei noch angemerkt, dass

$$\frac{\partial}{\partial \gamma} \Phi(x, \mu_{km}, \sigma_{km}) = -\dot{\mu}_{km} \mathcal{N}(x, \mu_{km}, \sigma_{km}) - \frac{\dot{\sigma}_{km}}{\sigma_{km}} \{x - \mu_{km}\} \mathcal{N}(x, \mu_{km}, \sigma_{km})$$

gilt.

A.6 Partielle Ableitungen des Gütemaßes

A.6.1 Nach w_i

$$\frac{\partial}{\partial w_i} G(\underline{\eta}, \gamma) \tag{Def.}$$

$$= \frac{\partial}{\partial w_i} \left\{ \int_{\underline{c}_1}^{\underline{c}_2} \left[\left\{ \tilde{F}(\underline{x}, \gamma) - F(\underline{x}, \underline{\eta}) \right\}^2 \right] d\underline{x} \right\} \tag{Differenziation unter dem Integral}$$

$$= \int_{\underline{c}_1}^{\underline{c}_2} \left[\frac{\partial}{\partial w_i} \left\{ \left\{ \tilde{F}(\underline{x}, \gamma) - F(\underline{x}, \underline{\eta}) \right\}^2 \right\} \right] d\underline{x} \tag{Kettenregel}$$

$$= 2 \int_{\underline{c}_1}^{\underline{c}_2} \left[\left\{ \tilde{F}(\underline{x}, \gamma) - F(\underline{x}, \underline{\eta}) \right\} \frac{\partial}{\partial w_i} \left\{ \tilde{F}(\underline{x}, \gamma) - F(\underline{x}, \underline{\eta}) \right\} \right] d\underline{x} \tag{ $\tilde{F}(\underline{x}, \gamma)$ frei von w_i }$$

$$= -2 \int_{\underline{c}_1}^{\underline{c}_2} \left[\left\{ \tilde{F}(\underline{x}, \gamma) - F(\underline{x}, \underline{\eta}) \right\} \frac{\partial}{\partial w_i} F(\underline{x}, \underline{\eta}) \right] d\underline{x} \tag{mit (A.22)}$$

$$= -2 \int_{\underline{c}_1}^{\underline{c}_2} \left[\left\{ \tilde{F}(\underline{x}, \gamma) - F(\underline{x}, \underline{\eta}) \right\} \mathbf{H}(\underline{x} - \underline{x}_i) \right] d\underline{x} \tag{Def., Einbringen in die Integrationsgrenzen}$$

$$= -2 \int_{\underline{x}_i}^{\underline{c}_2} \left[\tilde{F}(\underline{x}, \gamma) - F(\underline{x}, \underline{\eta}) \right] d\underline{x}$$

$$= 2 \int_{\underline{x}_i}^{\underline{c}_2} F(\underline{x}, \underline{\eta}) d\underline{x} - 2 \int_{\underline{x}_i}^{\underline{c}} \tilde{F}(\underline{x}, \gamma) d\underline{x} .$$

Mit den Definitionen (A.11) und (A.36) folgt

$$\boxed{\frac{\partial}{\partial w_i} G(\underline{\eta}, \gamma) = 2R(i, \underline{\eta}) - 2\tilde{R}(i, \underline{\eta}, \gamma)} \tag{A.49}$$

A.6.2 Nach x_{im}

$$\begin{aligned}
& \frac{\partial}{\partial x_{im}} G(\underline{\eta}, \gamma) && \text{Def.} \\
& = \frac{\partial}{\partial \mathbf{x}_{im}} \left\{ \int_{c_1}^{c_2} \left\{ \tilde{F}(\underline{\mathbf{x}}, \gamma) - F(\underline{\mathbf{x}}, \underline{\eta}) \right\}^2 d\underline{\mathbf{x}} \right\} && \text{(Differenziation unter dem Integral)} \\
& = \int_{c_1}^{c_2} \frac{\partial}{\partial \mathbf{x}_{im}} \left\{ \left\{ \tilde{F}(\underline{\mathbf{x}}, \gamma) - F(\underline{\mathbf{x}}, \underline{\eta}) \right\}^2 \right\} d\underline{\mathbf{x}} && \text{(Kettenregel)} \\
& = 2 \int_{c_1}^{c_2} \left\{ \tilde{F}(\underline{\mathbf{x}}, \gamma) - F(\underline{\mathbf{x}}, \underline{\eta}) \right\} \frac{\partial}{\partial \mathbf{x}_{im}} \left\{ \tilde{F}(\underline{\mathbf{x}}, \gamma) - F(\underline{\mathbf{x}}, \underline{\eta}) \right\} d\underline{\mathbf{x}} && (\tilde{F}(\underline{\mathbf{x}}, \gamma) \text{ frei von } x_{im}) \\
& = 2 \int_{c_1}^{c_2} \left\{ \tilde{F}(\underline{\mathbf{x}}, \gamma) - F(\underline{\mathbf{x}}, \underline{\eta}) \right\} \left\{ -\frac{\partial}{\partial \mathbf{x}_{im}} F(\underline{\mathbf{x}}, \underline{\eta}) \right\} d\underline{\mathbf{x}} && \text{(mit (A.23))} \\
& = 2 \int_{c_1}^{c_2} \left\{ \tilde{F}(\underline{\mathbf{x}}, \gamma) - F(\underline{\mathbf{x}}, \underline{\eta}) \right\} \left\{ -w_i \left\{ \prod_{p=1, p \neq m}^d H(x_p - x_{ip}) \right\} \delta(\mathbf{x}_m - \mathbf{x}_{im}) \right\} d\underline{\mathbf{x}} && \text{(Linearität, Fubini, Filtereigenschaft)} \\
& = 2w_i \int_{c_1}^{c_2} \left\{ \prod_{p=1, p \neq m}^d H(\mathbf{x}_p - \mathbf{x}_{ip}) \right\} \left\{ \tilde{F}(\underline{\mathbf{x}}|_{x_m \equiv x_{im}}, \gamma) - F(\underline{\mathbf{x}}|_{x_m \equiv x_{im}}, \underline{\eta}) \right\} d\mathbf{x}_q && \text{(Einbringen in die Integrationsgrenzen)} \\
& \quad \quad \quad q=1, \dots, d; q \neq m \\
& = 2w_i \int_{x_{iq}}^{c_2} \left[\tilde{F}(\underline{\mathbf{x}}|_{x_m \equiv x_{im}}, \gamma) - F(\underline{\mathbf{x}}|_{x_m \equiv x_{im}}, \underline{\eta}) \right] dx_q \cdot
\end{aligned}$$

Die Linearität des Integrals liefert zusammen mit den Definitionen (A.37) und (A.12)

$$\boxed{\frac{\partial}{\partial x_{im}} G(\underline{\eta}, \gamma) = 2w_i \left\{ \tilde{R}(i, m, \underline{\eta}, \gamma) - R(i, m, \underline{\eta}) \right\}} \quad (\text{A.50})$$

Literaturverzeichnis

- [1] BABOVSKY, HANS: *Mathematische Methoden in der Systemtheorie: Fourieranalysis*. Teubner, 1987.
- [2] BAUER, HEINZ: *Wahrscheinlichkeitstheorie*. de Gruyter, 5., durchgesehene und verbesserte Auflage, 2002.
- [3] BRONSTEJN, IL'JA N., GERHARD MUSIOL und HEINER MÜHLIG: *Taschenbuch der Mathematik*. Deutsch, 3., überarbeitete und erweiterte Auflage, 1997.
- [4] HÄGGSTRÖM, OLLE: *Finite Markov chains and algorithmic applications*. Cambridge University Press, 2002.
- [5] HANEBECK, UWE D., KAI BRIECHLE, and ANDREAS RAUH: *Progressive Bayes: A New Framework for Nonlinear State Estimation*. In *Proceedings of SPIE, AeroSense Symposium*, volume 5099, pages 256 – 267, Orlando, Florida, May 2003.
- [6] HANEBECK, UWE D. and OLIVER C. SCHREMPF: *Greedy Algorithms for Dirac Mixture Approximation of Arbitrary Probability Density Functions*. In *Proceedings of the 2007 IEEE Conference on Decision and Control (CDC 2007)*, New Orleans, Louisiana, December 2007.
- [7] HENZE, NORBERT: *Stochastik II*. Vorlesungsskript, 1997.
- [8] HEUSER, HARRO: *Lehrbuch der Analysis Teil 1*. B. G. Teubner, 11. Auflage, 1994.
- [9] OPPENHEIM, ALAN V. und ALAN S. WILLSKY: *Signale und Systeme*. VCH, 2. Auflage, 1992.
- [10] SCHREMPF, OLIVER C., DIETRICH BRUNN und UWE D. HANEBECK: *Density Approximation Based on Dirac Mixtures with Regard to Nonlinear Estimation and Filtering*. In: *Proceedings of the 2006 IEEE Conference on Decision and Control (CDC 2006)*, San Diego, California, Dezember 2006.
- [11] SCHREMPF, OLIVER C., DIETRICH BRUNN und UWE D. HANEBECK: *Dirac Mixture Density Approximation Based on Minimization of the Weighted Cramér-von Mises Distance*. In: *Proceedings of the 2006 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2006)*, Seiten 512–517, Heidelberg, Germany, September 2006.

- [12] UWE D. HANEBECK, OLIVER C. SCHREMPF: *Stochastische Informationsverarbeitung*.
Vorlesungsskript, Wintersemester 2006/07.