

Transparente Verhaltensentscheidungen automatisierter Fahrzeuge – Sensitivitätsanalyse von Planungsalgorithmen

Zur Erlangung des akademischen Grades einer

**DOKTORIN DER INGENIEURWISSENSCHAFTEN
(Dr.-Ing.)**

von der KIT-Fakultät für Maschinenbau des
Karlsruher Instituts für Technologie (KIT)

angenommene

DISSERTATION

von

M.Sc. Franziska Henze

geb. in Hildesheim

Tag der mündlichen Prüfung:

15. Februar 2023

Hauptreferent:
Korreferent:

Prof. Dr.-Ing. Christoph Stiller
Prof. Dr.-Ing. Matthias Althoff



Dieses Werk ist lizenziert unter einer Creative Commons
Namensnennung 4.0 International Lizenz (CC BY 4.0):
<https://creativecommons.org/licenses/by/4.0/deed.de>

Zusammenfassung

Im automatisierten Fahren arbeiten verschiedene Algorithmen zusammen, die Informationen übergeben, z. B. von der Wahrnehmung an die Planung. Dabei können sich Unsicherheiten fortpflanzen, die die Entscheidung des Planungsalgorithmus stark beeinflussen und möglicherweise zu Kollisionen führen, falls sie vom Planer nicht angemessen berücksichtigt werden. Aus diesem Grund wird die Qualität eines Planers häufig mit dem Auftreten von Kollisionen quantifiziert. Nutzer werden jedoch nicht nur ein sicheres, sondern aber auch ein ruhiges Fahrverhalten erwarten. Daher definiert diese Arbeit ein Maß für zulässige Planungsentscheidungen, das Schwankungen zwischen Planungsergebnissen betrachtet. Dies trägt dazu bei, dass unerwartete Fahrentscheidungen vermieden werden.

In dieser Arbeit wird die Sensitivitätsanalyse nach Morris verwendet, um den Einfluss von Unsicherheiten in der Wahrnehmung auf Planungsentscheidungen zu untersuchen. Morris' Analyse unterscheidet zwischen Inputs, die für die Planungsentscheidung relevant waren, und Inputs, die die Entscheidung nicht beeinflusst haben. Daraus wird abgeleitet, welcher Input zu unsicher war und welche Konsequenzen gezogen werden müssen: Einerseits sollte die Wahrnehmung speziell für diesen Input genauer arbeiten, andererseits sollte der Planer robuster gegenüber Unsicherheiten dieser Art werden. Aus diesen Ergebnissen kann leider noch nicht gefolgert werden, welche Menge an Wahrnehmungsunsicherheiten akzeptabel wäre. Dafür definiert ein stochastisches Optimierungsproblem die maximalen Unsicherheiten, die gerade noch zu Planungsergebnissen führen, die nicht stärker als ein Grenzwert schwanken. Dieses Vorgehen bietet für verschiedene Verkehrsszenarien die Möglichkeit, die Anforderungen an die Wahrnehmung genau mit den Voraussetzungen der Planung abzustimmen. Zusammengefasst können sowohl die Sensitivitätsanalyse als auch das stochastische Optimierungsproblem den Entwicklungsprozess unterstützen, indem sie z. B. beim Debugging helfen.

Diese neuen Ansätze tragen dazu bei, die Verarbeitung der unsicherheitsbehafteten Informationen nachvollziehbarer zu gestalten und die Entwicklung automatisierter Fahrzeuge zu unterstützen. Im automatisierten Fahren betrifft Nachvollziehbarkeit aber nicht nur den Entwicklungsprozess, sondern auch den Endnutzer der Fahrfunktion. Die aktuelle Forschung erwartet, dass die Skepsis, die in großen Teilen der Bevölkerung gegenüber dem automatisierten Fahren herrscht, u. a. durch das Aufbauen von angemessenem Vertrauen abgebaut werden kann. Eine Möglichkeit, um Vertrauen bei Nutzern aufzubauen, sind Erklärungen, die das Verständnis der Nutzer verbessern. Bevor angemessene Erklärungen vom automatisierten Fahrzeug bereit gestellt werden können, sind zunächst Informationen nötig, die für die Beantwortung möglicher Fragen notwendig sind. Dafür wird zunächst das umfangreiche Wissen von Entwicklern als Ausgangspunkt verwendet. Daher wird zunächst mithilfe von Fokusgruppen untersucht, welche Fragen die Experten zu Fahrentscheidungen stellen, die sie während Testfahrten im Realverkehr als unerwartet wahrnahmen. Die Ergebnisse deuten darauf hin, dass vorrangig unerwartete laterale Bewegungen erklärt werden sollten. Um derartige unerwartete Fahrentscheidungen auch für Nutzer ohne detailliertes Fachwissen zu erklären, wird erneut die Sensitivitätsanalyse verwendet: Mit ihrer Hilfe werden die Informationen, die für zentrale Entscheidungen des automatisierten Fahrzeugs relevant sind, identifiziert. Anschließend werden diese interpretiert und in Zusammenhang mit der Umgebung sowie den geltenden Verkehrsregeln gesetzt. Das Ergebnis sind automatisch generierte, textbasierte Beschreibungen, die das Verhalten des automatisierten Fahrzeugs in leicht verständlicher Sprache wiedergeben und die nicht von menschlichen Fehlinterpretationen beeinflusst sind. Die Ergebnisse einer weiteren Studie zeigen, dass die generierten Sätze zwar derzeit noch als zu lang wahrgenommen werden und daher noch nicht als Nutzeranzeigen im Realverkehr verwendet werden können. Dennoch sind sie ein wichtiger und vielversprechender Schritt, um die technischen Voraussetzungen für zukünftige, informationsreiche und leicht verständliche Nutzeranzeigen zu erfüllen, die die Nachvollziehbarkeit von automatisierten Verhaltensentscheidungen erhöhen.

Summary

In automated driving, different algorithms work together to pass information, e. g., from sensing to planning modules. In this process, uncertainties can propagate, which in turn influence the decision of the planner. If not adequately considered by the planner this may lead to collisions. Therefore, the quality of a planner is often quantified by the occurrence of collisions. However, for users of automated driving functions not only the absence of collisions will be important, but also the driving experience. To account for this problem, this work defines a measure for admissible planning decisions that considers variations between planning results. This contributes to avoiding unexpected driving decisions.

In this work, Morris' sensitivity analysis is used to examine the effect of uncertainties in sensing modules on planning decisions. It distinguishes between inputs that were relevant for the planning decision and inputs that did not influence the decision. Based on this, all inputs are identified that were too uncertain. Consequently, either the sensing modules need to process the respective input more accurately or the planner needs to become more robust to uncertainties of this kind. However, this finding is not yet sufficient to infer the maximum amount of perceptual uncertainties that will still result in admissible planning decisions. Therefore, a stochastic optimization problem defines the maximum amount of uncertainties permitted that will still lead to planning results that do not vary more than a predefined threshold. This approach offers the possibility that perception requirements can be precisely matched to planning assumptions in any traffic scenario. In summary, both sensitivity analysis and the stochastic optimization problem can help in the development, e. g. by supporting to debugging process.

These new approaches contribute to a more comprehensible processing of uncertain information and to supporting the development of automated vehicles. In automated driving, however, transparency concerns not only the development process but also the user of the automated function. Current research expects that

the skepticism that prevails in large parts of the population towards automated driving can, among other factors, be reduced through the establishment of adequate trust in automation. One way to build trust among users is to generate explanations that increase the users' understanding of the function. However, before adequate explanations can be provided by the automated vehicle, all the information needed to answer any questions must be available. For this, the developers' considerable knowledge can be used as basis: In a first step, focus groups are used to investigate the questions experts ask when perceiving an unexpected driving decision during test drives in real traffic. The results indicate that priority should be given to explaining unexpected lateral motion. In order to explain such unexpected behavior of automated vehicles to users without detailed knowledge, sensitivity analysis is again used: It identifies the information that is relevant for key decisions of the automated vehicle. This subset of information is then interpreted and put into context with the environment and the applicable traffic rules. The result is an automatically generated, text-based description of the behavior of the automated vehicle in easy-to-understand language. One major advantage is that these descriptions are not affected by potential human misinterpretations. The results of another user study show that even though the sentences are still perceived as too long and therefore cannot be used in real traffic yet, they are nevertheless an important and promising step towards fulfilling the technical requirements for future human-machine interfaces that increase the comprehensibility of automated driving decisions.

Vorwort

Bei der Erstellung dieser Arbeit wurde ich von einer Reihe von Menschen in vielfältigster Weise unterstützt, denen ich meinen Dank ausspreche. Zunächst danke ich Prof. Dr.-Ing. Christoph Stiller für seine Unterstützung und den wissenschaftlichen Freiraum, den er mir gelassen hat. Weiterhin möchte ich mich bei Prof. Dr.-Ing. Matthias Althoff für die Übernahme des Korreferats und das damit verbundene Interesse an meiner Arbeit bedanken. Außerdem danke ich Dennis Faßbender für seine Gedankenanstöße, sein Fachwissen und seine Offenheit gegenüber neuen Ideen.

Als ich mein Promotionsprojekt gestartet habe, habe ich nicht damit gerechnet mit Menschen so unterschiedlicher Fachrichtungen zusammenzuarbeiten. Die vielfältigen Gespräche, die ich mit Maschinenbauern, Informatikern, Mathematikern und Psychologen über die unterschiedlichsten Aspekte des automatisierten Fahrens geführt habe, haben mich und meine Arbeit enorm geprägt. Dafür möchte ich mich bei allen, die mir ihre Zeit geschenkt haben, danken. Vor allem möchte ich mich bei meinen Kolleg:innen des Audi AI-Labs und des Instituts für Mess- und Regelungstechnik (MRT) des KIT bedanken – ich hätte mir keine besseren Sicherheits(bei)fahrer:innen wünschen können!

Ich danke der AUDI AG für die Möglichkeit, dieses Promotionsprojekt im Rahmen des Doktorandenprogramms durchführen zu können. Allen AI-Lab-Kolleg:innen, vor allem Dr.-Ing. Karl-Heinz Siedersberger und Dr.-Ing. Ulrich Hofmann, danke ich für das tolle Umfeld, das ich in der Vorentwicklung erleben durfte. Allen Doktorand:innen danke ich dafür, dass wir gemeinsam gekämpft, gelitten und gelacht haben: Danke an Henrik Bey für seine Simulationsumgebung des IDM-MOBIL-Modells. Danke an Johannes Ossig für seine beruhigende Sichtweise auf die Dinge. Danke an Michael Eder und Dr. rer. nat. Oliver Sieber (auch wenn du nicht zu den AI-Lab-Doks gehörst) für ihr Korrekturlesen von allen mathematischen Herleitungen, bei denen die meisten anderen schreiend davon gelaufen sind. Danke an Moritz Sackmann für sein Interesse und sein Feedback. Danke an Pia Wald für unsere Gedankenf*rze. Danke an Sophie

Feinauer für ihre unbeirrten Nachfragen. Und danke an Karin Christensen, Markus Rußer und Natalie Magdalena Stasinski, die ich als Studierende betreuen durfte und die mich mit ihren Arbeiten tatkräftig unterstützten.

Als nächstes danke ich allen *mrt*: :*Dreamboyz* und *-girlz*. Es war großartig, wie ihr mich bei jedem Sommerseminar mit offenen Armen aufgenommen habt und wie ich dadurch nach jeder Woche mit einem vor neuen Ideen schwirrenden Kopf und einem Lachen auf dem Gesicht zurückgefahren bin. Stellvertretend für alle seien hier Annika Meyer, Florian Wirth, Jan-Hendrik Pauls, Johannes Fischer, Kevin Rösch und Dr.-Ing. Meike Rietdorf genannt. Außerdem danke ich Erna Nagler und Alexandra Stotz dafür, dass sie mich bei allen organisatorischen und bürokratischen Fragen kompetent unterstützt haben.

Ein Teil dieser Arbeit basiert auf qualitativen Ergebnissen, die nur möglich sind, weil so viele Menschen an der Fokusgruppen-Studie und der Online-Umfrage teilgenommen haben. Stellvertretend für alle, die sich dafür Zeit genommen haben, danke ich Kevin Rösch und Marie Fehrmann: Zum Glück seid ihr die einzigen, die nach der Umfrage die Erwähnung in der Danksagung eingefordert habt. Und nein, ich weiß immer noch nicht, wer die dummen Kommentare in der Online-Umfrage hinterlassen hat – das nennt man anonyme Umfrage und es funktioniert wirklich einwandfrei.

Als letztes danke ich allen, die mich auf dem Weg hierher gefördert und gefordert haben, die an mich geglaubt und mich zum Lachen gebracht haben. Vor allem meiner Familie danke ich von Herzen dafür, dass sie immer hinter mir steht, mich unterstützt, stolz auf mich ist und mir dadurch Kraft gibt.

Ingolstadt, Februar 2023

Franziska Henze

Inhaltsverzeichnis

Zusammenfassung und Summary	i
Vorwort	v
Abkürzungen und Symbole	ix
1 Einleitung	1
2 Stand der Wissenschaft und Beiträge der Arbeit	5
2.1 Unsicherheiten in der Wahrnehmung und Planung	6
2.2 Erklärungen algorithmischer Entscheidungen	9
3 Analyse von unsicherheitsbehafteten Informationen	15
3.1 Modellierung zulässiger Unsicherheiten	16
3.2 Abschätzung der Wahrscheinlichkeitsungleichung	26
3.2.1 Chebychev-Saw-Abschätzung	26
3.2.2 Bernoulli-a-posteriori-Abschätzung	33
3.3 Sensitivitätsanalyse nach Morris	40
4 Dynamisches Abbruchkriterium	51
4.1 Definition	51
4.2 Beispiele	56
4.2.1 Testfunktion: g -Funktion	56
4.2.2 Anwendung: Überholmanöver	59
5 Identifikation von zulässigen Unsicherheiten zwischen Wahrnehmung und Planung	63
5.1 Kontinuierliches Optimierungsproblem	63
5.2 Diskretes Optimierungsproblem	66

5.3	Lösungsverfahren mit Sensitivitätsanalyse	68
5.4	Anwendung: Überholmanöver	78
6	Informationsbedarf von HAF-Entwicklern	85
6.1	Fokusgruppe	86
6.2	Ergebnisse	89
6.3	Diskussion	93
7	Generierung erläuternder Beschreibungen	95
7.1	Textgenerierung	96
7.1.1	Annahmen	96
7.1.2	Grundlagen der Textgenerierung	99
7.1.3	Erläuternde Beschreibungen	101
7.2	Validierung: Online-Studie	105
7.2.1	Methode	105
7.2.2	Ergebnisse	109
7.3	Diskussion	112
8	Zusammenfassung und Ausblick	115
A	Anhang	119
A.1	Erwartungswert und Varianz für Beispiel 3.1.8.e	119
A.2	Sensitivitätsanalyse für Beispiel 3.2.6.c	121
A.3	Beweis: Abbruch des dynamischen Abbruchkriteriums	121
A.4	Ergänzende Abbildungen zum dynamischen Abbruchkriterium	125
A.5	Monotonie der Nebenbedingung für Beispiel 3.1.8.g	125
A.6	Optimierungsproblem für Beispiel 5.3.1	131
A.7	Optimierungsergebnisse für Ego-Position p, Beispiel 3.2.6.d	132
A.8	Maschinenlesbare Verkehrsregeln für die Textgenerierung	133
A.9	Grundlagen, Vor- und Nachteile der erläuternden Aussagen	134
	Literaturverzeichnis	137

Abkürzungen und Symbole

Abkürzungen

AF	automatisiertes Fahrzeug
ANOVA	Varianzanalyse (engl. <i>analysis of variance</i>)
DSC	dynamisches Abbruchkriterium (engl. <i>dynamic stop criterion</i>)
E	Experte
EE	Elementarer Effekt
Ego	Ego-Fahrzeug
Experte	Person mit ausgeprägtem fachlichen Hintergrundwissen
FG	Fokusgruppe
FSW	Fahrstreifenwechsel
HAF	hochautomatisiertes Fahren
HMI	Human-Machine-Interface (dt. <i>Mensch-Maschine-Schnittstelle</i>)
IDM	Intelligent Driver Model
MADS	Mesh-Adaptive Direct Search (dt. <i>Gitter-Adaptive direkte Suche</i>)
MOBIL	Minimizing Overall Braking Induced by Lane Changes
MCS	Multilevel Coordinate Search (dt. <i>Multilevel Koordinatensuche</i>)
N	Novize
Novize	Person ohne Fachwissen
NLG	Natural Language Generation (dt. <i>Natürliche Textgenerierung</i>)
POMDP	Partially Observable Markov Decision Process
RSS	Responsibility Sensitive Safety
SA	Sensitivitätsanalyse
StVO	Straßenverkehrsordnung
Ü	Überholmanöver
W-Dichte	Wahrscheinlichkeitsdichte

W-Maß	Wahrscheinlichkeitsmaß
W-Raum	Wahrscheinlichkeitsraum
W-Verteilung	Wahrscheinlichkeitsverteilung
ZV	Zufallsvariable

Symbole

Die folgenden Tabellen geben einen Überblick über die Symbole, die in der Arbeit häufig verwendet werden. Symbole, die hier nicht aufgeführt werden, werden nur in einem begrenzten Teil des Textes genutzt, daher sei der Leser auf die Definition am Anfang des jeweiligen Abschnitts verwiesen. Sofern nicht anders angegeben gelten die folgenden Konventionen für die Notation:

- kursive Buchstaben: keine Einheiten, z. B. k, m
- gerade Buchstaben: einheitenbehaftete Größe, z. B. Zeit $t(s)$
- eine Ausnahme davon bilden Tiefindizes: Tiefindizes sind kursiv, wenn sie den Wert einer Variablen meinen (z. B. einen Laufindex), und gerade, wenn sie etwas mit Worten erklären (Sonderfall: Zeit t , s. u.)
- kleine, nicht fett-gedruckte Buchstaben: Skalare, z. B. k, m
- kleine, fett-gedruckte Buchstaben: Vektoren z. B. \mathbf{M}, \mathbf{q} (einheitenbehaftet)
- große, nicht fett-gedruckte Buchstaben: eindimensionale ZV, z. B. X, \hat{V} (einheitenbehaftet)
- große, fett-gedruckte Buchstaben: mehrdimensionale ZV, z. B. $\hat{\mathbf{Q}}$ (einheitenbehaftet)

Funktionen- und Variablennamen

\mathcal{A}_M	Menge an aktiven Inputs beim DSC
a	longitudinale Beschleunigung des Ego-Fahrzeugs im IDM, vgl. (3.1)
$\bar{a} \in \mathbb{R}^+$	maximale komfortable Beschleunigung im IDM, i. d. R. $\bar{a} = 1.5 \text{ m/s}^2$

$\mathcal{B}(p)$	Bernoulliverteilung mit Erfolgswahrscheinlichkeit p
$\bar{\mathbf{b}} \in \mathbb{R}^{+m}$	akzeptable Abweichungen für zulässige Planungsentscheidungen
d_c	longitudinale Distanz des Ego-Fahrzeugs zum Fahrzeug c
$d_{o,i,t}(\hat{\mathbf{q}}_{t,j})$	EE (3.17) für Output o , Input i unter Messabweichung $\hat{\mathbf{q}}_{t,j}$ zum Zeitpunkt t
$d_w \in \mathbb{R}^+$	gewünschter Sicherheitsabstand, s. Seite 17
Δ_i	Differenz in Input i zwischen zwei Stützstellen eines EE, d. h. zur Berechnung einer Sekantensteigung in Richtung e_i für die SA
$\Delta v = v - v_{vs}$	Geschwindigkeitsdifferenz zwischen dem Ego und dem vorausfahrenden Fahrzeug auf demselben Fahrstreifen
$\delta \in \mathbb{R}^+$	freier Beschleunigungsexponent im IDM, i. d. R. $\delta = 4$
$\mathbb{E}(X)$	Erwartungswert einer eindimensionalen ZV X
$\mathbb{E}(\mathbf{X}) \in \mathbb{R}^k$	Erwartungswert einer k -dimensionalen ZV \mathbf{X}
$e_i \in \mathbb{R}^k$	Einheitsvektor mit Eintrag 1 an Komponente i und 0 sonst
$f_{\mathcal{N}(\mathbb{E}(X), \text{Var}(X))} : \Omega \rightarrow [0, 1]$	W-Dichte einer $\mathcal{N}(\mathbb{E}(X), \text{Var}(X))$ -verteilten ZV X auf Ω
$\mathbf{f} : \mathbb{R}^k \rightarrow \mathbb{R}^m$	Planungsfunktion mit k Inputs und m Outputs
$f_{\mathbf{a}}$	Berechnung der Beschleunigung a nach IDM, vgl. (3.1)
f_o	Berechnung einer einzelnen Planungsentscheidung, $o = 1, \dots, k$
$\mathbf{f}(\mathbf{q})$	Planungsentscheidung für Messgröße \mathbf{q}
$\mathbf{f}(\mathbf{q} + \hat{\mathbf{Q}})$	als ZV aufgefasste Planungsentscheidung unter Messergebnis $\mathbf{q} + \hat{\mathbf{Q}}$
$\mathbf{f}(\mathbf{q} + \hat{\mathbf{q}})$	unsichere Planungsentscheidung unter Messergebnis $\mathbf{q} + \hat{\mathbf{Q}}$, Realisierung der ZV $\mathbf{f}(\mathbf{q} + \hat{\mathbf{Q}})$

κ_{act}	Konstante im DSC, Unterscheidung zwischen aktiven und inaktiven Inputs
κ_{stop}	Konstante im DSC, Abbruch der Iteration
λ_{S}	Konstante (3.12) zur Berechnung der Chebychev-Saw-Ungleichung
$J : \mathbb{R}^k \rightarrow \mathbb{R}$	Zielfunktion des kontinuierlichen und numerischen Optimierungsproblems (5.1), (5.4)
$k \in \mathbb{N}$	Anzahl an Inputs, z. B. für eine Planungsfunktion
$M \in \mathbb{N}$	Stichprobengröße
$M_{\text{min}} \in \mathbb{N}$	Mindeststichprobengröße für das DSC
$m \in \mathbb{N}$	Anzahl an Outputs, z. B. von einer Planungsfunktion
μ_X	Stichprobenmittelwert (3.8) der ZV X mit unbekannter Verteilung
$\mu_{o,i,t}$	charakteristische Größe für unbekannte Verteilung der EE: Stichprobenmittelwert (3.18) der EE für Output o , Input i zum Zeitpunkt t
$\mu_{o,i,t}^*$	charakteristische Größe für unbekannte Verteilung der EE: Stichprobenmittelwert (3.20) der absoluten EE für Output o , Input i zum Zeitpunkt t
$\mathcal{N}(\mathbb{E}(X), \text{Var}(X))$	Normalverteilung der ZV X mit Erwartungswert $\mathbb{E}(X)$ und Varianz $\text{Var}(X)$
\mathbb{P}_X	W-Maß für ZV X
$p \in [0, 1]$	Wahrscheinlichkeit
$\bar{p} \in [0, 1]^m$	Wahrscheinlichkeiten für zulässige Planungsentscheidungen
p	longitudinale Position des Ego-Fahrzeugs, Messgröße
$\Phi_{f_a} : \Omega \rightarrow [0, 1]$	W-Verteilung der ZV $f_a(v + \hat{V})$

$\phi_{f_a} : \Omega \rightarrow \mathbb{R}$	W-Dichte der ZV $f_a(v + \hat{V})$
$\mathbf{q} \in \Omega$	Umgebungsbeschreibung, Messgröße (in der Simulation vollständig bekannt)
$\hat{\mathbf{Q}}$	als ZV aufgefasste Messabweichung
$\hat{\mathbf{q}} \in \Omega$	Messabweichung, Realisierung der ZV $\hat{\mathbf{Q}}$
$\mathbf{q} + \hat{\mathbf{Q}} \in \Omega$	als ZV aufgefasstes Messergebnis
$\mathbf{q} + \hat{\mathbf{q}} \in \Omega$	Messergebnis, Realisierung der ZV $\mathbf{q} + \hat{\mathbf{Q}}$
$\mathcal{R}_{o,t}$	Menge an Inputs (7.1), die zum Zeitpunkt t für Output o relevant sind
σ_i	bekannte Standardabweichung eines normalverteilten Inputs i
σ_X^2	Stichprobenvarianz (3.9) der ZV X mit unbekannter Verteilung
$\sigma_{o,i,t}^2$	charakteristische Größe für unbekannte Verteilung der EE: Stichprobenvarianz (3.19) der EE für Output o , Input i zum Zeitpunkt t
σ^*	analytische Lösung des kontinuierlichen Optimierungsproblems (5.1)
σ_C^*	numerische Lösung des kontinuierlichen Optimierungsproblems (5.1) mit Chebychev-Ungleichung zur Approximation der stochastischen Nebenbedingung (5.1.2)
σ_S^*	numerische Lösung des diskreten Optimierungsproblems (5.4) mit Chebychev-Saw-Ungleichung zur Approximation der stochastischen Nebenbedingung (5.4.2)
$t \in \mathbb{R}$	Zeit in Sekunden (s)
$\mathcal{U}([0, 1])$	Gleichverteilung im Intervall $[0, 1]$
$\mathbf{u} \in \mathbb{R}^{+k}$	Vektor mit maximalen Messunsicherheiten, vgl. Nebenbedingung (5.2)
u. d. N.	unter der Nebenbedingung

$\text{Var}(X) \in \mathbb{R}^+$	Varianz einer eindimensionalen ZV X
$\text{Var}(\mathbf{X}) \in \mathbb{R}^{+k \times k}$	Varianz einer k -dimensionalen ZV \mathbf{X} ; für paarweise unkorrelierte ZV $(X_1, \dots, X_k)^\top$ gilt $\text{Var}(\mathbf{X}) = \text{diag}(\text{Var}(X_1), \dots, \text{Var}(X_k))$
v	longitudinale Geschwindigkeit des Ego-Fahrzeugs, Messgröße
\hat{V}	als ZV aufgefasste Messabweichung für Ego-Geschwindigkeit v
\hat{v}	Messabweichung der longitudinalen Geschwindigkeit des Ego-Fahrzeugs, Realisierung der ZV \hat{V}
v_c	longitudinale Geschwindigkeit des Fahrzeugs c
v_w	Wunschgeschwindigkeit im IDM
$\mathbf{w} \in \mathbb{R}^{+k}$	Gewichtsvektor für Zielfunktion J des Optimierungsproblems
X	ZV
$z_{1-\alpha}$	$(1 - \alpha)$ -Quantil der Normalverteilung
$\Omega \subseteq \mathbb{R}^k$	Grundmenge für Umgebungsbeschreibungen

Mathematische Operatoren und Räume

$ \cdot $	Betragsfunktion
$\ \cdot\ _F$	Frobeniusnorm, für eine Matrix $A \in \mathbb{R}^{m \times n}$ gilt $\ A\ _F^2 := \sum_{i=1}^m \sum_{j=1}^n a_{i,j} ^2$
\wedge	logisches Und
\vee	logisches Oder
\cdot^\top	Transponieren eines Vektors
erf	Errorfunktion
exp	Exponentialfunktion
max	Maximum

\min	Minimum
\mathbb{N}	natürliche Zahlen
$\mathbb{P}(A B)$	bedingte Wahrscheinlichkeit von Ereignis A unter der Bedingung B
\mathbb{R}	reelle Zahlen
\mathbb{R}^+	reelle positive Zahlen ohne 0: $\{x \in \mathbb{R} \mid x > 0\}$
\mathbb{R}_0^+	reelle positive Zahlen mit 0: $\{x \in \mathbb{R} \mid x \geq 0\}$

Allgemeine Tiefindizes

a	Beschleunigung (engl. <i>acceleration</i>)
<i>i</i>	Laufindex, meist $i = 1, \dots, k$
lc	Fahrstreifenwechsel (engl. <i>lane change</i>)
<i>o</i>	Laufindex, meist $o = 1, \dots, m$
nl	nachfolgendes Fahrzeug hinter dem Ego auf dem linken Fahrstreifen
nr	nachfolgendes Fahrzeug hinter dem Ego auf dem rechten Fahrstreifen
ns	nachfolgendes Fahrzeug hinter dem Ego auf demselben Fahrstreifen
<i>t</i>	Zeitpunkt t , Notation: für eine Funktion $f : \mathbb{R}^+ \rightarrow \mathbb{R}$ gilt die abkürzende Schreibweise $f(t) := f_t$
vl	vorausfahrendes Fahrzeug vor dem Ego auf dem linken Fahrstreifen
vr	vorausfahrendes Fahrzeug vor dem Ego auf dem rechten Fahrstreifen
vs	vorausfahrendes Fahrzeug vor dem Ego auf demselben Fahrstreifen

1 Einleitung

Die Entwicklung automatisierter Fahrzeuge wird seit Jahren in vielen verschiedenen Bereichen vorangetrieben: Einerseits wird die Sensorik weiterentwickelt (Marti, de Miguel, Garcia & Perez, 2019) und neue algorithmische Ansätze getestet (Paden, Čáp, Yong, Yershov & Frazzoli, 2016). Andererseits werden neue Möglichkeiten erforscht, wie Nutzer mit einem automatisierten Fahrzeug (AF) interagieren können (Ayoub, Zhou, Bao & Yang, 2019). Bei ersterem wird das Augenmerk vorrangig auf die technische Funktionalität gelegt und beispielsweise versucht, dem AF beizubringen, den menschlichen Fahrstil zu imitieren (You, Lu, Filev & Tsiotras, 2019). Letzteres beschäftigt sich intensiv mit der Unterstützung des Nutzers in seiner neuen Rolle als Passagier (Kyriakidis et al., 2019). Sicher sind beide Forschungsschwerpunkte essentiell dafür, diese neue Technologie so zu implementieren, dass AF die Marktreife erhalten und nach dem Übergang von Pferdekutsche zum Automobil der nächste bedeutende Schritt in der Mobilität der Zukunft werden (Becker, 2022).

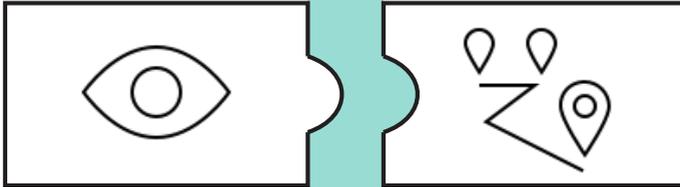
Bei diesem Übergang stellt sich jedoch mehr und mehr heraus, dass die dafür notwendigen Algorithmen sehr komplex sein werden. Es liegt in der Natur der Dinge, dass mit steigender Komplexität gleichzeitig das Verständnis darüber erschwert wird. Zusätzlich gehen Nutzer von AF davon aus, dass das Fahrzeug wie erwartet funktioniert und sind folglich überrascht, wenn eine Fahrfunktion sich nicht so verhält, wie sie vermuten. Daher ist es einerseits wichtig, den Nutzern schon vor der ersten Fahrt mit einem AF ausführlich zu erklären, welche Aufgaben das Fahrzeug in welchen Situationen übernehmen kann und welche Verantwortungsbereiche noch beim Fahrer liegen (Feinauer, Schuller, Groh, Huestegge & Petzoldt, 2022). Doch selbst wenn die Nutzer im Umgang mit AF gut ausgebildet sind, ist davon auszugehen, dass sie Situationen erleben, in denen das AF eine unerwartete Entscheidung trifft. Dies kann beispielsweise ein Fahrmanöver in einer urbanen Situation sein, in der neben vielen Fußgängern, Skateboardern und Fahrradfahrern die Straßenverhältnisse durch schlechte Sicht unübersichtlich sind und eine leere McDonald's-Tüte über die Fahrbahn weht.

Damit der Nutzer in solchen Situationen den Überblick behält und nicht durch das unvorhergesehene Verhalten seines AF verunsichert wird, benötigt es klare und eindeutige Human-Machine-Interface-Konzepte (HMI) (Beggiato et al., 2015).

In der HMI-Entwicklung gibt es unterschiedliche Schwerpunkte, die berücksichtigt werden, z. B. verschiedene Rückmeldemodalitäten (Wald, Haentjes, Albert, Cramer & Bengler, 2021) oder welche Informationen während der Fahrt bereit gestellt werden sollen (Beggiato et al., 2015). Tatsächlich nehmen die meisten dieser Forschungsarbeiten aber an, dass die Informationen, die vom Nutzer eingefordert werden, ohne Probleme zur Verfügung stehen. Praktisch ist es derzeit jedoch so, dass viele Fahrfunktionen die Entscheidungen auf einem technisch detaillierten Niveau berechnen. Dies führt dazu, dass die Informationen, die zur Entscheidung führten, zwar vorhanden sind, jedoch von Personen ohne Fachwissen (Novizen) nicht verstanden werden können – sie benötigen zusätzliche Hilfe bei der Interpretation der technischen Details. Hier offenbart sich eine Lücke zwischen technischer Funktionalität und Verständnis der Fahrfunktionen: Nachvollziehbare Verhaltensentscheidungen sind nötig, um das Gelingen dieser hochkomplexen, innovativen Technologie bis hin zur Einführung auf den Straßen zu unterstützen.

In der realen Welt gehören unsichere Informationen zum Alltag. Dies ist problematisch, wenn auf ihrer Grundlage wichtige Entscheidungen getroffen werden sollen: Meist ist nicht genau klar, welche Qualität die Berechnungen haben werden, d. h. wie sich unsichere Inputs auf die Zuverlässigkeit der Ergebnisse auswirken. Daher stellt diese Arbeit mit der Sensitivitätsanalyse (SA) von Planungsalgorithmen eine Methode vor, die in verschiedenen Verkehrssituationen untersucht, welche Eingaben für die Entscheidungsfindung wichtig sind (Kapitel 3). Dadurch kann aus einer Menge an Inputs eine Teilmenge identifiziert werden, die durch ihre Wahrnehmungsunsicherheit die Berechnung der Planungsfunktion maßgeblich beeinflusst haben und dadurch möglicherweise die Ursache für Fehlentscheidungen sein können. Da dieses Verfahren sampling-basiert ist, ist es rechenaufwändig. In der Regel wird a priori eine feste Anzahl an Stichproben ausgewertet. Um die Stichprobengröße zu reduzieren, stellt Kapitel 4 ein dynamisches Abbruchkriterium (DSC) vor, mit dem diese an die Komplexität der Situation angepasst werden kann: So werden für Situationen, in denen die SA eine eindeutige Klassifizierung bereits nach einer kleinen Stichprobengröße festlegen kann, nur wenige Stichprobenwerte (engl.

Sensitivitätsanalyse nach Morris (1991)
(Kapitel 3, Henze, Faßbender & Stiller,
2020; Henze, Faßbender & Stiller, 2021).



dynamisches Abbruchkriterium (Kapitel 4, Henze, Rußer, Faßbender, Stiller & Schneider, 2022).

Bedarf von HAF-Entwicklern
(Kapitel 6, Henze, Stasinski,
Faßbender & Stiller, 2021).

stochastisches Optimierungsproblem (Kapitel 5, Henze, Faßbender & Stiller, 2021).

Generierung von erläuternden Beschreibungen (Kapitel 7, Henze, Faßbender & Stiller, 2022).

Abbildung 1.1: Aufbau der Arbeit.

samples) ausgewertet, während in Situationen mit uneindeutiger Klassifizierung so lange weitere Stichprobenwerte bestimmt werden, bis die Ergebnisse der SA abgesichert sind.

Mit einer SA kann also untersucht werden, ob das Messrauschen einer Information zu groß ist und zu unsicheren Fehlentscheidungen führt. In diesem Fall gibt es zwei Möglichkeiten, die Ergebnisse der Wahrnehmung und die Verarbeitung in der Planung besser aufeinander abzustimmen: Einerseits ist es möglich, den Planungsalgorithmus robuster zu gestalten, damit er Unsicherheiten besser verarbeiten kann. Andererseits können auch die Informationen z. B. aus der Wahrnehmung oder Prädiktion verbessert werden, indem bessere Sensorik oder genauere Algorithmen eingesetzt werden. Um dafür einschätzen zu können, ob die Anforderungen der Planung an die Wahrnehmung erfüllt werden können, ist es wichtig, obere Schranken für die Unsicherheiten zu bestimmen. Dafür

wird in Kapitel 5 ein stochastisches Optimierungsproblem definiert und gelöst: Es bestimmt für eine Verkehrssituation, welche Unsicherheiten gerade noch akzeptabel sind, damit die Abweichung des Ergebnisses von der zu erwartenden Entscheidung nur geringfügig ist.

Eine möglichst passgenaue Abstimmung der Wahrnehmung auf die Planung ist eine Voraussetzung dafür, dass ein AF seine Verhaltensentscheidung ohne Störeinflüsse treffen kann und seine Entscheidungsfindung nachvollziehbar ist. Darauf aufbauend müssen Mittel und Wege gefunden werden, wie diese Transparenz auch an die Nutzer weitergegeben werden kann. In dieser Arbeit wird eine Methode vorgeschlagen, wie textbasierte Beschreibungen des automatisierten Fahrverhaltens erzeugt werden können. Dafür werden die Informationen, die in die Entscheidungsfindung eingeflossen sind, sowie ihre Effekte analysiert und in eine für den Nutzer leichter verständliche Textform überführt. Dadurch kann das AF seine Entscheidungen selbstständig erläutern. In Kapitel 6 wird zunächst untersucht, welche Anforderungen Entwickler von hochautomatisierten Fahrfunktionen an Erklärungen haben, wenn das AF während einer Testfahrt eine unerwartete Fahrentscheidung trifft. Dabei wird die Annahme getroffen, dass die Erklärungen, die den Entwicklern in ihrem Arbeitsalltag helfen, zusammen mit ihrem Expertenwissen die Grundlage legen können, um auch unwissendere Nutzer von AF zu unterstützen. Daher werden die daraus gewonnenen Erkenntnisse in der automatischen Generierung von erläuternden Beschreibungen umgesetzt: Kapitel 7 beschreibt, wie mithilfe der SA die Informationen ermittelt werden, die für eine Planungsentscheidung relevant waren. Für diese wird dann der Bezug zu der Umgebung und den geltenden Verkehrsregeln hergestellt, sodass schließlich textbasierte Aussagen das Verhalten des AF automatisch beschreiben.

Um diese Arbeit in Zusammenhang mit bestehenden Forschungsarbeiten zu setzen und ihren Beitrag herauszuarbeiten, gibt Kapitel 2 zunächst einen Überblick über relevante Arbeiten zu Unsicherheiten in Wahrnehmung und Planung einerseits und Erklärungen von Algorithmen andererseits.

2 Stand der Wissenschaft und Beiträge der Arbeit

Ein amerikanisches Meinungsforschungsinstitut untersuchte 2017 die Einstellung von US-Amerikanern über AF (Pew Research Center, 2017). Dabei fanden sie heraus, dass 56 % der Befragten nicht mit einem AF fahren würden. 42 % davon gaben an, dass sie kein Vertrauen hätten oder besorgt seien, die Kontrolle abzugeben, und weitere 30 % begründeten ihre Abneigung mit Sicherheitsbedenken. Hartwich, Schmidt, Gräfin und Krems (2020); Shariff, Bonnefon und Rahwan (2021) untersuchten das Vertrauen der Nutzer von AF gegenüber dem von Beifahrern von menschlichen Fahrern. In ihren Studien fanden sie heraus, dass Nutzer derzeit noch ein deutlich höheres Vertrauen in menschliche Fahrer setzen als in von Maschinen gesteuerte Fahrzeuge, selbst wenn beide Fahrten mit demselben Risiko behaftet sind (Shariff et al., 2021) bzw. die exakt selbe Fahrweise zeigen (Hartwich et al., 2020). Die Ethik-Kommission Automatisiertes und Vernetztes Fahren warf die Frage auf, wie viel sicherer AF im Vergleich zu menschlichen Fahrern sein müssen, um gesellschaftlich akzeptiert zu werden (Bundesministerium für Verkehr und digitale Infrastruktur, 2017). Daraus lassen sich zwei zentrale Handlungsfelder ableiten:

- technische Sicherheit (engl. *safety*).
- Maßnahmen, die das Vertrauen der Nutzer unterstützen.

Technische Sicherheit ist eine notwendige Voraussetzung für Vertrauen und ein sehr breit gefächertes Aspekt. In diesem Zusammenhang wird in der vorliegenden Arbeit vorrangig die Auswirkung von Wahrnehmungsunsicherheiten auf die Planungsentscheidung untersucht. Dabei steht zunächst die Frage im Vordergrund, wie groß diese sein dürfen, damit der Planungsalgorithmus noch robust genug ist und zulässige Entscheidungen treffen kann. Und auch wenn aktuelle Ansätze darauf abzielen, das Fahrverhalten eines menschlichen Nutzers möglichst genau abzubilden (Ossig & Cramer, 2020; Ossig, Cramer, Eckl &

Bengler, 2022; You et al., 2019), ist doch vorstellbar, dass sich das zukünftige Fahrverhalten eines AF genauso von der menschlichen Strategie lösen wird, wie sich die Spielstrategie von AlphaGo von der menschlichen Strategie gelöst hat, indem es für Menschen „unkonventionelle“, „überraschende“ oder „unerwartete“ Spielzüge ausführt (Sedol, 2016a, 2016b). Damit ist zu erwarten, dass Situationen entstehen, in denen das Fahrverhalten eines AF die Einschätzung der Nutzer in Bezug auf seine Kompetenz negativ beeinflusst (Wald, Henrich, Albert, Ossig & Bengler, 2022). In solchen Fällen ist es nicht nur wichtig, dass die Entscheidung technisch abgesichert ist, sondern auch, dass der Nutzer in die Entscheidungsfindung einbezogen und mit Erklärungen unterstützt wird (Omeiza, Webb, Jirotko & Kunze, 2022). Aus diesem Grund wird in einem zweiten Schritt eine Methode vorgestellt, wie die wichtigen Aspekte, die zur Fahrentscheidung geführt haben, so aus den Algorithmen extrahiert werden können, dass darauf aufbauende Erklärungen tatsächlich den Entscheidungsprozess beschreiben, ohne zu viele heuristische und damit fehleranfällige Annahmen zu verwenden.

Diese Einführung in die Literatur ist zweigeteilt: Abschnitt 2.1 stellt aktuelle Planungsansätze unter Unsicherheiten und Arbeiten zu Anforderungen zwischen Wahrnehmung und Planung mit Bezug auf Unsicherheiten vor. Abschnitt 2.2 gibt einen Überblick zu Erklärungen im Allgemeinen und von Planungsalgorithmen im automatisierten Fahren im Speziellen.

2.1 Unsicherheiten in der Wahrnehmung und Planung

Die komplexe Aufgabe des automatisierten Fahrens wird typischerweise in die drei Teilbereiche *sense – plan – act* unterteilt (Brooks, 1986; Kortenkamp & Simmons, 2008). Die Wahrnehmung nimmt die Informationen von Sensoren wie z. B. Lidar oder Radar auf (z. B. Krämer, Stiller & Bouzouraa, 2018), verarbeitet diese zusammen mit Informationen aus der Karte (z. B. Eder, Skibinski, Mickler & Ulbrich, 2022) und gibt sie anschließend mit der Prädiktion anderer Verkehrsteilnehmer an das Planungsmodul weiter (z. B. Sackmann, Leemann, Bey, Hofmann & Thielecke, 2021). Dabei gibt es verschiedene Quellen von Unsicherheiten, z. B. bei der Vorhersage des Verhaltens anderer Verkehrsteilnehmer (z. B. Vogl, Sackmann, Kürzinger & Hofmann, 2020) oder bei der

Lokalisierung des AF. So leiten Reid et al. (2019) auf Grundlage von aktuellen Unfallraten im Straßenverkehr Anforderungen an die Lokalisierungsgenauigkeit zukünftiger AF her.

Diese Unsicherheiten können sich über die Verarbeitungskette *sense – plan – act* hinweg fortpflanzen. Daher konzentriert sich ein Zweig der aktuellen Forschung zu Planungsalgorithmen auf Planung unter Unsicherheiten: Tas, Hauser und Stiller (2018) schlagen vor, die Fahrentscheidung des AF aufzuschieben, falls nicht genügend Informationen über die Umgebung abgesichert sind. Eine Möglichkeit dafür ist der Partially Observable Markov Decision Process (POMDP), der unsichere oder noch nicht bekannte Beobachtungen im Planungsprozess berücksichtigen und so mögliche Reaktionen des automatisierten Fahrzeugs (Ego) frühzeitig vorbereiten kann. Hubmann, Schulz, Becker, Althoff und Stiller (2018) stellen in einem Kreuzungsszenario eine neue Roll-out-Heuristik vor, mit deren Hilfe der POMDP online gelöst werden kann. Dafür berücksichtigen sie vor allem Unsicherheiten in der Prädiktion anderer Verkehrsteilnehmer, Messrauschen und der Interaktion zwischen dem Ego und der Umgebung. Darauf aufbauend beschreiben Bey, Tratz, Sackmann, Lange und Thielecke (2020) anhand eines unbekanntes Hindernisses, welche Schwierigkeiten bei der Feinabstimmung eines POMDP berücksichtigt werden müssen.

Faßbender, Heinrich und Wünsche (2016) nutzen ein A*-Planer zur Lösung eines numerischen Optimierungsproblems, um Trajektorien in anspruchsvollen Umgebungen, z. B. Straßen mit engen Kurven, zu berechnen. Dabei berücksichtigen sie zunächst jedoch keine anderen Verkehrsteilnehmer oder andere Unsicherheiten. Daher lösen W. Xu, Pan, Wei und Dolan (2014) ein numerisches Optimierungsproblem, das die unsicheren Trajektorien anderer Verkehrsteilnehmer mithilfe von Kalman-Filtern modelliert. Zusätzlich kann in einem optimierungsbasierten Planer auch die Wahrscheinlichkeit für eine Kollision beschränkt werden: Dafür stellen de Groot, Brito, Ferranti, Gavrila und Alonso-Mora (2021) ein Optimierungsproblem mit stochastischen Nebenbedingungen auf, das die Auswirkungen der unsicheren Wahrnehmungen auf das Planungsergebnis abschätzt.

Einen etwas anderen Ansatz, um die Zuverlässigkeit von Planungsentscheidungen zu charakterisieren, verfolgen Hoel, Driggs-Campbell, Wolff, Laine und Kochenderfer (2020): Sie entwickeln eine Reinforcement-Learning-Methode, die nicht nur eine Planungsentscheidung berechnet, sondern auch angibt, mit wie viel Unsicherheit diese behaftet ist. Schmidt, Kontes, Plinge und Mutschler

(2021) stellen einen weiteren Reinforcement-Ansatz vor, der zunächst sicheres Fahrverhalten bestimmt und darauf aufbauend sichere Entscheidungsbäume extrahiert. Das Ergebnis sind sichere und gleichzeitig interpretierbare Entscheidungsstrategien.

In vielen Ansätzen wird die Unsicherheit in der Wahrnehmung also getrennt von ihrer Verarbeitung in der Planung betrachtet. Stöckle, Utschick, Herrmann und Dirndorfer (2018, 2019) stellen daher ein Optimierungsproblem auf, das für ein Notbremssystem zulässige Unsicherheiten bestimmt, sodass eine Kollision vermieden werden kann. Dabei nutzen sie viele Modellannahmen, z. B. zu Fahrzeugdynamik, die die Modellierung des Messrauschens durch Linearisierung stark vereinfachen. Auch Rohde, Stellet, Mielenz und Zöllner (2015) beschäftigen sich mit der Lokalisierungsgenauigkeit eines AF: Sie weisen daraufhin, dass sich ein AF zur Lokalisierung u. a. an Landmarken orientiert und untersuchen daher, wie die verrauschte Positionswahrnehmung der Landmarken die Genauigkeit der Ego-Positionsschätzung beeinflusst. Sie leiten daraus ab, wie Sensoreigenschaften modellbasiert charakterisiert werden müssen, damit sie die Genauigkeitsanforderungen erfüllen.

Ein weiterer Grund, weshalb Wahrnehmung und Planung in Bezug auf Unsicherheiten aufeinander abgestimmt sein müssen, ist die Validierung der Fahrfunktionen. Philipp, Qian, Hartjen, Schuldt und Howar (2021) stellen dafür eine Methode vor, die aus einer gegebenen Planungskomponente ableitet, wie genau die Sensorinformationen sein müssen. Dafür verrauschen sie bei einem Linksabbiegevorgang mit entgegenkommendem Verkehr die Position des anderen Fahrzeugs mit systemischen Fehlern. Daraus leiten sie für dieses Beispiel konkrete Sensorspezifikationen ab, um eine Kollision zu vermeiden. Mit einem ähnlichen Ziel stellen Piazzoni, Cherian, Slavik und Dauwels (2021) ein Wahrnehmungs-Fehler-Modell vor und untersuchen die Auswirkungen verschiedener Fehlerquellen auf Fehlermetriken wie Anzahl an Kollisionen oder Mindestabstand. Mit ihrem Vorschlag können sie sowohl einflussreiche Fehlerfaktoren als auch Planungsschwachstellen identifizieren. Dabei diskutieren sie ihre Ergebnisse jedoch nur qualitativ und lassen offen, wie eine konkrete Definition von Planungsschwachstellen aussehen könnte.

Abgrenzend zu den beschriebenen Ansätzen liefert die vorliegende Arbeit folgende Beiträge in Bezug auf eine besser abgestimmte Schnittstelle zwischen Wahrnehmung und Planung:

- Definition von zulässigen Planungsentscheidungen: Viele der oben genannten Ansätze verwenden als Fehlermetrik, ob eine Kollision stattfindet. Tatsächlich sollte ein guter Planer aber z. B. auch am Fahrkomfort gemessen werden. Daher wird eine neue Definition von zulässigen Planungsentscheidungen eingeführt, die beurteilt, ob eine Fahrentscheidung starken Schwankungen unterliegt (vgl. Kapitel 3).
- Unterscheidung zwischen einflussreichen und irrelevanten Informationen: Wenn Fahrfehler auftreten, gibt es häufig vielfältige potentielle Fehlerursachen. Mithilfe einer Sensitivitätsanalyse (SA) werden die Informationen identifiziert, die großen Einfluss auf die Planungsentscheidung hatten und damit möglicherweise als Fehlerquelle in Frage kommen (vgl. Kapitel 3 und 4).
- Herleitung von Anforderungen an die Wahrnehmung: In der Entwicklung von Fahrfunktionen sind die Anforderungen, die der Planer an die Wahrnehmung stellt, eine zentrale Schnittstelle, um die Systeme passgenau aufeinander abzustimmen. Mithilfe der zulässigen Planungsentscheidungen wird ein Optimierungsproblem aufgestellt und ein dazu passender numerischer Lösungsansatz vorgeschlagen, der für eine Verkehrssituation aus einem gegebenen Planer die maximal zulässigen Unsicherheiten der Wahrnehmungsinformationen ableitet (vgl. Kapitel 5).

2.2 Erklärungen algorithmischer Entscheidungen

Die Europäische Kommission (2021) legt in ihrem Gesetz über künstliche Intelligenz fest, dass „Hochrisiko-KI-Systeme [. . .] so konzipiert und entwickelt [werden], dass ihr Betrieb hinreichend transparent ist, damit die Nutzer die Ergebnisse des Systems angemessen interpretieren und verwenden können“ (§ 13 Abs. 1 S. 1). Dies bedeutet, dass es allein schon aus rechtlicher Sicht im Interesse von Automobilherstellern liegt, ihre Fahrfunktionen transparent auszulegen. Aus wirtschaftlicher Sicht werden das Nutzervertrauen und die gesellschaftliche Akzeptanz einen großen Einfluss auf den Erfolg von AF (Omeiza et al., 2022) und damit auch einen maßgeblichen Anteil an der Rentabilität der Technik haben. Daher ist es ratsam, schon während des Entwicklungsprozesses

ses die Anforderungen der Nutzer zu berücksichtigen und so die technischen Weichen für transparente Fahrentscheidungen zu stellen.

Es gibt verschiedene Definitionen dafür, was eine Erklärung auszeichnet. Grundsätzlich unterscheiden Doran, Schulz und Besold (2017) zwischen interpretierbaren und verständlichen Systemen: Ersteres sind Systeme, die auf einer technisch-mathematischen Ebene transparent sind, z. B. Regressionsmodelle, aber noch ein gewisses Expertenwissen verlangen. Letzteres bezieht sich auf Systeme, die zusätzlich zum Ergebnis noch Symbole erzeugen, z. B. Wörter oder Visualisierungen, die es ermöglichen, Inputs mit Outputs zu verknüpfen. Für ein bereits gegebenes System, das post-hoc erklärt werden soll, differenzieren Montavon, Bach, Binder, Samek und Müller (2017) zwischen einer Interpretation und einer Erklärung: Eine Interpretation ist die Abbildung eines abstrakten Konzepts in ein Gebiet, das der Mensch verstehen kann, z. B. ein Bild. Eine Erklärung hingegen ist eine Teilmenge dieses Gebiets, die in einem konkreten Beispiel einen Beitrag zur Entscheidungsfindung geleistet hat. Diese Definition trifft also keine Aussagen zu Ursache-Wirkung-Beziehungen. Auch Rudin (2019) führt an, dass zwischen post-hoc Erklärungen von Black-Box-Modellen und inhärent interpretierbaren Modellen unterschieden werden sollte: Erklärungen können immer nur eine Approximation der Funktionsweise sein und daher zu Fehlinterpretationen führen. Daher schlägt sie vor, bei der Formulierung vorsichtig zu sein und statt von „Erklärungen“ von „Erläuterungen“ zu sprechen, wenn die Funktionsweise eines Algorithmus konkret beschrieben wird. Im restlichen Teil dieses Abschnitts wird der Begriff „erklären“ in der Bedeutung des Dudens verwendet: „so erläutern, dass der bzw. die andere die Zusammenhänge versteht“ (Dudenredaktion, 2022).

Im automatisierten Fahren haben Erklärungen verschiedene Dimensionen (Omeiza et al., 2022). Diejenigen, die für die Erklärung des Fahrverhaltens am wichtigsten sind, werden im Folgenden kurz erläutert.

Empfänger

Gute Erklärungen berücksichtigen in ihrem Aufbau u. a. den Empfänger (Samek & Müller, 2019). Im Zusammenhang mit AF werden für gewöhnlich drei potentielle Empfängergruppen genannt (Omeiza et al., 2022): Nutzer, Entwickler und Regulierungsbehörden. Es wird die Aufgabe von Regulierungsbehörden

sein, zu überprüfen, ob sich ein Fahrzeug regelkonform verhalten hat oder wie die Schuldfrage nach einem Unfall geklärt werden kann. Nolte, Ernst, Richelmann und Maurer (2018) geben dabei zu bedenken, dass die Erklärungen einer Handlung des AF, z. B. die Rechtfertigung nach einem Unfall, nicht auf technisch-mathematischer Ebene gegeben werden sollte, sondern auf einem abstrakteren Level, z. B. auf der taktischen Planungsebene. Es ist zu erwarten, dass dies verschiedenste juristische Fragen umfassen wird, z. B. die DSGVO-konforme Datenspeicherung (Vogel & Althoff, 2020), und ist damit nicht Teil dieser Arbeit.

Für Entwickler, die zumeist in großen Teams an den komplexen Systemen arbeiten, ist es im Entwicklungsprozess zunehmend schwieriger, auftretende Fehlfunktionen im Detail zu verstehen. Nahata, Omeiza, Howard und Kunze (2021) stellen daher einen Ansatz mit Supervised Learning vor, der die Wahrscheinlichkeit für eine Kollision vorhersagt und gleichzeitig die Informationen extrahiert, die zu diesem Risiko führen. Die dabei entstehenden Aussagen sind sehr detailliert und erfordern ein genaues Verständnis der Funktionalität. Daher sind diese noch nicht geeignet, um Nutzer ohne Fachwissen in ihrem Verständnis zu unterstützen. Im Vergleich dazu lassen Borgo, Cashmore und Magazzeni (2018) den Nutzer eine Alternative vorschlagen und vergleichen diese anschließend mit dem generierten Plan um zu zeigen, dass der Roboter-Plan nicht schlechter als der Plan des Menschen ist. Damit integrieren sie den Nutzer in den Erklärungsprozess. Sie schlagen außerdem vor, dass nur die Entscheidungen von Nutzern hinterfragt und damit erklärt werden müssen, die nicht ihren Erwartungen entsprechen. Dieses Vorgehen verlangt eine sehr klar formulierte Mensch-Maschine-Schnittstelle, bei der der Vorschlag des Menschen vom AF nicht fehlinterpretiert werden kann.

Zeitpunkt

Erklärungen können angezeigt werden, bevor eine Fahrentscheidung umgesetzt wird, oder nachdem das Verhalten beobachtet wurde (Koo et al., 2015). Für vorausschauende Ankündigungen gibt es verschiedene Möglichkeiten, z. B. schlagen Wald et al. (2021) vor, eine Wankbewegung des AF als Ankündigung für einen Fahrstreifenwechsel (FSW) zu nutzen. Koo et al. (2015) untersuchen für eine SAE-Level 2 Notbremsfunktion (On-Road Automated Driving (ORAD) Committee, 2021), wie eine Erklärung kurz vor einem Notbremsmanöver

gestaltet sein sollte. Da in dieser Arbeit angenommen wird, dass Erklärungen von Nutzern vorrangig dann in Anspruch genommen werden, wenn das Fahrverhalten nicht ihren Erwartungen entspricht (Borgo et al., 2018), werden im weiteren Verlauf nur post-hoc Erklärungen betrachtet.

Inhalt

Im Kontext des automatisierten Fahrens werden verschiedene Fragetypen untersucht: *was?*, *warum?*, *warum nicht?*, *was und warum?*, *was wäre, wenn?* (z. B. Fox, Long & Magazzeni, 2017; Koo et al., 2015; Omeiza, Kollnig, Web, Jirotko & Kunze, 2021). Daneben gibt es verschiedene weitere Fragen, die vom Nutzer gestellt werden können, z. B. *Warum ist dieser Plan effizienter/sicherer/. . . als ein anderer Plan?* (Borgo et al., 2018; Fox et al., 2017; Korpan & Epstein, 2018). Je nachdem, welche Frage eine Erklärung adressiert, haben die Informationen verschiedene Auswirkungen auf den Nutzer: In verschiedenen Studien und in unterschiedlichen Anwendungsbereichen konnte gezeigt werden, dass Erklärungen zu *was und warum?* bzw. *was und warum nicht?* sowohl das Verständnis als auch das Vertrauen ins System verbessern (Goldman & Bustin, 2022; Lim, Dey & Avrahami, 2009; Omeiza, Kollnig et al., 2021). Daher liegt der Fokus dieser Arbeit auf der Beantwortung der Frage *was und warum?*.

Neben den oben erwähnten Ansätzen für die Generierung von Erklärungen speziell für Entwickler oder Nutzer (Borgo et al., 2018; Nahata et al., 2021) gibt es noch weitere Vorschläge, die weniger stark an die konkreten Anforderungen einzelner Empfänger angepasst sind: Wang, Pynadath und Hill (2016) stellen vor, wie textbasierte Erklärungen von Entscheidungen, die mithilfe von POMDPs getroffen wurden, generiert werden können. Kim, Rohrbach, Darrell, Canny und Akata (2018) trainieren die Ausgabe von textbasierten Erklärungen auf der Basis von mehreren Kamerabildern, die mit menschlichen Erklärungen versehen sind. Der Nachteil ihres Vorschlags ist, dass die Erklärungen auf den Annahmen über das Verhalten basieren, die die Menschen während der Beschreibung der Trainingsdaten treffen, und nicht darauf, wie das System seine Entscheidung tatsächlich getroffen hat. Diese Trennung zwischen Entscheidungsprozess und Erklärungsgenerierung ist problematisch, weil sie z. B. in Corner Cases zu fehlerhaften Erklärungen des Systemverhaltens führen kann. Omeiza, Webb, Jirotko und Kunze (2021) bauen Entscheidungsbäume auf, die aus Verkehrsregeln, Beobachtungen und Aktionen bestehen, und leiten daraus

Aussagen ab, die das Verhalten des AF erklären. Dieser Ansatz ist sehr leicht auf weitere Anwendungsfälle und zusätzliche Inhalte erweiterbar. Auch er birgt jedoch den Nachteil, dass Corner Cases vergessen werden können und dann Erklärungen entweder fehlerhaft sind oder nicht existieren.

Im Vergleich zu diesen Arbeiten liefert die vorliegende Arbeit die nachfolgenden Beiträge:

- Expertenwissen und Expertenfragen als Ausgangspunkt: Wenn alle Fragen zum Verhalten von AF, die Experten stellen, beantwortet werden können, so sind auch alle Informationen vorhanden, die für die Erklärung des Verhaltens gegenüber Nutzern ohne Fachwissen benötigt werden. Daher wird mithilfe einer Fokusgruppenstudie untersucht, welchen Informationsbedarf Entwickler in Bezug auf unerwartete Verhaltensentscheidungen eines AF haben und wie dieser gedeckt werden kann.
- Generierung von Erläuterungen auf Basis des Planers: Gute Erklärungen sind inhaltlich auf den Empfänger abgestimmt. Daher wird eine Methode vorgestellt, die mithilfe der SA erläuternde Beschreibungen erzeugt, die in ihrem Informationsgehalt leicht auf die Bedürfnisse des Empfängers abgestimmt werden können. Im Vergleich zu existierenden Ansätzen basieren diese lediglich auf dem Planungsalgorithmus und den geltenden Verkehrsregeln und beschreiben das Verhalten, ohne von menschlichen Fehlinterpretation beeinflusst zu werden.

3 Analyse von unsicherheitsbehafteten Informationen

Der Literaturüberblick hat gezeigt, dass es in der Forschung verschiedene Ansätze zu Planungsalgorithmen gibt, die Unsicherheiten berücksichtigen. Bei der Evaluierung eines Planers wird häufig das Auftreten von Kollisionen überprüft, was jedoch nur einen Teil einer guten Planungsentscheidung beschreibt und nichts über den Einfluss von Unsicherheiten auf die Zuverlässigkeit der Entscheidung aussagt. Daher stellt dieses Kapitel einen neuen Ansatz vor, der den Einfluss von unsicherheitsbehafteten Informationen auf die Planungsentscheidung charakterisiert. Dafür führt Abschnitt 3.1 eine Definition für zulässige Planungsergebnisse ein, die die Qualität von Verhaltensentscheidungen unter Unsicherheiten misst. Die dafür notwendigen stochastischen Grundlagen werden analog zu Klenke (2013) verwendet. Abschnitt 3.2 stellt zwei Abschätzungen vor, um die Zulässigkeit von Planungsergebnissen zu überprüfen: Für Entscheidungen mit kontinuierlichem Ergebnisraum, z. B. die Berechnung der Beschleunigung, wird die Chebychev-Saw-Ungleichung verwendet (vgl. Abschnitt 3.2.1), während für Entscheidungen zwischen zwei Homotopieklassen, z. B. FSW nach links oder nicht, die Bernoulli-a-posteriori-Abschätzung vorgestellt wird (vgl. Abschnitt 3.2.2). Für den Fall, dass eine Fahrentscheidung als nicht zuverlässig eingestuft wird, wird in Abschnitt 3.3 die SA nach Morris vorgestellt, mit deren Hilfe die Informationen identifiziert werden, deren Unsicherheiten die Fahrentscheidung am stärksten beeinflusst haben.

Vorläufige Ergebnisse wurden in Henze et al. (2020); Henze, Faßbender und Stiller (2021) veröffentlicht. Abbildungen, Tabellen und andere Ergebnisse wurden für eine konsistente Darstellung angepasst.

3.1 Modellierung zulässiger Unsicherheiten

Typischerweise nutzt das automatisierte Fahrzeug zur Planung der Fahrentscheidung eine Umgebungsbeschreibung, die Informationen aus der Wahrnehmung, Karte oder Prädiktion enthält. Mathematisch kann ein solcher Planungsalgorithmus durch eine Funktion modelliert werden, wie die folgende Definition beschreibt.

Definition 3.1.1 (Planungs- und Umgebungsbeschreibung). Seien $\Omega \subseteq \mathbb{R}^k$, $k, m \in \mathbb{N}$. Der Vektor $\mathbf{q} \in \Omega$ heißt Umgebungsbeschreibung und die Funktion $\mathbf{f} : \Omega \rightarrow \mathbb{R}^m$, $\mathbf{q} \mapsto \mathbf{f}(\mathbf{q})$ heißt Planungsfunktion für einen Planungsalgorithmus mit k Inputs \mathbf{q} und m Outputs $\mathbf{f}(\mathbf{q})$.

Eine Planungsfunktion bildet also die Umgebungsbeschreibung \mathbf{q} auf die Planungsentscheidung $\mathbf{f}(\mathbf{q})$ ab. Die Inputs und Outputs der Planungsfunktion können verschiedene Informationen beinhalten, wie das folgende Beispiel zeigt:

Beispiel 3.1.2 (Umgebungsbeschreibung). Der Inputvektor $\mathbf{q} \in \Omega$ kann verschiedene Informationen beinhalten, z. B. den Zustand $(\mathbf{x}, \mathbf{v}, \mathbf{a})^\top$ des Egos mit der Position $\mathbf{x} = (x_x, x_y)^\top$, der Geschwindigkeit $\mathbf{v} = (v_{\text{long}}, v_{\text{lat}})^\top$ und der Beschleunigung $\mathbf{a} = (a_{\text{long}}, a_{\text{lat}})^\top$. Zur Vereinfachung der Notation wird im Rest der Arbeit die Notation $v := v_{\text{long}}$ und $a := a_{\text{long}}$ für die longitudinale Geschwindigkeit und Beschleunigung des Egos verwendet. Außerdem bezeichnet p die Position des Egos auf der Fahrstreifenmitte. Zusätzlich umfasst der Vektor \mathbf{q} die Beobachtungen anderer Objekte, z. B. die Distanz d_{vs} zum vorausfahrenden Fahrzeug auf demselben Fahrstreifen und dessen longitudinale Geschwindigkeit v_{vs} .

Beispiel 3.1.3 (Planungsergebnis). Die Ergebnisse $\mathbf{f}(\mathbf{q}) = (f_1(\mathbf{q}), \dots, f_o(\mathbf{q}), \dots, f_m(\mathbf{q}))^\top$ können z. B. eine Beschleunigung $o = a$ mit Werten in \mathbb{R} oder einen FSW $o = \text{lc}$ mit Werten in $\{1, 0, -1\} \simeq \{\text{links, keiner, rechts}\}$ als Output anzeigen.

Zu bemerken ist, dass Definition 3.1.1 keine Voraussetzungen an den Planungsalgorithmus stellt. Daher können die in den folgenden Kapiteln beschriebenen Methoden für jeden Planungsansatz verwendet werden, z. B. für optimierungs-

(Ajanovic, Lacevic, Shyrokau, Stolz & Horn, 2018; Bey et al., 2019) oder sampling-basierte Planer (Bey et al., 2020; Hubmann, Becker, Althoff, Lenz & Stiller, 2017).

Beispiel 3.1.4 (Planungsalgorithmus). Das Intelligent Driver Model (IDM) nach Treiber, Hennecke und Helbing (2000) ist ein Fahrzeugfolgemodell, das zu einem Zeitpunkt $t \in \mathbb{R}_0^+$ die longitudinale Geschwindigkeit eines Fahrzeugs an die Distanz d_{vs} des vorausfahrenden Fahrzeugs und dessen Geschwindigkeit v_{vs} anpasst. Zum Zeitpunkt $t \in \mathbb{R}_0^+$ berechnet es die Beschleunigung $a_t \in \mathbb{R}$ durch¹

$$a_t = \bar{a} \left(2 - \left(\frac{v_t}{v_w} \right)^\delta - \left(\frac{d_w(v_t, \Delta v_t(v_t, v_{vs}, t))}{\min(d_{vs}, d_w(v_t, \Delta v_t(v_t, v_{vs}, t)))} \right)^2 \right), \quad (3.1)$$

mit der maximalen komfortablen Beschleunigung $\bar{a} \in \mathbb{R}^+$, der longitudinalen Ego-Geschwindigkeit $v : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$, der Wunschgeschwindigkeit $v_w \in \mathbb{R}^+$, dem freien Beschleunigungsexponenten $\delta \in \mathbb{R}^+$, der Geschwindigkeitsdifferenz $\Delta v : \mathbb{R}_0^+ \times \mathbb{R}_0^+ \times \mathbb{R}_0^+ \rightarrow \mathbb{R}$, $\Delta v_t := \Delta v_t(v_t, v_{vs}, t) = v_t - v_{vs, t}$ zwischen dem Ego und dem vorausfahrenden Fahrzeug und der aktuellen Distanz $d_{vs} : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$ zum vorausfahrenden Fahrzeug. Der gewünschte Sicherheitsabstand $d_w : \mathbb{R}_0^+ \times \mathbb{R} \rightarrow \mathbb{R}^+$ zum vorausfahrenden Fahrzeug hängt von der aktuellen Geschwindigkeit v_t sowie der Geschwindigkeitsdifferenz Δv_t zwischen dem Ego und dem vorausfahrenden Fahrzeug ab:

$$d_w(v_t, \Delta v_t) = d_{\text{Stau}} + \max \left(0, v_t \Delta T + \frac{v_t \Delta v_t}{2\sqrt{\bar{a}b}} \right).$$

Dabei werden der minimale Abstand im Stau $d_{\text{Stau}} \in \mathbb{R}^+$, die gewünschte Zeitlücke $\Delta T \in \mathbb{R}^+$ und die gewünschte Verzögerung $b \in \mathbb{R}^+$ berücksichtigt.

¹In dieser Arbeit wird für zeitabhängige Funktionen $f : \mathbb{R}_0^+ \rightarrow \mathbb{R}$ die abkürzende Schreibweise $f_t := f(t)$ verwendet.

Die vorgestellte Formulierung ist eine leichte Modifikation des Originals von Treiber et al. (2000), da dieses für $v_t = v_w$ eine negative Beschleunigung

$$a_t = \underset{>0}{\bar{a}} \left(1 - \underbrace{\left(\frac{v_w}{v_w} \right)^\delta}_{=1} - \underbrace{\left(\frac{d_w(v_w, \Delta v_t)}{d_{vs,t}} \right)^2}_{>0} \right) < 0$$

berechnet und das Ego daher bremst, auch wenn das vorausfahrende Fahrzeug sehr weit entfernt ist. Wenn stattdessen $d_{vs,t}$ im Nenner des zweiten Bruchs durch das Minimum $\min(d_{vs,t}, d_w(v_t, \Delta v_t))$ ersetzt wird, wird jedes Fahrzeug, das weiter als der gewünschte Sicherheitsabstand d_w entfernt ist, ignoriert. Das leicht veränderte Modell (3.1) wird so zu $a_t = 0 \text{ m/s}^2$ für $v_t = v_w$ und $d_{vs,t} \geq d_w(v_w, \Delta v_t)$, d. h. die Geschwindigkeit v_w wird gehalten und das Ego bremst nicht. Abschließend werden die Geschwindigkeit und Position des Fahrzeugs entsprechend der kinematischen Differentialgleichungen $\dot{v}_t = a_t$ und $\dot{p}_t = v_t$ bestimmt, z. B. durch numerische Integration.

Durch seine einzelne Gleichung (3.1) zur Berechnung der Beschleunigung ist das IDM ein sehr simples, aber effizientes Modell, das zur Vorhersage von Verhaltensentscheidungen verwendet werden kann (Sackmann, Bey, Hofmann & Thielecke, 2020). Auch wenn es Planungsansätze gibt, die für den Einsatz im Realverkehr besser geeignet sind, wird es in dieser Arbeit verwendet, weil es gut nachvollziehbar ist und schnell ausgewertet werden kann. Diese Eigenschaften sind wichtig, um die Funktionsweise der im Folgenden vorgestellten Ideen zu verdeutlichen.

Im Realverkehr gibt es verschiedenste Ursachen für Unsicherheiten, z. B. aufgrund von Messrauschen der Sensoren, der Schätzung von Messgrößen oder der Prädiktion anderer Verkehrsteilnehmer (vgl. Abschnitt 2.1). Diese verschlechtern für gewöhnlich die Qualität der Umgebungsbeschreibung, die als Basis für die Planungsentscheidung dient. Im weiteren Verlauf der Arbeit wird jedoch nicht zwischen den einzelnen Ursachen unterschieden und Informationen, die an den Planungsalgorithmus weitergegeben werden, werden allgemein als Informationen aus der Wahrnehmung bezeichnet. Daraus abgeleitet bezeichnet der Begriff *Messunsicherheiten* die Unsicherheiten, die aus verschiedenen Informationsquellen resultieren können. Wie im weiteren Verlauf deutlich

wird, ist dies jedoch keine Einschränkung, da die vorgestellten Methoden aufgrund der schwachen Annahmen auch auf andere als die hier vorgestellten Messunsicherheiten erweitert werden können.

In der Messtechnik gibt es unterschiedliche Ansätze zur Modellierung derartiger Messabweichungen, z. B. mithilfe von Toleranzen. In dieser Arbeit werden sie stochastisch beschrieben.

Definition 3.1.5 (stochastische Beschreibung von Messabweichungen). Sei eine Messgröße $\mathbf{q} \in \Omega$ bekannt und ihre Messabweichung gegeben durch eine Zufallsvariable $\hat{\mathbf{Q}}$ auf dem Wahrscheinlichkeitsraum (W-Raum) $(\Omega, \mathcal{A}, \mathbb{P})$ mit der σ -Algebra \mathcal{A} über Ω und dem Wahrscheinlichkeitsmaß (W-Maß) $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$. Dann heißt die Zufallsvariable $\mathbf{q} + \hat{\mathbf{Q}}$ Messergebnis der Messgröße \mathbf{q} mit \mathbb{P} -verteilter Messabweichung $\hat{\mathbf{Q}}$.

Die als Zufallsvariable (ZV) aufgefasste Messabweichung $\hat{\mathbf{Q}}$ kann durch verschiedene W-Maße beschrieben werden. In der Messtechnik wird für $\Omega = \mathbb{R}$ häufig die Normalverteilung verwendet, sodass diese Modellierung auch in dieser Arbeit für kontinuierliche Inputs genutzt wird.

Beispiel 3.1.6 (Normalverteilung). Die eindimensionale Normalverteilung einer ZV X ist für gegebenen Erwartungswert $\mathbb{E}(X) \in \mathbb{R}$ und Varianz $\text{Var}(X) \in \mathbb{R}^+$ durch die Wahrscheinlichkeitsdichte (W-Dichte) $f_{\mathcal{N}(\mathbb{E}(X), \text{Var}(X))} : \Omega \rightarrow [0, 1]$ eindeutig bestimmt durch

$$f_{\mathcal{N}(\mathbb{E}(X), \text{Var}(X))}(x) = \frac{1}{\sqrt{2\pi\text{Var}(X)}} \exp\left(-\frac{(x - \mathbb{E}(X))^2}{2\text{Var}(X)}\right).$$

Wenn der Erwartungswert $\mathbb{E}(X) = 0$ gegeben ist, so heißt die ZV X zentriert (Klenke, 2013). Für eine normalverteilte, mehrdimensionale ZV mit paarweise stochastisch unabhängigen Komponenten schreiben wir den Erwartungswert $\mathbb{E}(\mathbf{X}) = \mathbf{0} \in \mathbb{R}^k$ und die Varianz $\text{Var}(\mathbf{X}) = \text{diag}(\text{Var}(X_1), \dots, \text{Var}(X_k)) \in \mathbb{R}_0^{+k \times k}$. Die W-Dichte einer normalverteilten Komponente $i = 1, \dots, k$ wird mit dem Erwartungswert $\mathbb{E}(X_i) = 0$ und der Varianz $\text{Var}(X_i) \in \mathbb{R}^+$ analog zu oben definiert.

Die Beschreibung der Messabweichungen durch andere Wahrscheinlichkeitsverteilungen (W-Verteilungen) oder für diskrete Größen ist auch möglich, da

keine weiteren Annahmen an die zugrundeliegende W -Verteilung vorausgesetzt werden, erfordert jedoch weiteren Modellierungsaufwand. Um nun Messergebnisse der Umgebungsbeschreibung zu simulieren, wird zunächst angenommen, dass die Messgröße $\mathbf{q} \in \Omega \subseteq \mathbb{R}^k$ bekannt ist. Die Messabweichung wird durch die ZV $\hat{\mathbf{Q}}$ als zentrierte, normalverteilte ZV mit Erwartungswert $\mathbb{E}(\hat{\mathbf{Q}}) = \mathbf{0}$ modelliert. Indem die Summe $\mathbf{q} + \hat{\mathbf{q}}$ aus der bekannten Messgröße \mathbf{q} und einer Realisierung $\hat{\mathbf{q}}$ der ZV $\hat{\mathbf{Q}}$ gebildet wird, kann die Planungsfunktion \mathbf{f} sowohl für die Messgröße \mathbf{q} als auch das Messergebnis $\mathbf{q} + \hat{\mathbf{q}}$ ausgewertet werden. Dies liefert den entscheidenden Vorteil der Simulation gegenüber Experimenten in der realen Welt, da nur in der Simulation die Messgröße vollständig bekannt ist und dadurch Vergleiche zwischen den Planungsentscheidungen basierend auf der Messgröße $\mathbf{f}(\mathbf{q})$ und dem Messergebnis $\mathbf{f}(\mathbf{q} + \hat{\mathbf{q}})$ möglich sind. Da die Planungsentscheidung $\mathbf{f}(\mathbf{q} + \hat{\mathbf{Q}})$ auf Grundlage der ZV $\hat{\mathbf{Q}}$ getroffen wird, bildet diese ebenfalls eine weitere ZV auf \mathbb{R}^m . Ihr W -Maß ist durch das Bildmaß gegeben:

Definition 3.1.7 (angelehnt an Klenke (2013), Def. 1.98 Bildmaß). Seien (Ω, \mathcal{A}) und (Ω', \mathcal{A}') Messräume und \mathbb{P} ein W -Maß auf (Ω, \mathcal{A}) . Außerdem sei $X : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$ eine ZV. Das durch

$$\mathbb{P}_X : \mathcal{A}' \rightarrow [0, \infty], \quad A' \mapsto \mathbb{P}_X(A') := \mathbb{P}(X^{-1}(A'))$$

definierte Maß auf (Ω', \mathcal{A}') heißt Bildmaß von \mathbb{P} unter X .

Wie in Definition 3.1.1 beschrieben, werden keine zusätzlichen Annahmen an die Planungsfunktion \mathbf{f} getroffen, sodass sich auf dieser Grundlage zunächst keine Aussagen über die W -Verteilung der ZV $\mathbf{f}(\mathbf{q} + \hat{\mathbf{Q}})$ ableiten lassen. Es sind also in der Regel keine Aussagen darüber möglich, wie stark die Fahrentscheidung unter Unsicherheiten von der in einer Welt mit perfekten Informationen abweicht. Das folgende Beispiel zeigt jedoch, dass die Formulierung der W -Verteilung bereits für einen einfachen, bekannten Planungsalgorithmus komplex werden kann.

Beispiel 3.1.8.a. In Beispiel 3.1.4 wurde das Längsführungsmodell IDM eingeführt, das die Beschleunigung eines AF bestimmt. Wenn dieses ohne Einwirkung anderer Fahrzeuge geradeaus fährt (vgl. Abb. 3.1), so reduziert sich die Umgebungsbeschreibung \mathbf{q} auf die longitudinale Geschwindigkeit $q = v$

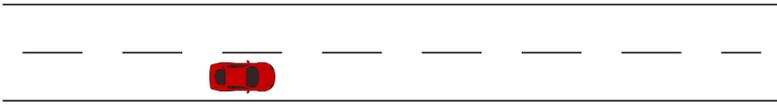


Abbildung 3.1: Freifahrt. Das rote Ego fährt ohne den Einfluss anderer Verkehrsteilnehmer geradeaus mit einer Geschwindigkeit $v = 25 \text{ km/h}$ und hat die Wunschgeschwindigkeit $v_w = 30 \text{ km/h}$. Zu diesem Zeitpunkt beschleunigt es daher mit $a = 0.7766 \text{ m/s}$.

des Egos ($k = 1$). Damit vereinfacht sich die Berechnung der Beschleunigung (3.1) mit $\Delta v_t = 0$ und $s_t = \infty$ zu

$$f_a(v_t) = a_t = \bar{a} \left(2 - \left(\frac{v_t}{v_w} \right)^\delta - 1 \right) = \bar{a} \left(1 - \left(\frac{v_t}{v_w} \right)^\delta \right). \quad (3.2)$$

Für $\delta \in \mathbb{R}^+$ ist die Beschleunigung $f_a(v_t) \leq \bar{a}$ für alle v_t nach oben durch die maximale komfortable Beschleunigung \bar{a} beschränkt. Die normalverteilte ZV $\hat{V} \sim \mathcal{N}(0, \sigma_v^2)$ mit Erwartungswert $\mathbb{E}(\hat{V}) = 0$ und Varianz $\mathbb{V}\text{ar}(\hat{V}) = \sigma_v^2$ beschreibt die Messabweichung der Ego-Geschwindigkeit v . Sei \hat{v} eine Realisierung der ZV \hat{V} . Dann gilt

$$f_a(v_t + \hat{v}) = \bar{a} \left(1 - \left(\frac{v_t + \hat{v}}{v_w} \right)^\delta \right) \leq C \Leftrightarrow \hat{v} \geq \sqrt[\delta]{1 - \frac{C}{\bar{a}}} v_w - v_t$$

und für die W-Verteilung $\Phi_{f_a} : \Omega \rightarrow [0, 1]$ der ZV $f_a(v_t + \hat{V})$ ergibt sich

$$\Phi_{f_a}(C) = \mathbb{P}(f_a(v_t + \hat{v}) \leq C) \quad (3.3)$$

$$\begin{aligned} &= \mathbb{P}\left(\hat{v} \geq \sqrt{\delta} \sqrt{1 - \frac{\min(C, \bar{a})}{\bar{a}}} v_w - v_t\right) \\ &= 1 - \mathbb{P}\left(\hat{v} < \sqrt{\delta} \sqrt{1 - \frac{\min(C, \bar{a})}{\bar{a}}} v_w - v_t\right) \\ &= 1 - \int_{-\infty}^{\sqrt{\delta} \sqrt{1 - \frac{\min(C, \bar{a})}{\bar{a}}} v_w - v_t} \frac{1}{\sqrt{2\pi\sigma_v^2}} \exp\left(-\frac{(\hat{v} - \mathbb{E}(\hat{V}))^2}{2\sigma_v^2}\right) d\hat{v} \\ &= 1 - \frac{1}{2} \left(\operatorname{erf}\left(\frac{\sqrt{\delta} \sqrt{1 - \frac{\min(C, \bar{a})}{\bar{a}}} v_w - v_t}{\sqrt{2\sigma_v^2}}\right) + 1 \right). \end{aligned} \quad (3.4)$$

Die letzte Gleichung folgt mit dem Erwartungswert $\mathbb{E}(\hat{V}) = 0$. Die Ableitung einer W-Verteilung Φ_X ist die W-Dichte ϕ_X der ZV X (Klenke, 2013, Def. 1.106), sodass

$$\begin{aligned} \phi_{f_a}(C) &= \frac{d\Phi_{f_a}}{dC}(C) \\ &= \begin{cases} \frac{v_w}{\sqrt{2\pi\sigma_v^2\bar{a}\delta}} \exp\left(-\left(\frac{\sqrt{\delta} \sqrt{1 - \frac{C}{\bar{a}}} v_w - v_t}{\sqrt{2\sigma_v^2}}\right)^2\right) \left(1 - \frac{C}{\bar{a}}\right)^{\frac{1-\delta}{\delta}} & C \leq \bar{a} \\ 0 & C > \bar{a} \end{cases} \end{aligned}$$

gilt. Die obere Schranke \bar{a} der Beschleunigung während einer Freifahrt (vgl. Gleichung (3.2)) erklärt die Fallunterscheidung $C \leq \bar{a}$: Da für alle Geschwindigkeiten v_t die Beschleunigung $f_a(v_t)$ kleiner ist als die maximale komfortable Beschleunigung \bar{a} , verschwindet die Wahrscheinlichkeit $\mathbb{P}(f_a(v_t + \hat{v}) > \bar{a})$ und daher gilt auch für die W-Dichte $\phi_{f_a}(C) = 0$ für $C > \bar{a}$.

Dieses einfache Beispiel einer Längsführung mit nur einem Input $k = 1$ verdeutlicht, dass die analytische Beschreibung der W-Dichte eines Planungsalgorithmus komplex sein kann. Die Herleitung von W-Dichten in Verkehrssituationen mit weiteren Inputs und mit komplexeren Planungsalgorithmen ist für Planer ohne geschlossene Formulierung unter Umständen unmöglich,

z. B. für Planer mit numerischer Optimierung. Dennoch ist es wichtig, die Unsicherheiten der Planungsergebnisse, die durch Messabweichungen entstehen, zu betrachten. Im Idealfall soll die Fahrentscheidung möglichst robust gegenüber Messabweichungen sein. Daher wird angenommen, dass für jede Fahrentscheidung f_o ($o = 1, \dots, m$) eine obere Schranke \bar{b}_o existiert, sodass zwei Fahrentscheidungen $f_o(\mathbf{q}_1)$, $f_o(\mathbf{q}_2)$ ähnlich sind, wenn

$$|f_o(\mathbf{q}_1) - f_o(\mathbf{q}_2)| \leq \bar{b}_o$$

für alle $o = 1, \dots, m$ gilt. Da die Fahrentscheidung $\mathbf{f}(\mathbf{q} + \hat{\mathbf{Q}})$ jedoch stochastisch ist, können wir diese Ungleichung nur für jeweils zwei Realisierungen des Messergebnis $\mathbf{q} + \hat{\mathbf{Q}}$ überprüfen. In einer Simulationsumgebung ist es am naheliegendsten, eine Realisierung von $\mathbf{f}(\mathbf{q} + \hat{\mathbf{Q}})$ mit der „wahren“ Entscheidung $\mathbf{f}(\mathbf{q})$ zu vergleichen, also $|f_o(\mathbf{q} + \hat{\mathbf{q}}) - f_o(\mathbf{q})|$. Da das Ergebnis $\mathbf{f}(\mathbf{q})$ jedoch in der Realität unbekannt ist, ist es besser, eine Realisierung mit der unter den gegebenen Unsicherheiten zu erwartenden Fahrentscheidung $\mathbb{E}(\mathbf{f}(\mathbf{q} + \hat{\mathbf{Q}}))$ zu vergleichen, also $|f_o(\mathbf{q} + \hat{\mathbf{q}}) - \mathbb{E}(f_o(\mathbf{q} + \hat{\mathbf{Q}}))|$. Dies stimmt mit der ISO-Richtlinie 5725-1:1994 überein (International Organization for Standardization, 1994): Darin wird Genauigkeit (engl. *accuracy*) als Grad der Übereinstimmung zwischen Testergebnis ($f_o(\mathbf{q}_1)$) und akzeptiertem Referenzwert definiert. Als akzeptierter Referenzwert kann z. B. der Erwartungswert einer ZV verwendet werden ($\mathbb{E}(\mathbf{f}(\mathbf{q} + \hat{\mathbf{Q}}))$), weil er theoretisch hergeleitet werden kann. Damit werden nun zulässige Planungsentscheidungen definiert:

Definition 3.1.9 (zulässige Planungsentscheidungen). Sei $\mathbf{f} : \Omega \rightarrow \mathbb{R}^m$, $\Omega \subseteq \mathbb{R}^k$, ein Planungsalgorithmus, $\mathbf{q} \in \Omega$ eine Messgröße und $\hat{\mathbf{Q}}$ die als ZV aufgefasste Messabweichung mit Erwartungswert $\mathbb{E}(\hat{\mathbf{Q}}) = \mathbf{0}$ und Varianzen $\text{Var}(\hat{\mathbf{Q}}) = (\sigma_1^2, \dots, \sigma_k^2)^\top \in \mathbb{R}^{+k}$. Seien außerdem zulässige Abweichungen $\bar{\mathbf{b}} = (\bar{b}_1, \dots, \bar{b}_m)^\top \in \mathbb{R}^{+m}$ und Wahrscheinlichkeiten $\bar{p} \in [0, 1]^m$ gegeben. Dann heißt eine Planungsentscheidung $f_o(\mathbf{q} + \hat{\mathbf{Q}})$ zulässig, wenn sie sich mit Wahrscheinlichkeit \bar{p}_o um nicht mehr als \bar{b}_o von der zu erwartenden Entscheidung $\mathbb{E}(f_o(\mathbf{q} + \hat{\mathbf{Q}}))$ unterscheidet, d. h. wenn gilt

$$\mathbb{P}_{f_o(\mathbf{q} + \hat{\mathbf{Q}})} \left(|f_o(\mathbf{q} + \hat{\mathbf{q}}) - \mathbb{E}(f_o(\mathbf{q} + \hat{\mathbf{Q}}))| > \bar{b}_o \right) \leq \bar{p}_o. \quad (3.5)$$

Dabei sei das W-Maß $\mathbb{P}_{f_o(\mathbf{q} + \hat{\mathbf{q}})}$ durch das Bildmaß definiert (vgl. Definition 3.1.7).

Insgesamt hängt also die Genauigkeit der Planungsentscheidung von der Messunsicherheit des Messergebnisses ab, die wiederum durch die k Standardabweichungen $\sqrt{\text{Var}(\hat{\mathbf{Q}})} = \sigma_{\hat{\mathbf{Q}}} = (\sigma_{\hat{Q}_1}, \dots, \sigma_{\hat{Q}_k})^\top \in \mathbb{R}^{+k}$ definiert ist.

Beispiel 3.1.8.b (Freifahrt: zulässige Planungsentscheidungen). Aus Beispiel 3.1.8.a ist die W-Dichte der ZV $f_a(v_t + \hat{V})$ bekannt und damit kann mit Ungleichung (3.5) überprüft werden, ob für eine Messunsicherheit σ_v zulässige Planungsentscheidungen bestimmt werden können. Dafür werde angenommen, ein AF fahre mit der Geschwindigkeit $v_t = 25 \text{ km/h} = 6.94 \text{ m/s}$ und habe eine Wunschgeschwindigkeit $v_w = 30 \text{ km/h} = 8.3 \text{ m/s}$. Die weiteren Parameter des IDM seien gegeben durch $\bar{a} = 1.5 \text{ m/s}^2$ und $\delta = 4$ (vgl. Beispiel 3.1.4). Die zulässige Abweichung der Beschleunigung sei $\bar{b}_a = 0.8 \text{ m/s}^2$ und die Messabweichung sei durch eine $\mathcal{N}(0, \sigma_v^2)$ -verteilte ZV \hat{V} beschrieben. Dann gilt mit den Momenten der ZV \hat{V}

$$\mathbb{E}(\hat{V}^0) = 1, \quad \mathbb{E}(\hat{V}^1) = 0, \quad \mathbb{E}(\hat{V}^2) = \sigma_v^2, \quad \mathbb{E}(\hat{V}^3) = 0, \quad \mathbb{E}(\hat{V}^4) = 3\sigma_v^4,$$

und der Linearität des Erwartungswertes für den Erwartungswert der ZV $f_a(v_t + \hat{V})$ und die Messunsicherheit $\sigma_v = 1.3 \text{ m/s}$

$$\begin{aligned} \mathbb{E}(f_a(v_t + \hat{V})) &= \mathbb{E}\left(\bar{a} \left(1 - \left(\frac{v_t + \hat{V}}{v_w}\right)^4\right)\right) \\ &= \bar{a} - \frac{\bar{a}}{v_w^4} \mathbb{E}\left(\sum_{k=0}^4 \binom{4}{k} v_t^{4-k} \hat{V}^k\right) \\ &= \bar{a} - \frac{\bar{a}}{v_w^4} \sum_{k=0}^4 \binom{4}{k} v_t^{4-k} \mathbb{E}(\hat{V}^k) \\ &= \bar{a} - \frac{\bar{a}}{v_w^4} \left(v_t^4 + 6v_t^2 \sigma_v^2 + 3\sigma_v^4\right) = 0.6219 \text{ m/s}^2, \end{aligned}$$

vgl. Abb. 3.2, oben. Damit gilt also für die Wahrscheinlichkeit für unzulässige

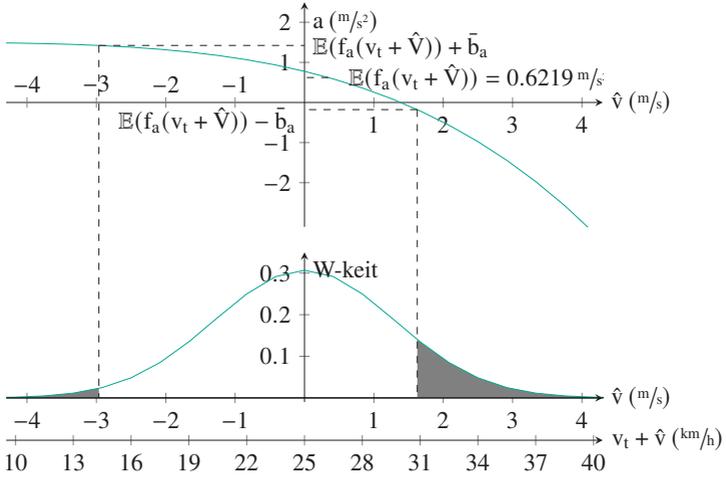


Abbildung 3.2: Zusammenhang zwischen Beschleunigung a , W-Verteilung der Messabweichung \hat{V} und Definition 3.1.9 für zulässige Planungsentscheidungen. Oben: – Beschleunigung a nach IDM in Abhängigkeit von der Messabweichung \hat{v} . Unten: – W-Verteilung der Messabweichung \hat{V} ; ■ ungültige Messabweichungen nach (3.6). Messgröße $v_t = v_t + \hat{v} = 25 \text{ km/h}$ für $\hat{v} = 0 \text{ m/s}$.

Planungsentscheidungen mit der W-Verteilung Φ_{f_a} , vgl. (3.4),

$$\begin{aligned}
 & \mathbb{P}_{f_a(v_t + \hat{v})} \left(\left| f_a(v_t + \hat{v}) - \mathbb{E}(f_a(v_t + \hat{V})) \right| > \bar{b}_a \right) \\
 &= \mathbb{P}_{f_a(v_t + \hat{v})} \left(f_a(v_t + \hat{v}) > \bar{b}_a + \mathbb{E}(f_a(v_t + \hat{V})) \right) \\
 & \quad + \mathbb{P}_{f_a(v_t + \hat{v})} \left(f_a(v_t + \hat{v}) < \mathbb{E}(f_a(v_t + \hat{V})) - \bar{b}_a \right) \\
 &= 1 - \mathbb{P}_{f_a(v_t + \hat{v})} \left(f_a(v_t + \hat{v}) \leq \mathbb{E}(f_a(v_t + \hat{V})) + \bar{b}_a \right) \\
 & \quad + \mathbb{P}_{f_a(v_t + \hat{v})} \left(f_a(v_t + \hat{v}) < \mathbb{E}(f_a(v_t + \hat{V})) - \bar{b}_a \right) \\
 &= 1 - \Phi_{f_a}(\mathbb{E}(f_a(v_t + \hat{V})) + \bar{b}_a) + \Phi_{f_a}(\mathbb{E}(f_a(v_t + \hat{V})) - \bar{b}_a) \\
 &= 0.0113 + 0.1055 = 0.1168, \tag{3.6}
 \end{aligned}$$

vgl. Abb. 3.2, unten. Für eine Wahrscheinlichkeitsschranke $p_a = 0.1$ ist also die Messungenauigkeit $\sigma_v = 1.3 \text{ m/s}$ zu groß oder das IDM nicht robust genug, um in dieser Situation zulässige Planungsentscheidungen treffen zu können.

Die Abschätzung in Definition 3.1.9 ist analytisch schwer auszuwerten, da anders als in Beispiel 3.1.8 für gewöhnlich keine geschlossene Form der W-Verteilung von $f_o(\mathbf{q} + \hat{\mathbf{Q}})$ existiert. Stattdessen wird die Wahrscheinlichkeit mithilfe von stochastischen Ungleichungen nach oben abgeschätzt.

3.2 Abschätzung der Wahrscheinlichkeitsungleichung

Das Ergebnis eines Planungsalgorithmus kann verschiedene Formen annehmen, dies kann u. a. vom verwendeten Ansatz abhängen, z. B. numerische Optimierung oder sampling-basierte Verfahren. Außerdem sind kontinuierliche Ergebnisse, z. B. Beschleunigung oder Lenkwinkel, und diskrete Ergebnisse, z. B. FSW nach rechts, links oder nicht, denkbar. In diesem Abschnitt werden Abschätzungen für die Wahrscheinlichkeitsungleichung aus Definition 3.1.9 für kontinuierliche und binäre Planungsentscheidungen vorgestellt.

3.2.1 Chebychev-Saw-Abschätzung für allgemeine Planungsentscheidungen

Wahrscheinlichkeiten der Form $\mathbb{P}_X (|x - \mathbb{E}(X)| > \bar{b})$ können mithilfe der Chebychev-Ungleichung nach oben abgeschätzt werden:

Satz 3.2.1 (Tchébychef und Khanikof (1867), Klenke (2013), Satz 5.11 Chebychev-Ungleichung). *Sei X eine ZV mit Erwartungswert $\mathbb{E}(X)$ und endlicher Varianz $\text{Var}(X) \in \mathbb{R}^+$. Dann gilt für jedes $\bar{b} \in \mathbb{R}^+$ die Chebychev-Ungleichung*

$$\mathbb{P}_X (|x - \mathbb{E}(X)| > \bar{b}) \leq \frac{\text{Var}(X)}{\bar{b}^2} \quad (3.7)$$

Die Chebychev-Ungleichung setzt keine Annahmen zur W-Verteilung der ZV voraus. Sie kann daher für beliebig verteilte ZV angewendet werden, z. B. unabhängig davon, ob die Ergebnismenge Ω kontinuierlich oder diskret ist. Der Nachteil dieser Freiheit liegt darin, dass die obere Schranke sehr grob ist und daher die abgeleiteten Ergebnisse sehr konservativ bleiben. Das folgende Beispiel verdeutlicht dies.

Beispiel 3.1.8.c (Freifahrt: Chebychev-Ungleichung). Für die Freifahrt eines AF wurde in Beispiel 3.1.8.b der Erwartungswert berechnet. Analog dazu gilt mit den Momenten höherer Ordnung der $\mathcal{N}(0, \sigma_v^2)$ -verteilten ZV \hat{V}

$$\mathbb{E}(\hat{V}^5) = 0, \quad \mathbb{E}(\hat{V}^6) = 15\sigma_v^6, \quad \mathbb{E}(\hat{V}^7) = 0, \quad \mathbb{E}(\hat{V}^8) = 105\sigma_v^8$$

für die Varianz

$$\begin{aligned} \text{Var}(f_a(v_t + \hat{V})) &= \mathbb{E}\left(\left[f_a(v_t + \hat{V}) - \mathbb{E}(f_a(v_t + \hat{V}))\right]^2\right) \\ &= \mathbb{E}\left(\left[\bar{a}\left(1 - \left(\frac{v_t + \hat{V}}{v_w}\right)^4\right) - \bar{a} + \frac{\bar{a}}{v_w^4}\left(v_t^4 + 6v_t^2\sigma_v^2 + 3\sigma_v^4\right)\right]^2\right) \\ &= \frac{\bar{a}^2}{v_w^8}\mathbb{E}\left(\left[-v_t^4 - 4v_t^3\hat{V} - 6v_t^2\hat{V}^2 - 4v_t\hat{V}^3 - \hat{V}^4 + v_t^4 + 6v_t^2\sigma_v^2 + 3\sigma_v^4\right]^2\right) \\ &= \frac{\bar{a}^2}{v_w^8}\left(\left(36v_t^4\sigma_v^4 + 36v_t^2\sigma_v^6 + 9\sigma_v^8\right) + \mathbb{E}(\hat{V})(-48v_t^5\sigma_v^2 - 24v_t^3\sigma_v^4)\right. \\ &\quad + \mathbb{E}(\hat{V}^2)(16v_t^6 - 72v_t^4\sigma_v^2 - 36v_t^2\sigma_v^4) \\ &\quad + \mathbb{E}(\hat{V}^3)(48v_t^5 - 48v_t^3\sigma_v^2 - 24v_t\sigma_v^4) \\ &\quad + \mathbb{E}(\hat{V}^4)(68v_t^4 - 12v_t^2\sigma_v^2 - 6\sigma_v^4) \\ &\quad \left.+ \mathbb{E}(\hat{V}^5)56v_t^3 + \mathbb{E}(\hat{V}^6)28v_t^2 + \mathbb{E}(\hat{V}^7)8v_t + \mathbb{E}(\hat{V}^8)\right) \\ &= \frac{\bar{a}^2}{v_w^8}\left(16v_t^6\sigma_v^2 + 168v_t^4\sigma_v^4 + 384v_t^2\sigma_v^6 + 96\sigma_v^8\right). \end{aligned}$$

Daher hat die ZV $f_a(v_t + \hat{V}) = \bar{a}\left(1 - (v_t + \hat{V}/v_w)^\delta\right)$ mit den Parametern $\bar{a} = 1.5 \text{ m/s}^2$, $v_t = 25 \text{ km/h} = 6.94 \text{ m/s}$, $v_w = 30 \text{ km/h} = 8.3 \text{ m/s}$, $\delta = 4$ für die normalverteilte ZV

$\hat{V} \sim \mathcal{N}(0, \sigma_v^2)$ mit der Messunsicherheit $\sigma_v = 1.3 \text{ m/s}$ den Erwartungswert und die Varianz

$$\begin{aligned}\mathbb{E}(f_a(v_t + \hat{V})) &= 0.6219 \text{ m/s}^2, \\ \mathbb{V}\text{ar}(f_a(v_t + \hat{V})) &= 0.4101 \text{ m}^2/\text{s}^4.\end{aligned}$$

Durch Einsetzen in die Chebychev-Ungleichung (3.7) gilt für $\bar{b}_a = 0.8 \text{ m/s}^2$

$$\mathbb{P}_{f_a(v_t + \hat{V})} \left(\left| f_a(v_t + \hat{v}) - \mathbb{E}(f_a(v_t + \hat{V})) \right| > 0.8 \text{ m/s}^2 \right) \leq \frac{0.4101 \text{ m}^2/\text{s}^4}{0.8 \text{ m}^2/\text{s}^4} = 0.6408.$$

Dies bedeutet, dass die Wahrscheinlichkeit, dass die Beschleunigung um mehr als $\bar{b}_a = 0.8 \text{ m/s}^2$ von der erwarteten Beschleunigung abweicht, kleiner als 64.08 % ist. Beispiel 3.1.8.b hat gezeigt, dass dies zwar richtig, aber nur eine sehr grobe Abschätzung ist.

Um die Chebychev-Ungleichung anzuwenden, müssen der Erwartungswert und die Varianz der ZV bekannt sein. In der Regel sind die Verteilungen der Fahrentscheidungen – anders als in Beispiel 3.1.8.c – nicht bekannt (s. Abschnitt 3.1). Daher können im Allgemeinen keine Aussagen über die W-Verteilung der Planungsergebnisse sowie deren Erwartungswert und Varianz getroffen werden, sodass die Chebychev-Ungleichung (3.7) nicht für Aussagen über die Zuverlässigkeit der Fahrentscheidungen verwendet werden kann. Stattdessen wird eine Variante der Chebychev-Ungleichung nach Saw, Yang und Chin Mo (1984) für den geschätzten Erwartungswert und die geschätzte Varianz verwendet.

Definition 3.2.2. Sei $\{x_j\}_{j=1}^M$ eine Stichprobe der ZV X . Dann sind der Stichprobenmittelwert und die Stichprobenvarianz gegeben durch

$$\mu_X := \frac{1}{M} \sum_{j=1}^M x_j, \quad (3.8)$$

$$\sigma_X^2 := \frac{1}{M-1} \sum_{j=1}^M (x_j - \mu_X)^2. \quad (3.9)$$

Satz 3.2.3 (Saw et al. (1984), Chebychev-Saw-Ungleichung). Sei $\{x_j\}_{j=1}^M$ eine unabhängig identisch verteilte Stichprobe der ZV X . Dann gilt für den Stichprobenmittelwert μ_X und die Stichprobenvarianz σ_X^2 die Chebychev-Saw-Ungleichung

$$\mathbb{P}_X \left(|x - \mu_X| > \lambda_S \sqrt{\frac{M+1}{M}} \sigma_X \right) \leq \frac{1}{M+1} g_{M+1} \left(\frac{M \lambda_S^2}{M-1 + \lambda_S^2} \right), \quad (3.10)$$

mit der Funktion $g_{M+1} : \mathbb{R} \rightarrow \mathbb{R}$ für $x \geq 1$

$$g_{M+1}(x) = \begin{cases} v(x), & \text{falls } v(x) \text{ gerade ist,} \\ v(x), & \text{falls } v(x) \text{ ungerade ist und } x < a^2(x), \\ v(x) - 1, & \text{falls } v(x) \text{ ungerade ist und } x \geq a^2(x), \end{cases}$$

$$v(x) = \max \left\{ w \in \mathbb{N}_0 \mid w < \frac{(M+1)}{x} \right\},$$

$$a^2(x) = \frac{(M+1)(M+1-v(x))}{1+v(x)(M+1-v(x))}.$$

Bemerkung 3.2.4. In der Simulation ist es akzeptabel, die Messergebnisse als unabhängig anzunehmen, auch wenn diese Annahme in der Realität nicht zwangsläufig erfüllt ist.

Die Chebychev-Saw-Ungleichung gibt also eine obere Schranke für die Wahrscheinlichkeit an, dass eine Realisierung der ZV $f_a(v_t + \hat{V})$ weiter als \bar{b}_a vom Stichprobenmittelwert $\mu_{f_a(v_t + \hat{V})}$ entfernt ist. Im Gegensatz zur Chebychev-Ungleichung (3.7) reichen dafür der Stichprobenmittelwert und die Stichprobenvarianz als Schätzer für den Erwartungswert und die Varianz aus. Das führt dazu, dass der Stichprobenmittelwert $\mu_{f_a(v_t + \hat{V})}$ anstelle des Erwartungswerts $\mathbb{E}(f_a(v_t + \hat{V}))$ als Bezugspunkt verwendet wird (vgl. Definition 3.1.9). Da der Erwartungswert in der Umsetzung immer unbekannt ist und durch den Stichprobenmittelwert geschätzt werden muss, hat die Chebychev-Ungleichung nur akademischen Wert. In der praktischen Umsetzung und im weiteren Verlauf der Arbeit werden wir auf die sampling-basierte Definition 3.2.5 der zulässigen Planungsentscheidungen zurückgreifen:

Definition 3.2.5 (zulässige Planungsentscheidungen für Stichproben). Sei $\mathbf{f} : \Omega \rightarrow \mathbb{R}^m$, $\Omega \subseteq \mathbb{R}^k$, ein Planungsalgorithmus, $\mathbf{q} \in \Omega$ eine Messgröße und $\hat{\mathbf{Q}}$ die zugehörige, als ZV aufgefasste Messabweichung. Seien außerdem zulässige Abweichungen $\bar{\mathbf{b}} = (\bar{b}_1, \dots, \bar{b}_m)^\top \in \mathbb{R}^{+m}$ und Wahrscheinlichkeiten $\bar{p} \in [0, 1]^m$ gegeben. Dann heißt eine Planungsentscheidung $f_o(\mathbf{q} + \hat{\mathbf{Q}})$ zulässig, wenn sie sich mit Wahrscheinlichkeit \bar{p}_o um nicht mehr als eine zulässige obere Schranke \bar{b}_o vom Stichprobenmittelwert $\mu_{f_o(\mathbf{q} + \hat{\mathbf{Q}})}$ unterscheidet, d. h. wenn gilt

$$\mathbb{P}_{f_o(\mathbf{q} + \hat{\mathbf{Q}})} \left(\left| f_o(\mathbf{q} + \hat{\mathbf{q}}) - \mu_{f_o(\mathbf{q} + \hat{\mathbf{Q}})} \right| > \bar{b}_o \right) \leq \bar{p}_o. \quad (3.11)$$

Diese Definition steht immer noch in Einklang mit der ISO-Richtlinie 5725-1:1994, da neben dem Erwartungswert einer ZV auch der Mittelwert einer ausreichend großen Stichprobe als akzeptierter Referenzwert genutzt werden kann ($\mu_{f_o(\mathbf{q} + \hat{\mathbf{Q}})}$, International Organization for Standardization, 1994). Sei $\{f_o(\mathbf{q} + \hat{\mathbf{q}}_j)\}_{j=1}^M$ eine Stichprobe der ZV $f_o(\mathbf{q} + \hat{\mathbf{Q}})$ und $\sigma_{f_o(\mathbf{q} + \hat{\mathbf{Q}})}^2$ die zugehörige Stichprobenvarianz. Um die Wahrscheinlichkeit (3.11) mit der Ungleichung von Chebychev-Saw (3.10) nach oben abzuschätzen, wird die Konstante $\lambda_{S,o}$ durch

$$\lambda_{S,o} := \sqrt{\frac{M}{M+1}} \frac{\bar{b}_o}{\sigma_{f_o(\mathbf{q} + \hat{\mathbf{Q}})}} \quad (3.12)$$

definiert. Nach Auflösen nach \bar{b}_o und Einsetzen in (3.11) wird die Chebychev-Saw-Ungleichung (3.10) angewendet und somit gilt

$$\begin{aligned} \mathbb{P}_{f_o(\mathbf{q} + \hat{\mathbf{Q}})} \left(|f_o(\mathbf{q} + \hat{\mathbf{q}}) - \mu_{f_o(\mathbf{q} + \hat{\mathbf{Q}})}| > \bar{b}_o \right) &\leq \frac{1}{M+1} g^{M+1} \left(\frac{M \lambda_{S,o}^2}{M-1 + \lambda_{S,o}^2} \right) \\ &\leq \bar{p}_o \quad \forall o = 1, \dots, m \end{aligned} \quad (3.13)$$

Beispiel 3.1.8.d (Freifahrt: Chebychev-Saw-Ungleichung). Für Beispiel 3.1.8 kann für Messabweichungen $\{\hat{v}_j\}_{j=1}^M$ in der Simulation die Stichprobenvarianz einer Stichprobe $\{f_a(v_t + \hat{v}_j)\}_{j=1}^M$ berechnet werden, sodass diese in die Chebychev-Saw-Ungleichung (3.13) eingesetzt werden kann. Mit dem

Stichprobenmittelwert $\mu_{f_a(v_t+\hat{v})} = 0.6149 \text{ m/s}^2$ und der Stichprobenvarianz $\sigma_{f_a(v_t+\hat{v})}^2 = 0.4132 \text{ m}^2/\text{s}^4$ ergibt sich für die Stichprobe $\{\hat{v}_j\}_{j=1}^M$ mit $M = 300$ Stichprobenwerten die obere Schranke

$$\lambda_{S,a} = \sqrt{\frac{300}{301}} \frac{0.8 \text{ m/s}^2}{\sqrt{0.4132 \text{ m}^2/\text{s}^4}} = 1.2445,$$

$$\frac{1}{M+1} g_{M+1} \left(\frac{M \lambda_{S,a}^2}{M-1 + \lambda_{S,a}^2} \right) = 0.6445,$$

d. h. die obere Schranke nach der Chebychev-Saw-Ungleichung für zulässige Planungsentscheidungen liegt bei 64.45 %. Dies ist genauso wie die Abschätzung für bekannte Erwartungswert und Varianz sehr grob und die Diskussion dafür kann analog zu Beispiel 3.1.8.c geführt werden.

Zusammengefasst bedeutet dies, dass die Ungleichung (3.5) für zulässige Planungsentscheidungen nur exakt ausgewertet werden kann, wenn Planer \mathbf{f} bekannt ist, und auch dann eine geschlossene Form möglicherweise nicht existiert. Stattdessen wird Approximation (3.13) mit Stichprobenmittelwert (3.8) und Stichprobenvarianz (3.9) für die ZV $f_o(\mathbf{q} + \hat{\mathbf{Q}})$ verwendet, um zu überprüfen, ob die Messunsicherheiten $\sigma_{\hat{\mathbf{Q}}}$ eine nicht-tolerierbare Abweichung des Planungsergebnisses verursachen. Glücklicherweise nutzen die Ungleichungen keine Annahmen an die zugrundeliegende W -Verteilungen, sodass beliebige W -Dichten zur Beschreibung der Messabweichungen verwendet werden können. Bisher haben wir nur W -Dichten für kontinuierliche Planungsentscheidungen wie die Beschleunigung getroffen, die für die Trajektorienplanung berechnet werden. Auf der taktischen Planungsebene werden jedoch häufig zunächst diskrete Entscheidungen getroffen, wie das nächste Beispiel beschreibt.

Beispiel 3.2.6.a (Überholmanöver: IDM-MOBIL). Das Minimizing Overall Braking Induced by Lane Changes (MOBIL) Modell ist ein FSW-Modell von Kesting, Treiber und Helbing (2007). In dieser Arbeit wird ein asymmetrisches FSW-Kriterium genutzt, das das Rechtsfahrgebot der meisten europäischen Ländern mit Rechtsverkehr widerspiegelt (vgl. z. B. StVO, 2013, § 2): Es

bestimmt die FSW-Entscheidung von links nach rechts ($L \rightarrow R$) und umgekehrt ($R \rightarrow L$) durch

$$\begin{aligned}
 L \rightarrow R: \quad & \text{EUR}(a^P(vr)) - a + h(a_{ns}^P(v_s) - a_{ns}) > \Delta a_{th} - \Delta a_{bias} \\
 & \text{mit } \text{EUR}(a^P(vr)) := \begin{cases} \min(a^P(vr), a), & v > v_{vl} > v_{krit} \\ a^P(vr), & \text{sonst,} \end{cases} \\
 R \rightarrow L: \quad & a_{nl}^P - \text{EUR}(a) + h(a_{nl}^P - a_{nl}) > \Delta a_{th} + \Delta a_{bias} \\
 & \text{mit } \text{EUR}(a) := \begin{cases} \min(a, a^P(vl)), & v > v_{vs} > v_{krit} \\ a, & \text{sonst.} \end{cases}
 \end{aligned}$$

Dabei ist a die Beschleunigung des Egos, v seine Geschwindigkeit und v_{krit} die maximale Geschwindigkeit, für die das Rechtsfahrgebot nicht mehr gilt. Außerdem wird die Beschleunigung a_{ns} des Fahrzeugs auf demselben Fahrstreifen hinter dem Ego, die Beschleunigung a_{nl} des nachfolgenden Fahrzeugs auf dem linken Fahrstreifen hinter dem Ego, die Geschwindigkeit v_{vl} des vorausfahrenden Fahrzeugs auf dem linken Fahrstreifen vor dem Ego sowie die Geschwindigkeit v_{vs} des Fahrzeugs auf demselben Fahrstreifen vor dem Ego verwendet. Die Prädiktion $a_{c_1}^P(c_2)$ bezeichnet die vorhergesagte Beschleunigung des Fahrzeugs c_1 , wenn ihm das Fahrzeug c_2 vorausfährt. Außerdem modelliert der Höflichkeitsparameter $h \in [0, 1]$ die Kooperationsbereitschaft des Egos ($h = 0$ egoistisch, $h = 1$ kooperativ), die Schwelle Δa_{th} verhindert, dass das Ego den Fahrstreifen zu häufig wechselt und der Wert Δa_{bias} schafft Anreize für einen FSW nach rechts anstelle eines FSWs nach links. Zusammengefasst bestimmt das MOBIL also die Richtung des FSW als diskrete Variable

$$lc = \begin{cases} -1, & \text{falls } L \rightarrow R, \\ 1, & \text{falls } R \rightarrow L, \\ 0, & \text{sonst.} \end{cases}$$

Zur Berechnung und Prädiktion der Beschleunigungen können verschiedene Längsführungsmodelle verwendet werden. In dieser Arbeit wird das MOBIL-Modell analog zu Kesting et al. (2007) mit dem IDM gekoppelt und so das Verhalten eines AF simuliert. Dieser Ansatz hat einige Nachteile, z. B. dass das Modell für die Entscheidung über einen FSW nur wenige Verkehrsteilnehmer beachtet, während im Realverkehr zuvor noch strategische und taktische Aspekte

zur Routenplanung berücksichtigt werden (Kesting et al., 2007). Außerdem kann es so zu Beschleunigungen führen, die nicht stetig sind. Da es jedoch ein einfaches Modell mit wenigen Gleichungen und Annahmen ist, ist seine Nachvollziehbarkeit sehr gut und es bietet damit ideale Voraussetzungen für das Testen neuer Methoden. Daher wird im weiteren Verlauf das IDM mit dem MOBIL verwendet, um das Verhalten eines AF in Überhol szenarios zu simulieren.

Die Berechnung eines FSW ist eine diskrete Entscheidung zwischen Homotopieklassen (Bhattacharya, Kumar & Likhachev, 2010). Dadurch ist die Ergebnismenge diskret und nicht mehr kontinuierlich, sodass andere W -Verteilungen zur Modellierung der Messabweichungen verwendet werden müssen. Das folgende Kapitel beschreibt, wie die Abschätzung (3.13) nach Chebychev und Saw verbessert werden kann, wenn bekannt ist, dass es nur zwei Homotopieklassen gibt.

3.2.2 Bernoulli-a-posteriori-Abschätzung für zwei Planungsentscheidungen

Falls eine Fahrentscheidung f_o nur zwischen zwei Homotopieklassen unterscheidet, können diese in $\Omega = \{0, 1\} \subset \mathbb{R}$ kodiert werden. Eine Zweipunktverteilung auf $\{0, 1\}$ kann mit der Bernoulliverteilung $\mathcal{B}(p_o)$ mit Erfolgswahrscheinlichkeit $p_o = \mathbb{P}(f_o(\mathbf{q} + \hat{\mathbf{Q}}) = 1) \in [0, 1]$ beschrieben werden. Diese hat dann den Erwartungswert und die Varianz

$$\begin{aligned}\mathbb{E}(f_o(\mathbf{q} + \hat{\mathbf{Q}})) &= p_o, \\ \text{Var}(f_o(\mathbf{q} + \hat{\mathbf{Q}})) &= p_o(1 - p_o).\end{aligned}$$

Leider ist die Wahrscheinlichkeit p_o i. d. R. unbekannt und kann auch nicht mit akzeptablem Rechenaufwand bestimmt werden. Stattdessen wird sie mithilfe einer a posteriori W -Dichte e geschätzt. Dafür wird angenommen, dass in einer

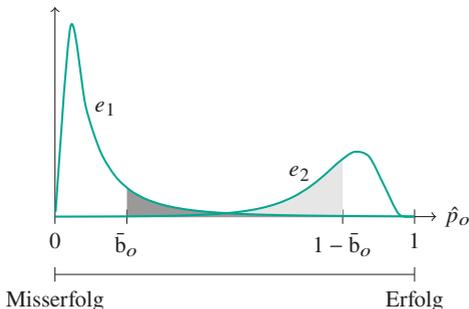


Abbildung 3.3: Zwei Schätzer e_1, e_2 für jeweils eine unbekannte Erfolgswahrscheinlichkeit p_1, p_2 in einem Bernoulli-Experiment. ■ $\mathbb{P}_{e_1}(p_o \geq \bar{b}_o)$; ■ $\mathbb{P}_{e_2}(p_b \leq 1 - \bar{a}_b)$, vgl. Ungleichung (3.15) für zulässige Planungsentscheidungen.

Stichprobe mit M Stichprobenwerten $h \leq M$ Erfolge und $t = M - h$ Misserfolge beobachtet wurden. Dann ist die Schätzfunktion e von p_o gegeben durch

$$\begin{aligned}
 e(\hat{p}_o \mid H = h, T = t) &= \frac{\mathbb{P}(H = h \mid \hat{p}_o, M = h + t)g(\hat{p}_o)}{\int_0^1 \mathbb{P}(H = h \mid x, M = h + t)g(x) \, dx} \\
 &\stackrel{g \equiv 1}{=} \frac{1}{B(h + 1, t + 1)} \hat{p}_o^h (1 - \hat{p}_o)^t \\
 &= \frac{(h + t + 1)!}{h!t!} \hat{p}_o^h (1 - \hat{p}_o)^t, \tag{3.14}
 \end{aligned}$$

wobei die Funktion $g : [0, 1] \rightarrow [0, 1]$ eine a priori W-Dichte des Schätzers e und B die Betafunktion ist. Falls keine Beobachtungen vorhanden sind, die für die Wahl von g verwendet werden könnten, wird $g(\hat{p}_o) \equiv 1$ gewählt. Zwei Schätzer e_1, e_2 , d. h. W-Dichten für zwei Wahrscheinlichkeiten p_1, p_2 nach zwei verschiedenen Beobachtungen sind in Abb. 3.3 dargestellt. Für die Berechnung der Wahrscheinlichkeit, dass ein Parameter p_o in einem Intervall $[a, b]$ oder (a, b) liegt, gilt mit der W-Dichte e also

$$\mathbb{P}(a \leq \hat{p}_o \leq b) = \mathbb{P}(a < \hat{p}_o < b) = \int_a^b e(x \mid H = h, T = t) \, dx.$$

Zusammengefasst liefert e einen Intervallschätzer für die unbekannte Erfolgswahrscheinlichkeit p_o für die $\mathcal{B}(p_o)$ -verteilte Fahrentscheidung $f_o(\mathbf{q} + \hat{\mathbf{Q}})$,

wenn unter M Messergebnissen $\{\mathbf{q} + \hat{\mathbf{q}}_j\}_{j=1}^M$ h -mal die Fahrentscheidung zugunsten einer Homotopieklasse getroffen und $(M - h)$ -mal zugunsten der anderen Homotopieklasse. Ein Beispiel ist, dass sich ein AF bei h Messergebnissen für den FSW nach links entschieden hat, während es in den anderen $M - h$ Fällen weiterhin dem Fahrstreifen folgt. Damit kann nun Abschätzung (3.11) vereinfacht werden. Dafür wird zunächst eine Fallunterscheidung mit der zulässigen oberen Schranke $\bar{b}_o < 0.5$ betrachtet:

- für $p_o \leq \bar{b}_o$ gilt:

$$|f_o(\mathbf{q} + \hat{\mathbf{q}}) - p_o| = \begin{cases} |0 - p_o| = p_o \leq \bar{b}_o, & f_o(\mathbf{q} + \hat{\mathbf{q}}) = 0, \\ |1 - p_o| = 1 - p_o \geq 1 - \bar{b}_o \\ \geq 0.5 \geq \bar{b}_o, & f_o(\mathbf{q} + \hat{\mathbf{q}}) = 1 \end{cases}$$

- für $\bar{b}_o < p_o < 1 - \bar{b}_o$ gilt:

$$|f_o(\mathbf{q} + \hat{\mathbf{q}}) - p_o| = \begin{cases} p_o \geq \bar{b}_o, & f_o(\mathbf{q} + \hat{\mathbf{q}}) = 0, \\ 1 - p_o \geq 1 - (1 - \bar{b}_o) = \bar{b}_o, & f_o(\mathbf{q} + \hat{\mathbf{q}}) = 1 \end{cases}$$

- für $1 - \bar{b}_o \leq p_o$ gilt:

$$|f_o(\mathbf{q} + \hat{\mathbf{q}}) - p_o| = \begin{cases} p_o \geq 1 - \bar{b}_o \geq \bar{b}_o, & f_o(\mathbf{q} + \hat{\mathbf{q}}) = 0, \\ 1 - p_o \leq 1 - (1 - \bar{b}_o) = \bar{b}_o, & f_b(\mathbf{q} + \hat{\mathbf{q}}) = 1 \end{cases}$$

Mit diesen Vorüberlegungen und mit der Definition der bedingten Wahrscheinlichkeit gilt für Abschätzung (3.11) für eine Fahrentscheidung o mit zwei möglichen Ergebnissen $\Omega = \{0, 1\}^2$

$$\begin{aligned}
 & \mathbb{P}(|f_o(\mathbf{q} + \hat{\mathbf{q}}) - \mathbb{E}(f_o(\mathbf{q} + \hat{\mathbf{Q}}))| \geq \bar{b}_o) \\
 &= \mathbb{P}(|f_o(\mathbf{q} + \hat{\mathbf{q}}) - p_o| \geq \bar{b}_o) \\
 &= \mathbb{P}(|f_o(\mathbf{q} + \hat{\mathbf{q}}) - p_o| \geq \bar{b}_o \wedge p_o \leq \bar{b}_o) \\
 &\quad + \mathbb{P}(|f_o(\mathbf{q} + \hat{\mathbf{q}}) - p_o| \geq \bar{b}_o \wedge \bar{b}_o < p_o < 1 - \bar{b}_o) \\
 &\quad + \mathbb{P}(|f_o(\mathbf{q} + \hat{\mathbf{q}}) - p_o| \geq \bar{b}_o \wedge 1 - \bar{a}_b \leq p_o) \\
 &\stackrel{\text{s.o.}}{=} \mathbb{P}(f_o(\mathbf{q} + \hat{\mathbf{q}}) = 1 \wedge p_o \leq \bar{b}_o) \\
 &\quad + \mathbb{P}(\bar{b}_o < p_o < 1 - \bar{b}_o) \\
 &\quad + \mathbb{P}(f_o(\mathbf{q} + \hat{\mathbf{q}}) = 0 \wedge 1 - \bar{b}_o \leq p_o) \\
 &= \mathbb{P}(f_o(\mathbf{q} + \hat{\mathbf{q}}) = 1 \mid p_o \leq \bar{b}_o) \mathbb{P}(p_o \leq \bar{b}_o) \\
 &\quad + \mathbb{P}(\bar{b}_o < p_o < 1 - \bar{b}_o) \\
 &\quad + \mathbb{P}(f_o(\mathbf{q} + \hat{\mathbf{q}}) = 0 \mid 1 - \bar{b}_o \leq p_o) \mathbb{P}(1 - \bar{b}_o \leq p_o) \\
 &\leq \bar{b}_o \mathbb{P}(p_o \leq \bar{b}_o) \\
 &\quad + \mathbb{P}(\bar{b}_o < p_o < 1 - \bar{b}_o) \\
 &\quad + \bar{b}_o \mathbb{P}(1 - \bar{b}_o \leq p_o) \\
 &= \bar{b}_o (\mathbb{P}(p_o \leq \bar{b}_o) + \mathbb{P}(1 - \bar{b}_o \leq p_o)) \\
 &\quad + \mathbb{P}(\bar{b}_o < p_o < 1 - \bar{b}_o). \tag{3.15}
 \end{aligned}$$

Die Ungleichung folgt aus

$$\begin{aligned}
 & \mathbb{P}(f_o(\mathbf{q} + \hat{\mathbf{q}}) = 1 \mid p_o \leq \bar{b}_o) = p_o \Big|_{p_o \leq \bar{b}_o} \leq \bar{b}_o \\
 & \mathbb{P}(f_o(\mathbf{q} + \hat{\mathbf{q}}) = 0 \mid 1 - \bar{b}_o \leq p_o) = 1 - p_o \Big|_{p_o \geq 1 - \bar{b}_o} \leq \bar{b}_o.
 \end{aligned}$$

²Zugunsten der Übersichtlichkeit wird hier auf die konkrete Notation des W-Maßes als Subindex verzichtet, dieses ergibt sich aus den jeweils betrachteten W-Räumen und mit Definition 3.1.7 des Bildmaßes $\mathbb{P}_{\hat{\mathbf{Q}}}(\hat{\mathbf{q}} \mid f_o(\mathbf{q} + \hat{\mathbf{q}}) = 1) = \mathbb{P}_{f_o(\mathbf{q} + \hat{\mathbf{Q}})}(f_o(\mathbf{q} + \hat{\mathbf{q}}) = 1) = p_o$ und analog $\mathbb{P}_{\hat{\mathbf{Q}}}(\hat{\mathbf{q}} \mid f_o(\mathbf{q} + \hat{\mathbf{q}}) = 0) = \mathbb{P}_{f_o(\mathbf{q} + \hat{\mathbf{Q}})}(f_o(\mathbf{q} + \hat{\mathbf{q}}) = 0) = 1 - p_o$.

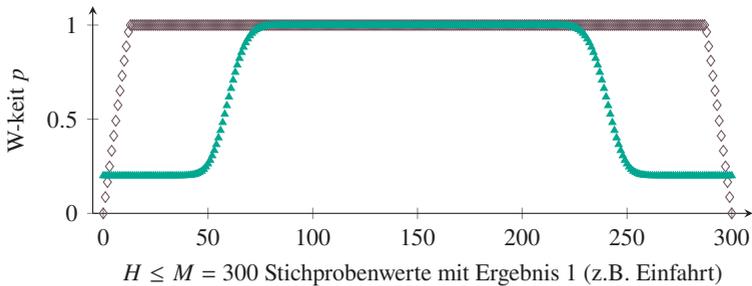


Abbildung 3.4: Wahrscheinlichkeit dafür, dass die Planungsentscheidung I_c nach Definition 3.2.5 für $\bar{b}_{1c} = 0.2$ zulässig ist: \diamond Abschätzung mit Chebychev-Saw-Ungleichung (3.13); \blacktriangle Abschätzung mit Bernoulli-a-posteriori-Ungleichung (3.15).



Abbildung 3.5: Überholmanöver. Das rote Fahrzeug ist das automatisierte Ego, das das gelbe vorausfahrende Fahrzeug überholen wird. Exemplarische Darstellung der Situation zum Zeitpunkt $t = 0$ s (vgl. Abb. 3.6).

Einsetzen der W-Dichte (3.14) in Abschätzung (3.15) ergibt für diskrete Fahrentscheidungen f_o eine bessere Abschätzung von (3.11) als durch die allgemeine Abschätzung (3.13). Abb. 3.4 zeigt die Verbesserung durch die Bernoulli-a-posteriori-Abschätzung (3.15) im Vergleich zur Chebychev-Saw-Abschätzung (3.13) mit $M = 300$ Stichprobenwerten. Wie in der Abbildung angedeutet, ist Ungleichung (3.15) mit (3.14) nach unten beschränkt, sodass sie selbst im günstigsten Fall, in der die Entscheidung eindeutig ist (dh. $\mathbb{P}(p_o \leq \bar{b}_o) = 1$ oder $\mathbb{P}(p_o \geq 1 - \bar{b}_o) = 1$), nach unten durch \bar{b}_o beschränkt ist. Daher sollte immer $\bar{b}_o \leq \bar{p}$ gelten. Andernfalls kann diese Ungleichung nie erfüllt werden. Es wird jedoch deutlich, dass die Bernoulli-a-posteriori-Abschätzung für einen größeren Bereich eine bessere Abschätzung als die konservative Chebychev-Saw-Abschätzung liefert.

Beispiel 3.2.6.b (Überholmanöver: Vergleich zwischen Chebychev-Saw- und Bernoulli-a-posteriori-Ungleichung). Während eines Überholmanövers trifft das rote Ego die Entscheidung, ein langsamerer vorausfahrendes Fahrzeug zu

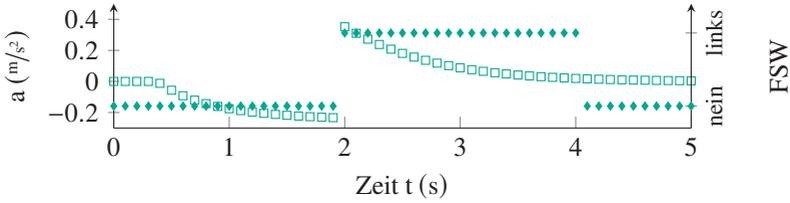


Abbildung 3.6: Fahrentscheidungen im Überholmanöver-Beispiel 3.2.6 gemäß IDM-MOBIL-Modell. □ Beschleunigung nach IDM; ◆ FSW.

überholen (vgl. Abb. 3.5). Entsprechend dem IDM-MOBIL-Modell wird die Geschwindigkeit angepasst und der FSW initiiert (Beispiel 3.2.6.a, Henze et al., 2020; Kesting et al., 2007). Abbildung 3.6 zeigt den Beschleunigungsverlauf sowie die binäre FSW-Entscheidung (FSW nach links oder kein FSW) in der ersten Phase des Überholmanövers. Für $t \leq 0.3$ s verändert das Ego seine Beschleunigung nicht, dann bremst es leicht. Zum Zeitpunkt $t = 2$ s, wenn die Entscheidung für einen FSW getroffen wird, ist die Berechnung der Beschleunigung unstetig, weil das IDM-MOBIL-Modell keine stetigen Beschleunigungsverläufe abbildet. Während des Überholmanövers beschleunigt das Ego auf seine Wunschgeschwindigkeit und schließt den FSW nach links ab.

In jedem Zeitschritt wird die Planungsentscheidung für $M = 300$ Messergebnisse ausgewertet und mithilfe der Chebychev-Saw- und der BAP-Ungleichungen die Zuverlässigkeit der Planungsentscheidung getestet. Dafür werden die Messunsicherheiten $\sigma_v = 0.1$ m/s (Ego-Geschwindigkeit), $\sigma_{d_{nl}} = 0.5$ m (Distanz zum nachfolgenden Fahrzeug auf dem linken Fahrstreifen), $\sigma_{d_{vs}} = 1$ m (Distanz zum vorausfahrenden Fahrzeug auf demselben Fahrstreifen), $\sigma_{v_{nl}} = \sigma_{v_{vs}} = 1$ m/s (Geschwindigkeit des nachfolgenden bzw. vorausfahrenden Fahrzeugs) angenommen. Für $t = 0.5$ s gibt es $h = 71$ Messergebnisse, für die der Planungsalgorithmus f sich für einen FSW nach links entscheiden würde. Daraus ergibt sich der Schätzer e

$$e(\hat{p}_{lc} \mid H = 71, T = 229) = \frac{301!}{71! 229!} \hat{p}_{lc}^{71} (1 - \hat{p}_{lc})^{229}.$$

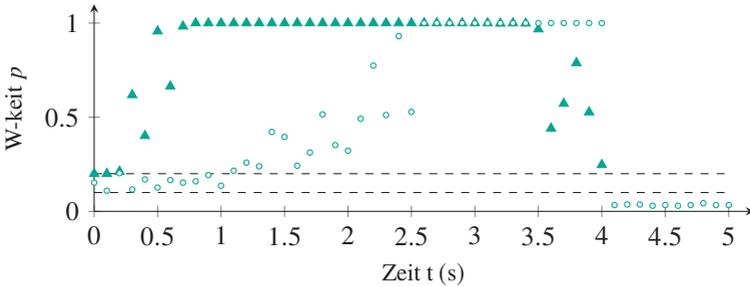


Abbildung 3.7: Obere Schranken für die Wahrscheinlichkeiten für zulässige Planungsergebnisse. ○ Wahrscheinlichkeit für Beschleunigung mit Chebyshev-Saw-Ungleichung (3.13) für Schranke $\bar{b}_a = 0.8 \text{ m/s}$; ▲ Wahrscheinlichkeit für FSW mit Bernoulli-a-posteriori-Ungleichung (3.15) für Schranke $\bar{b}_{lc} = 0.2$.

Mit der oberen zulässigen Schranke $\bar{b}_{lc} = 0.2$ ergibt sich mit Abschätzung (3.15) die obere Schranke

$$\begin{aligned}
 & \mathbb{P}_{f_{lc}(\mathbf{q}+\hat{\mathbf{Q}})} \left(\left| f_{lc}(\mathbf{q} + \hat{\mathbf{q}}) - \mathbb{E}(f_{lc}(\mathbf{q} + \hat{\mathbf{Q}})) \right| > 0.2 \right) \\
 &= \mathbb{P} \left(\left| f_{lc}(\mathbf{q} + \hat{\mathbf{q}}) - \mathbb{E}(f_{lc}(\mathbf{q} + \hat{\mathbf{Q}})) \right| \geq 0.2 \right) \\
 &= 0.2 \left(\mathbb{P}(p_{lc} \leq 0.2) + \mathbb{P}(0.8 \leq p_{lc}) \right) + \mathbb{P}(0.2 < p_{lc} < 0.8) \\
 &= 0.9568
 \end{aligned}$$

für die Wahrscheinlichkeit von unzulässigen Planungsentscheidungen ergibt. Der Planungsalgorithmus berechnet in dieser Situation also keine zulässigen Planungsergebnisse. Analog dazu wurden die Wahrscheinlichkeiten für unzulässige Beschleunigungs- und FSW-Entscheidungen mit der Ungleichung (3.13) und (3.15) berechnet. Die Ergebnisse für den Zeitraum $[0 \text{ s}, 5 \text{ s}]$ sind in Abb. 3.7 dargestellt. Es wird deutlich, dass für $t \in [0 \text{ s}, 4 \text{ s}]$, d. h. im Zeitraum vor und während des FSW, sowohl für die Beschleunigung als auch für den FSW die Wahrscheinlichkeit dafür, dass die Planungsentscheidung mehr als $\bar{b}_a = 0.8 \text{ m/s}$ bzw. $\bar{b}_{lc} = 0.2$ von der erwarteten Entscheidung abweicht, größer ist als $\bar{p}_a = 0.1$ bzw. $\bar{p}_{lc} = 0.2$. Für $t > 4 \text{ s}$ wird die Wahrscheinlichkeit für einen FSW nicht mehr angezeigt, da das Ego auf dem linken Fahrstreifen lokalisiert wird und ein weiterer FSW nach links nicht möglich ist, weil kein weiterer Fahrstreifen existiert.

Durch die Ungleichungen (3.13) und (3.15) gelingt es nun, in einer Simulation abzuschätzen, ob die Messunsicherheiten zu starken Schwankungen in den Planungsergebnissen führen. Falls dies nur mit maximaler Wahrscheinlichkeit \bar{p} geschieht, so sind die Messunsicherheiten akzeptabel. Andernfalls muss genauer analysiert werden, welche Messunsicherheit den größten Einfluss auf die Planungsabweichungen hat (vgl. Beispiel 3.2.6.b). Dafür wird in Abschnitt 3.3 Morris' Methode der elementaren Effekte vorgestellt.

3.3 Sensitivitätsanalyse nach Morris

Sensitivitätsanalysen (SA) werden dafür genutzt, um den Einfluss eines Inputs auf das Ergebnis einer Funktion zu beschreiben, d. h. es wird untersucht, wie stark die Änderung eines Inputs das Ergebnis verändert. Im Allgemeinen werden SA in 2×2 Gruppen unterteilt: global vs. lokal und quantitativ vs. qualitativ. Lokale SA konzentrieren sich auf die Approximation von partiellen Ableitungen in einem Punkt (Iooss & Lemaître, 2015). Sie bieten daher nur eine begrenzte Aussagekraft über die Auswirkungen größerer Schwankungen eines Inputs. Globale Methoden untersuchen den Definitionsbereich der Inputs in dem gegebenen stochastischen Zusammenhang weiträumig (Iooss & Lemaître, 2015) und sind daher belastbarer. Iooss und Lemaître (2015) geben einen umfassenden Überblick über verschiedene globale SA. Der Unterschied zwischen quantitativen und qualitativen Ansätzen liegt in der Interpretation ihrer Ergebnisse: Quantitative Methoden zerlegen die Unsicherheit des Ergebnisses und ordnen diese den einzelnen unsicheren Inputs anteilig zu (Campolongo, Cariboni & Saltelli, 2007). Damit lässt sich der Einfluss aller Inputs klar quantifizieren, d. h. genau aussagen, wie viel größer der Einfluss eines Inputs als der eines anderen ist (Campolongo, Tarantola & Saltelli, 1999). Im Gegensatz dazu bieten qualitative Methoden eine weniger rechenintensive Möglichkeit, um einflussreiche von unwichtigen Inputs zu trennen und die Inputs entsprechend ihres grob geschätzten Einflusses zu sortieren Campolongo et al. (1999).

Da in dieser Arbeit mit dem automatisierten Fahren eine Anwendung betrachtet wird, bei der Rechenzeit eine knappe Ressource ist, wird Morris' Methode der Elementaren Effekte ausgewählt (Morris, 1991): Diese SA ist eine globale, qualitative Methode, die die Inputs einer Funktion entweder als relevant oder irrelevant klassifiziert. Morris' Methode wird in verschiedenen Bereichen

angewendet, z. B. um den Einfluss chemischer Reaktionen auf den Klimawandel zu untersuchen (Campolongo et al., 2007) oder in der Simulation von Energieverteilung in Gebäuden (Menberg, Heo, Augenbroe & Choudhary, 2016). Fraikin, Funk, Frey und Gauterin (2019, 2020) nutzen Morris' Methode, um die Parameter einer automatisierten Fahrfunktion zu kalibrieren: Mithilfe der SA identifizieren sie die Inputs, die keinen Einfluss auf die Fahrfunktion haben, fixieren diese und reduzieren so den Suchraum für die anschließende Optimierung, die die Parameter kalibrieren soll.

Die Grundlage der SA bildet eine einfache Überlegung: Für eine stetige Funktion $f : [0, 1]^k \rightarrow \mathbb{R}$ seien die Inputs paarweise unabhängig $\mathcal{U}([0, 1])$ -verteilt. Dann gibt die Sekantensteigung $(f(\mathbf{x} + \Delta_i \mathbf{e}_i) - f(\mathbf{x})) / \Delta_i$ zwischen zwei Punkten \mathbf{x} und $\mathbf{x} + \Delta_i \mathbf{e}_i$ die Änderungsrate der Funktion in Richtung i zwischen den Punkten \mathbf{x} und $\mathbf{x} + \Delta_i \mathbf{e}_i$ an. Durch die Verteilung der Inputs hat auch die Sekantensteigung eine Verteilung, die mithilfe von stochastischen Kenngrößen charakterisiert werden kann. Bevor dargestellt wird, wie diese zur Unterscheidung zwischen relevanten und irrelevanten Inputs verwendet wird, werden einige Anpassungen erläutert, die für die Anwendung auf Planungsalgorithmen notwendig sind. Rutjens, Band, Jones und Owen (2022) stellen vor, wie Morris' Methode auf Modelle angewendet werden kann, deren Inputs nicht $\mathcal{U}([0, 1])$ -verteilt sind. Für andere W-Verteilungen schlagen sie vor, dass zwei Stichprobenwerte $\mathbf{x}, \mathbf{x} + \Delta_i \mathbf{e}_i$ mithilfe von z. B. der inversen Verteilungsfunktion in andere W-Räume projiziert werden können (Kolonko, 2008, Kapitel 8 Inversionsmethode). Der Nachteil dieses Vorschlags ist, dass die Differenz der beiden projizierten Stichprobenwerte stark schwanken kann, obwohl sie im Ausgangsstadium nur einen Unterschied von Δ aufweisen. Insbesondere am Rand der Verteilung kann es dadurch zu Verzerrungen kommen, vgl. Abb. 3.8. Daher wird in dieser Arbeit auf die Transformation einer $\mathcal{U}([0, 1])$ -verteilten Stichprobe in eine normalverteilte Stichprobe verzichtet. Stattdessen werden direkt $\mathcal{N}(0, \sigma_i^2)$ -verteilte Stichprobenwerte gezogen. Als Abstand für die Sekantensteigung wird die Standardabweichung $\Delta_i = \sigma_i$ gewählt.

Eine weitere Ergänzung von Rutjens et al. (2022) zu Morris' Methode ist die Skalierung der Inputs. Sie schlagen vor, dass für Inputs, deren Größenordnungen sich stark unterscheiden, eine lineare Skalierung von $[\min x_i, \max x_i] \rightarrow [0, 1]$ sinnvoll sein kann, um die Größenordnungen der Elementaren Effekte (EE) besser vergleichen zu können. In dieser Anwendung ist dies aber nicht zwingend notwendig, da für die Messabweichungen für Geschwindigkeiten und Distanzen

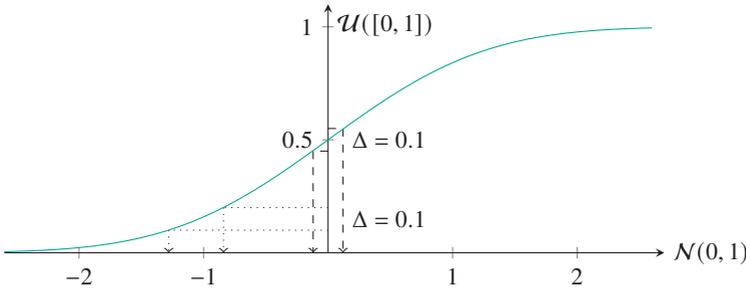


Abbildung 3.8: Inversionsmethode für Standardnormalverteilung $\mathcal{N}(0, 1)$. Es werden Stichprobenwerte nach $\mathcal{U}([0, 1])$ gezogen, z. B. $\{0.1, 0.45\}$ und die zugehörigen Werte mit Differenz $\Delta = 0.1$ berechnet: $\{0.1 + 0.1, 0.45 + 0.1\}$. Anschließend werden diese mit der Verteilungsfunktion $\Phi_{\mathcal{N}(0,1)}$ der Standardnormalverteilung $\mathcal{N}(0, 1)$ in den anderen W-Raum projiziert: $\Phi_{\mathcal{N}(0,1)}(-1.2816) = 0.1$, $\Phi_{\mathcal{N}(0,1)}(-0.8416) = 0.2$, $\Phi_{\mathcal{N}(0,1)}(-0.1257) = 0.45$, $\Phi_{\mathcal{N}(0,1)}(0.1257) = 0.55$. Die Differenzen zwischen den projizierten Werten unterscheiden sich also je nach Lage: $-0.8416 + 1.2816 = 0.4399 > 0.2514 = 0.1257 + 0.1257$. Die EE werden also am Rand der Verteilung verzerrt.

dieselben Größenordnungen angenommen werden. Sollte sich das ändern, könnte eine detailliertere Betrachtung der Skalierung und Entdimensionalisierung des Modells sinnvoll sein (vgl. dazu auch Sin & Germaey, 2009).

In der Herleitung der Methode nach Morris (1991) wird die Stetigkeit der zu analysierenden Funktion vorausgesetzt. Dieses ist bei Planungsfunktionen meistens nicht global gegeben. Da es aber das Ziel ist, dass Planungsentscheidungen lokal nur kleinen Schwankungen unterliegen, kann angenommen werden, dass es für jeden Punkt \mathbf{q} eine Umgebung $U(\mathbf{q})$ gibt, auf der die Planungsfunktion lokal stetig ist und innerhalb derer die SA nach Morris angewendet werden kann. Zusammengefasst werden für die Analyse eines Planungsalgorithmus \mathbf{f} zum Zeitpunkt t für die Messgröße \mathbf{q}_t Messabweichungen verwendet, die durch die normalverteilte ZV $\hat{\mathbf{Q}}_t$ modelliert werden. Daher kann für eine Planungsentscheidung f_o die ZV

$$d_{o,i,t}(\hat{\mathbf{Q}}_t) := \frac{f_o(\mathbf{q}_t + \hat{\mathbf{Q}}_t + \Delta_i \mathbf{e}_i) - f_o(\mathbf{q}_t + \hat{\mathbf{Q}}_t)}{\Delta_i}$$

definiert werden. Um nun relevante Inputs von irrelevanten Inputs zu unterscheiden, soll die W -Verteilung dieser ZV beschrieben werden, z. B. durch ihren Erwartungswert und ihre Varianz.

Beispiel 3.1.8.e (Freifahrt: Erwartungswert und Varianz der EE in der Sensitivitätsanalyse (analytische Herleitung)). Für Beispiel 3.1.8 kann dank der bekannten W -Dichte die Verteilung der ZV $d_{a,v,t}(\hat{V})$ für die Beschleunigung $o = a$ und den Input Ego-Geschwindigkeit $i = v$ charakterisiert werden. Für die ZV $\hat{V} \sim \mathcal{N}(0, \sigma_v^2)$ und eine Realisierung \hat{v} gilt

$$\begin{aligned} d_{a,v,t}(\hat{v}) &= \frac{f_a(v_t + \hat{v} + \Delta_v) - f_a(v_t + \hat{v})}{\Delta_v} \\ &= \frac{1}{\Delta_v} \left(\bar{a} \left(1 - \frac{(v_t + \hat{v} + \Delta_v)^4}{v_w^4} \right) - \bar{a} \left(1 - \frac{(v_t + \hat{v})^4}{v_w^4} \right) \right) \\ &= \frac{\bar{a}}{\Delta_v v_w^4} \left((v_t + \hat{v})^4 - (v_t + \hat{v} + \Delta_v)^4 \right), \end{aligned} \quad (3.16)$$

sodass für den Erwartungswert und die Varianz der EE für $\bar{a} = 1.5 \text{ m/s}^2$, $v_w = 30 \text{ km/h}$, $v_t = 25 \text{ km/h}$ und $\Delta_v = \sigma_v = 1.3 \text{ m/s}^2$

$$\begin{aligned} \mathbb{E}(d_{a,v,t}(\hat{V}_{t,j})) &= -\frac{\bar{a}}{v_w^4} \left(4v_t^3 + 16v_t\sigma_v^2 \right) = -0.4751 \text{ s}^{-1} \\ \mathbb{V}\text{ar}(d_{a,v,t}(\hat{V}_{t,j})) &= \frac{\bar{a}^4}{v_w^8} \left(180v_t^4\sigma_v^2 + 900v_t^2\sigma_v^4 + 473\sigma_v^6 \right) = 0.0807 \text{ s}^{-2} \end{aligned}$$

gilt (siehe Anhang A.1 für eine Herleitung der beiden Größen). Das bedeutet, dass das AF im Mittel -0.4751 m/s^2 weniger beschleunigt, wenn seine Geschwindigkeit v um 1 m/s höher ist und in diesem Fall die Abweichungen im Mittel um 0.2840 m/s^2 um den Erwartungswert streuen (Standardabweichung).

Wie bereits in Abschnitt 3.2 diskutiert, ist die analytische Bestimmung des Erwartungswertes und der Varianz der ZV $d_{o,i,t}(\hat{\mathbf{Q}}_t)$ im Allgemeinen nicht möglich, weil i. d. R. zu wenig über die Planungsfunktion \mathbf{f} bekannt ist. Stattdessen werden der Erwartungswert und die Varianz mithilfe einer Stichprobe approximiert. Da eine einzelne Stützstelle $\mathbf{q}_t + \hat{\mathbf{q}}_t$ nur lokale Aussagen zur Änderungsrate treffen kann, bestimmt Morris' Methode die Sekantensteigung an $M \geq 30$ Stützstellen (vgl. Abb. 3.9). Dadurch wird das Verfahren eine globale SA (Kucherenko, Rodriguez-Fernandez, Pantelides & Shah, 2009; Saltelli et

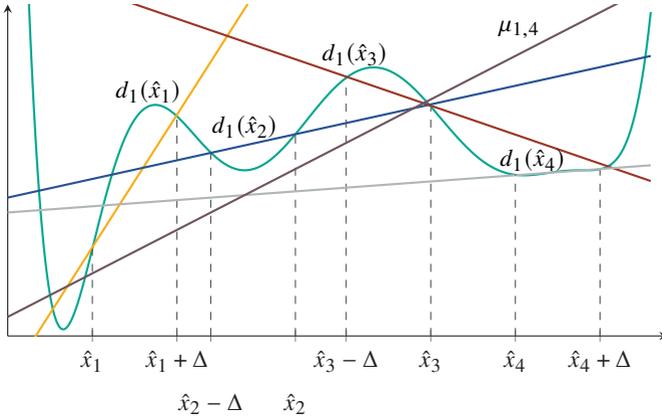
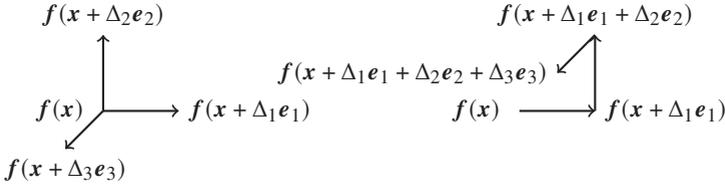


Abbildung 3.9: Beispiel für EE. – Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$; – Stichprobenmittelwert $\mu_{1,4}$ für $M = 4$ Stichprobenwerte; die restlichen Sekanten sind die EE für die Stichprobe $\{\hat{x}_j\}_{j=1}^4$ mit Abstand Δ .



(a) Radiales Design (Campolongo et al., 2011).

(b) Wendeltreppen-Design (Saltelli et al., 2010).

Abbildung 3.10: Trajektoriengenerierung für Morris' Methode der EE für $k = 3$. Insgesamt wird die Funktion f für die Berechnung aller EE $(k + 1)$ -mal für jeden Stichprobenwert \mathbf{x} ausgewertet.

al., 2010). In der Literatur werden zahlreiche Sampling-Strategien vorgeschlagen (Campolongo et al., 2007; Morris, 1991; Rutjens et al., 2022), die das gemeinsame Ziel teilen, den Suchraum mithilfe von quasi-zufälligen Sequenzen möglichst effizient abzutasten. Untersuchungen haben jedoch gezeigt, dass die Unterschiede zwischen den Methoden zur Stichprobengenerierung nicht konsistent für verschiedene Testfunktionen beobachtbar sind (Rußer, 2021; Rutjens et

al., 2022). Daher sind die vorläufigen Ergebnisse nicht belastbar genug und es wird auf eine Monte-Carlo-Strategie zurückgegriffen, um die Stichprobenwerte zu erzeugen. Ausgehend von der so erzeugten Stichprobe $\{\mathbf{x}_j\}_{j=1}^M$ werden nun die zugehörigen Stützstellen $\{\mathbf{x}_j + \Delta_i \mathbf{e}_i\}_{j=1}^M$ für $i = 1, \dots, k$ erzeugt. Dafür kann das radiale Design verwendet werden, bei dem ausgehend von einem Basispunkt \mathbf{x}_j immer nur eine Komponente in eine Richtung verändert wird. So entsteht ein stern-förmiges Muster (vgl. Abb. 3.10(a)). Laut Saltelli et al. (2010) ist diese Vorgehensweise eine angemessene Alternative zum Wendeltreppen-Design, bei dem immer eine Komponente nach der nächsten verändert wird, sodass sich ein kreisförmiges Muster ergibt (vgl. Abb. 3.10(b)). Dies wird für $M \geq 30$ Stichproben durchgeführt. Für die Analyse einer Planungsfunktion f_o wird nun die Menge $\{\mathbf{q}_t + \hat{\mathbf{q}}_{t,j}\}_{j=1}^M$ an Messergebnissen verwendet, sodass eine Realisierung $d_{o,i,t}(\hat{\mathbf{q}}_{t,j})$ der ZV $d_{o,i,t}(\hat{\mathbf{Q}}_t)$ definiert ist durch

$$d_{o,i,t}(\hat{\mathbf{q}}_{t,j}) := \frac{f_o(\mathbf{q}_t + \hat{\mathbf{q}}_{t,j} + \Delta_i \mathbf{e}_i) - f_o(\mathbf{q}_t + \hat{\mathbf{q}}_{t,j})}{\Delta_i}. \quad (3.17)$$

Im Folgenden bezeichnet der Begriff EE immer eine Realisierung (3.17) der ZV und nicht die zugehörige ZV selbst. Um die Sensitivität der Planungsfunktion nicht nur lokal in einem einzelnen Punkt zu messen, werden die EE $\{d_{o,i,t}(\hat{\mathbf{q}}_{t,j})\}_{j=1}^M$ durch die Größen $\mu_{o,i,t}$ und $\sigma_{o,i,t}^2$

$$\mu_{o,i,t} := \frac{1}{M} \sum_{j=1}^M d_{o,i,t}(\hat{\mathbf{q}}_{t,j}), \quad (3.18)$$

$$\sigma_{o,i,t}^2 := \frac{1}{M-1} \sum_{j=1}^M \left(d_{o,i,t}(\hat{\mathbf{q}}_{t,j}) - \mu_{o,i,t} \right)^2. \quad (3.19)$$

zusammengefasst, die die Menge $\{d_{o,i,t}(\hat{\mathbf{q}}_{t,j})\}_{j=1}^M$ charakterisieren. Die mittlere Änderungsrate $\mu_{o,i,t}$ beschreibt, wie stark sich die Funktion f_o im Mittel in Richtung i verändert. Die Streuung $\sigma_{o,i,t}$ der Sekantensteigung um ihren Mittelwert $\mu_{o,i,t}$ gibt an, wie groß der Einfluss anderer Inputs oder durch Nichtlinearitäten in der Funktion f_o ist. Mit ihrer Hilfe kann zwischen relevanten und irrelevanten Inputs unterschieden werden:

- Falls $\mu_{o,i,t} \approx \sigma_{o,i,t} \approx 0$, so beeinflusst Input i den Output o nicht.

- Falls $\mu_{o,i,t} \neq 0$ und $\sigma_{o,i,t} \approx 0$, so beeinflusst Input i den Output o linear, z. B. steigt die Beschleunigung f_a (d. h. $o = a$) zum Zeitpunkt t linear mit Input i , falls $\mu_{a,i,t} > 0$ größer 0 ist und $\sigma_{o,i,t} \approx 0$.
- Falls $\mu_{o,i,t} \neq 0$ und $\sigma_{o,i,t} > 0$, so beeinflusst Input i das Ergebnis nicht-linear oder es fließen zusätzlich auch noch andere Inputs $\tilde{i} \neq i$ in die Berechnung ein.

Zusammengefasst beschreibt die erste Kategorie also irrelevante Inputs, während die zweite und dritte Kategorie relevante Inputs umfasst. Dies führt dazu, dass die Methode der EE niemals Fehler 1. Art begeht, also Inputs als einflussreich einstuft, die tatsächlich keinen Einfluss haben (falsch-positiv Saltelli et al., 2010). Fehler 2. Art, also dass Inputs als irrelevant gewertet werden, obwohl sie tatsächlich einflussreich sind, können hingegen bei zu geringer Stichprobengröße auftreten.

Beispiel 3.1.8.f (Freifahrt: Elementarer Effekt in der Sensitivitätsanalyse (Simulation)). In Beispiel 3.1.8.e konnten der Erwartungswert und die Varianz für die Freifahrt analytisch berechnet werden. Für eine Simulation mit $M = 300$ Stichprobenwerten können die mittlere Änderungsrate $\mu_{a,v,t}$ und die Streuung $\sigma_{a,v,t}$ mit den Definitionen (3.18) und (3.19) auch approximiert werden durch

$$\begin{aligned}\mu_{a,v,t} &= -0.4654 \text{ s}^{-1}, \\ \sigma_{a,v,t} &= 0.0770 \text{ s}^{-2}.\end{aligned}$$

Die Simulation mit $M = 300$ Messergebnissen approximiert die charakteristischen Größen $\mathbb{E}(d_{a,v,t}(\hat{V}_{t,j})) = -0.4751 \text{ s}^{-1}$ und $\mathbb{V}\text{ar}(d_{a,v,t}(\hat{V}_{t,j})) = 0.0807 \text{ s}^{-2}$ also ausreichend gut.

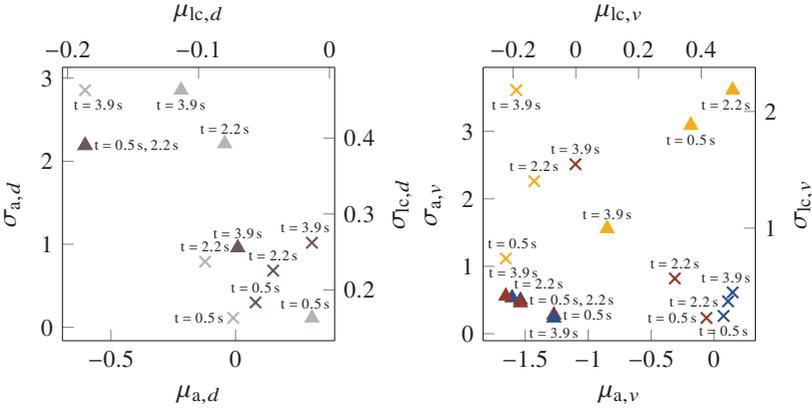
Campolongo et al. (2007) schlagen eine weitere charakteristische Größe $\mu_{o,i,t}^*$ vor, die anstelle der Streuung $\sigma_{o,i,t}$ anzeigt, ob ein Parameter einen linearen oder nicht-linearen bzw. interagierenden Einfluss ausübt. Er ist definiert durch

$$\mu_{o,i,t}^* := \frac{1}{M} \sum_{j=1}^M |d_{o,i,t}(\hat{\mathbf{q}}_{t,j})|, \quad (3.20)$$

d. h. als Stichprobenerwartungswert der absoluten EE. Relevante Inputs haben einen linearen Einfluss, wenn $\mu_{o,i,t}^* \approx |\mu_{o,i,t}|$ und einen nicht-linearen oder

interagierenden Einfluss, wenn $\mu_{o,i,t}^* > |\mu_{o,i,t}|$ gilt. Wenn nur zwischen relevanten und irrelevanten Parametern unterschieden werden soll, ohne ihren Einfluss genauer zu charakterisieren, ist die Berechnung von $\mu_{o,i,t}^*$ rechengünstiger als die von $\sigma_{o,i,t}^2$, da für die Stichprobenstandardabweichung σ_t^2 zunächst noch der Stichprobenerwartungswert $\mu_{o,i,t}$ bestimmt werden muss (Campolongo et al., 2007).

Beispiel 3.2.6.c (Überholmanöver: Elementarer Effekt in der Sensitivitätsanalyse (Simulation)). Beispiel 3.2.6.b hat gezeigt, dass die Messunsicherheiten $\sigma_v = 0.1 \text{ m/s}$, $\sigma_{d_{nl}} = 0.5 \text{ m}$, $\sigma_{d_{vs}} = 1 \text{ m}$, $\sigma_{v_{nl}} = \sigma_{v_{vs}} = 1 \text{ m/s}$ im Zeitraum $[0 \text{ s}, 5 \text{ s}]$ zu groß sind. Daher wird für dieses Überholmanöver eine SA mit $M = 300$ Stichprobenwerten gerechnet (Henze et al., 2020), die die folgenden Planungsinputs analysiert: die Ego-Geschwindigkeit v , die Geschwindigkeit v_{vs} des vorausfahrenden Fahrzeugs und die Distanz d_{vs} zum vorausfahrenden Fahrzeug auf demselben Fahrstreifen, die Geschwindigkeit v_{nl} des nachfolgenden Fahrzeugs und Distanz d_{nl} zum nachfolgenden Fahrzeug auf dem linken Fahrstreifen des Egos. Abb. 3.11 zeigt die charakteristischen Größen $\mu_{o,i}$ und $\sigma_{o,i}$ für den Einfluss der o. g. Planungsinputs i auf die Planungsergebnisse Beschleunigung oder FSW $o \in \{a, lc\}$ exemplarisch für die Zeitpunkte $t \in \{0.5 \text{ s}, 2.2 \text{ s}, 3.9 \text{ s}\}$. Da sich das Szenario mit der Zeit entwickelt, verändert sich auch der Einfluss der Messunsicherheiten und damit die charakteristischen Größen, eine Übersicht für den gesamten Zeitraum findet sich in Anhang A.2. Wie oben bereits beschrieben, haben Inputs mit $\mu_{o,i,t} \approx \sigma_{o,i,t} \approx 0$ keinen Einfluss, während alle anderen als relevant eingestuft werden. Die Distanzen haben für $t = 0.5 \text{ s}$ einen geringeren Einfluss auf die Berechnung der Ego-Beschleunigung, mit fortschreitender Zeit ($t = 2.2 \text{ s}, 3.9 \text{ s}$) nimmt dieser jedoch weiter zu: Die negativen Vorzeichen von $\mu_{lc,d_{nl}}$ und $\mu_{a,d_{nl}}$ zeigen an, dass das Ego den FSW abbrechen und damit schwächer beschleunigen würde, falls die Distanz zum nachfolgenden Fahrzeug kleiner wäre (vgl. z. B. $t = 3.9 \text{ s}$), während das nachfolgende Fahrzeug zum Zeitpunkt $t = 0.5 \text{ s}$ weder die Berechnung der Beschleunigung noch des FSWs beeinflusst (vgl. $\mu_{lc,d_{nl}}, \mu_{a,d_{nl}}, \mu_{lc,v_{nl}}$ und $\mu_{a,v_{nl}}$). Im Vergleich dazu hat die Ego-Geschwindigkeit v zu allen Zeitpunkten $t \in \{0.5 \text{ s}, 2.2 \text{ s}, 3.9 \text{ s}\}$ einen großen Einfluss sowohl auf seine Beschleunigung als auch auf den FSW: Das negative Vorzeichen von $\mu_{a,v}$ bedeutet, dass das Ego schwächer beschleunigen würde, falls es eine größere Geschwindigkeit hätte ($t \in \{0.5 \text{ s}, 0.5 \text{ s}, 0.5 \text{ s}\}$). Insgesamt kann die SA also herausarbeiten, welche Inputs zu welchem Zeitpunkt relevant sind.



(a) SA-Ergebnisse für Distanzen $d \in \{d_{vs}, d_{nl}\}$. (b) SA-Ergebnisse für Geschwindigkeiten $v \in \{v, v_{vs}, v_{nl}\}$.

Abbildung 3.11: Ergebnisse der SA für Beispiel 3.2.6.c (Überholmanöver) zu den Zeitpunkten $t \in \{0.5\text{ s}, 2.1\text{ s}, 3.9\text{ s}\}$ für $M = 300$ Stichprobenwerten. Einfluss der Inputs auf die Beschleunigung $o = a$ (Achsen unten und links): \times d_{vs} , \times d_{nl} , \times v , \times v_{vs} , \times v_{nl} ; Einfluss der Inputs auf den FSW $o = lc$ (Achsen oben und rechts) FSW $o = lc$: \blacktriangle d_{vs} , \blacktriangle d_{nl} , \blacktriangle v , \blacktriangle v_{vs} , \blacktriangle v_{nl} .

Wie Beispiel 3.2.6.c gezeigt hat, kann sich ein Szenario mit fortschreitender Zeit verändern und damit ändert sich i. d. R. auch die Relevanz der Inputs. Daher können Messungenauigkeiten zu einem Zeitpunkt akzeptabel sein, zu einem anderen Zeitpunkt jedoch zu unzulässigen Planungsentscheidungen führen. In diesen Fällen muss untersucht werden, wie die Schnittstelle zwischen Wahrnehmung und Planung verbessert werden kann. Dafür gibt es zwei Möglichkeiten: Einerseits kann die Wahrnehmung verbessert werden, andererseits der Planer robuster werden gegenüber den Messabweichungen. In beiden Fällen muss genau herausgefunden werden, welche Inputs den größten Teil an Abweichungen in den Ergebnissen verursachen und daher verbessert werden sollten.

Zusammengefasst hat Kapitel 3 dargestellt, wie der Einfluss von unsicherheitsbehafteten Informationen auf das Planungsergebnis eines AF analysiert werden kann. Dafür wurden zunächst zulässige Planungsergebnisse so definiert, dass sie nur mit einer geringen Wahrscheinlichkeit stärker als ein Grenzwert von der

erwarteten Fahrentscheidung abweichen. Da diese stochastische Ungleichung aufgrund der komplexen Planungsfunktion in der Regel nicht analytisch ausgewertet werden kann, wurden zwei Abschätzungen hergeleitet, die sowohl für kontinuierliche Fahrentscheidungen, z. B. Beschleunigung, als auch für diskrete Entscheidungen, z. B. FSW, eine obere Schranke angeben. Falls diese nicht erfüllt werden, die Planungsentscheidung also nicht zulässig ist, kann die SA nach Morris genutzt werden, um die Informationen zu identifizieren, die durch ihre Messunsicherheiten den größten Einfluss auf die Planungsentscheidung haben. Da dies ein rechenintensives Verfahren sein kann, stellt Kapitel 4 ein dynamisches Abbruchkriterium vor, das den Rechenaufwand reduziert.

4 Dynamisches Abbruchkriterium

Wie Kapitel 3 aufgezeigt hat, ist die verwendete SA nach Morris sampling-basiert. Dies führt dazu, dass die Analyse eines Planungsmoduls sehr rechenintensiv sein kann. Um diesen Nachteil abzumildern, stellt Abschnitt 4.1 das dynamische Abbruchkriterium (DSC, dynamic stop criterion) vor, mit dem die Anzahl der ausgewerteten Stichprobenwerte iterativ an die Komplexität der Situation angepasst werden kann. Abschnitt 4.2 zeigt die Anwendung des DSC auf ein akademisches Beispiel (Abschnitt 4.2.1) und im Kontext des automatisierten Fahrens (Überholmanöver, Abschnitt 4.2.2).

Teile der Ergebnisse¹ wurden vorläufig im Rahmen der 10th International Conference on Sensitivity Analysis and Model Output 2022 (Henze, Rußer et al., 2022) vorgestellt. Abbildungen, Tabellen und andere Ergebnisse wurden für eine konsistente Darstellung angepasst.

4.1 Definition

Die zentrale Idee von Morris' Methode der Elementaren Effekte besteht darin, die Sekantensteigung an zufällig gewählten Punkten in Richtung der Einheitsvektoren e_i ($i = 1, \dots, k$) zu berechnen (vgl. Gleichung (3.17)) und anhand dieser den Einfluss der Inputs als relevant oder irrelevant einzustufen. Abbildung 3.10 veranschaulicht für $k = 3$ Inputs, dass für die Berechnung aller k EE $\{d_{o,i}\}_{i=1}^k$ eines zufällig gewählten Punktes x mindestens $k + 1$ Funktionsauswertungen notwendig sind². Für eine Stichprobe mit $M \in \mathbb{N}$ Stichprobenwerten ergeben sich also $M(k + 1)$ Berechnungen von f , um die

¹Die Idee für das DSC ist in Zusammenarbeit mit Markus Rußer im Rahmen seiner Masterarbeit (Rüßer, 2021) entstanden.

²Im radialen Design (vgl. Abb. 3.10) ist mit $f(x)$ eine Funktionsauswertung der SA fixiert, während alle anderen k Auswertungen mit einem Input i und dem EE in Richtung e_i assoziiert sind.

charakteristischen Größen $\mu_{o,i}$, $\mu_{o,i}^*$ und $\sigma_{o,i}$ für ein Ergebnis o zu bestimmen. Da die Stichprobengröße $M \in \mathbb{N}$ a priori ohne Wissen über die zu untersuchende Funktion festgelegt wird, kann die Sensitivitätsanalyse für eine große Anzahl $k \gg 0$ an Inputs rechenintensiv sein. Daher soll das DSC die Stichprobengröße an die Komplexität der Situation anpassen, die analysiert wird. Der Vorteil daran ist, dass nur in den Fällen, in denen es notwendig ist, viele EE ausgewertet werden, während in allen anderen Fällen möglichst viel Rechenzeit eingespart wird. Dafür wird ein Abbruchkriterium definiert, das die Schwankung der charakteristischen Größen $\mu_{o,i}$, $\mu_{o,i}^*$ und $\sigma_{o,i}$ misst: Touzani und Busby (2014) führen dazu ein

Algorithm 4.1 dynamisches Abbruchkriterium (DSC)

Input: Mindeststichprobengröße M_{\min} , Konstanten $\kappa_{\text{stop}} > \kappa_{\text{act}} > 0$, Δ_i , Funktion $f : \mathbb{R}^k \rightarrow \mathbb{R}^m$, Stichprobe $\{\hat{x}_j\}_{j=1}^{M_{\min}}$

Output: relatives Residuum $\epsilon_M^2 \leq \kappa_{\text{stop}}$

- 1: EE $d_i(\hat{x}_j) \leftarrow (f(\hat{x}_j + \Delta_i e_i) - f(\hat{x}_j)) / \Delta_i \forall j = 1, \dots, M_{\min}, i = 1, \dots, k$
 - 2: $M \leftarrow M_{\min}$
 - 3: $\mu_{o,i,M} \leftarrow (3.8)$, $\mu_{o,i,M}^* \leftarrow (3.20)$, $\sigma_{o,i,M}^2 \leftarrow (3.9) \forall o = 1, \dots, m, \forall i$
 - 4: $\epsilon_{o,i,M}^2 \leftarrow (4.1)$, $\epsilon_M^2 \leftarrow (4.3) \forall o, i$
 - 5: $\mathcal{A}_M \leftarrow \{i \in \{1, \dots, k\} \mid \epsilon_{o,i,M}^2 > \kappa_{\text{act}} \forall o = 1, \dots, m\}$
 - 6: $M_i \leftarrow M_{\min}$ für alle $i \in \{1, \dots, k\} \setminus \mathcal{A}_M$
 - 7: **while** $\epsilon_M^2 > \kappa_{\text{stop}}$ **do**
 - 8: ziehe neuen Stichprobenwert \hat{x}_{M+1}
 - 9: **for** $i \in \mathcal{A}_M$ **do**
 - 10: $d_i(\hat{x}_{M+1}) \leftarrow f(\hat{x}_{M+1})$
 - 11: aktualisiere $\mu_{o,i,M+1}$, $\mu_{o,i,M+1}^*$, $\sigma_{o,i,M+1}$, $\epsilon_{o,i,M+1}^2$, ϵ_M^2
 - 12: **if** $\epsilon_{o,i,M+1}^2 \leq \kappa_{\text{act}} \forall o = 1, \dots, m$ **then**
 - 13: $\mathcal{A}_{M+1} \leftarrow \mathcal{A}_M \setminus \{i\}$
 - 14: $M_i \leftarrow M + 1$
 - 15: **end if**
 - 16: **end for**
 - 17: $M \leftarrow M + 1$
 - 18: **end while**
-

Aufgrund dieses linearen Zusammenhangs werden die Begriffe *Anzahl der Funktionsauswertungen* und *Anzahl der EE* in diesem Kapitel synonym verwendet.

globales Fehlerkriterium ein, um die Konvergenz des Sensitivitätsindex eines quantitativen gradienten-basierten globalen Sensitivitätsmaßes (engl. *derivative-based global sensitivity measure*, DGSM) zu untersuchen. Im Folgenden wird analog dazu die Konvergenz der qualitativen SA-Maße $\mu_{o,i,M}$, $\mu_{o,i,M}^*$ und $\sigma_{o,i,M}$ genutzt, um die Stichprobengröße an die Schwierigkeit der Funktion und Situation anzupassen. Der Pseudocode ist in Algorithmus 4.1 dargestellt. Das DSC basiert auf der Annahme, dass es in jeder Situation drei Teilmengen an Inputs gibt:

1. eindeutig relevante Inputs: Inputs, die schon nach der Auswertung weniger EE eindeutig als relevant für die Planungsentscheidung klassifiziert werden
2. eindeutig irrelevante Inputs: Inputs, die schon nach der Auswertung weniger EE eindeutig als irrelevant für die Planungsentscheidung klassifiziert werden
3. nicht eindeutige Inputs: Inputs, deren Relevanz gering ist und die daher bei einer geringen Stichprobengröße möglicherweise falsch kategorisiert werden

Für Inputs der ersten beiden Kategorien reichen bereits wenige EE aus, um die Inputs eindeutig einer der Kategorien relevant oder irrelevant zuzuordnen. Lediglich für die dritte ist es sinnvoll, mehr Stichprobenwerte auszuwerten, um dadurch Fehlinterpretationen zu vermeiden. Daher kann die Anzahl an Funktionsauswertungen reduziert werden, indem nur für die letzte Kategorie viele EE ausgewertet werden (aktive Inputs), während für die ersten beiden Kategorien die Auswertung gestoppt wird, sobald sie zweifellos als relevant bzw. irrelevant eingestuft wurden (inaktive Inputs). Da a priori unbekannt ist, in welche der drei Kategorien ein Input gehört, wird die Konvergenz seiner

zugehörigen charakteristischen Größen $\mu_{o,i,M}$, $\mu_{o,i,M}^*$ und $\sigma_{o,i,M}$ betrachtet. Dafür werden die Matrizen $\nu_{\mu,M,M-l}$, $\nu_{\mu^*,M,M-l}$, $\nu_{\sigma,M,M-l} \in \mathbb{R}^{m \times k}$ definiert:

$$\nu_{\mu,M_i,M_i-l,o,i} = \begin{cases} 0 & \mu_{o,i,M_i} = 0, \\ \frac{|\mu_{o,i,M_i-l} - \mu_{o,i,M_i}|}{|\mu_{o,i,M_i}|} & \text{sonst,} \end{cases}$$

$$\nu_{\mu^*,M_i,M_i-l,o,i} = \begin{cases} 0 & \mu_{o,i,M_i}^* = 0, \\ \frac{|\mu_{o,i,M_i-l}^* - \mu_{o,i,M_i}^*|}{|\mu_{o,i,M_i}^*|} & \text{sonst,} \end{cases}$$

$$\nu_{\sigma,M_i,M_i-l,o,i} = \begin{cases} 0 & \sigma_{o,i,M_i} = 0, \\ \frac{|\sigma_{o,i,M_i-l} - \sigma_{o,i,M_i}|}{|\sigma_{o,i,M_i}|} & \text{sonst.} \end{cases}$$

Diese berechnen die relative Differenz einer charakteristischen Größe μ_{o,i,M_i} , μ_{o,i,M_i}^* oder σ_{o,i,M_i} nach $M_i > M_{\min}$ und $M_i - l$ Stichprobenwerte. Für die minimale Stichprobengröße soll dabei $M_{\min} \geq 10$ gelten. Die relativen Differenzen der letzten 10 hinzugefügten Stichproben werden arithmetisch gemittelt:

$$\epsilon_{M_i,o,i} = \begin{pmatrix} \epsilon_{\mu,M_i,o,i} \\ \epsilon_{\mu^*,M_i,o,i} \\ \epsilon_{\sigma,M_i,o,i} \end{pmatrix} := \frac{1}{10} \sum_{l=1}^{10} \begin{pmatrix} |\nu_{\mu,M_i,M_i-l,o,i}| \\ |\nu_{\mu^*,M_i,M_i-l,o,i}| \\ |\nu_{\sigma,M_i,M_i-l,o,i}| \end{pmatrix}. \quad (4.1)$$

Um einen Input i als inaktiv einzustufen, weil bereits genügend Informationen zu seiner Klassifizierung als relevant bzw. irrelevant vorliegen, werden diese arithmetischen Mittel mit einer Konstante κ_{act} komponentenweise für jeden Output o verglichen:

$$\begin{pmatrix} \epsilon_{\mu,M_i,o,i} \\ \epsilon_{\mu^*,M_i,o,i} \\ \epsilon_{\sigma,M_i,o,i} \end{pmatrix} < \kappa_{\text{act}}. \quad (4.2)$$

Diese Bedingung ist also erfüllt, wenn die Veränderungen in Output o , die durch die letzten 10 Stichprobenwerte von Input i verursacht wurden, nur geringfügig waren, d. h. wenn durch das Hinzufügen der letzten 10 Stichprobenwerte nur wenig neue Informationen hinzugefügt wurden und die drei charakteristischen

Größen μ_{o,i,M_i} , μ_{o,i,M_i}^* und σ_{o,i,M_i} nur wenig geschwankt sind. In diesem Fall ist die Kategorisierung von Input i als relevant bzw. irrelevant belastbar. Folglich wird Input i nach M_i Stichprobenwerten als inaktiv eingestuft und es werden keine weiteren EE in diese Richtung ausgewertet. Daher werden im weiteren Verlauf nach M Stichprobenwerten nur noch EE für aktive Inputs

$$\mathcal{A}_M := \{i = 1, \dots, k \mid \epsilon_{\mu,M,o,i} > \kappa_{\text{act}} \vee \epsilon_{\mu^*,M,o,i} > \kappa_{\text{act}} \vee \epsilon_{\sigma,M,o,i} > \kappa_{\text{act}}\}$$

ausgewertet. Dieser erste Schritt des DSC dient dazu, die Anzahl der EE für jeden Input i einzeln an die Relevanz des Inputs i anzupassen. Um nun die gesamte Iteration zu stoppen, wird die Konvergenz der verbliebenen aktiven Inputs in \mathcal{A}_M betrachtet. Dazu sei $\mathbf{M} = (M_1, \dots, M_k)^\top$, $M_i = M$ für $i \in \mathcal{A}_M$ und $M_i < M$ für $i \in \{1, \dots, k\} \setminus \mathcal{A}_M$ der Vektor an bereits ausgewerteten EE für Input i : Für einen aktiven Input $i \in \mathcal{A}_M$ wurden M Stichprobenwerte ausgewertet, während für inaktive Inputs $i \in \{1, \dots, k\} \setminus \mathcal{A}_M$ die Auswertung nach $M_i < M$ Stichprobenwerten vorzeitig beendet wurde. Damit werden die globalen Fehlerkonstanten

$$\boldsymbol{\epsilon}_M = \begin{pmatrix} \epsilon_{\mu,M} \\ \epsilon_{\mu^*,M} \\ \epsilon_{\sigma,M} \end{pmatrix} := \frac{1}{10} \sum_{l=1}^{10} \begin{pmatrix} \|\mathbf{v}_{\mu,M,M-l}\|_F \\ \|\mathbf{v}_{\mu^*,M,M-l}\|_F \\ \|\mathbf{v}_{\sigma,M,M-l}\|_F \end{pmatrix} \quad (4.3)$$

eingeführt. Um die gesamte globale Iteration zu stoppen, wird der globale Fehler $\boldsymbol{\epsilon}_M$ komponentenweise mit einer weiteren Konstante κ_{stop} verglichen

$$\begin{pmatrix} \epsilon_{\mu,M} \\ \epsilon_{\mu^*,M} \\ \epsilon_{\sigma,M} \end{pmatrix} \leq \kappa_{\text{stop}} \quad (4.4)$$

und die Sensitivitätsanalyse wird gestoppt, sobald alle drei Ungleichungen erfüllt sind. Der folgende Satz zeigt, dass die globalen Konstanten $\boldsymbol{\epsilon}_M$ konvergieren:

Satz 4.1.1. *Für eine Funktion $f : \mathbb{R}^k \rightarrow \mathbb{R}^m$ bricht das DSC mit den Konstanten $\kappa_{\text{act}}, \kappa_{\text{stop}} > 0$ ab, falls $\sqrt{km}\kappa_{\text{act}} \leq \kappa_{\text{stop}}$ gilt.*

Der Beweis für Satz 4.1.1 ist in Anhang A.3 angegeben. Um den Abbruch des DSC zu garantieren, muss also die Schranke κ_{act} deutlich kleiner als κ_{stop} gewählt werden. Für eine große Anzahl an Inputs $k \gg 0$ ist dies eine

Einschränkung für die Reduktion an zu berechnende EE. In der Praxis kann zugunsten einer verbesserten Rechenzeit aber $\kappa_{\text{act}} > 1/\sqrt{km}\kappa_{\text{stop}}$ gewählt und damit auf den garantierten Abbruch verzichtet werden (vgl. Abschnitt 4.2.2), da häufig völlig irrelevante Inputs mit z. B. $\mu_{o,i,M}^* = 0$ einen gewissen Spielraum liefern.

Morris' Methode ist eine qualitative Sensitivitätsanalyse (Campolongo et al., 2007; Morris, 1991) und als solche nicht dafür gedacht, den exakten Einfluss eines Inputs zu berechnen. Manchmal kann es aber sehr wichtig sein, die Größenordnung der charakteristischen Größen gut abzuschätzen: beispielsweise wenn erläutert werden soll, warum ein Fahrzeug eine Verkehrsregel verletzt hat (vgl. Abschnitt 7.1). In diesen Anwendungsfällen mit vielen Inputs $k \gg 0$ ist die Einordnung der meisten Inputs als relevant oder irrelevant bereits für wenige ausgewertete EE eindeutig und nur wenige Inputs müssen genauer analysiert werden. Das oben vorgestellte DSC erhöht die benötigte Stichprobengröße und damit verbundene Rechenzeit nur dann, wenn es nötig ist.

4.2 Beispiele

In diesem Abschnitt wird das DSC auf zwei Beispiele angewendet: Zunächst wird die g -Funktion analysiert (Abschnitt 4.2.1), die in der Literatur zur Validierung von SA verwendet wird, z. B. Ge und Menendez (2017); Saltelli et al. (2010); Touzani und Busby (2014). Anschließend wird das Überholmanöver aus Beispiel 3.2.6 analysiert (Abschnitt 4.2.2).

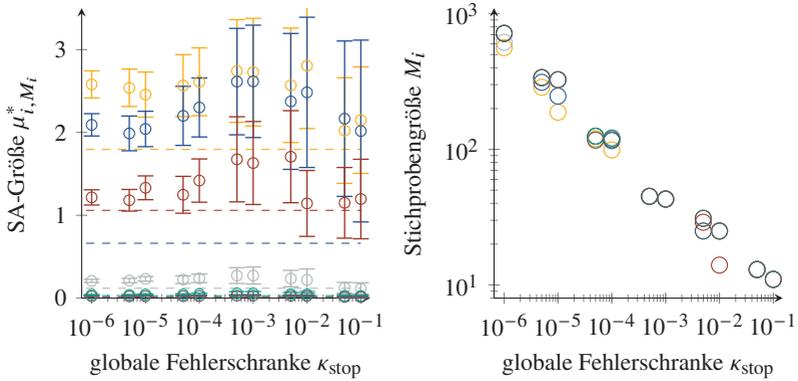
4.2.1 Testfunktion: g -Funktion

Im Folgenden wird die Funktion

$$g : [0, 1]^k \rightarrow \mathbb{R},$$

$$g(\mathbf{x}) := \prod_{i=1}^k g_i(x_i) = \prod_{i=1}^k \frac{|4x_i - 2| + a_i}{1 + a_i} \quad (4.5)$$

mit den $k = 6$ Parametern $[a_1, \dots, a_6]^\top = [0, 0.2, 0.9, 9, 50, 99]^\top$ genutzt (Campolongo & Saltelli, 1997). Alle Inputs sind $\mathcal{U}([0, 1])$ -gleichverteilt und



(a) Absoluter Stichprobenmittelwert μ_{i, M_i}^* der SA mit DSC nach M_i EE, Fehlerbalken zeigen 95 % Konfidenzintervalle, gestrichelte Linien geben Ergebnisse analog zu Campolongo und Saltelli (1997) an.

(b) Anzahl M_i der EE, die bei der SA mit DSC zur Berechnung der charakteristischen Größen verwendet wurden.

Abbildung 4.1: Vergleich SA mit und ohne DSC (vgl. Satz 4.1.1) für die g -Funktion (4.5) mit $k = 6$ Inputs. $M_{\min} = 10$, $\kappa_{\text{act}} = 1/\sqrt{km}\kappa_{\text{stop}}$. $\circ a_1 = 0$; $\circ a_2 = 0.2$; $\circ a_3 = 0.9$; $\circ a_4 = 9$; $\circ a_5 = 50$; $\circ a_6 = 99$.

ein Input i hat großen Einfluss, d. h. ist relevant, wenn der zugehörige Parameter a_i klein ist, und kleinen Einfluss, d. h. ist irrelevant, andernfalls. Für die SA mit DSC seien $\Delta_i = 0.25$ für alle $i = 1, \dots, 6$, $M_{\min} = 10$, und $\kappa_{\text{act}} = 6^{-1/2}\kappa_{\text{stop}}$. Als Referenz wird eine weitere SA ohne DSC analog zu Campolongo und Saltelli (1997) mit $M_i = 10$ und $\Delta_i = 0.25$ für alle $i = 1, \dots, 6$ berechnet.

Abbildung 4.1(a) zeigt die charakteristische Größe μ_{i, M_i}^* für alle Inputs $i = 1, \dots, 6$ in Abhängigkeit von der globalen Abbruchschranke κ_{stop} : Die Größen, die analog zu Campolongo et al. (2007) als Referenz berechnet wurden (gestrichelte Linien), weichen teilweise stark von denen ab, die mit dem DSC berechnet wurden (siehe z. B. Input $i = 2$ für $a_2 = 0.2$). Dies ist darauf zurückzuführen, dass die Stichprobengröße $M_i = 10$ in der SA ohne DSC zu klein ist und die charakteristische Größe μ_{i, M_i}^* noch nicht konvergiert ist (Campolongo & Saltelli, 1997). Bei der SA mit DSC werden die Konfidenzintervalle mit abnehmendem κ_{stop} schmaler, was wie erwartet eine bessere Präzision der Berechnungen mit DSC anzeigt. Weitere Untersuchungen zeigen,

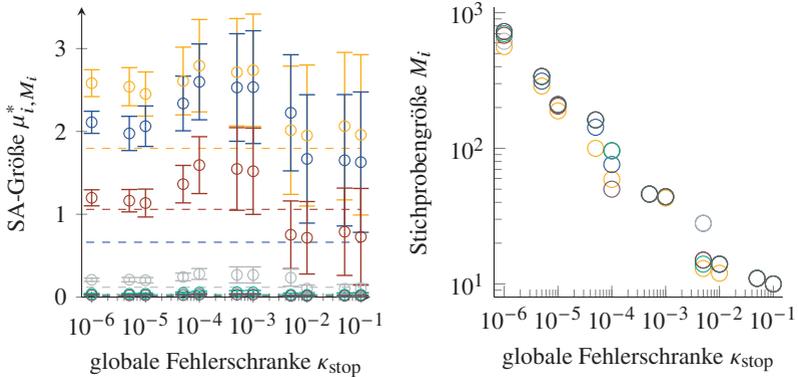
dass diese Beobachtungen auch für die Stichprobenerwartungswerte μ_{i,M_i} und die Stichprobenvarianzen σ_{i,M_i} gelten, vgl. Abbildungen A.2(a) und A.2(b).

Wie erwartet ist der Nachteil dieser verbesserten Genauigkeit eine größere Anzahl an Funktionsauswertungen, vgl. Abb. 4.1(b) und Tabelle 4.1. Dabei wird deutlich, dass durch das Aussortieren von inaktiven Inputs $\{1, \dots, k\} \setminus \mathcal{A}_M$ nur wenig Rechenaufwand eingespart werden kann, weil das Kriterium $\sqrt{km}\kappa_{\text{act}} = \kappa_{\text{stop}}$ sehr streng ist (jeweils linke Einträge der Zellen). Stattdessen wird dasselbe Experiment mit $\kappa_{\text{act}} = 0.5\kappa_{\text{stop}}$ wiederholt. Dann zeigt Abb. 4.2(a) analog zu Abb. 4.1(a), dass mit abnehmendem κ_{stop} die Konfidenzintervalle immer kleiner und die charakteristischen Größen μ_{i,M_i}^* immer genauer bestimmt werden. Abb. 4.2(b) und vor allem Tabelle 4.1 verdeutlichen, dass durch diese Abschwächung die Anzahl an Funktionsauswertungen weiter reduziert werden kann und das DSC dennoch abbricht.

Verglichen mit dem SA-Ansatz ohne DSC aus Campolongo et al. (2007), der bereits annehmbare Klassifizierungsergebnisse liefert, wirken die vorgeschlagenen Verbesserungen teuer, ohne dass sie erhebliche Laufzeitverbesserungen erwirken würden. Allerdings geben Campolongo und Saltelli (1997); Menberg et al. (2016) auch an, dass eine Stichprobengröße $M \in \{4, \dots, 10\}$ nur schlechte Schätzungen der charakteristischen Größe μ^* liefert, was durch Abb. 4.1(a) bestätigt wird. Außerdem sind mindestens $M \geq 30$ Stichprobenwerte nötig, um einige stochastische Ergebnisse anzuwenden, z. B. der zentrale Grenzwertsatz in Abschnitt 5.2. Daher wird im weiteren Verlauf die minimale Stichprobengröße

Tabelle 4.1: Stichprobengrößen nach Abbruch durch DSC für ausgewählte κ_{stop} . $M_{\text{min}} = 10$, $\kappa_{\text{act}} \leq 1/\sqrt{km}\kappa_{\text{stop}}$ (vgl. Abb. 4.1(b) und Satz 4.1.1) / $\kappa_{\text{act}} \leq 0.5\kappa_{\text{stop}}$ (vgl. Abb. 4.2(b)). Minimum $\min M_i$ und Maximum $\max M_i$ sind für jedes κ_{stop} farblich markiert.

κ_{stop}	Stichprobengröße M_i für Inputs $i \in \{1, \dots, 6\}$						$\frac{\sum M_i}{6 \max M_i}$
	$a_1 = 0$	$a_2 = 0.2$	$a_3 = 0.9$	$a_4 = 9$	$a_5 = 50$	$a_6 = 99$	
10^{-1}	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10	10 / 10	1 / 1
10^{-2}	24 / 12	24 / 14	13 / 14	24 / 14	24 / 14	24 / 14	0.92 / 0.98
10^{-3}	42 / 43	42 / 44	42 / 44	42 / 44	42 / 44	42 / 44	1 / 1
10^{-4}	98 / 59	115 / 76	120 / 96	120 / 96	120 / 50	117	0.96 / 0.82
10^{-5}	187 / 189	246 / 210	325 / 205	325 / 210	325 / 210	325 / 210	0.89 / 0.98
10^{-6}	561 / 563	715 / 726	715 / 681	614 / 616	715 / 700	715 / 726	0.94 / 0.92



(a) Absoluter Stichprobenmittelwert μ_{i, M_i}^* der SA mit DSC nach M_i EE, Fehlerbalken zeigen 95 % Konfidenzintervalle, gestrichelte Linien geben Ergebnisse analog zu Campolongo und Saltelli (1997) an.

(b) Anzahl M_i der EE, die bei der SA mit DSC zur Berechnung der charakteristischen Größen verwendet wurden.

Abbildung 4.2: Vergleich SA mit und ohne DSC für die g -Funktion (4.5) mit $k = 6$ Inputs.

$M_{\min} = 10$, $\kappa_{\text{act}} = 0.5\kappa_{\text{stop}}$. $\circ a_1 = 0$; $\circ a_2 = 0.2$; $\circ a_3 = 0.9$; $\circ a_4 = 9$; $\circ a_5 = 50$; $\circ a_6 = 99$ (angelehnt an Henze, Rufer et al. (2022)).

auf $M_{\min} = 50$ angehoben. Dennoch können weiterhin Fehlklassifizierungen auftreten, wenn für die automatische Klassifizierung von Inputs eine Schranke $\epsilon > 0$ verwendet wird: Für $\mu^* \approx \epsilon$ ist die Entscheidung, ob Input i relevant ist, für kleinere Sampling-Größen sehr unsicher. In diesen Fällen verbessert das DSC die Zuverlässigkeit der Abschätzungen vor allem für Inputs mit großen Fluktuationen. Dadurch können zuverlässigere Informationen erreicht und trotzdem die Stichprobengröße gering gehalten werden.

4.2.2 Anwendung: Überholmanöver

Beispiel 3.2.6.c hat die Ergebnisse der SA ohne DSC für das Überholmanöver vorgestellt. Dieselbe Analyse kann nun auch mit einer SA mit DSC durchgeführt werden. Dafür seien $\kappa_{\text{stop}} = 0.05$ und $\kappa_{\text{act}} = 10^{-3}$ die Abbruchkonstanten und $M_{\min} = 50$ die minimale Stichprobengröße. Abb. 4.4 zeigt die Ergebnisse der SA für $M = 300$ (Henze et al., 2020, , vgl. Abb. 3.11) und

mit DSC exemplarisch für die Zeitpunkte $t \in \{0.5 \text{ s}, 2.2 \text{ s}, 3.9 \text{ s}\}$ (analog zu Beispiel 3.2.6.c): Die Ergebnisse beider Methoden ähneln sich, vor allem für die Parameter mit $\mu_{o,i,t} \equiv \sigma_{o,i,t} \equiv 0$ liegen die Ergebnisse nahe beieinander. Dies ist wichtig, da die SA nach Morris als qualitative Methode vor allem zwischen relevanten und irrelevanten Inputs unterscheidet und die exakte Bestimmung der charakteristischen Größen nicht im Vordergrund steht.

Die Anzahl an ausgewerteten EE ist in Abb. 4.3 dargestellt: Zu allen Zeitpunkten außer $t = 4.0 \text{ s}$ bricht das DSC die Auswertungen mit weniger als $M = 300$ EE ab und berücksichtigt so die Komplexität der Situation angemessen. Insgesamt werden 51 884 EE berechnet, was einer Ersparnis von 32.18 % gegenüber der SA mit $M = 300$ EE entspricht ($M = 300$ für $k = 5$ Inputs und 51 Zeitpunkten). Im Vergleich zur g -Funktion (Abschnitt 4.2.1) wird auch hier deutlich, dass das zweite Ziel des DSC, der Abbruch für jeden inaktiven Input einzeln und unabhängig von den anderen Inputs, in dieser Anwendung nur zu einer schwachen Reduktion des Rechenaufwands führt. Lediglich für die Distanz d_{nl} zum Fahrzeug links hinter dem Ego werden zu Beginn weniger EE berechnet: Beispielsweise bestimmt die SA mit DSC für $t = 0.5 \text{ s}$ die charakteristischen Parameter $\mu_{a,d_{nl}} = \mu_{lc,d_{nl}} = \sigma_{a,d_{nl}} = \sigma_{lc,d_{nl}} = 0$, die Distanz wird daher als irrelevant eingestuft und somit reichen $M = 51$ EE aus (zum Vergleich die Ergebnisse der SA ohne DSC zum Zeitpunkt $t = 0.5 \text{ s}$: $\mu_{a,d_{nl}} = -0.0092$, $\mu_{lc,d_{nl}} = -0.0133$, $\sigma_{a,d_{nl}} = 0.1131$, $\sigma_{lc,d_{nl}} = 0.1630$). Damit ist die SA mit DSC genauso gut dafür geeignet, irrelevante Inputs von relevanten Inputs zu unterscheiden, und reduziert gleichzeitig den Rechenaufwand, indem die

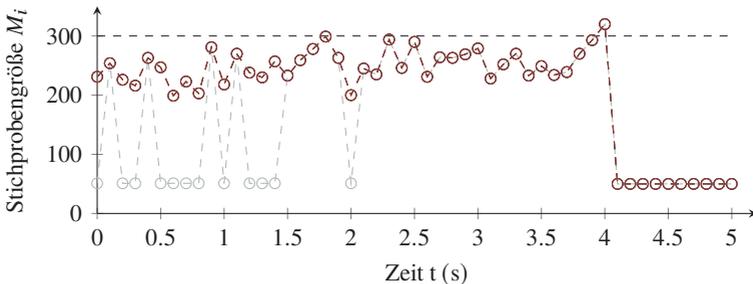


Abbildung 4.3: Anzahl an ausgewerteten EE für die SA mit DSC für Beispiel 3.2.6.c (Überholmanöver). \circ d_{nl} , \square v_{nl} . Für die anderen Inputs d_{vs} , v und v_{vs} wurden genauso viele EE ausgewertet wie für v_{nl} .

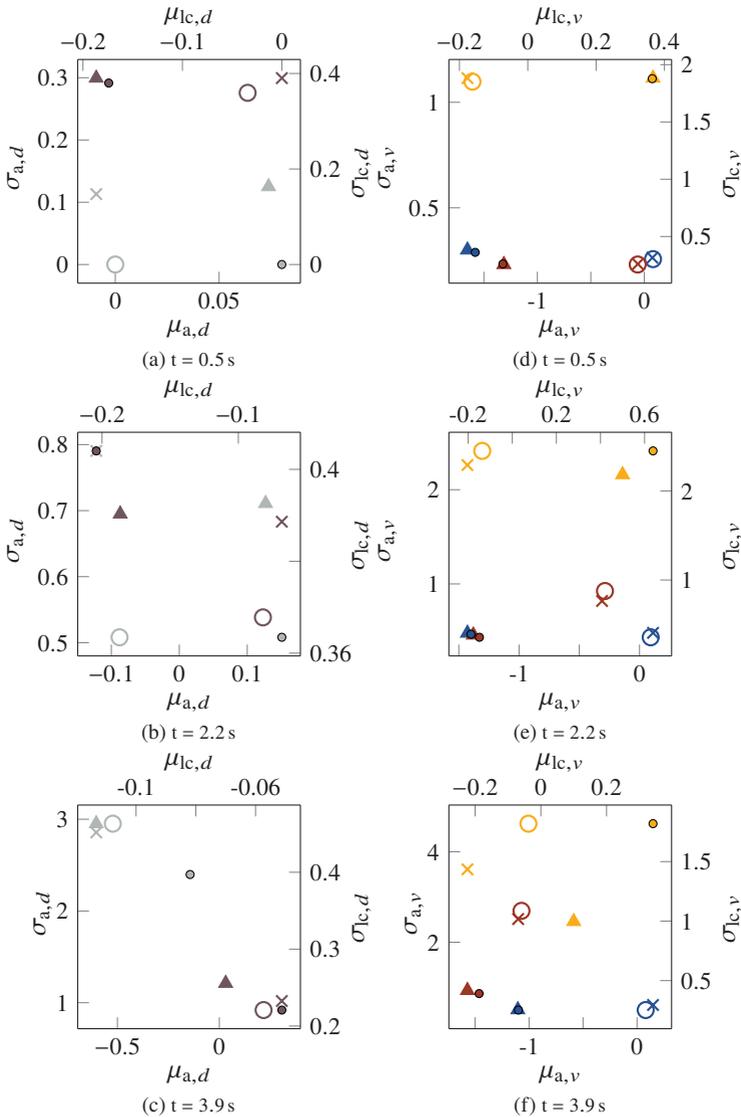


Abbildung 4.4: Ergebnisse der SA mit und ohne DSC für Beispiel 3.2.6.c (Überholmanöver). Einfluss der Inputs i auf die Beschleunigung $o = a$, SA mit $M = 300$: \times d_{vs} , \times d_{nl} , \times v , \times v_{vs} , \times v_{nl} ; SA mit DSC: \circ d_{vs} , usw. Einfluss der Inputs i auf den FSW $o = lc$, SA mit $M = 300$: \blacktriangle d_{vs} , usw.; SA mit DSC: \bullet d_{vs} , usw.

Anzahl an ausgewerteten EE an die Komplexität der Fahrsituation angepasst wird.

5 Identifikation von zulässigen Unsicherheiten zwischen Wahrnehmung und Planung

Kapitel 3 konnte anhand der Definition 3.1.9 zu zulässigen Planungsentscheidungen darstellen, dass Messunsicherheiten zu starken Schwankungen im Planungsergebnis führen können. Mit den dort vorgestellten Abschätzungen der stochastischen Ungleichung kann allerdings nur untersucht werden, ob gegebene Unsicherheiten zu zulässigen Planungsentscheidungen führen, aber nicht, welche Unsicherheiten im Allgemeinen akzeptabel wären. Auch die SA konnte nur dabei helfen, die Informationen zu identifizieren, die den größten Einfluss hatten, aber nicht aussagen, wie viel kleiner die Messunsicherheiten sein sollen. Dafür wird in Abschnitt 5.1 ein stochastisches Optimierungsproblem definiert, das die Messunsicherheiten möglichst maximiert und gleichzeitig die Wahrscheinlichkeit für eine unzulässige Planungsentscheidung beschränkt. In Abschnitt 5.2 wird dieses durch ein diskretes Problem approximiert und anschließend numerisch gelöst: Dafür stellt Abschnitt 5.3 eine Modifikation der Multilevel-Koordinatensuche (engl. *Multilevel Coordinate Search (MCS)*) vor, die mit der SA kombiniert wird. Abschließend wendet Abschnitt 5.4 die MCS mit SA auf ein Überholmanöver an, um die akzeptablen Messunsicherheiten für das IDM-MOBIL-Modell zu bestimmen.

Vorläufige Ergebnisse wurden in Henze, Faßbender und Stiller (2021) veröffentlicht. Abbildungen, Tabellen und andere Ergebnisse wurden für eine konsistente Darstellung angepasst.

5.1 Kontinuierliches Optimierungsproblem

Mit den oben vorgestellten Ansätzen zu Wahrscheinlichkeitsabschätzungen und zur SA kann bisher nur bestimmt werden, ob die Messunsicherheiten

für zulässige Planungsentscheidungen ausreichend gering sind (vgl. Definition 3.1.9) und welche Unsicherheiten großen Einfluss auf das Planungsergebnis haben. Interessant ist aber insbesondere, wie groß die maximal zulässigen Messunsicherheiten σ^* sein dürfen, mit denen der Planungsalgorithmus noch akzeptable Ergebnisse bestimmen kann. Daher wird ein Optimierungsproblem mit stochastischen Nebenbedingungen definiert. Dafür wird Definition 3.1.9 von zulässigen Planungsentscheidungen genutzt:

Definition 5.1.1. Sei $\mathbf{f} : \mathbb{R}^k \rightarrow \mathbb{R}^m$ eine Planungsfunktion, $\mathbf{q} \in \Omega$ eine Umgebungsbeschreibung und $\mathbf{u} \in \mathbb{R}^{+k}$ der Vektor mit maximalen Messunsicherheiten. Seien weiterhin $\bar{b}_o \in \mathbb{R}$ zulässige Abweichungen und $\bar{p}_o \in [0, 1]$ zulässige Wahrscheinlichkeiten für alle $o = 1, \dots, m$ (vgl. Definition 3.1.9). Dann wird mit der Zielfunktion $J : \mathbb{R}^k \rightarrow \mathbb{R}$, $\sigma \mapsto J(\sigma) := \mathbf{w} \cdot \sigma$, ein Maximierungsproblem mit stochastischen Nebenbedingungen definiert durch

$$\max_{\sigma \in (\mathbf{0}, \mathbf{u}]} J(\sigma) = \max_{\sigma \in (\mathbf{0}, \mathbf{u}]} \mathbf{w} \cdot \sigma = \max_{\sigma \in (\mathbf{0}, \mathbf{u}]} \sum_{i=1}^k w_i \sigma_i, \quad (5.1.1)$$

$$\text{u. d. N. } \mathbb{P}_{f_o(\mathbf{q} + \hat{\mathbf{Q}})} \left(\left| f_o(\mathbf{q} + \hat{\mathbf{Q}}) - \mathbb{E}(f_o(\mathbf{q} + \hat{\mathbf{Q}})) \right| > \bar{b}_o \right) \leq \bar{p}_o \\ \forall o = 1, \dots, m, \quad (5.1.2)$$

$$\hat{Q}_i \sim \mathcal{N}(0, \sigma_i^2) \quad \forall i = 1, \dots, k. \quad (5.1.3)$$

Das Maximierungsproblem (5.1) bestimmt also die maximal zulässigen Messunsicherheiten σ^* , sodass die Planungsergebnisse nach Definition 3.1.9 noch zulässig sind. Dabei sorgt es dafür, dass die Anforderungen σ^* an die Wahrnehmungsmodule so schwach wie möglich sind (vgl. Zielfunktion (5.1.1)). Der Gewichtsvektor $\mathbf{w} \in \mathbb{R}^+$ kann alle Inputs unterschiedlich stark gewichten: Je größer das Gewicht w_i ist, desto schwächer ist die Einschränkung für Input i . Damit können z. B. Kosten für einen Sensor oder Rechenzeit für die Weiterverarbeitung von Informationen berücksichtigt werden. Außerdem können Schranken für σ angegeben werden, die auf technische Grenzen zurückzuführen sind, z. B. auf das Wissen, dass die Messungenaugigkeit eines Sensors i nicht schlechter als $u_i > 0$ ist, also

$$\sigma \in (\mathbf{0}, \mathbf{u}]. \quad (5.2)$$

Die stochastische Nebenbedingung (5.1.2) fordert, dass die resultierenden Planungsentscheidungen zulässig gemäß Definition 3.1.9 sind, während Nebenbedingung (5.1.3) die W -Verteilungen der Messabweichungen aller Inputs $i = 1, \dots, k$ definiert. Zusammenfassend ist das Optimierungsproblem (5.1) ein *decision-dependent single chance-constraint maximization problem*, d. h. die W -Verteilung der ZV hängt von der Wahl des Optimierungsparameters σ ab (engl. *decision dependent*) und die stochastischen Ungleichungen für zulässige Planungsergebnisse werden voneinander entkoppelt erfüllt (engl. *single chance constraint*) (Lejeune, Margoty & Oliveira, 2018; Li, Arellano-Garcia & Wozny, 2008). Das folgende Beispiel zeigt ein solches Optimierungsproblem für $k = 1$.

Beispiel 3.1.8.g (Freifahrt: kontinuierliches Optimierungsproblem). Beispiele 3.1.8.b und 3.1.8.d haben gezeigt, dass die Messunsicherheit $\sigma_v = 1.3 \text{ m/s}$ mit Wahrscheinlichkeit $p > \bar{p}_a = 0.1$ zu unzulässigen Planungsabweichungen führen kann, weil die Messunsicherheit zu groß oder der Planungsalgorithmus nicht robust genug ist. Das zugehörige kontinuierliche Optimierungsproblem mit $k = 1$ ist

$$\max_{\sigma_v \in (0,2]} J(\sigma_v) = \max_{\sigma_v \in (0,2]} \sigma_v, \quad (5.3.1)$$

$$\text{u. d. N. } \mathbb{P}_{f_a(v+\hat{V})} \left(\left| f_a(v + \hat{v}) - \mathbb{E}(f_a(v + \hat{V})) \right| > \bar{b}_a \right) \leq \bar{p}_a, \quad (5.3.2)$$

$$\hat{V} \sim \mathcal{N}(0, \sigma_v^2). \quad (5.3.3)$$

Das Ziel ist, die maximal zulässige Unsicherheit σ_v^* zu berechnen, sodass Nebenbedingung (5.3.2) für die Schranken $\bar{b}_a = 0.8 \text{ m/s}^2$ und $\bar{p}_a = 0.1$ gerade noch erfüllt ist. Für dieses eindimensionale Beispiel ist es möglich, die Lösung des kontinuierlichen Problems genau zu berechnen, weil die W -Verteilung der Messabweichungen und der Zusammenhang f_a zwischen dem verrauschten Input und dem Ergebnis bekannt sind (vgl. Beispiel 3.1.8.a). Da die Zielfunktion J streng monoton steigend ist und die zulässige Menge durch die kontinuierlichen Nebenbedingungen abgeschlossen und beschränkt ist, nimmt die Zielfunktion J ihr Maximum auf dem Rand der zulässigen Menge an: Dafür wird zuerst gezeigt, dass die Funktion $\sigma_v \mapsto \mathbb{P}_{f_a(v_t+\hat{V})} \left(\left| f_a(v_t + \hat{v}) - \mathbb{E}(f_a(v_t + \hat{V})) \right| > \bar{b}_a \right)$

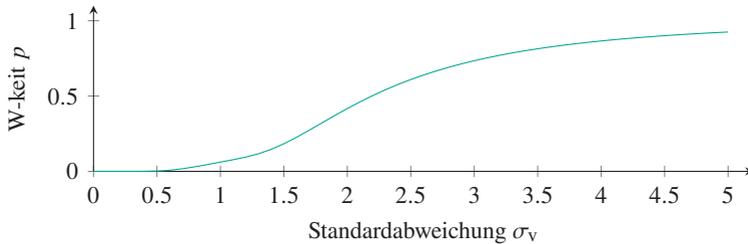


Abbildung 5.1: Die Funktion $\sigma_v \mapsto \mathbb{P}_{f_a(v_t+\hat{V})} \left(|f_a(v_t + \hat{v}) - \mathbb{E}(f_a(v_t + \hat{V}))| > \bar{b}_a \right)$ ist im Intervall $(0, 2]$ streng monoton steigend, vgl. (5.3.2). Der Beweis ist in Anhang A.5 dargestellt.

streng monoton steigt. Anschaulich kann dies in Abb. 5.1 aus dem Verlauf der Nebenbedingung (5.3.2) abgelesen werden. Der Beweis, dass die Ableitung

$$\frac{d \mathbb{P}_{f_a(v_t+\hat{V})} \left(|f_a(v_t + \hat{v}) - \mathbb{E}(f_a(v_t + \hat{V}))| > \bar{b}_a \right)}{d \sigma_v} > 0$$

strikt positiv ist, ist in Anhang A.5 dargestellt. Zur Berechnung der optimalen Standardabweichung σ_v^* wird Nebenbedingung (5.3.2) abschließend aktiv gesetzt, d. h. aus der Ungleichung wird eine Gleichung, die nach σ_v^* aufgelöst werden kann. So ergibt sich als optimale Standardabweichung $\sigma_v^* = 1.2266 \text{ m/s}$ als Lösung des kontinuierlichen Problems (5.3) für Beispiel 3.1.8.

5.2 Diskretes Optimierungsproblem

Für gewöhnlich sind *decision dependent single chance constraint* Optimierungsprobleme wie Problem (5.1) analytisch nur unter starken Voraussetzungen zu lösen, da keine geschlossene Form des zulässigen Suchraums existiert. Um die Nebenbedingung (5.1.2) numerisch umzusetzen, wird sie daher mit der Chebychev-Saw-Ungleichung approximiert (vgl. Abschnitt 3.2). Damit erhalten wir die folgende diskrete Approximation des kontinuierlichen Optimierungsproblems (5.1):

Definition 5.2.1. Sei $\mathbf{f} : \mathbb{R}^k \rightarrow \mathbb{R}^m$ eine Planungsfunktion, $\mathbf{q} \in \Omega$ eine Umgebungsbeschreibung und $\mathbf{u} \in \mathbb{R}^{+k}$ der Vektor mit maximalen Messunsicherheiten. Seien weiterhin $\bar{\mathbf{b}}_o \in \mathbb{R}$ zulässige Abweichungen und $\bar{p}_o \in [0, 1]$ zulässige Wahrscheinlichkeiten für alle $o = 1, \dots, m$ (vgl. Definition 3.1.9). Sei $\{\hat{\mathbf{q}}_j\}_{j=1}^M = \{(\hat{q}_{j,1}, \dots, \hat{q}_{j,k})^\top\}_{j=1}^M$ eine Stichprobe der ZV $\hat{\mathbf{Q}}_i \sim \mathcal{N}(0, \sigma_i^2)$, $i = 1, \dots, k$ mit paarweise unabhängigen, identisch verteilten Elementen. Dann wird mit der Zielfunktion $J : \mathbb{R}^k \rightarrow \mathbb{R}$, $\boldsymbol{\sigma} \mapsto J(\boldsymbol{\sigma}) := \mathbf{w} \cdot \boldsymbol{\sigma}$, ein Maximierungsproblem mit stochastischen Nebenbedingungen definiert durch

$$\max_{\boldsymbol{\sigma} \in (0, \mathbf{u}]} J(\boldsymbol{\sigma}) = \max_{\boldsymbol{\sigma} \in (0, \mathbf{u}]} \mathbf{w} \cdot \boldsymbol{\sigma} = \max_{\boldsymbol{\sigma} \in (0, \mathbf{u}]} \sum_{i=1}^k w_i \sigma_i, \quad (5.4.1)$$

$$\text{u. d. N. } \frac{1}{M+1} g_{M+1} \left(\frac{M \lambda_{S,o}^2}{M-1 + \lambda_{S,o}^2} \right) \leq \bar{p}_o \quad \forall o = 1, \dots, m \quad (5.4.2)$$

$$\hat{q}_{j,i} \sim \mathcal{N}(0, \sigma_i^2) \quad \forall j = 1, \dots, M, i = 1, \dots, k. \quad (5.4.3)$$

Die Funktion $g_{M+1} : [1, \infty] \rightarrow \mathbb{R}$ und die Parameter $\lambda_{S,o}$ sind definiert wie in Satz 3.2.3 und Gleichung (3.12).

Unter weiteren Annahmen an die Funktion \mathbf{f} kann gezeigt werden, dass für das Optimierungsproblem (5.4) eine Lösung existiert (Henze, Faßbender & Stilller, 2021). Ihre Eindeutigkeit kann jedoch nicht bewiesen werden, da die stochastischen Nebenbedingungen (5.4.2) im Allgemeinen keine konvexe zulässige Menge beschreiben: Beispielsweise könnte bei mehreren Sensoren ein sehr genauer Sensor die schlechte Performance eines anderen ausgleichen und umgekehrt, während durchschnittliche Ergebnisse beider Sensoren womöglich nicht ausreichen, um zuverlässige Planungsentscheidungen zu treffen.

Im restlichen Teil dieses Abschnittes betrachten wir die Beziehung zwischen den Lösungen für das kontinuierliche und das diskrete Optimierungsproblem. Wie in Abschnitt 3.2 beschrieben, ist es nicht möglich, ohne zusätzliche Annahmen über die Planungsfunktion \mathbf{f} weitere Aussagen über die Varianzen $\text{Var}(\mathbf{f}(\mathbf{q} + \hat{\mathbf{Q}}))$ der ZV $\mathbf{f}(\mathbf{q} + \hat{\mathbf{Q}})$ zu treffen. Daher werden die Stichprobenvarianzen $\sigma_{\mathbf{f}(\mathbf{q}+\hat{\mathbf{Q}})}^2$ verwendet, um die wahren Varianzen $\text{Var}(\mathbf{f}(\mathbf{q} + \hat{\mathbf{Q}}))$ abzuschätzen. Die Realisierungen $\{\mathbf{f}(\mathbf{q} + \hat{\mathbf{q}}_j)\}_{j=1}^M$ der ZV $\mathbf{f}(\mathbf{q} + \hat{\mathbf{Q}})$ sind in der Simulation paarweise unabhängig, gleichverteilt und haben die Erwartungswerte $\mathbb{E}(\mathbf{f}(\mathbf{q} + \hat{\mathbf{Q}}))$ und Varianzen $\text{Var}(\mathbf{f}(\mathbf{q} + \hat{\mathbf{Q}}))$. Damit sind die Voraussetzungen des zentralen

Grenzwertsatzes erfüllt: Die Verteilungen der Stichprobenvarianzen $\sigma_{f_o(\mathbf{q}+\hat{\mathbf{Q}})}^2$, $o = 1, \dots, m$, können für $M \geq 30$ durch eine Normalverteilung approximiert werden und das Konfidenzintervall

$$\left[0, \frac{M-1}{\chi_{M-1,1-\beta}^2} \sigma_{f_o(\mathbf{q}+\hat{\mathbf{Q}})}^2 \right]$$

enthält mit Wahrscheinlichkeit $\beta \in [0, 1]$ die Varianz $\text{Var}(f_o(\mathbf{q} + \hat{\mathbf{Q}}))$, wobei $\chi_{M-1,1-\beta}^2$ das $(1 - \beta)$ -Quantil der χ^2 -Verteilung mit $M - 1$ Freiheitsgraden ist. Mit Chebychevs Ungleichung (3.7) gilt für die stochastische Nebenbedingung (5.1.2) des kontinuierlichen Optimierungsproblems mit Wahrscheinlichkeit β

$$\begin{aligned} \mathbb{P}_{f_o(\mathbf{q}+\hat{\mathbf{Q}})} \left(\left| f_o(\mathbf{q} + \hat{\mathbf{q}}) - \mathbb{E}(f_o(\mathbf{q} + \hat{\mathbf{Q}})) \right| > \bar{a}_o \right) &\leq \frac{\text{Var}(f_o(\mathbf{q} + \hat{\mathbf{Q}}))}{\bar{a}_o^2} \\ &\leq \frac{1}{\bar{a}_o^2} \frac{M-1}{\chi_{M-1,1-\beta}^2} \sigma_{f_o(\mathbf{q}+\hat{\mathbf{Q}})}^2. \end{aligned} \quad (5.5)$$

Für $(5.5) \leq \bar{p}_o$ erfüllt die Lösung des diskreten Optimierungsproblems (5.4) mit Wahrscheinlichkeit β auch die stochastische Nebenbedingung (5.1.2) des kontinuierlichen Problems (5.1).

5.3 Lösungsverfahren mit Sensitivitätsanalyse

Das Lösen von Optimierungsproblemen mit stochastischen Nebenbedingungen ist anspruchsvoll (Lejeune et al., 2018), da der Zusammenhang zwischen den zu optimierenden Variablen und der zulässigen Menge häufig nicht-linear ist und damit zu einem nicht-konvexen Problem führen kann. Wenn die Verteilung der stochastischen Nebenbedingung unabhängig von den Suchvariablen σ ist, ändert sich diese nicht durch die Wahl der Suchvariablen. Bei Optimierungsproblem (5.4) beeinflussen die Variablen σ jedoch nicht nur den Wert der Zielfunktion J , sondern sind auch an die Unsicherheiten gekoppelt, die durch die stochastischen Nebenbedingungen den zulässigen Suchraum beschreiben, vgl. Abb. 5.2. Dadurch wird nicht nur die Entscheidung, ob eine Wahl σ zulässig ist, erschwert (Lejeune et al., 2018), sondern auch die Auswahl an potentiellen Lösungsverfahren reduziert.

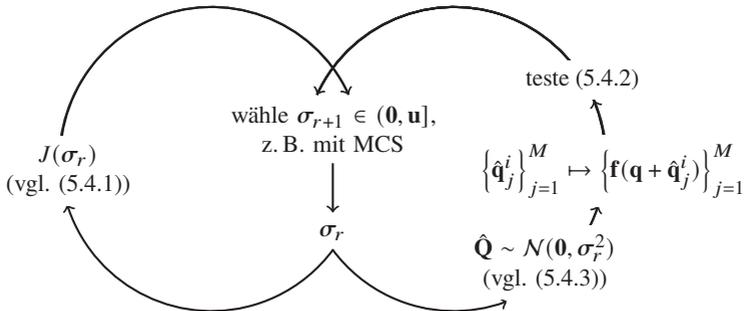


Abbildung 5.2: Schematische Darstellung allgemeiner numerischer Lösungsverfahren für Optimierungsprobleme mit stochastischen Nebenbedingungen am Beispiel (5.4) (engl. *decision dependent single chance constraint optimization problem*). Die Folge $\{\sigma_r\}_{r \in \mathbb{N}}$ von Suchvariablen wird durch ein numerisches Optimierungsverfahren erzeugt, z. B. mit einer MCS, bis das Verfahren mit einer Lösung σ_S^* abbricht.

Gewöhnliche Lösungsverfahren der restringierten Optimierung nutzen häufig u. a. den Gradienten der Nebenbedingungen, um das Optimum zu finden, z. B. Verfahren mit Lagrange-Multiplikatoren oder SQP-Verfahren (Reinhardt, Hoffmann & Gerlach, 2013). Bei den stochastischen Nebenbedingungen (5.1.2) ist das i. d. R. nicht umsetzbar, da durch die unbekannte Planungsfunktion \mathbf{f} auch der Zusammenhang zur W -Verteilung nicht bekannt und so die Bestimmung der Ableitung analytisch nicht möglich ist. Stattdessen kann die Nebenbedingung numerisch approximiert werden, z. B. mithilfe von Differenzenverfahren. Aufgrund ihres stochastischen Charakters reicht jedoch eine Approximation in einem einzelnen Punkt nicht aus. Daher wird in dieser Arbeit stattdessen die Multilevel-Koordinatensuche nach Huyer und Neumaier (1999) (engl. *MCS*) vorgeschlagen, die keine Gradienten nutzt.

Der zentrale Punkt dieses Verfahrens ist die Unterteilung des Suchraums in kleinere Gebiete: In jedem Schritt wird ein Teilgebiet ausgewählt und entlang einer einzelnen Koordinate in zwei neue Teilgebiete aufgeteilt. Bei hinreichend langer Rechenzeit wird so das globale Optimum der Zielfunktion gefunden. Um die Koordinate auszuwählen, entlang derer ein Gebiet unterteilt wird, werden die Informationen aus der SA verwendet. Das nachfolgende zweidimensionale Beispiel veranschaulicht die Funktionsweise der MCS mit SA, die anschließend detaillierter erklärt wird.

Beispiel 5.3.1 (MCS-Optimierung für $k = 2$). Seien $k = m = 2$, die oberen Schranken $\bar{\mathbf{b}} = (3, 0.329)^\top$ und die Wahrscheinlichkeiten $\bar{p}_1 = \bar{p}_2 = 0.1$ gegeben. Dann wird das diskrete Optimierungsproblem für eine Stichprobe $\{\mathbf{x}_j\}_{j=1}^M$ der ZV \mathbf{X} mit $\lambda_{S,i}$ aus Gleichung (3.12) definiert durch

$$\begin{aligned} \max_{\substack{\sigma_1 \in (0,1], \\ \sigma_2 \in (0,0.5]}} J(\boldsymbol{\sigma}) &:= \max_{\substack{\sigma_1 \in (0,1], \\ \sigma_2 \in (0,0.5]}} 1/\sqrt{2}(\sigma_1 + \sigma_2) \\ \text{u. d. N. } \frac{1}{M+1} g_{M+1} \left(\frac{M\lambda_{S,i}^2}{M-1+\lambda_{S,i}^2} \right) &\leq 0.1 \quad \forall i = 1, 2 \\ x_{j,i} &\sim \mathcal{N}(0, \sigma_i^2) \quad \forall j = 1, \dots, M, i = 1, 2. \end{aligned}$$

Die analytische Lösung des zugehörigen kontinuierlichen Optimierungsproblems ist durch die Standardabweichungen $\boldsymbol{\sigma}^* = (1, 0.2)^\top$ gegeben (vgl. Anhang A.6). Das diskrete Optimierungsproblem kann mit dem im Folgenden vorgestellten MCS-Algorithmus 5.1 gelöst werden. Für $s_{\max} = 15$ und $\bar{l} = 8$ bricht sie mit dem Ergebnis $\boldsymbol{\sigma}_S^* = (0.89, 0.07)^\top$ ab. Zwischenergebnisse des Lösungsprozesses sind in Abb. 5.3 bis 5.5 dargestellt.

Nachfolgend wird anhand von Beispiel 5.3.1 die Funktionsweise der MCS mit SA dargestellt. Algorithmus 5.1 gibt einen Überblick (s. S. 72), für weitere Details zum Originalalgorithmus sei auf Huyer und Neumaier (1999); Poštk, Huyer und Pál (2012) verwiesen.

Zu Beginn ist die zulässige Menge $[\mathbf{l}, \mathbf{u}] \subseteq \Omega$ durch ein Hyperrechteck $\mathbf{b} := [\mathbf{l}, \mathbf{u}]$ definiert, das iterativ in kleinere Hyperrechtecke unterteilt wird. Während der Initialisierung wertet der Algorithmus mehrere Punkte in jeder Koordinate des Suchraums aus (vgl. Algorithmus 5.1, Zeilen 1 und 2, Abb. 5.3; s. Huyer und Neumaier (1999) für mehr Details). Damit sammelt er Informationen, um den Suchraum in kleinere Hyperrechtecke zu unterteilen, denen in Abhängigkeit von z. B. ihrer Größe ein Level $s \in \{0, \dots, s_{\max}\}$ zugewiesen wird (vgl. Tabelle 5.1, Zeile 3). Für jedes Level wird die entsprechend der Zielfunktion vielversprechendste Box markiert, d. h. für ein Maximierungsproblem das Hyperrechteck mit dem größten lokalen Maximum (Box 4 für Level $s = 1$ in Abb. 5.3). Dass eine Box vielversprechend ist, bedeutet, dass sie noch in weitere Teilboxen unterteilt werden kann und es möglich ist, dort ein neues (lokales) Maximum zu finden.

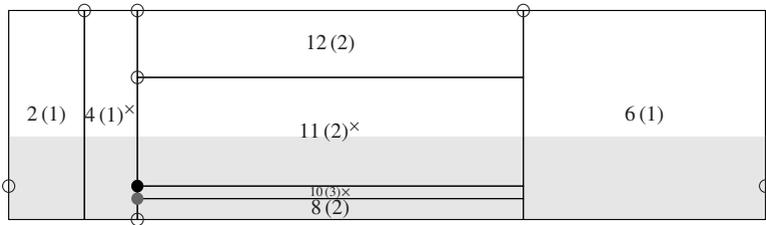


Abbildung 5.3: Nach der Initialisierung von Beispiel 5.3.1. Maximales Level $s_{\max} = 15$, Konvergenzparameter $\bar{l} = 8$, Zielfunktion $J(\sigma) = 1/\sqrt{2} \sum_i \sigma_i$, obere Schranke $u = (1, 0.5)^T$. Nummern zeigen Box-ID an (vgl. Tabelle 5.1), Zahlen in Klammern geben das Level der Box an, \times markiert die jeweils vielversprechendste Box. \circ unzulässiger Punkt; \bullet zulässige Lösung; \bullet aktuelles Optimum; \blacksquare analytisch zulässiger Suchraum $(0, 1] \times (0, 0.2]$.

In der zweiten Phase besteht ein Durchgang aus einer Iteration für jedes Level $s = 0, \dots, s_{\max}$: Für jedes Level s wird die vielversprechendste Box b ausgewählt (vgl. Algorithmus 5.1, Zeile 6) und entlang einer Koordinate geteilt (vgl. Tabelle 5.1, Zeilen 4 und 5; die Heuristiken für die Auswahl der zu teilenden Koordinate werden weiter unten erläutert). So entstehen zwei neue Hyperrechtecke b_1, b_2 mit $b = b_1 \cup b_2$ und Leveln $s_1 = s + 1, s_2 = s + 2$ (vgl. Algorithmus 5.1, Zeilen 22 und 23) sowie ein neuer, ausgewerteter Punkt $\sigma \in \overline{b_1} \cap \overline{b_2}$ (vgl. Algorithmus 5.1, Zeile 24). Das niedrigere Level $s + 1$ wird dabei der vielversprechenderen der beiden neuen Boxen b_1, b_2 zugewiesen.

Tabelle 5.1: Eigenschaften der Boxen für das MCS-Beispiel 5.3.1, $k = 2$. Die Vorgänger-Box wurde geteilt, um die aktuelle Box zu erhalten, z. B. sind die Boxen 13, 14 aus Box 4 entstanden. Level s sortiert alle Boxen für das Aufteilen. Die unteren beiden Zeilen geben die Anzahl der Teilungen für die Koordinaten $i = 1, 2$ an. Beispielsweise wurde Box 94 in Koordinate 2 geteilt, um Boxen 103, 104 zu erhalten, und damit erhöht sich ihre Anzahl in der unteren Zeile um 1.

Box ID	4	13	14	94	103	104
Vorgänger ID	1	4	4	54	94	94
Level s	1	4	3	5	6	7
# Teilungen in Koordinate $i = 1$	2	2	2	5	5	5
# Teilungen in Koordinate $i = 2$	0	1	1	0	1	1

Algorithm 5.1 MCS-Algorithmus mit SA (angelehnt an Henze, Faßbender und Stiller (2021)).

Input: Suchraum $[\mathbf{0}, \mathbf{u}]$, Planungsalgorithmus $f : \mathbb{R}^k \rightarrow \mathbb{R}^m$, $s_{\max}, \bar{l} \in \mathbb{N}$

Output: Optimum σ_S^* von Problem (5.4)

```

1: teile  $(\mathbf{0}, \mathbf{u}]$  in Boxen mit Leveln  $s \leq s_{\max}$ 
   ▶ Initialisierung, vgl. Abb. 5.3 (Huyer & Neumaier, 1999)
2:  $\forall$  Boxen  $\mathbf{b}$ : werte einen Punkt  $\sigma \in \mathbf{b}$  aus und bestimme Optimum  $\sigma_{S,\mathbf{b}} \in \mathbf{b}$ 
3: Optimum  $\sigma_S^* \leftarrow \max_{\mathbf{b}} \sigma_{S,\mathbf{b}}$ 
4: while  $\exists$  neues Optimum  $\sigma_S^*$  in den letzten  $\bar{l}$  Iterationen do ▶ Phase 2
5:   for  $s = 0, \dots, s_{\max}$ :  $\exists$  Boxen  $\{\mathbf{b}\}_s$  mit Level  $s$  do
6:      $\mathbf{b} \leftarrow$  Box mit dem größten lokalen Optimum aller  $\{\mathbf{b}\}_s$ 
7:      $\{i_l\} \leftarrow \arg \min_{i=1, \dots, k} \# \text{Teilungen der Box } \mathbf{b} \text{ in Koordinate } i$ 
8:     if  $s_b > 2k(i_l + 1)$  then ▶ Teilen nach Rang, vgl. Abb. 5.4(b)
9:        $i_{\text{split}} \leftarrow \arg \max_{i \in \{i_l : s_b > 2k(i_l + 1)\}} \|\mu_{\cdot, i}^*\|$ 
10:    else if Box  $\mathbf{b} = [l_b, u_b]$  vielversprechend then
11:      ▶ Teilen nach Gewinn, vgl. Abb. 5.4(a)
12:      if  $\sigma_{S,\mathbf{b}}$  zulässig then
13:         $\eta \leftarrow -1, l_b \leftarrow \{i | \sigma_i < u_{b,i}\}$  ▶ vgl. (5.6)
14:      else if  $\sigma_{S,\mathbf{b}}$  verletzt (5.4.2) then
15:         $\eta \leftarrow 1, l_b \leftarrow \{i | \sigma_i > l_{b,i}\}$  ▶ vgl. (5.6)
16:      else if  $\sigma_{S,\mathbf{b}}$  verletzt (5.2) then
17:         $\eta \leftarrow 0, l_b \leftarrow \{i | \sigma_i = 0\}$  ▶ vgl. (5.6)
18:      else
19:        continue
20:      end if
21:       $i_{\text{split}} \leftarrow \arg \max_{i \in l_b} 0.5(1 - \eta)w_i + \eta \|\mu_{\cdot, i}^*\|$  ▶ vgl. (5.6)
22:    end if
23:     $\mathbf{b}_1, \mathbf{b}_2 \leftarrow$  teile Box  $\mathbf{b}$  entlang Koordinate  $i_{\text{split}}$ 
24:     $s_{b_i} \leftarrow s + i$  (je nach dem, wie vielversprechend  $\mathbf{b}_i$  ist)  $\forall i = 1, 2$ 
25:    werte Punkt  $\sigma \in \overline{\mathbf{b}_1} \cap \overline{\mathbf{b}_2}$  aus für (5.4.1), (5.4.2)
26:     $\sigma_{S,\mathbf{b}_1}, \sigma_{S,\mathbf{b}_2} \leftarrow$  Optimum der zulässigen Punkte in den Boxen  $\mathbf{b}_1, \mathbf{b}_2$ 
27:     $\sigma_S^* \leftarrow \max\{\sigma_S^*, \sigma_{S,\mathbf{b}_1}, \sigma_{S,\mathbf{b}_2}\}$ 
28:     $\bar{l} \leftarrow \bar{l} + 1$ 
29:  end for
30: end while

```

Anschließend wird die vielversprechendste Box in den Leveln $s + 1$, $s + 2$ aktualisiert, falls der neue Punkt σ besser ist als die Punkte der bisher bekannten Boxen in diesen Leveln. Die Hyperrechtecke, die nicht vielversprechend sind, z. B. weil sie zu klein sind, werden in den danach folgenden Iterationen nicht mehr geteilt.

Im Originalalgorithmus schließt sich eine dritte Phase an, die für alle Boxen mit maximalem Level $s = s_{\max}$ eine lokale Suche durchführt. Da dies optional ist, wird hier jedoch darauf verzichtet. Die MCS ist beendet, sobald das Extremum sich wie in Abb. 5.5 während der letzten \bar{l} Durchgänge der zweiten Phase nicht verändert hat oder wenn keine vielversprechenden Hyperrechtecke verbleiben.

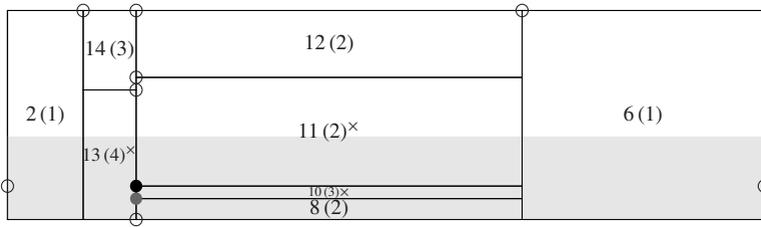
Huyer und Neumaier (1999) stellen zwei Möglichkeiten vor, die Koordinate zu bestimmen, entlang derer eine Box aufgeteilt wird: entweder entsprechend dem erwarteten Gewinn für die Zielfunktion oder nach Rang. Wenn *Teilen nach Rang* ausgewählt wird, d. h. wenn die Box ein hohes Level s hat, wird die Koordinate ausgewählt, die bisher am wenigsten aufgeteilt wurde (vgl. Box 94 in Abb. 5.4(b) und Boxen 103, 104 in Abb. 5.5). Somit werden auch Bereiche unterteilt, die bisher nur wenig untersucht wurden, d. h. das Suchgebiet wird global erkundet. Diese Heuristik *Teilen nach Rang* (engl. *split by rank*) wird übernommen (vgl. Algorithmus 5.1, Zeile 7).

Das *Teilen nach Gewinn* wird im Original durch ein Ersatzmodell abgeschätzt, um die Koordinate mit dem vermutlich größten Zuwachs zu identifizieren. Anstelle des von Huyer und Neumaier (1999) vorgeschlagenen Ersatzmodells wird in dieser Arbeit eine Kombination aus dem Gradienten der Zielfunktion $\nabla J = \mathbf{w}$ und den Informationen, die durch die SA ermittelt werden, genutzt (vgl. Algorithmus 5.1, Zeile 10; Box 4 in Abb. 5.3 und Boxen 13, 14 in Abb. 5.4(a)). Da das Optimierungsproblem das Maximum von J bestimmen soll, wird die zu teilende Koordinate i_{split} für eine Box \mathbf{b} durch

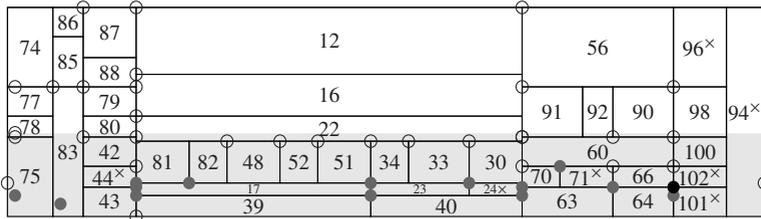
$$i_{\text{split}} := \arg \max_{i \in I_{\mathbf{b}}} 0.5(1 - \eta)w_i + \eta \|\boldsymbol{\mu}_{o,i}^*\| \quad (5.6)$$

bestimmt. Die Indexmenge $I_{\mathbf{b}}$ einer Box \mathbf{b} und das Gewicht η werden dafür in Abhängigkeit von den Nebenbedingungen des Optimierungsproblems und des besten Punkts $\sigma_{\mathbf{b}}$ der Box \mathbf{b} folgendermaßen definiert

- der beste Punkt $\sigma_{\mathbf{b}}$ ist zulässig $\Rightarrow I_{\mathbf{b}} := \{i \mid \sigma_{\mathbf{b},i} < u_{\mathbf{b},i}\}$ und $\eta = -1$ (Algorithmus 5.1, Zeile 11)



(a) *Teilen nach Gewinn*: Box 4 wurde entsprechend der SA mit $\eta = -1$, $I = \{2\}$, $i_{\text{split}} = 2$ geteilt. Zahlen in Klammern geben das Level der Box an.



(b) *Teilen nach Rang*: Box 94 wird entlang der am wenigsten geteilten Koordinate entsprechend der Heuristik *teile falls* $s > 2k$ ($\min_i \{\# \text{Teilungen der Box in Koordinate } i + 1\}$) in $i_{\text{split}} = 2$ geteilt. Dabei entstehen die Boxen 103, 104 (vgl. Abb. 5.5).

Abbildung 5.4: Heuristiken zum Teilen der Boxen in Beispiel 5.3.1. Maximales Level $s_{\text{max}} = 15$, Konvergenzparameter $\bar{l} = 8$, Zielfunktion $J(\sigma) = 1/\sqrt{2} \sum_i \sigma_i$, obere Schranke $\mathbf{u} = (1, 0.5)^T$. Nummern zeigen Box-ID an (vgl. Tabelle 5.1), \times markiert die jeweils vielversprechendste Box. \circ unzulässiger Punkt; \bullet zulässige Lösung; \bullet aktuelles Optimum; \blacksquare analytisch zulässiger Suchraum $(0, 1] \times (0, 0.2]$.

- der beste Punkt σ_b ist aufgrund der stochastischen Nebenbedingung (5.4.2) unzulässig $\Rightarrow I_b := \{i \mid \sigma_{b,i} > l_{b,i}\}$ und $\eta = 1$ (Algorithmus 5.1, Zeile 13)
- der beste Punkt σ_b ist wegen der Nebenbedingung (5.2) unzulässig $\Rightarrow I_b := \{i \mid \sigma_{b,i} = 0\}$ und $\eta = 0$ (Algorithmus 5.1, Zeile 15)

Dabei wird angenommen, dass kleine Standardabweichungen σ im Vergleich zu großen Standardabweichungen bessere Kandidaten dafür sind, die stochastische Nebenbedingung (5.4.2) zu erfüllen. Das Gewicht η in Gleichung (5.6) wird dafür genutzt, um die Ergebnisse der SA gegenüber dem Gradienten $\nabla J = \mathbf{w}$ der Zielfunktion abzuwägen. Es sorgt dafür, dass zunächst ein zulässiger Punkt gefunden und darauf aufbauend die Zielfunktion maximiert wird. Falls $I_b = \emptyset$

Beispiel 3.1.8.h (Freifahrt: numerische Optimierung). Das diskrete Optimierungsproblem für Beispiel 3.1.8 ist für die Stichprobe $\{\hat{v}_j\}_{j=1}^M$ und $\lambda_{S,a} = \sqrt{M/(M+1)}\bar{b}_a/\sigma_{\hat{v}_a(v+\hat{v})}$ (vgl. (3.12)) definiert durch

$$\max_{\sigma_v \in (0,2]} J(\sigma_v) = \max_{\sigma_v \in (0,2]} \sigma_v, \tag{5.7.1}$$

$$\text{u. d. N. } \frac{1}{M+1} g_{M+1} \left(\frac{M\lambda_{S,a}^2}{M-1+\lambda_{S,a}^2} \right) \leq \bar{p}_a, \tag{5.7.2}$$

$$\hat{v}_j \sim \mathcal{N}(0, \sigma_v^2) \quad \forall j = 1, \dots, M. \tag{5.7.3}$$

Seien $\bar{b}_a = 0.8 \text{ m/s}^2$ und $\bar{p}_a = 0.1$. Das Optimierungsproblem wird mit $M = 300$ Stichprobenwerten mit dem MCS-Algorithmus gelöst. Wie oben beschrieben wird dafür der Suchraum $(0, 2]$ iterativ in kleinere Intervalle unterteilt, vgl. Abb. 5.6. Wegen $k = 1$ gibt es in jeder Iteration nur ein Intervall, das vielversprechend ist und damit weiter unterteilt werden kann. Alle anderen Intervalle sind als nicht-vielversprechend markiert. Dafür gibt es zwei Gründe: Erstens, weil die untere Intervallgrenze nicht zulässig ist, d. h. die stochas-

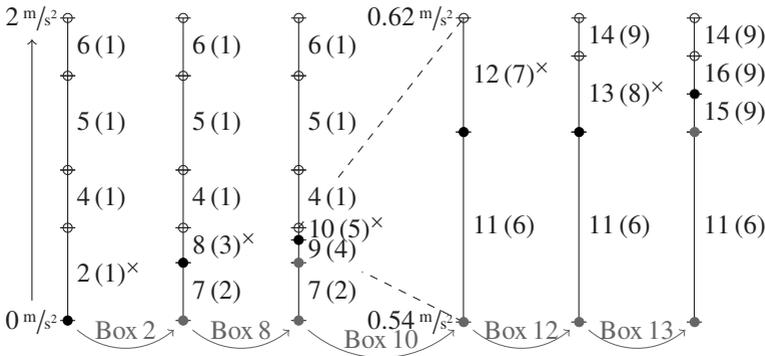


Abbildung 5.6: MCS für Beispiel 3.1.8.h (Freifahrt) mit $k = 1$, maximales Level $s_{\max} = 15$, Konvergenzparameter $\bar{l} = 8$, Optimierungsfunktion $J(\sigma) = \sigma$ im Suchraum $[0.01 \text{ m/s}^2, 2 \text{ m/s}^2]$. Das linke Intervall zeigt die Unterteilung nach der Initialisierung, die anderen die Unterteilung nach der Teilung von jeweils einer weiteren Box. Optimum: $\sigma_{S,v}^* = 0.6037 \text{ m/s}^2$. Nummern zeigen Box-ID an; Zahlen in Klammern das Level; \times vielversprechendste Box; \circ unzulässiger Punkt; \bullet zulässige Lösung; \bullet (aktuelles) Optimum.

tische Nebenbedingung nicht erfüllt. Es wird angenommen, dass kleinere Standardabweichungen bessere Kandidaten für die Erfüllung der stochastischen Nebenbedingung sind als größere. Daher ist zu erwarten, dass auch alle anderen Punkte, die größer sind als die untere Intervallgrenze, die stochastische Nebenbedingung verletzen und damit keine zulässigen Punkte sind (vgl. z. B. Intervall 5 in Abb. 5.6). Zweitens werden Intervalle als nicht-vielversprechend markiert, wenn die obere Grenze zulässig ist. Dann ist zwar anzunehmen, dass auch kleinere Standardabweichungen die Nebenbedingungen erfüllen und zulässig sind, aber wegen der Monotonie der Zielfunktion können sie kein neues Maximum sein (vgl. z. B. Intervall 7 in Abb. 5.6). Daher bleibt für $k = 1$ in jedem Schritt nur ein Intervall mit einer möglichen Koordinate zum Teilen und eine SA bringt in diesem Beispiel keine zusätzlichen Informationen.

Die MCS bricht mit dem Optimum $\sigma_{S,v}^* = 0.6037 \text{ m/s}$ ab. Mit $\sigma_{f_a(v+\hat{v})} = 0.2381 \text{ m/s}^2$ und $\lambda_{S,a} = 3.3543$ gilt $0.0897 < 0.1 = \bar{p}_a$ und die Nebenbedingung (5.7.2) ist erfüllt. Im Vergleich mit der Lösung $\sigma_{S,v}^* = 1.2266 \text{ m/s}^2$ des kontinuierlichen Problems (5.3), Beispiel 3.1.8.g, zeigt dies, dass die Lösung des diskreten Problems die analytische Lösung deutlich unterschätzt, aber alle Nebenbedingungen erfüllt.

Die MCS wertet für Beispiel 5.3.1 48 Punkte aus, bevor sie abbricht. Dies unterstützt die Aussage von Frandi und Papini (2014), dass Koordinatensuchen i. d. R. langsam konvergieren. Die MCS birgt jedoch durch seine Suchstrategie *Teilen durch Gewinn* den Vorteil, dass sie vor allem am Anfang schnellen Fortschritt erzielt und somit weniger Funktionsauswertungen benötigt, wenn sie das Optimum nur grob approximieren soll (Pošík et al., 2012). Außerdem ist das MCS-Verfahren gut zur Kopplung mit der SA nach Morris geeignet: Wenn in einer Iteration mehrere Koordinaten für die Unterteilung einer Box zur Auswahl stehen, wird die zu teilende Koordinate in Abhängigkeit des absoluten Stichprobenerwartungswerts μ^* ausgewählt, der von der SA berechnet wird. Somit wird die stochastische Approximation des Gradienten der Nebenbedingung verwendet, auch wenn der exakte Gradient fehlt.

Diese Kopplung der beiden Verfahren ist ein Vorteil gegenüber anderen Optimierungsalgorithmen wie z. B. der direkten gitteradaptiven Suche (engl. *Mesh-Adaptive Direct Search (MADS)*) (Audet, Custódio & Dennis, 2008): Bei der Verknüpfung von MADS mit einer SA kann die Konvergenz der MADS nicht mehr garantiert werden, da diese SA eine Schritt-für-Schritt-Methode

(engl. *One-At-A-Time*, OAT) ist, die in jeder Iteration nur eine Koordinate verändert. Das Konvergenzresultat der MADS setzt jedoch eine dichte Menge an Suchrichtungen voraus und somit kann die SA mit k Suchrichtungen nicht zur Approximation genutzt werden.

5.4 Anwendung: Überholmanöver

Beispiel 3.2.6.d (Überholmanöver: numerische Optimierung). Für Beispiel 3.2.6 wird ein diskretes Optimierungsproblem mit $k = 6$ Inputs und dem Output Beschleunigung a berechnet, d. h. $m = 1$. Die stochastische Nebenbedingung für die FSW-Entscheidung lc wird nicht in dem Optimierungsproblem berücksichtigt, weil Beispiel 3.2.6.b bereits zeigen konnte, dass die FSW-Entscheidung nach dem MOBIL-Modell gegenüber Messunsicherheiten nicht robust genug ist.

Das diskrete Optimierungsproblem wird zunächst mit MCS und SA ohne DSC gelöst ($s_{\max} = 15$, $\bar{l} = 8$). Die optimalen Standardabweichungen σ_s^* sind in Abb. 5.7 dargestellt. Abb. 5.7(a) und 5.7(b) zeigen, dass bis $t \leq 1.6$ s die zulässige Messabweichung der Distanz zum vorausfahrenden Fahrzeug und die Geschwindigkeit des vorausfahrenden Fahrzeugs immer weiter abnehmen, weil das Ego den Abstand zum vorausfahrenden Fahrzeug verringert. In diesem Zeitraum hat das Fahrzeug auf dem linken Fahrstreifen hinter dem Ego keinen Einfluss auf die Berechnung der Beschleunigung, da es noch zu weit weg ist und das Ego noch keinen FSW in Betracht zieht.

Für $t \in [1.7 \text{ s}, 2.0 \text{ s}]$ werden alle Standardabweichungen an den unteren zulässigen Rand 0.01 m/s bzw. 0.01 m optimiert, was darauf hindeutet, dass der zulässige Suchraum in diesem Zeitraum sehr klein ist. Dies ist darauf zurückzuführen, dass das Ego den Wechsel zwischen den Homotopieklassen *kein FSW* und *FSW nach links* abwägt. Hier ist das IDM-MOBIL-Modell nicht robust genug, um mit den unsicheren Umgebungsinformationen umgehen zu können. Die Wahrscheinlichkeitsnebenbedingung zeigt jedoch, dass nur für $t \in [1.9 \text{ s}, 2.0 \text{ s}]$ die Nebenbedingung mit $p > \bar{p}_a = 0.1$ verletzt wird, vgl. Abb. 5.7(c).

Sobald das Ego den FSW zum Zeitpunkt $t = 2.0$ s beginnt, vergrößern sich die zulässigen Messungenauigkeiten. Dabei fällt auf, dass sich die Anforderung

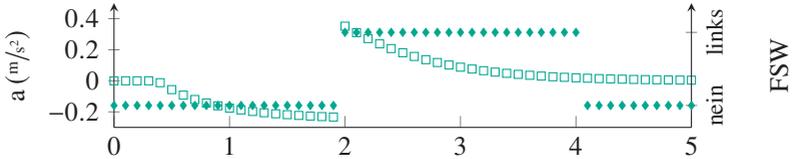
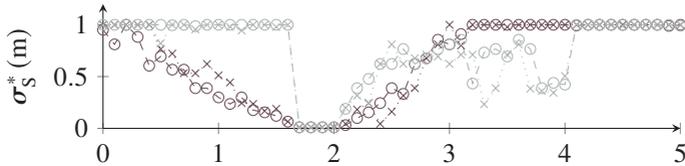
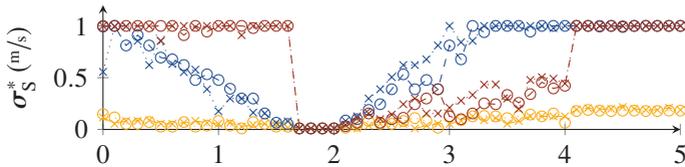


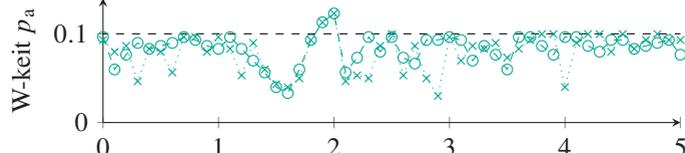
Abb. 3.6: Fahrentscheidung $\mathbf{f}(\mathbf{q}) = (f_a(\mathbf{q}), f_{ic}(\mathbf{q}))^T$ nach IDM-MOBIL-Modell ohne veräuschte Wahrnehmung. \square Beschleunigung $f_a(\mathbf{q})$; \blacklozenge FSW $f_{ic}(\mathbf{q})$.



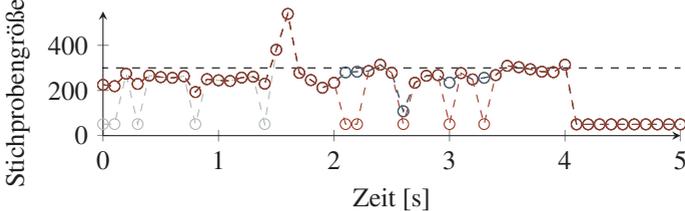
(a) Standardabweichungen $\sigma_{S,dvs}^*$, $\sigma_{S,dnl}^*$ für Distanzen.



(b) Standardabweichungen $\sigma_{S,v}^*$, $\sigma_{S,vvs}^*$, $\sigma_{S,vnl}^*$ für Geschwindigkeiten.



(c) Wahrscheinlichkeit für unzulässige Planungsentscheidungen mit Saw-Ungleichung (3.13).



(d) Anzahl an ausgewerteten EE für die Optima σ_S^* für die SA mit DSC.

Abbildung 5.7: Optimierungsergebnisse für Beispiel 3.2.6.d (Überholmanöver). Standardabweichungen σ_S^* mit SA mit $M = 300$ $\cdot \times \cdot$ d_{vs} , $\cdot \times \cdot$ d_{nl} , $\cdot \times \cdot$ v , $\cdot \times \cdot$ v_{vs} , $\cdot \times \cdot$ v_{nl} ; Standardabweichungen σ_S^* mit SA und DSC \circ d_{vs} , usw.

an das Fahrzeug vor dem Ego schneller abschwächen als für das Fahrzeug hinter dem Ego. Nachdem das Ego dem linken Ziel-Fahstreifen zugeordnet ist ($t \geq 4.1$ s), haben die Inputs aufgrund der vorliegenden Modellierung keinen Einfluss mehr auf die Ego-Beschleunigung, weil es kein vorausfahrendes Fahrzeug auf demselben Fahstreifen bzw. kein nachfolgendes Fahrzeug auf dem linken Fahstreifen mehr gibt. Daher ist lediglich die Ego-Geschwindigkeit v noch relevant.

Zusätzlich zu den Distanzen und Geschwindigkeiten der anderen Fahrzeuge wurde auch die Standardabweichung für die Position des Egos in der Optimierung berücksichtigt. Abgesehen vom Zeitintervall $[1.7$ s, 2.0 s] gilt für die Standardabweichungen jedoch $\sigma_p > 40$ m (vgl. Abb. A.4 in Anhang A.7). Dies bedeutet, dass dieser Parameter keinen nennenswerten Einfluss auf das Planungsergebnis hat. Dies stimmt mit den Informationen aus Beispiel 3.1.4 und Beispiel 3.2.6.a überein: Die Position des Egos wird nur zur Berechnung der Abstände zu anderen Fahrzeugen genutzt und hat auf einer geraden Strecke keinen Einfluss auf das Planungsergebnis, da die Ego-Position und die Distanzen zu den anderen Fahrzeugen in der vorliegenden Implementierung unabhängig von einander verrauscht werden. Daher wird dieser Parameter hier nicht weiter diskutiert.

Anschließend wurde das Optimierungsproblem auch mit MCS, SA und DSC gelöst. Dafür wurden dieselben Parameter verwendet wie in Abschnitt 4.2.2: $\kappa_{\text{stop}} = 0.05$, $\kappa_{\text{act}} = 10^{-3}$, $M_{\text{min}} = 50$. Der Vergleich der Optimierungsergebnisse mit und ohne DSC zeigt, dass die MCS mit SA und DSC ähnliche Ergebnisse liefert wie der Algorithmus ohne DSC und die Tendenzen der optimalen Standardabweichungen übereinstimmen (vgl. Abb. 5.7(a) und 5.7(b)). Die Unterschiede, die sich z. B. für die Geschwindigkeit des vorausfahrenden Fahrzeugs v_{vs} im Zeitraum $t \in [2.7$ s, 3.4 s] zeigen, sind im Rahmen der im Modell enthaltenen Stochastik akzeptabel. Die Auswertung der Wahrscheinlichkeitsnebenbedingung zeigt, dass beide Algorithmen in der Lage sind, zulässige Standardabweichungen so zu bestimmen, dass die stochastische Nebenbedingung nicht verletzt wird.

Der Vorteil des DSC liegt darin, dass es die Stichprobengröße wie gewünscht dynamisch an die Komplexität der Situation anpasst und somit den Rechenaufwand reduziert. Abb. 5.7(d) zeigt beispielhaft die Anzahl der ausgewerteten EE für die optimale Standardabweichung σ_S^* für alle $k = 6$ Inputs. Für den gesamten simulierten Zeitraum $[0$ s, 5 s] wurden für die SA des jeweiligen Optimums σ_S^*

36.76 % weniger EE ausgewertet als für die Referenz mit Stichprobengröße $M = 300$ (58 058 mit DSC vs. 91 800 ohne DSC, $M = 300$ EE fix für alle $k = 6$ Inputs und alle 51 Zeitschritte).

Um die Optimierungsergebnisse zu validieren, wird der Planungsalgorithmus für Umgebungsbeschreibungen ausgewertet, die mit einer $\mathcal{N}(\mathbf{0}, \sigma_S^{*2})$ -verteilten Unsicherheit behaftet sind. Eine Option sind Closed-Loop-Simulationen: Dabei nutzen alle Simulationen dieselben Startinformationen zum Zeitpunkt $t = 0$ s, aber entwickeln sich über den Zeithorizont unterschiedlich weiter. Dies liegt daran, dass verschiedene Planungsentscheidungen in einem Zeitpunkt zu unterschiedlichen Zuständen in nachfolgenden Zeitpunkten führen. Diese Korrelation zwischen den Zeitschritten wurde im Optimierungsproblem nicht berücksichtigt. Daher können die Ergebnisse einer solchen Validierungsstudie nur für kleine Zeitintervalle eine belastbare Aussage über die Güte der Optimierungsergebnisse treffen (Henze, Faßbender & Stiller, 2021).

Stattdessen wird das Optimierungsergebnis σ_S^* für jeden Zeitpunkt $t \in [0 \text{ s}, 5 \text{ s}]$ unabhängig von den vorangegangenen, unsicherheitsbehafteten Planungsentscheidungen validiert: Für jeden Zeitschritt $t \in [0 \text{ s}, 5 \text{ s}]$ wird der Zustand der Umgebung als Ausgangszustand verwendet, der ohne Unsicherheiten $\hat{v}_t = \mathbf{0}$ für alle $\tilde{t} < t$ erzeugt wurde. Daraufhin wird eine Stichprobe $\{\hat{v}_{t,j}\}_{j=1}^M$ mit $M = 300$ Stichprobenwerten gemäß der Verteilung $\mathcal{N}(\mathbf{0}, \sigma_S^{*2})$ gezogen und die Stichprobe $\{f_a(v_t + \hat{v}_{t,j})\}_{j=1}^M$ für den Planungsalgorithmus ausgewertet. Durch dieses Vorgehen werden möglicherweise auftretende Effekte durch Korrelationen zwischen den Zeitschritten vermieden und es wird untersucht, ob das Ergebnis σ_S^* dem Optimierungsproblem angemessen ist.

Abbildungen 5.8 bis 5.10 zeigen die Ergebnisse der Validierung für verschiedene Unsicherheiten $\sigma \in \{0.5\sigma_S^*, \sigma_S^*, 1.5\sigma_S^*\}$. Die oberen Abbildungen stellen jeweils den Beschleunigungsverlauf ohne Rauschen sowie den erwarteten Verlauf $\mu_{f_a(v_t + \hat{v}_t)}$ dar ($t \in [0 \text{ s}, 5 \text{ s}]$), der sich auf Grundlage der von der MCS und SA mit bzw. ohne DSC berechneten maximalen Unsicherheit σ_S^* ergibt. Außerdem sind für jeden Zeitschritt exemplarisch fünf der $M = 300$ berechneten Beschleunigungswerte dargestellt. Die unteren Abbildungen zeigen jeweils Abschätzungen dafür, wie viele Stichprobenwerte die stochastische Nebenbedingung (5.4.2) verletzen: Einerseits ist der Anteil der Stichprobenwerte dargestellt, die weiter von der erwarteten Entscheidung $\mu_{f_a(v_t + \hat{v}_t)}$ abweichen als die akzeptable Schranke $\bar{b}_a = 0.8 \text{ m/s}^2$. Andererseits ist die Abschätzung nach

Chebychev-Saw (3.13) angegeben. Diese obere, konservative Schranke liegt deutlich über den Beobachtungen der Stichprobe.

Für die Unsicherheit $\sigma = 0.5\sigma_S^*$ wird deutlich, dass die Abweichungen von der unter diesen Unsicherheiten zu erwartenden Beschleunigung $\mu_{f_a(v_t+\hat{v}_t)}$ unter der zulässigen Schranke $\bar{b} = 0.8 \text{ m/s}^2$ bleiben (vgl. Abb. 5.8(b)). Dies wird auch von der Chebychev-Saw-Abschätzung unterstützt. Lediglich für $t = 4.1 \text{ s}$ wird die Chebychev-Saw-Ungleichung verletzt. Dies ist darauf zurückzuführen, dass das zugrundeliegende IDM-MOBIL-Modell keine komfortablen Wechsel zwischen den Entscheidungen *FSW nach links* und *kein FSW* bestimmt. Der Anteil der Stichprobenwerte, die tatsächlich zu einer Beschleunigung führen, die weiter als \bar{b}_a von der Erwartung abweicht, beträgt jedoch lediglich 2.33 %, sodass die stochastische Nebenbedingung (5.4.2) eingehalten wird.

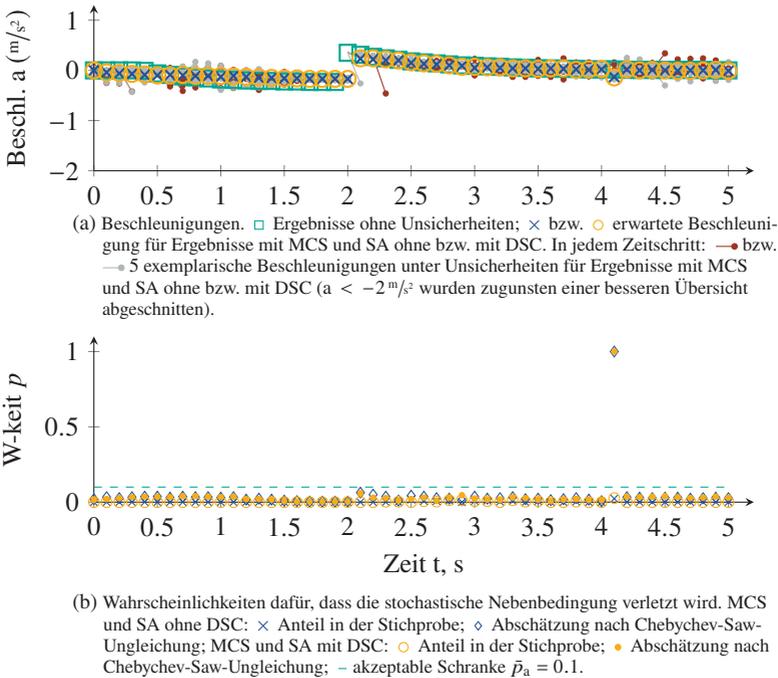


Abbildung 5.8: Validierungsergebnisse für Unsicherheiten $\sigma = 0.5\sigma_S^*$.

Für die Unsicherheit $\sigma = \sigma_S^*$, die durch Lösen des Optimierungsproblems berechnet wurde, zeigt Abb. 5.9(a), dass größere Abweichungen auftreten als für die Unsicherheiten $\sigma = 0.5\sigma_S^*$ (vgl. Abb. 5.8(a)). Abb. 5.9(b) zeigt an, dass lediglich die Chebychev-Saw-Ungleichung verletzt ist und auch in dieser Stichprobe der Anteil der unzulässigen Planungsentscheidungen kleiner als die akzeptable Schranke $\bar{p}_a = 0.1$ bleibt. Daher ist die Lösung σ_S^* des Optimierungsproblems angemessen.

Abbildung 5.10(a) zeigt, dass die Unsicherheiten $\sigma = 1.5\sigma_S^*$ zu teils sehr starken Abweichungen von der erwarteten Beschleunigung führen. Die Chebychev-Saw-Ungleichung ist für viele Zeitpunkte verletzt (vgl. Abb. 5.10(b)), auch wenn der Anteil der tatsächlich unzulässigen Planungsentscheidungen nur für die Zeitpunkte $t \in \{3.6 \text{ s}, 3.8 \text{ s}, 3.9 \text{ s}, 4.0 \text{ s}, 4.1 \text{ s}\}$ ($M = 300$) bzw. $t \in \{4.0 \text{ s}, 4.1 \text{ s}\}$

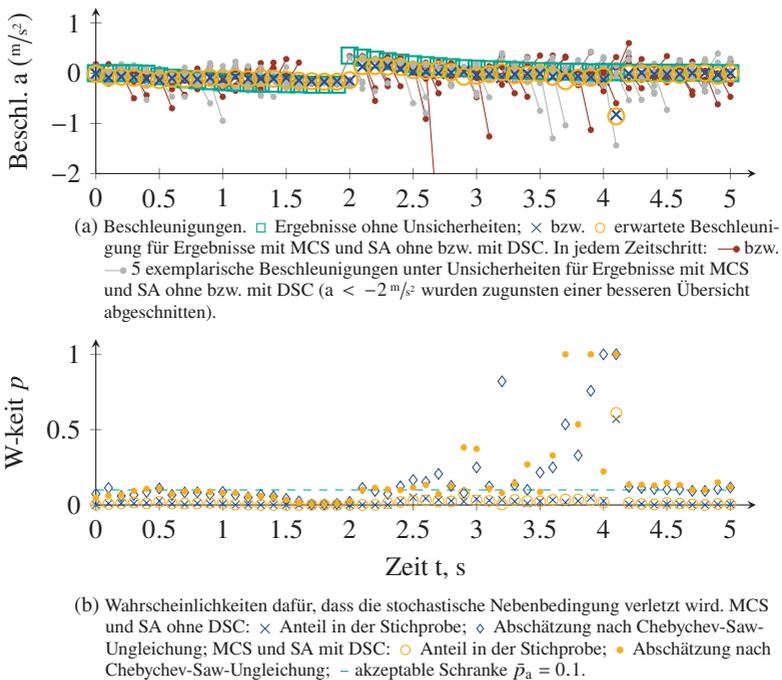


Abbildung 5.9: Validierungsergebnisse für Unsicherheiten $\sigma = \sigma_S^*$.

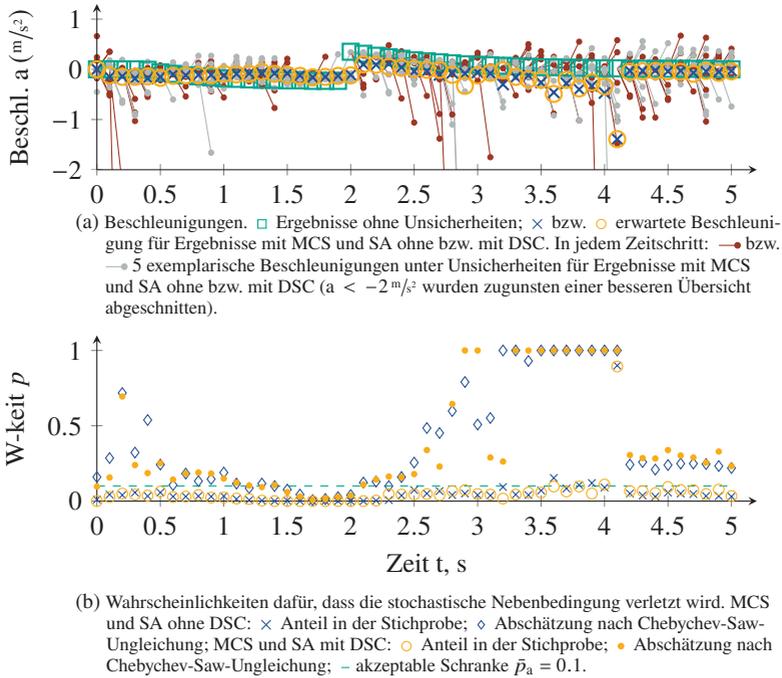


Abbildung 5.10: Validierungsergebnisse für Unsicherheiten $\sigma = 1.5\sigma_S^*$.

(mit DSC) den zulässigen Anteil von 10 % überschreitet. Dies bestätigt, dass die MCS mit SA sowohl mit als auch ohne DSC wie erwartet eher konservative Unsicherheiten bestimmt, diese aber zu zulässigen Planungsentscheidungen gemäß Definition 3.2.5 führen und größere Unsicherheiten diese nicht mehr erfüllen können.

6 Informationsbedarf von HAF-Entwicklern

Die bisher vorgestellte Arbeit beschäftigte sich hauptsächlich mit der technischen Nachvollziehbarkeit an der Schnittstelle zwischen Wahrnehmung und Planung. Dieses Kapitel wendet sich dem zweiten Aspekt von transparenten Verhaltensentscheidungen zu: Maßnahmen, die das Vertrauen der Nutzer unterstützen.

Die folgenden drei Studien zeigen, warum das Vertrauen von Nutzern in AF essentiell ist: Eine Befragung durch das amerikanische Meinungsforschungsinstitut Pew Research Center (2017) ergab, dass 56 % der Amerikaner kein AF nutzen würden, 42 % davon, weil sie einem AF nicht vertrauen oder besorgt sind, die Kontrolle abzugeben, und 30 %, weil sie Sicherheitsbedenken haben. 2021 zeigte eine weitere Studie von Shariff et al., dass Amerikaner lieber ein Taxi nutzen würden, das von einem Menschen gefahren wird, als ein AF, wenn beide Fahrzeuge dieselbe Risikobewertung haben. Hartwich et al. (2020) konnten diese Ergebnisse in einer Simulatorstudie bestätigen. Die Probanden erlebten zweimal die exakt selbe, vorher aufgezeichnete Fahrt: einmal in dem Glauben, dass ein Mensch die Simulatorfahrt durchführe, und einmal in dem Glauben, dass die Fahrt automatisiert sei. Die Ergebnisse zeigen, dass die Probanden ein AF als weniger sicher und weniger komfortabel wahrnehmen und es mit einem schlechter verständlichen Fahrverhalten sowie weniger Fahrspaß in Zusammenhang bringen. Dies deutet daraufhin, dass das Vertrauen von Nutzern in AF noch nicht groß genug ist.

Eine Möglichkeit, um das Vertrauen von Nutzern zu steigern, sind Erklärungen, die das Verhalten eines AF erläutern (Kim et al., 2018; Omeiza et al., 2022). Dafür wird im Rahmen dieser Arbeit angenommen, dass Fragen, die sich Novizen stellen, beantwortet werden können, wenn zuvor auch alle Fragen der Experten beantwortet werden konnten. Daher stellt Abschnitt 6.1 eine Fokusgruppenstudie vor, die den Informationsbedarf von Entwicklern

hochautomatisierter Fahrfunktionen untersucht, indem sie Fragen aufdeckt, die durch eine unerwartete Handlung des AF auftreten. Abschnitte 6.2 und 6.3 beschreiben und diskutieren die Ergebnisse.

Vorläufige Ergebnisse dieses Kapitels wurden in Henze, Stasinski et al. (2021) veröffentlicht. Abbildungen, Tabellen und andere Ergebnisse wurden für eine konsistente Darstellung angepasst.

6.1 Fokusgruppe

Fokusgruppen sind moderierte Gruppendiskussionen und gehören zu den qualitativen, explorativen Forschungsmethoden (Krueger & Casey, 2014; Mishra, 2016). Sie zielen darauf ab, Personen mit reichhaltigem Erfahrungs- oder Wissensschatz gezielt in einer Gruppe zusammenzubringen und sie über ein Thema diskutieren zu lassen, sodass neue Gedanken und Ideen entstehen (Krueger & Casey, 2014; Ryan, Gandha, Culbertson & Carlson, 2014). Mit der hier vorgestellten Fokusgruppenstudie wird der Informationsbedarf von Personen mit ausgeprägtem fachlichen Hintergrundwissen (Experten) während Testfahrten im Realverkehr untersucht, um so die Transparenz automatisierter Fahrentscheidungen zu verbessern. Dabei geht es nur um hochautomatisiertes Fahren (HAF)-Entwickler, die die Fahrfunktion während der Testfahrten vom Beifahrersitz aus analysieren. Die Bedürfnisse von Entwicklern, die das Fahrzeug und die Fahrfunktion als Sicherheitsfahrer vom Fahrersitz aus überwachen, wurden nicht untersucht.

Interviewleitfaden

Der Interviewleitfaden, der in jeder Gruppe besprochen wurde, ist in Tabelle 6.1 dargestellt. Er umfasste fünf offene und drei Single-Choice-Fragen. Die ersten beiden Fragen Q1 und Q2 dienten als Einstiegsfragen in das Thema, damit sich die Teilnehmer kennenlernen und von positiven Erfahrungen während Testfahrten berichten konnten. Übergangsfrage Q3 fragte nach Beschreibungen von Situationen, in denen sich das Fahrzeug anders als erwartet verhielt. Die Teilnehmer beschrieben Situationen, die sie in ihrem Testalltag erlebt hatten. Das Ziel dieser Diskussion war die Identifikation von Situationen, in denen

eine bessere Erklärung nötig ist. Diese wurden gesammelt und anschließend anhand der Adjektive *häufig*, *überraschend* und *verunsichernd* bewertet (vgl. Fragen Q3a – Q3c mit den zugehörigen 5-Punkt-Likert-Skalen). Indem alle Teilnehmer die Situationen bewerten konnten, ergab sich ein detaillierteres Bild von Beispielen, die zuvor möglicherweise von nur einem Teilnehmer beschrieben worden waren.

Die erste Hauptfrage Q4 bezog sich auf Informationsquellen, die die Entwickler derzeit verwenden, um Unterschiede zwischen durchgeführter und erwarteter Aktion nachzuvollziehen. Dadurch wurden Strategien identifiziert, die sich bereits bewährten und die für zukünftige Ansätze weiterentwickelt werden können. Für die zweite Hauptfrage Q5 sollten die Teilnehmer Vorschläge beschreiben, die das Verständnis der Unterschiede noch weiter vereinfachen könnten. Dadurch sollten die Experten zu einem kreativen Brainstorming angeregt werden, um Verbesserungen zu sammeln, die in der Zukunft hilfreich sein könnten. Die Wünsche wurden wie zuvor gesammelt und die Teilnehmer bewerteten, wie *wünschenswert* sie wären (siehe Fragen Q5a mit 5-Punkt-Likert-Skala).

Tabelle 6.1: Interviewleitfaden für die Fokusgruppenstudie in Diskussionsreihenfolge. Nach den Fragen Q3, Q5, wurden die Teilnehmer darum gebeten, die Single-Choice-Fragen Q3a-Q3c bzw. Q5a für die zuvor genannten Situationen bzw. Vorschläge anonym und ohne weitere Diskussion zu beantworten. Fortsetzung auf der nächsten Seite.

Q1 Eröffnungsfrage

Stellt euch bitte kurz vor, indem ihr euren Namen nennt, an welchem Themengebiet ihr arbeitet und wie viele Stunden im Monat ihr in Erprobungsfahrzeugen im Realverkehr verbringt.

Q2 Einführungsfrage

Denkt mal bitte zurück an die letzten automatisierten Fahrten im Erprobungsfahrzeug: Was war denn so eine Situation, in der etwas besser geklappt hat, als ihr erwartet habt?

Q3 Übergangsfrage

Wenn ihr weiter an die letzten Erprobungsfahrten denkt: Beschreibt bitte ein paar Situationen, in denen das Erprobungsfahrzeug eine andere Aktion ausführt als die, die ihr erwartet habt.

Q3a Umfrage

Welche Situation tritt wie häufig auf? (Single Choice, Antwortmöglichkeiten: 1 – *nie*, 2 – *selten*, 3 – *gelegentlich*, 4 – *oft*, 5 – *immer*)

Q3b Umfrage

Welche Situation empfindet ihr als wie überraschend? (Single Choice, Antwortmöglichkeiten: 1 – *überhaupt nicht überraschend*, 2 – *eher nicht überraschend*, 3 – *neutral*, 4 – *eher überraschend*, 5 – *überraschend*)

Q3c Umfrage

Welche Situation wirkt wie verunsichernd auf euch als Mitfahrer? (Single Choice, Antwortmöglichkeiten: 1 – *überhaupt nicht verunsichernd*, 2 – *eher nicht verunsichernd*, 3 – *neutral*, 4 – *eher verunsichernd*, 5 – *verunsichernd*)

Q4 Hauptfrage

Welche Informationen nutzt ihr aktuell, um Unterschiede zwischen ausgeführten und erwarteten Aktionen nachzuvollziehen?

Q5 Hauptfrage

Welche zusätzlichen Informationen wünscht ihr euch, um die Unterschiede zwischen ausgeführten und erwarteten Aktionen noch besser verstehen zu können?

Q5a Umfrage

Für wie wünschenswert haltet ihr diese Information im Allgemeinen? (Single Choice, Antwortmöglichkeiten: 1 – *überhaupt nicht wünschenswert*, 2 – *eher nicht wünschenswert*, 3 – *neutral*, 4 – *eher wünschenswert*, 5 – *wünschenswert*)

Stichprobe

Für die Studie wurden insgesamt 16 Experten aus der universitären Forschung sowie der Vor- und Serienentwicklung für HAF der AUDI AG mit verschiedenen Entwicklungsschwerpunkten befragt. Alle hatten Erfahrung mit dem Testen von AF im Realverkehr auf der Autobahn (11/16 Experten) oder im urbanen Bereich (7/16 Experten), wobei die Anzahl der Testfahrten von einer Fahrt in zwei Monaten bis zu mehreren Fahrten pro Woche variierte.

Durchführung

Insgesamt wurden zwei Online-Fokusgruppen mit jeweils fünf und eine mit sechs Experten durchgeführt, um eine Informationssättigung zu erreichen (Krueger & Casey, 2014). Jede Diskussion wurde mithilfe einer Videokonferenz-Software durchgeführt und dauerte etwa 90 Minuten (Allen, 2014; Sweet, 2001; Turney & Pocknee, 2005). Die Wortäußerungen jeder Sitzung wurden von drei unabhängigen stillen Beobachtern mitgeschrieben und anschließend in einem Protokoll zusammengeführt, das abschließend alle Teilnehmer bestätigten. Die qualitativen Daten wurden mithilfe der Methode des permanenten Vergleichs (engl. *constant comparative method*) ausgewertet (Boeije, 2002; Krueger & Casey, 2014): Bei diesem Vorgehen werden alle Aussagen aus den drei Fokusgruppen systematisch verglichen und in Gruppen eingeteilt. Die quantitativen Daten, die mithilfe der Single-Choice-Fragen Q3a-Q3c und Q5a über die Situationen und Informationswünsche gesammelt wurden (vgl. Abschnitt 6.1), wurden mithilfe des arithmetischen Mittels zusammengefasst: Die Wertungen m_h , $m_{\ddot{u}}$, m_v und m_w geben an, wie *häufig*, *überraschend* oder *verunsichernd* eine Situation bzw. wie *wünschenswert* eine Information im Mittel von den Teilnehmern einer Fokusgruppe bewertet wurde.

6.2 Ergebnisse

Die Äußerungen der Teilnehmer lassen sich in drei Hauptthemen einteilen: 1. Situationen mit unerwarteten Fahrentscheidungen, 2. Informationen über das Ego und seine Umgebung und 3. Informationspräsentation. Diese werden im Weiteren beschrieben und diskutiert.

1. Situationen mit unerwarteten Fahrentscheidungen

Insgesamt beschrieben die Entwickler 21 Situationen, in denen das AF anders als erwartet handelte. In neun der Situationen passte das AF seine Geschwindigkeit nicht an die Umgebung an, z. B. beim Annähern an einen Kreisverkehr oder an ein vorausfahrendes Fahrzeug. Ein Experte beschrieb eine Situation, in der das Fahrzeug auf eine Verkehrsinsel zufuhr, die gerade von einem Fahrradfahrer überquert wurde. Anstatt zu bremsen, entschied sich das Fahrzeug dazu,

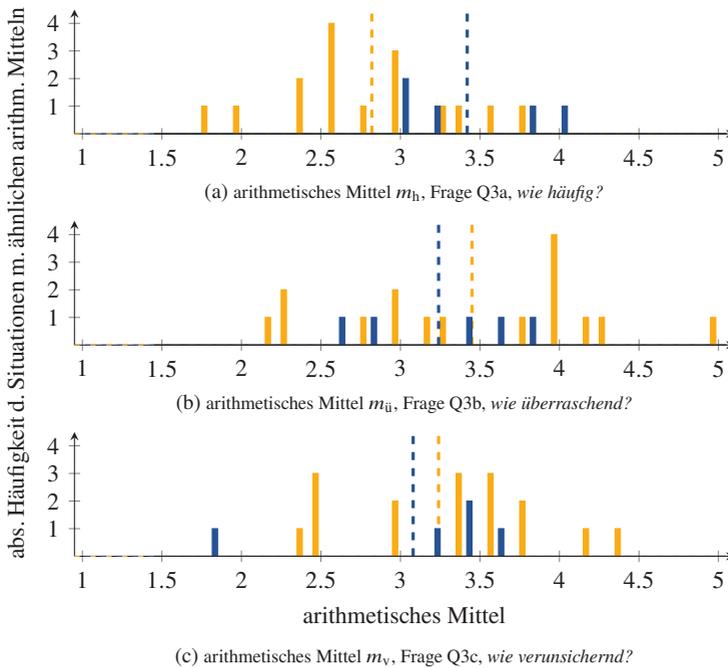


Abbildung 6.1: Absolute Häufigkeiten für Situationen mit longitudinaler bzw. lateraler Bewegung. Situationen mit ähnlichen arithmetischen Mitteln werden in Intervallen mit Breite 0.1 zusammengefasst und gezählt. ■ 16 Situationen mit lateraler Bewegung (■ arithmetisches Mittel aller Antworten: häufig $M_{h,lat} = 2.82$, überraschend $M_{\ddot{u},lat} = 3.45$, verunsichernd $M_{v,lat} = 3.24$), ■ 5 Situationen mit longitudinaler Bewegung (■ arithmetisches Mittel aller Antworten: häufig $M_{h,long} = 3.42$, überraschend $M_{\ddot{u},long} = 3.24$, verunsichernd $M_{v,long} = 3.08$), (angelehnt an Henze, Stasinski et al. (2021)).

dem Radfahrer auszuweichen und steuerte auf die Verkehrsinsel zu. Die meisten Teilnehmer bewerteten diese Situation als *eher verunsichernd* oder *verunsichernd* (arithmetisches Mittel $m_v = 3.8$, 1. Fokusgruppe (FG)).

Die Experten berichteten außerdem, dass ein Fahrstreifenwechsel auf der Autobahn unverständlich sein könnte: Zum Beispiel wechselte das AF den Fahrstreifen, während Experten keine offensichtlichen Gründe für diese Entscheidung ausmachen konnten. Dies wurde als *eher verunsichernd* oder *verunsichernd*

($m_v = 3.6$, 3. FG) und als *eher überraschend* oder *überraschend* wahrgenommen ($m_{\ddot{u}} = 4.25$, 3. FG). Eine andere Gruppe beschrieb eine ähnliche Situation, welche zwar *selten* aufträte ($m_h = 2$, 2. FG), dann jedoch meist *eher verunsichernd* ($m_v = 3.8$, 2. FG) und *überraschend* ($m_{\ddot{u}} = 5$, 2. FG) sei. Einer der Teilnehmer fasste seine Erfahrungen dazu folgendermaßen zusammen:

„Grundsätzlich kann man sagen, dass querführungsrelevante Aktionen meistens überraschender und kritischer sind als längsführende.“

Um diese Beobachtung zu untersuchen, wurden alle 21 Situationen in die Gruppen unerwartete *longitudinale* oder *laterale* Bewegung eingeteilt. Die Histogramme in Abb. 6.1 zeigen die absoluten Häufigkeiten von Situationen mit ähnlichen arithmetischen Mitteln für jedes der drei Attribute *häufig*, *überraschend*, *verunsichernd*. Wenn beispielsweise zwei laterale Situationen für das Attribut *häufig* ähnliche arithmetische Mittel im Intervall [2.35, 2.45) haben, dann zeigt Abb. 6.1 für 2.4 einen orangenen Balken mit Höhe 2.

Von den insgesamt 21 beschriebenen Situationen wurden 16 der lateralen und 5 der longitudinalen Bewegung zugeordnet. Abb. 6.1 (a) zeigt, dass Situationen mit unerwarteter lateraler Bewegung seltener auftreten als mit longitudinaler Bewegung (*häufig*: arithmetische Mittel aller Antworten für laterale bzw. longitudinale Bewegungen $M_{h,lat} = 2.82 < 3.42 = M_{h,long}$ mit Standardabweichungen 0.96 bzw. 0.78). Dies deutet darauf hin, dass sich die Teilnehmer an deutlich mehr unterschiedliche Fahrentscheidungen mit lateraler Bewegung erinnern ($16 > 5$), wenn sie nach unerwarteten Aktionen gefragt werden, auch wenn laterale Aktionen insgesamt seltener aufzutreten scheinen. Die Mittelwerte der Antworten für die Attribute *überraschend* und *verunsichernd* sind jeweils ähnlich groß (*überraschend*: $M_{\ddot{u},lat} = 3.45 > 3.24 = M_{\ddot{u},long}$; *verunsichernd*: $M_{v,lat} = 3.31 > 3.08 = M_{v,long}$; alle Standardabweichungen im Intervall [1.2, 1.4]).

2. Informationen über das Ego und seine Umgebung

Im Rahmen der Fokusgruppen beschrieben die Entwickler die Visualisierungen, die sie während der Testfahrten vorrangig benutzen, um unerwartete Aktionen des AF zu verstehen. Diese beinhalten eine Darstellung der Umgebung, die

sowohl statische Objekte, z. B. Lichtsignalanlagen oder Fahrstreifen, als auch dynamische Objekte, z. B. andere Fahrzeuge und deren Objekt-Fahrstreifen-Zuordnung, umfasst. Sie erklärten, dass weitere Informationen zu anderen Objekten hilfreich wären, z. B. ob genügend Zeit bliebe um eine Trajektorie anzupassen, falls ein neues Objekt entdeckt würde. Außerdem sprachen die Experten über die Darstellung des Egos und seine geplante Trajektorie: Sie fanden eine abstrakte, intuitive Darstellung des aktuell geplanten Manövers *wünschenswert* ($m_w = 5, 3$. FG) und würden diese gerne durch eine noch abstraktere Verknüpfung zwischen dem Plan des AF und relevanten Umgebungsmerkmalen erweitern ($m_w = 3.4, 3$. FG).

Ein Teilnehmer schlug zusätzlich vor, dass eine Situation, in der Fehlverhalten beobachtet wurde, mit ähnlichen Situationen, die das AF zuvor erfolgreich durchfahren hatte, verglichen werden sollte. Er versprach sich dadurch die Identifikation von Faktoren, die unterschiedlich waren und dadurch möglicherweise den Fehler verursacht hatten ($m_w = 4.8, 1$. FG). Auch wenn diese Vorschläge jeweils nur in einer Fokusgruppe beschrieben wurden, sind sie dennoch vielversprechend, da sie sehr detailliert dargestellt wurden und die Umfrageergebnisse die Unterstützung innerhalb der Fokusgruppe zeigten.

3. Informationspräsentation

In jeder der drei Gruppen wurde deutlich, dass das Visualisierungskonzept sehr wichtig ist. Beispielsweise beschrieben die Teilnehmer, dass ein Entwickler auf dem Beifahrersitz verschiedene Informationen auf mehreren Monitoren beobachten, aber dennoch von der Anzahl an Fehlermeldungen überwältigt sein kann. Daher äußerten sie den Wunsch, einen solchen

„information overkill“

zu vermeiden und die Möglichkeit zu haben, auf die verschiedenen Informationsbedürfnisse der Entwickler eingehen zu können, indem z. B. nur Teilmengen an Informationen ausgewählt und angezeigt werden können.

6.3 Diskussion

Die Ergebnisse liefern Hinweise darauf, wie die Transparenz von unerwarteten Fahrentscheidungen verbessert werden kann. Ein wichtiger Punkt ist, dass Informationen präzise und leicht zugänglich dargestellt werden sollen, um eine Informationsflut (engl. *information overload*) und die damit einhergehende Überforderung der Nutzer zu vermeiden (vgl. Amditis et al., 2010; Stoter, Cornelisse & Dalmolen, 2011). Nichtsdestotrotz sollten Erklärungen die relevanten Umgebungsmerkmale beschreiben. Es wurde außerdem deutlich, dass eine unerwartete Fahrentscheidung vorrangig durch laterale Bewegung entsteht. Daher sollten primär diese erklärt werden, ohne jedoch die longitudinale Bewegung zu sehr zu vernachlässigen.

Den Eindruck eines Entwicklers, laterale Bewegung sei häufig überraschender und kritischer als longitudinale, ließ sich aufgrund der kleinen Stichprobe und der unterschiedlichen Stichprobengrößen von 16 Situationen mit lateraler und 5 Situationen mit longitudinaler Bewegung weder unterstützen noch widerlegen. Allerdings fällt auf, dass sich die Entwickler an eine größere Vielfalt an Situationen mit lateraler Bewegung erinnern, auch wenn diese jeweils weniger häufig aufzutreten scheinen. Ein Grund dafür könnte sein, dass die Entwickler Aktionen mit lateraler Bewegung als vielfältiger wahrnehmen und dadurch jede einzeln beschriebene Situation per se als weniger häufig wahrnehmen. Außerdem hatten die meisten Entwickler, die teilgenommen haben, hauptsächlich Erfahrungen mit Testfahrten auf Autobahnen, wo die Reaktionszeit für longitudinale Bewegung aufgrund der größeren Abstände größer ist als für laterale und diese dadurch möglicherweise als kritischer wahrgenommen werden. Allerdings zeigt das Beispiel mit der Verkehrsinsel, dass detailliertere Erklärungen von lateraler Bewegung das Verständnis auch in urbanen Umgebungen unterstützen können. Daraus lässt sich schlussfolgern, dass sich Erklärungen vorrangig auf laterale Bewegung fokussieren und dabei Umgebungsmerkmale einschließen sollten.

7 Generierung erläuternder Beschreibungen

Die Ergebnisse der Fokusgruppenstudie in Kapitel 6 zeigen, dass Entwickler vor allem in Situationen mit unerwarteter lateraler Bewegung zusätzliche Informationen zur Umgebung gutheißen. Allerdings ist zu erwarten, dass die Informationen, die einen Experten in seinem Verständnis unterstützen, für Laien größtenteils unverständlich sind und daher weitere Erklärungen notwendig werden. Aus diesem Grund stellt das folgende Kapitel eine Methode vor, die die Informationen, die die Fahrentscheidung beeinflusst haben, identifiziert, in Zusammenhang mit der Umgebung setzt und so aufbereitet, dass sie auch für Menschen ohne Fachwissen leicht verständlich sind. Das Ziel ist dabei nicht, ein vollwertiges HMI-Konzept vorzuschlagen, sondern vielmehr einen technischen Ansatz zu bieten, der die für zukünftige HMIs notwendigen Informationen zur Verfügung stellt. Dafür werden die Entscheidungen des AF mithilfe von automatisch generierten Aussagen beschrieben, um die Transparenz zu verbessern und die Fahrentscheidungen für den Nutzer nachvollziehbarer zu machen.

In Abschnitt 7.1 werden die benötigten Annahmen (Abschnitt 7.1.1) sowie die Grundlagen der Textgenerierung (Abschnitt 7.1.2) beschrieben. Außerdem werden in Abschnitt 7.1.3 drei Ansätze vorgestellt, die sich in ihrem Informationsgehalt unterscheiden. Diese werden in Abschnitt 7.2 mithilfe einer Online-Befragung von Novizen und Experten validiert. Daraus kann abgeleitet werden, welche Eigenschaften Nutzer für die Erläuterungen von automatisierten Fahrentscheidungen bevorzugen und welche Konzepte die Grundlage für zukünftige Nutzeranzeigen bilden sollten.

Dieses Kapitel basiert auf den vorläufigen Ergebnissen, die in Henze, Faßbender und Stiller (2022) veröffentlicht wurden. Abbildungen, Tabellen und andere Ergebnisse wurden für eine konsistente Darstellung angepasst.

7.1 Textgenerierung

Wie in Kapitel 3 dargestellt, liefert die SA nach Morris eine Möglichkeit, um zwischen relevanten und irrelevanten Informationen zu unterscheiden. Dies ist ein guter Ausgangspunkt für die Erklärung von Fahrentscheidungen, weil präzise Erklärungen nur die Informationen benötigen, die auch tatsächlich die Fahrentscheidung beeinflusst haben. Im Folgenden wird daher vorgestellt, wie mit den Ergebnissen der SA Aussagen erzeugt werden können, die die Fahrentscheidungen eines AF erläutern.

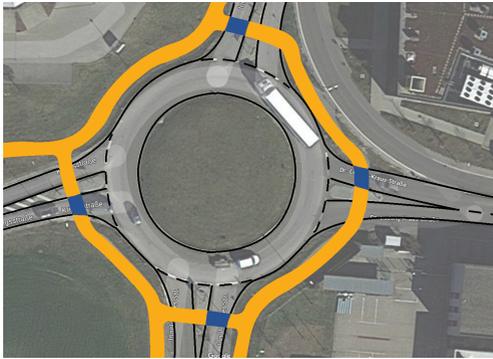
Wie in Abschnitt 3.3 bereits beschrieben, werden der Stichprobenmittelwert μ_t und die Stichprobenstandardabweichung σ_t benutzt, um zum Zeitpunkt t zwischen relevanten und irrelevanten Inputs zu entscheiden. Für einen Planungsalgorithmus, der $k \in \mathbb{N}$ Inputs zur Berechnung des Ergebnisses verwendet, ist die Menge an relevanten Inputs zum Zeitpunkt t gegeben durch

$$\mathcal{R}_{o,t}(\epsilon) := \{i = 1, \dots, k \mid |\mu_{o,i,t}| > \epsilon_i \vee \sigma_{o,i,t} > \epsilon_i\}, \quad (7.1)$$

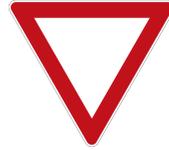
wobei die Elemente des Vektors $\epsilon \in \mathbb{R}^k$, $\epsilon_i \geq 0$ für alle $i = 1, \dots, k$ situationsabhängig und klein gewählt werden. Eine Aussage, die die Fahrentscheidung o erläutert, sollte also eine Teilmenge der relevanten Inputs $\mathcal{R}_{o,t}(\epsilon)$ berücksichtigen. Wie Kapitel 3 gezeigt hat, erfordert die Einschätzung und Interpretation der Ergebnisse einer SA jedoch ein großes Fachwissen sowohl über die SA als auch über den Planungsalgorithmus selbst (vgl. Abschnitt 3.3). Daher müssen die Informationen erst interpretiert werden, bevor sie den Nutzer in seinem Verständnis der Fahrentscheidungen unterstützen können. Dafür sind einerseits Annahmen über die Darstellung von Verkehrsregeln und andererseits leicht verständliche Beschreibungen der vom Planungsalgorithmus verwendeten Inputs nötig. Diese werden im Folgenden beschrieben.

7.1.1 Annahmen

Es werden zwei zentrale Annahmen als Voraussetzung für die Generierung von textbasierten Erläuterungen getroffen: die Existenz von maschinenlesbaren Verkehrsregeln und Informationen über Inputs und Outputs, die ohne Fachwissen leicht verständlich sind.



(a) Schematische Darstellung eines Kreisverkehrs in Gaimersheim, Deutschland (48° 47' 23.5"N 11° 22' 36.7"E), Bild © 2022 GeoBasis-DE/BKG, GeoContent, Maxar Technologies, Kartendaten. ● Radwege, ■ Verhandlungszonen über Vorfahrt zwischen Radfahrer und Autofahrer.



(b) Verkehrszeichen 205, Vorfahrt gewähren (StVO, 2013).



(c) Verkehrszeichen 215, Kreisverkehr (StVO, 2013).

Abbildung 7.1: Schematische Darstellung eines Kreisverkehrs.

Maschinenlesbare Verkehrsregeln

Die erste Annahme der Methode ist die Existenz von mathematischen Beschreibungen der Verkehrsregeln, die das AF befolgen muss. Diese werden verwendet, um zu testen ob das AF die Regeln einhält. Derzeit wird daran gearbeitet, wie die juristischen Regeln in maschinenlesbare Regeln umgewandelt werden können: Esterle, Gressenbuch und Knoll (2020); Maierhofer, Rettinger, Mayer und Althoff (2020) formalisierten die deutsche Straßenverkehrsordnung (StVO) auf Autobahnen bzw. Schnellstraßen und Rizaldi et al. (2017) beschrieben ein Überholmanöver. Zusätzlich zu den länderspezifischen Verkehrsregeln gibt es auch noch kulturelle Konventionen, die zwar nicht juristisch vorgeschrieben sind, die die Teilnehmer am Straßenverkehr jedoch als gegeben annehmen (Shalev-Shwartz, Shammah & Shashua, 2017). Beispielsweise ist es in Deutschland in unübersichtlichen Situationen üblich, einem anderen Verkehrsteilnehmer die Vorfahrt zu geben, aber nicht, sie ihm zu nehmen. Solche Verhandlungen können z. B. zwischen Autofahrern, die in einen Kreisverkehr einfahren wollen, und Radfahrern, die sich auf einem angrenzenden Radweg neben der Kreisfahrbahn befinden, auftreten, vgl. Abb. 7.1(a): Wenn die Zufahrtsstraße zum Kreisverkehr mit den Zeichen *Vorfahrt gewähren* (Abb. 7.1(b)) und *Kreisverkehr* (Abb. 7.1(c))

gekennzeichnet ist, so sind die Radfahrer gegenüber dem einfahrenden Fahrzeug wartepflichtig (OLG Hamm, 2012). In der Realität verhandeln die beteiligten Auto- und Radfahrer die Vorfahrt jedoch meist individuell durch Mimik und Gestik, d. h. Vorfahrt wird von einem der Verkehrsteilnehmer freiwillig gegeben, bevor sie der andere annimmt. Diese kulturell sehr verschiedenen Konventionen muss ein AF insbesondere im gemischten Verkehr mit Autofahrern befolgen und sie sind daher z. B. in dem Vorschlag zu Responsibility Sensitive Safety (RSS) mathematisch formalisiert (Shalev-Shwartz et al., 2017).

Insgesamt ist also die erste Voraussetzung für die Generierung von Erläuterungen die Existenz von maschinenlesbaren, verifizierbaren Regeln, mit denen ausgewertet werden kann, gegen welche Verkehrsregeln das AF verstößt. Diese können dann als mögliche Gründe für sein Verhalten angeführt werden. Um diesen Ansatz zunächst so einfach und damit so transparent wie möglich zu halten, nutzt es eine Auswahl an *if-else*-Regeln anstelle der beschriebenen State-of-the-Art-Ansätze. Mit den in Anhang A.8 dargestellten Regeln wird in der Simulation überprüft, ob sich das AF z. B. an eine Geschwindigkeitsbegrenzung hält: $\text{Geschwindigkeit } v < v_{\text{limit}}$?

Informationen über Inputs und Outputs

Die zweite Annahme richtet sich an das Wissen der Funktionsentwickler: Um den Zusammenhang zwischen den Inputs und Outputs eines Planungsalgorithmus verständlich beschreiben zu können, müssen sie versprachlicht werden. Dafür wird angenommen, dass die Entwickler für jeden In- und Output die Attribute „Name“, „Verb“, „Regeln“ und „Objekt“ zusammenstellen können. Für die Beschleunigung a_c eines Objekts c könnte die Beschreibung beispielsweise sein:

$$a_c \rightarrow \left\{ \begin{array}{ll} \text{„Name“:} & \text{„Beschleunigung“} \\ \text{„Verb“:} & \text{„beschleunigen“ falls } a > 0, \text{ „bremsen“ sonst} \\ \text{„Regeln“:} & \{ \text{„getToDestinationFast“}, \text{ „speedLimit“} \} \\ \text{„Objekt“:} & \text{Beschreibung des Objekts } c, \text{ z. B. „Fahrzeug vor} \\ & \text{dem Ego auf demselben Fahrstreifen“}. \end{array} \right.$$

Dabei wird deutlich, dass der Wert eines Attributs vom Input bzw. Output selbst abhängen kann, wie hier „Verb“: Es beschreibt das Verhalten von Objekt c ,

z. B. was es mit Bezug auf das Attribut „Name“ tut, und ist wichtig, um den Effekt eines Inputs oder Outputs zu versprachlichen. Durch die Verwendung von Intervallen sind noch mehr Fallunterscheidungen möglich, um z. B. zu beschreiben, ob das Fahrzeug stark, moderat oder schwach beschleunigt.

Die Liste „Regeln“ enthält mathematische Formulierungen von Verkehrsregeln, d. h. Regeln, die das Fahrzeug befolgen muss und eine Erklärung ihrer Bedeutung. Es ist möglich, dort alle Regeln aufzuführen. Da dies jedoch bei der Überprüfung einen größeren Rechenaufwand nach sich zieht, enthält die Liste in dieser Arbeit nur eine Auswahl an Regeln, die in Anhang A.8 dargestellt werden: Die Liste eines Inputs enthält eine Regel dann, wenn die Änderung des zugehörigen Wertes eine notwendige Bedingung dafür ist, dass eine verletzte Regel wieder eingehalten werden kann. Wenn ein AF z. B. die Regel „speedLimit“ verletzt, dann muss seine Beschleunigung verändert werden, damit die Geschwindigkeit die Grenzwerte der Verkehrsregel einhält.

Abschließend kann das Attribut „Objekt“ bei Bedarf das Verhältnis des Objekts zum Ego durch z. B. eine Positionsbeschreibung angeben, da die Inputs auch Eigenschaften anderer Objekte bezeichnen können. Dadurch soll der Nutzer unterstützt werden, das Objekt bei Bedarf auch während der Fahrt schnell zu identifizieren.

Insgesamt ermöglicht die modulare Struktur, weitere Attribute schnell hinzuzufügen oder sie in andere Sprachen zu übersetzen (vgl. Henze, Faßbender & Stillner, 2022, für eine englische Übersetzung des oben diskutierten Beispiels).

7.1.2 Grundlagen der Textgenerierung

Eine Möglichkeit, Text zu generieren, ist die automatische Weiterverarbeitung von Daten in natürliche Sprache (engl. *data-to-text Natural Language Generation (NLG)*). Eine typische NLG-Architektur besteht aus drei Teilen (Gatt & Kraemer, 2018; Perera & Nand, 2017):

1. der Textplaner, d. h. was gesagt wird
2. der Satzplaner, d. h. wie etwas gesagt wird
3. der Realisierer, d. h. etwas tatsächlich sagen.

Algorithm 7.1 Allgemeine Implementierungsdetails für alle Aussagentypen.

Input: Beschreibung $\mathbf{q}_t \in \mathbb{R}^k$ der Umgebung zum Zeitpunkt t , Planungsalgorithmus $\mathbf{f} : \mathbb{R}^k \rightarrow \mathbb{R}^m$, Fahrentscheidung $\mathbf{f}(\mathbf{q}_t) \in \mathbb{R}^m$

Output: erläuternde Beschreibung \mathcal{S}

- 1: $\mu_t, \sigma_t \leftarrow$ char. Größen der SA von Planungsalgorithmus \mathbf{f} \triangleright Kapitel 3
 - 2: $\mathcal{R}_{o,t}(\epsilon) \leftarrow \{i \in \{1, \dots, k\} \mid |\mu_{o,i,t}| > \epsilon_i \vee \sigma_{o,i,t} > \epsilon_i\}$ \triangleright vgl. (7.1)
 - 3: $\mathcal{V}_i \leftarrow$ verletzte Regeln für Input $i \in \mathcal{R}_{o,t}(\epsilon)$
 \triangleright überprüfe Regeln des Attributs “Regeln”, vgl. Seite 97
 - 4: Handlung-Satz $A \leftarrow$ beschreibt Ego-Aktion basierend auf Ergebnis $\mathbf{f}(\mathbf{q}_t)$
 \triangleright verwende Attribute der Outputs, vgl. Seite 98
 - 5: Wichtig-Satz $W_i \leftarrow$ benennt wichtige Inputs $i \in \mathcal{R}_{o,t}(\epsilon)$ \triangleright Binns et al. (2018)
 - 6: Effekt-Satz $E_i \leftarrow$ beschreibt Effekt von $i \in \mathcal{R}_{o,t}(\epsilon)$ basierend auf $\mu_{i,t}, \mathbf{f}(\mathbf{q}_t)$
 \triangleright verwende Attribute der Inputs, vgl. Seite 98
 - 7: Regel-Satz $R_i \leftarrow$ verspricht die Regeln $v \in \mathcal{V}_i$ für alle $i \in \mathcal{R}_{o,t}(\epsilon)$
 \triangleright erkläre Verkehrsregeln, vgl. Seite 97
 - 8: S .Hauptsatz \leftarrow Handlung-Satz A
 \triangleright dieselbe Formulierung für alle drei Aussagentypen
 - 9: **for** $i \in \mathcal{R}_{o,t}(\epsilon)$ **do**
 \triangleright füge Details für alle Aussagentypen zu allen relevanten Inputs hinzu
 - 10: S .Kontext $_i, S$.Verkehrsregeln $_i \leftarrow \dots (W_i, E_i, R_i)$
 \triangleright rufe Algorithmen 7.2 bis 7.4 auf, um Nebensätze zu konstruieren
 - 11: **end for**
 - 12: $S \leftarrow$ überarbeiten(S)
 \triangleright Sätze zusammenfügen, Wiederholungen durch Pronomen ersetzen
-

Diese Arbeit fokussiert sich auf den ersten Teil, da die Darstellung des Inhalts ein HMI-Aspekt ist, der außerhalb des Umfangs dieser Arbeit liegt.

Der Inhalt wird aufgrund der SA des Planungsalgorithmus ausgewählt (Algorithmus 7.1, Zeilen 1 bis 3) und deren Interpretation hängt vom Kommunikationsziel ab: Der Hauptsatz beschreibt die Handlung des AF (Handlung-Satz, vgl. Algorithmus 7.1) und die nachgestellten Nebensätze geben Informationen zur Umgebung oder zu Verkehrsregeln an. Systematische Satzbauregeln werden angewendet, um die Aussagen zu formulieren (Satzplaner), da der Anwendungsbereich sehr speziell ist und keine variantenreichen Satzstrukturen nötig

sind (vgl. Algorithmus 7.1, Zeilen 4 bis 8). Abschließend fügen vordefinierte Satzstrukturen die Haupt- und Nebensätze zusammen, fügen Pronomen ein und verbinden alle Sätze zu grammatikalisch korrekten Aussagen (Realisierer, vgl. Algorithmus 7.1, Zeile 12).

7.1.3 Erläuternde Beschreibungen

Im Fokus dieser Arbeit steht die Erzeugung des Inhalts von erläuternden Aussagen, die dem Nutzer von AF helfen, die Fahrentscheidungen unabhängig von seinem Wissen zu verstehen, und nicht deren konkrete Umsetzung. Diese ist ein HMI-Aspekt, der im Weiteren nicht behandelt wird. Es gibt viele verschiedene Möglichkeiten, aus den Informationen der SA und den Annahmen Aussagen zu formulieren, die sich in ihrem Informationsgehalt unterscheiden. Im Folgenden werden drei davon exemplarisch vorgestellt: Wichtige-Inputs-Aussage, regelbasierte und schlussfolgernde Aussagen. Alle nutzen die Fahrentscheidung $\mathbf{f}(\mathbf{q})$, den Stichprobenmittelwert μ_t und die Stichprobenstandardabweichung σ_t , die Menge an relevanten Inputs $\mathcal{R}_{o,t(\epsilon)}$ sowie die in Abschnitt 7.1.1 beschriebenen Annahmen, allerdings in unterschiedlichem Ausmaß. Alle Aussagen starten mit einem Hauptsatz, der die Fahrentscheidung verspricht (vgl. Algorithmus 7.1, Zeile 8), jedoch unterscheiden sich die nachfolgenden Sätze in Informationstyp und Informationsdichte (vgl. Algorithmus 7.1, Zeile 10, Übersicht: Tabelle A.2):

1. Wichtige-Inputs-Aussage (WI)

Der Wichtige-Inputs-Ansatz wurde von Binns et al. (2018) vorgeschlagen (engl. *input influence*) und listet die Namen aller Inputs $i \in \mathcal{R}_{o,t(\epsilon)}$ auf (vgl. Algorithmus 7.2), z. B. für die Beschleunigung $o = a$: „das Ego bremst und dafür ist die Geschwindigkeit des Fahrzeugs vor dem Ego wichtig“ (vgl. Algorithmus 7.2, Zeile 1). Dieser Ansatz liefert die wenigsten weiter bearbeiteten Informationen, weil kein Effekt der Inputs auf den Output hergestellt wird, und ist daher am wenigsten anfällig für Fehlinterpretationen.

Algorithm 7.2 Wichtige-Inputs-Aussage.

Input: Wichtig-Satz W , Effekt-Satz E , Regel-Satz R **Output:** Nebensätze Kontext, Verkehrsregeln

- 1: Kontext \leftarrow Wichtig-Satz W
 - 2: Verkehrsregeln \leftarrow None
-

Algorithm 7.3 Schlussfolgernde Aussage.

Input: Wichtig-Satz W , Effekt-Satz E , Regel-Satz R **Output:** Nebensätze Kontext, Verkehrsregeln

- 1: **if** $|\mu_{i,t}| - z_{1-\alpha}\sigma_{i,t} < 0$ **then** \triangleright vgl. (7.2)
 - 2: Kontext \leftarrow Wichtig-Satz W
 - 3: Verkehrsregeln \leftarrow Regel-Satz R + “wahrscheinlich”
 - 4: **else**
 - 5: Kontext \leftarrow Effekt-Satz E
 - 6: Verkehrsregeln \leftarrow Regel-Satz R
 - 7: **end if**
-

2. Schlussfolgernde Aussage (SF)

Der schlussfolgernde Ansatz beschreibt, auf welche Art und Weise die Inputs $i \in \mathcal{R}_{o,t}(\epsilon)$ die Fahrentscheidung o beeinflussen (vgl. Algorithmus 7.3): Indem das Vorzeichen des Stichprobenmittelwerts $\mu_{o,i,t}$ mit der Fahrentscheidung in Verbindung gebracht wird, wird der Effekt des Inputs interpretiert, z. B. „das Ego bremst, weil das Fahrzeug vor ihm langsamer fährt“ (Ego-Beschleunigung $a < 0$ und $\mu_{o,v_{vs},t} < 0$, wenn Input v_{vs} die Geschwindigkeit des vorausfahrenden Fahrzeugs auf demselben Fahrstreifen bezeichnet). Allerdings ist der Stichprobenmittelwert $\mu_{o,i,t}$ manchmal klein, d. h. $\mu_{o,i,t} \approx 0$, und sein Vorzeichen daher aufgrund der stochastischen Eigenschaften der SA nicht eindeutig (vgl. Algorithmus 7.3, Zeile 1) (Campolongo et al., 2007; Henze et al., 2020; Morris, 1991). Um in diesen Fällen zu testen, ob das Vorzeichen $\text{sign}(\mu_{o,i,t})$ mit Wahrscheinlichkeit $(1 - \alpha)$ eindeutig ist, wird das einseitige Konfidenzintervall

$$C := \left[\left| \mu_{o,i,t} \right| - z_{1-\alpha}\sigma_{o,i,t}, \infty \right) \quad (7.2)$$

mit dem $(1 - \alpha)$ -Quantil $z_{1-\alpha}$ der Normalverteilung verwendet. Diese Annahme ist akzeptabel, weil die SA mindestens 30 Stichprobenwerte zieht (vgl. Kapitel 3

und 4). Falls das Intervall $0 \in C$ einschließt, so kann das Vorzeichen mit Wahrscheinlichkeit $p > \alpha$ nicht zweifelsfrei bestimmt werden und dadurch können keine zuverlässigen Rückschlüsse auf den Effekt des Inputs i gezogen werden. Daher werden in diesen Fällen die Namen der betreffenden Inputs analog zum Wichtige-Inputs-Ansatz nur genannt, ohne ihren Einfluss genauer zu beschreiben (vgl. Algorithmus 7.3 und Zeilen 1 und 2). Falls stattdessen das Konfidenzintervall $0 \notin C$ nicht enthält, so ist das Vorzeichen mit Wahrscheinlichkeit $p > 1 - \alpha$ eindeutig und die Schlüsse, die auf Grundlage von Input $i \in \mathcal{R}_{o,t}(\epsilon)$ gezogen wurden, treffen mit Wahrscheinlichkeit $(1 - \alpha)$ zu (vgl. Algorithmus 7.3 und Zeilen 4 und 5).

Damit die Entscheidung des AF nicht nur mit Beobachtungen der Umgebung verknüpft wird, sondern auch mit den geltenden Verkehrsregeln, nennt der schlussfolgernde Ansatz auch die Regeln, die zum Zeitpunkt t verletzt werden. Falls das AF beispielsweise bremst, können sowohl die Regeln „speedLimit“ als auch „safeLongDistance“ verletzt sein: Aufgrund der Liste „Regeln“ des Inputs $i \in \mathcal{R}_{o,t}(\epsilon)$ ist es möglich zu unterscheiden, ob ein Fahrzeug bremst, weil es eine Geschwindigkeitsbegrenzung einhalten muss ($v < v_{\text{limit}}$ verletzt?) oder weil vor ihm ein langsameres Fahrzeug fährt (z. B. sicherer Abstand zum vorausfahrenden Fahrzeug nach RSS überprüfen (Shalev-Shwartz et al., 2017)). Da diese Unterscheidung ebenfalls durch das Vorzeichen $\text{sign}(\mu_{o,i,t})$ beeinflusst werden kann, ist auch hier die Begründung durch Regeln anfällig für Missinterpretationen (vgl. Konfidenzintervall (7.2) zur Interpretation des Effekts eines Input i ; Algorithmus 7.3 und Zeile 3). Um diese Unsicherheit zu verdeutlichen, wird dem Nebensatz das Wort „wahrscheinlich“ zugesetzt, falls mit Wahrscheinlichkeit $p > \alpha$ eine Fehlinterpretation des Vorzeichens $\text{sign}(\mu_{o,i,t})$ vorliegt (vgl. Algorithmus 7.3 und Zeile 3).

Da dieser Ansatz die meisten Informationen zu den relevanten Inputs, deren Effekt auf die Fahrentscheidung und die verletzten Verkehrsregeln angibt, generiert er die längsten Sätze mit der höchsten Informationsdichte.

3. Regelbasierte Aussagen (RB)

Im Gegensatz zu den anderen beiden Ansätzen gibt der regelbasierte Ansatz keine Hinweise auf die Umgebung des AF (vgl. Algorithmus 7.4). Stattdessen führt er analog zum schlussfolgernden Ansatz die Verkehrsregeln auf, die das

Algorithm 7.4 Regelbasierte Aussage.**Input:** Wichtig-Satz W , Effekt-Satz E , Regel-Satz R **Output:** Nebensätze Kontext, Verkehrsregeln

- 1: Kontext \leftarrow None
- 2: **if** $|\mu_{i,t}| - z_{1-\alpha}\sigma_{i,t} < 0$ **then** ▷ vgl. (7.2)
- 3: Verkehrsregeln \leftarrow Regel-Satz R + “wahrscheinlich”
- 4: **else**
- 5: Verkehrsregeln \leftarrow Regel-Satz R
- 6: **end if**



Abbildung 7.2: Das rote AF überholt das gelbe Fahrzeug (Geschwindigkeit: 90 km/h) mit einer Geschwindigkeit 120 km/h. Am Ende des Überholmanövers trifft es die Entscheidung, den Fahrstreifen nach rechts zu wechseln und die Geschwindigkeit beizubehalten. Die Geschwindigkeitsbegrenzung liegt bei 120 km/h. Transparente rote und gelbe Fahrzeuge zeigen die vergangenen Positionen mit einem Abstand von jeweils 3 s an. Die verblassenden Linien stellen die geplanten Trajektorien dar (angelehnt an Henze, Faßbender und Stiller (2022)).

AF verletzt und deutet durch den Zusatz „wahrscheinlich“ möglicherweise unsichere Interpretationen an (vgl. Algorithmus 7.4, Zeilen 2 und 3). Daher ist dies der einzige der drei Ansätze, der keine detaillierten Informationen über Beobachtungen des Umfelds gibt und ausschließlich abstrakte, sehr stark weiterverarbeitete Informationen liefert, um das Verhalten des AF zu erläutern.

Tabelle 7.1 gibt jeweils einen Beispielsatz für die drei Aussagentypen für das Überholmanöver an: Diese erläutern das Verhalten eines AF, das nach einem Überholvorgang wieder auf den rechten Fahrstreifen zurückkehrt (vgl. Abb. 7.2). Es wird insbesondere deutlich, dass sich die Informationen teilweise überschneiden, wobei der schlussfolgernde Ansatz am längsten ist. Im nächsten Abschnitt wird die Online-Studie beschrieben, mit der die drei Ansätze für zwei Verkehrssituationen verglichen und validiert wurden.

Tabelle 7.1: Erläuternde Aussagen für das Überholmanöver, vgl. Abb. 7.2 (angelehnt an Henze, Faßbender und Stiller (2022)).

menschliche Referenzaussage

Das rote Fahrzeug wechselt den Fahrstreifen nach rechts, weil der Überholvorgang abgeschlossen ist und es sich an das Rechtsfahrgebot halten muss. (auf der Grundlage von Koo et al., 2015; Lim et al., 2009; Meck & Precht, 2021; Y. Xu et al., 2020; Zhou & Danks, 2020)

Wichtige-Inputs-Aussage

Das rote Fahrzeug wechselt den Fahrstreifen nach rechts und dafür ist die Distanz zum Fahrzeug rechts hinter dem roten Fahrzeug wichtig.

Schlussfolgernde Aussage

Das rote Fahrzeug wechselt den Fahrstreifen nach rechts, weil das Fahrzeug rechts hinter dem roten Fahrzeug weiter entfernt ist, andere Fahrzeuge nicht behindert werden sollen und das Rechtsfahrgebot in Deutschland gilt.

Regelbasierte Aussage

Das rote Fahrzeug wechselt den Fahrstreifen nach rechts, weil das Rechtsfahrgebot in Deutschland gilt und andere Fahrzeuge nicht behindert werden sollen.

7.2 Validierung: Online-Studie

Wie beschrieben gibt es vielfältige Möglichkeiten, erläuternde Aussagen zu formulieren. Allerdings sind nicht alle dafür geeignet, das Verhalten eines AF befriedigend zu beschreiben. Daher ist das primäre Ziel der Studie der Vergleich zwischen den drei vorgestellten Ansätzen und einer Referenzaussage, um zu testen ob eine der Aussagen vergleichbar gut ist wie diese.

7.2.1 Methode

Zur Validierung der Aussagen wurde eine Online-Umfrage verwendet, die die Zufriedenheit der Teilnehmer mit den verschiedenen erläuternden Aussagen abfragte.

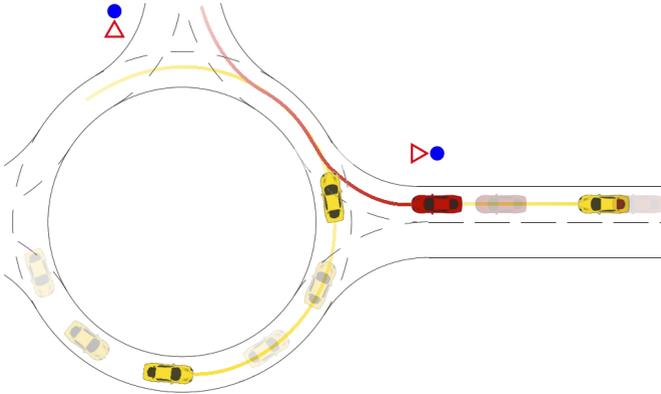


Abbildung 7.3: Das rote AF nähert sich dem Kreisverkehr mit der Geschwindigkeit 4 km/h und bewertet, ob die Lücke zwischen den beiden gelben Fahrzeugen im Kreisverkehr groß genug ist, um sicher einzufahren. Beide gelben Fahrzeuge fahren mit einer Geschwindigkeit von 15 km/h . Anschließend entscheidet das AF, in den Kreisverkehr einzufahren und fängt an, zu beschleunigen. Die Geschwindigkeitsbegrenzung liegt bei 30 km/h . Transparente rote und gelbe Fahrzeuge zeigen die vergangenen Positionen mit einem Abstand von jeweils 2 s an. Die verblassenden Linien stellen die geplanten Trajektorien dar (angelehnt an Henze, Faßbender und Stiller (2022)).

Material

Um die Ansätze in verschiedenen Verkehrssituationen zu testen, zeigten zwei Videos jeweils eine Simulation des Verhaltens eines AF während eines Überholmanövers und in einem Kreisverkehr¹. Wie in Abschnitt 7.1 beschrieben, analysierte eine SA jeweils eine Situation (vgl. Abb. 7.2 und 7.3). Die Positionen der Fahrzeuge in der Situation, die zum Zeitpunkt t analysiert wurde, sind jeweils durch die vollen Abbildungen der Fahrzeuge dargestellt, vergangene Positionen sind durch transparente Abbildungen und zukünftige durch Linien angedeutet.

Die Ergebnisse der SA wurden genutzt, um die Aussagen entsprechend der Wichtige-Inputs-Aussage, schlussfolgernden und regelbasierten Ansätze zu

¹Die Videos sind hier verfügbar: https://www.mrt.kit.edu/z/publ/download/Henze2022_takeOver_small.mp4 (Überholmanöver), https://www.mrt.kit.edu/z/publ/download/Henze2022_singleGap3_small.mp4 (Kreisverkehr).

Tabelle 7.2: Erläuternde Aussagen für die Kreisverkehr-Situation, vgl. Abb. 7.3 (angelehnt an Henze, Faßbender und Stiller (2022)).

menschliche Referenzaussage

Das rote Fahrzeug beschleunigt, um in den Kreisverkehr einzufahren, der nun frei ist. Außerdem fährt das rote Fahrzeug in den Kreisverkehr ein, weil der Sicherheitsabstand groß genug ist, um das andere Fahrzeug nicht zu behindern.

Wichtige-Inputs-Aussage

Das rote Fahrzeug beschleunigt und dafür sind die Geschwindigkeit des roten Fahrzeugs, die Position des roten Fahrzeugs, die Distanz zum Fahrzeug links im Kreisverkehr und die Geschwindigkeit des Fahrzeugs links im Kreisverkehr wichtig. Außerdem fährt das rote Fahrzeug in den Kreisverkehr ein und dafür sind die Position des roten Fahrzeugs, die Distanz zum Fahrzeug links im Kreisverkehr und die Geschwindigkeit des Fahrzeugs links im Kreisverkehr wichtig.

Schlussfolgernde Aussage

Das rote Fahrzeug beschleunigt und dafür sind die Geschwindigkeit des roten Fahrzeugs, die Distanz zum Fahrzeug links im Kreisverkehr und die Geschwindigkeit des Fahrzeugs links im Kreisverkehr wichtig, sodass wahrscheinlich das rote Fahrzeug das Ziel möglichst schnell erreichen möchte oder andere Fahrzeuge nicht behindert werden soll. Außerdem fährt das rote Fahrzeug in den Kreisverkehr ein und dafür sind die Distanz zum Fahrzeug links im Kreisverkehr und die Geschwindigkeit des Fahrzeugs links im Kreisverkehr wichtig, sodass wahrscheinlich andere Fahrzeuge nicht behindert werden sollen.

Regelbasierte Aussage

Das rote Fahrzeug beschleunigt, weil es das Ziel möglichst schnell erreichen möchte und andere Fahrzeuge nicht behindert werden sollen. Außerdem fährt das rote Fahrzeug in den Kreisverkehr ein, weil andere Fahrzeuge nicht behindert werden sollen.

erzeugen (vgl. Tabellen 7.1 und 7.2). Zusätzlich arbeitete eine Forscherin, die nicht mit den drei Ansätzen vertraut war, sich in die Literatur zu Erklärungen automatisierter Fahrfunktionen ein, identifizierte Empfehlungen zur Formulierung (vgl. Tabelle 7.3) und nutzte diese, um jeweils eine Erklärung für beide Situationen zu formulieren. Anschließend wurden einige Formulierungen durch charakteristische maschinen-generierte Ausdrücke ersetzt, um die Effek-

Tabelle 7.3: Literatur, die als Grundlage für die Formulierung der menschlichen Referenz dienten.

Literatur	Ergebnis
Koo et al. (2015)	<i>was und warum?</i> Informationen erzeugen das sicherste Fahrverhalten, aber überfordern Nutzer womöglich
Zhou und Danks (2020)	Erklärungen sollten an Nutzer und ihre Ziele angepasst sein
Lim et al. (2009)	<i>warum?</i> und <i>warum nicht?</i> Erklärungen erzeugen ein besseres Nutzerverständnis als <i>was?</i> und <i>was wäre, wenn?</i> Erklärungen
Meck und Precht (2021)	Aktiv ist besser als Passiv, Präsens ist besser als Vergangenheit, Indikativ, hypotaktische und längere Sätze (mehr als 15 Wörter) sind für Informationshinweise akzeptabel, wenige Adjektive/Adverbien/ Interjektionen

te von natürlicher klingenden Wortlauten zu minimieren. Die resultierenden Referenzaussagen sind in den Tabellen 7.1 und 7.2 abgedruckt.

Stichprobe

Insgesamt beendeten $N = 62$ Teilnehmer die Studie, die jeweils mindestens drei Jahre den Führerschein besaßen (15 Frauen, Mittelwert Alter $M_{\text{Alter}} = 32.29$ Jahre, Standardabweichung $SD_{\text{Alter}} = 11.00$ Jahre). Es gab zwei Gruppen Experten und Novizen mit jeweils 31 Teilnehmern. Alle Experten hatten jeweils mindestens ein Jahr Erfahrung in der Forschung und/oder Entwicklung automatisierter Fahrfunktionen (Mittelwert Erfahrung $M_{\text{Erf}} = 6.31$ Jahre, Standardabweichung $SD_{\text{Erf}} = 8.41$ Jahre).

Abhängige Variable

Um die Qualität der Aussagen zu messen, wurde der Explanation Satisfaction Fragebogen von Hoffmann, Mueller, Klein und Litman (2019) verwendet. Da die

Befragung auf Deutsch durchgeführt wurde, wurde der Fragebogen vorher mit der Forward-Backward Übersetzungsmethode nach Tsang, Royse und Terkawi (2017) übersetzt (Brislin, 1970).

Studiendesign

Die Online-Studie verwendete ein drei-faktorielles Design mit den zwei Innersubjektfaktoren Situation (zwei Ausprägungen) und Aussagentyp (vier Ausprägungen) sowie einem Zwischensubjektfaktor Erfahrung (zwei Ausprägungen). Nachdem die Teilnehmer in die Studienteilnahme eingewilligt hatten, wurde ihnen mitgeteilt, dass ein Automobilhersteller ihnen zwei Videos eines selbstfahrenden Fahrzeugs mit neuen Fahrfunktionen zeigen würde, aber noch nicht zu viel über die Fahrfunktionen verraten wolle. Die Teilnehmer wurden dazu aufgefordert, das Verhalten des Fahrzeugs trotzdem so gut wie möglich zu verstehen und dafür die angezeigten Aussagen zu verwenden. Die Datenerfassung bestand aus zwei Teilen, die jeweils eine Verkehrssituation abdeckte: Nachdem die Teilnehmer das Video angesehen hatten, wurde ihnen ein Bild der Situation angezeigt, die durch die Aussagen erläutert wurde (vgl. Abb. 7.2 und 7.3 ohne transparente Fahrzeuge und Linien). Sie füllten für jede der Aussagen den Fragebogen aus und hatten abschließend die Möglichkeit, in einem Freifeldtext die Aussagen zu kommentieren.

7.2.2 Ergebnisse

Die Daten wurden mit einer Mixed ANOVA (Varianzanalyse) mit den beiden Innersubjektfaktoren Situation (Überholmanöver, Kreisverkehr) und Aussagentyp (menschliche Referenz – m, Wichtige-Inputs-Aussage – WI, schlussfolgernde Aussage – SF, regelbasierte – RB) sowie dem Zwischensubjektfaktor Erfahrung (Experten, Novizen) bei einem Signifikanzniveau $\alpha = 0.05$ ausgewertet. Die Tests nach Mauchly und Levene zeigten keine Verletzung der Sphärizität bzw. Homogenität der Fehlervarianzen ($p > 0.05$), auch die Homogenität der Kovarianzmatrizen war gemäß Box-Test gegeben ($p = 0.276$).

Die Analyse zeigte einen signifikanten Haupteffekt der Erfahrung ($F(1, 60) = 11.01, p = 0.002, \eta_p^2 = 0.16$) und Interaktionseffekte zwischen Situation \times Aussagentyp sowie zwischen Situation \times Aussagentyp \times Erfahrung ($F(3, 60) =$

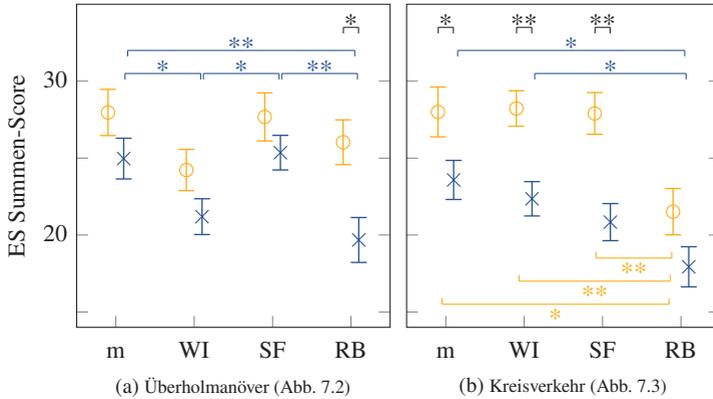


Abbildung 7.4: Explanation satisfaction (ES) für erläuternde Aussagen: m – menschliche Referenz, WI – Wichtige-Inputs-Aussage, SF – schlussfolgernde Aussage, RB – regelbasierte Aussage. Daten werden als Mittelwert \pm Standardfehler des ES Summen-Scores angezeigt. \times Experten, \circ Novizen; * $p < 0.05$, ** $p < 0.001$ (angelehnt an Henze, Faßbender und Stiller (2022)).

9.14, $p < 0.001$, $\eta_p^2 = 0.13$ und $F(3, 180) = 3.74$, $p = 0.012$, $\eta_p^2 = 0.06$). Insgesamt bewerteten die Experten alle Aussagen schlechter als die Novizen, wobei statistisch signifikante Unterschiede beim Überholmanöver (\ddot{U}) für die regelbasierte Aussage (Mittelwertdifferenz $M_{E-N,\ddot{U},RB} = M_{E,\ddot{U},RB} - M_{N,\ddot{U},RB} = -6.36$, $p = 0.003$ zwischen Experten (E) und Novizen (N)) und im Kreisverkehr für die menschliche Referenz sowie für die Wichtige-Inputs-Aussage und schlussfolgernde Aussage auftraten ($M_{E-N,KV,m} = -4.42$, $p = 0.036$; $M_{E-N,KV,WI} = -5.87$, $p < 0.001$; $M_{E-N,KV,SF} = -7.07$, $p < 0.001$). In den anderen Fällen waren die Unterschiede nicht statistisch signifikant (vgl. Abb. 7.4).

Um die Effekte der Innersubjektfaktoren Situation und Aussagentyp in beiden Erfahrungsgruppen zu untersuchen, wurden zwei 2 (Situation) \times 4 (Aussagentyp) ANOVAs mit Messwiederholungen durchgeführt (Signifikanzniveau $\alpha = 0.05$). Erneut lag keine Verletzung der Sphärizität vor. In beiden Gruppen wurde ein signifikanter Effekt des Aussagentyps (Experten $F(3, 90) = 9.32$, $p < 0.001$, $\eta_p^2 = 0.24$; Novizen: $F(3, 90) = 7.19$, $p < 0.001$, $\eta_p^2 = 0.19$) und eine signifikante Interaktion zwischen Situation \times Aussagentyp (Experten: $F(3, 90) = 4.31$, $p = 0.007$, $\eta_p^2 = 0.13$; Novize: $F(3, 90) = 8.26$, $p < 0.001$, $\eta_p^2 = 0.22$) entdeckt. Post-Hoc Tests mit Bonferroni-Korrektur wurden

für paarweise Vergleiche verwendet. Die Mittelwertunterschiede zwischen den drei automatisch generierten Ansätzen und der Referenzaussage sind für das Überholmanöver und den Kreisverkehr in Tabelle 7.4 dargestellt. Die Ergebnisse zeigen, dass Experten die regelbasierten Aussagen in beiden Situationen signifikant schlechter bewerteten als die Referenzaussage (vgl. Tabelle 7.4) und auch signifikant schlechter als die schlussfolgernde Beschreibung beim Überholmanöver ($M_{E,Ü,SF-RB} = 5.68, p < 0.001$) sowie die Wichtige-Inputs-Aussage im Kreisverkehr ($M_{E,KV,WI-RB} = 4.42, p = 0.028$, cf. Fig. 7.4). Zwischen der Wichtige-Inputs-Aussage und der menschlichen Aussage gab es im Kreisverkehr keinen signifikanten Unterschied und auch zwischen der schlussfolgernden und der menschlichen Beschreibung gab es in keiner der beiden Situationen Unterschiede (vgl. Tabelle 7.4).

Analog dazu bewerteten die Novizen die schlussfolgernden Aussagen in beiden Situationen genauso gut wie die Referenzbeschreibung (vgl. Tabelle 7.4). Für den Kreisverkehr zeigen die Post-Hoc Tests keine signifikanten Unterschiede zwischen der Wichtige-Inputs-Aussage und der menschlichen Referenz, erneut vergleichbar zu den Experten. Schlussendlich bewerteten auch die Novizen die regelbasierten Aussagen im Kreisverkehr signifikant schlechter als die anderen ($M_{N,R,WI-RB} = 6.71, p < 0.001$; $M_{N,R,SF-RB} = 6.39, p < 0.001$, vgl. Abb. 7.4).

Zusätzlich zu den Fragebögen hatten die Teilnehmer noch die Möglichkeit, die Aussagen zu kommentieren. Zum Beispiel erklärten einige Teilnehmer, der Planungsalgorithmus sollte zusätzliche Informationen berücksichtigen, die die SA als nicht relevant eingestuft hatte und die daher auch nicht in den Aussagen verwendet wurden (7 Experten, 2 Novizen): Sie fügten die Informationen hinzu, weil sie auf Grundlage ihrer Erfahrungen erwarteten, dass diese relevant sein sollten. Insgesamt 7 Teilnehmer merkten an, dass Informationen, die sie für unnötig („nicht nötig“, „unnötig“), erwartbar („selbstverständlich“, „offensichtlich“) oder „überflüssig“, hielten, nicht explizit erwähnt werden sollten (3 Experten, 4 Novizen). Mehrere Teilnehmer beschrieben, dass sie das Wort „wahrscheinlich“ in den Aussagen zum Kreisverkehr verunsicherte (7 Experten, 2 Novizen), und einige hinterfragten die Begründung „möglichst schnell“, weil sie ein „möglichst sicheres“ Fahrverhalten bevorzugten (7 Experten, 2 Novizen).

Tabelle 7.4: Paarweiser Vergleich zwischen erläuternden Aussagen und menschlicher Referenz m: WI – Wichtige-Inputs-Aussage, SF – schlussfolgernde Aussage, RB – regelbasierte Aussage. Mittelwerte $M_{m-n} = M_m - M_n$ ($n \in \{WI, SF, RB\}$) und Wahrscheinlichkeiten p mit Bonferroni-Korrektur; * $p < 0.05$, ** $p < 0.001$ (angelehnt an Henze, Faßbender und Stiller (2022)).

		Überholmanöver		Kreisverkehr	
n		M_{m-n}	p	M_{m-n}	p
Experten	m WI	3.774*	0.010	1.226	1.000
	SF	-0.387	1.000	2.742	0.417
	RB	5.290**	< 0.001	5.645*	0.005
Novizen	m WI	3.742	0.070	-0.226	1.000
	SF	0.290	1.000	0.097	1.000
	RB	1.935	0.344	6.484*	0.001

7.3 Diskussion

Dieser Abschnitt bespricht die Vor- und Nachteile des beschriebenen Vorgehens zur automatisierten Generierung erläuternder Beschreibungen und diskutiert die Ergebnisse und Einschränkungen der Onlinestudie.

Eine Stärke der vorgestellten Methode ist ihre Modularität, die das Hinzufügen von weiteren Sprachen oder zusätzlichen Informationen erleichtert. Zusätzlich nutzt sie sehr wenige Annahmen über den Planungsalgorithmus des AF, da die SA unabhängig vom Planungsansatz funktioniert. Dennoch gehen alle Aussagen ausschließlich aus dem Planungsalgorithmus selbst hervor, ohne dass sie von menschlichen Interpretationen beeinflusst sind, wie es z. B. durch das Labeln in Ansätzen mit maschinellem Lernen möglich wäre (Rudin, 2019; Y. Xu et al., 2020). Ein anderer Vorteil ist die Anpassungsfähigkeit der Methode an Nutzeranforderungen, indem z. B. die Art der Informationen verändert wird. Dies ist wichtig, weil die Umfrage bereits bei einer kleinen Stichprobengröße sehr unterschiedliche Informationsbedarfe gezeigt hat. Vor allem Nutzer mit Vorwissen scheinen Erwartungen an die Erläuterungen zu haben, die bisher nicht erfüllt werden und bewerten diese daher deutlich schlechter als Novizen. Dies wurde vor allem deutlich, als die Experten die Begründung „möglichst

schnell“ kritisierten und stattdessen Sicherheit als Handlungsmaxime anführten, während Novizen stattdessen erwarteten, dass das AF sicher und regelkonform fahren kann. Diese Beobachtung unterstützt den Vorschlag, die Gründe für Fahrentscheidungen auf die Bedürfnisse der Nutzer anzupassen.

Mehrere Teilnehmer merkten an, dass sie das Wort „wahrscheinlich“ verunsicherte, weil sie seinen Ursprung nicht nachvollziehen konnten. Einerseits können Anzeigen von Unsicherheiten das Vertrauen in nicht sicherheitskritische Anwendungen verbessern, z. B. ungewisse Reichweiteninformationen bei elektrischen Fahrzeugen (Jung, Sirkin, Gür & Steinert, 2015). Andererseits kann die Kommunikation von Unsicherheiten mithilfe von Wörtern das Vertrauen im Vergleich zur Kommunikation mithilfe von Intervallen oder nur mit Punktschätzern (d. h. ohne Angabe der Unsicherheit) verringern (van der Bles, van der Linden, Freeman & Spiegelhalter, 2020). Daher stellt sich die Frage, wie und in welchen Situationen Unsicherheiten kommuniziert werden sollten.

Sowohl Experten als auch Novizen bewerteten die regelbasierten Aussagen als unzufriedenstellend, weil Verweise auf die Umgebung fehlten. Sie bewerteten die schlussfolgernden Beschreibungen ähnlich gut wie die menschliche Referenz, auch wenn sie aus den längsten Sätzen bestanden. Das ist überraschend, weil die qualitativen Daten darauf hindeuten, dass viele Teilnehmer diese Beschreibungen als zu lang empfanden und kürzere Alternativen vorschlugen. Dies spricht dafür, dass Teilnehmer Aussagen mit Informationen sowohl zur Umgebung als auch zu Verkehrsregeln bevorzugen, auch wenn ihre Bewertung der Aussagen von der Aussagenlänge und der Zeit, bis sie die Informationen verstanden haben, möglicherweise negativ beeinflusst wurden. Außerdem deuteten die Teilnehmer an, dass keine offensichtlichen Tatsachen erläutert werden sollten, z. B. die Beschleunigung kurz vor der Einfahrt in den Kreisverkehr. Diese Beobachtungen unterstützen die Ergebnisse aus der Fokusgruppenstudie, eher laterale als longitudinale Bewegung zu erläutern (vgl. Kapitel 6, Beggiato et al., 2015; Henze, Faßbender & Stiller, 2021).

Die vorgestellte Textgenerierung basiert auf einer SA, die durch ihren Samplignansatz rechenintensiv ist. Daher ist eine Reduktion der Laufzeit nötig, bevor die Methode auch online im Realverkehr eingesetzt werden kann. Dafür wurde in Kapitel 4 das DSC vorgestellt (Henze, Faßbender & Stiller, 2022). Auch wenn die Anwendung noch im Realverkehr getestet werden muss, wird die Textgenerierung dennoch auch dann von Nutzen sein, wenn die Laufzeit zu lang sein sollte: Mögliche Anwendungsgebiete wären z. B. bei Debugging-Aufgaben

im Entwicklungsprozess, bei der Testautomatisierung oder als Analyse-Tool nach einem kritischen Zwischenfall, der weitere Untersuchungen verlangt.

Eine wesentliche Einschränkung der Validierungsstudie ist die Instruktion der Teilnehmer: Die Aufgabe, fundierte Vermutungen über die automatisierte Fahrfunktion anzustellen, war sehr speziell und hatte wenig Bezug zum Realverkehr. Die Videos zeigten die Verkehrssituationen aus der Vogelperspektive, was möglicherweise für einige Teilnehmer ungewohnt war. Der Vorteil dieser Darstellung ist, dass sich die Videos möglichst ähnlich sehen und unabhängig von diversen anderen Verkehrseinflüssen waren. Nichtsdestotrotz ist anzunehmen, dass die individuelle Wahrnehmung von kritischen Verkehrssituationen sowie die Visualisierung der Informationen einen großen Einfluss auf die Einstellung der Nutzer haben wird. Da die Videos außerdem nur Situationen zeigten, die nicht sicherheitskritisch waren und keine gefährdeten Verkehrsteilnehmer (z. B. Radfahrer oder Fußgänger) oder unerwarteten Fahrentscheidungen involvierten, ist die Übertragbarkeit der Ergebnisse auf den Realverkehr begrenzt. Nichtsdestotrotz sollten HMIs auf Grundlage des Wichtige-Inputs-Ansatzes oder schlussfolgernden Ansatzes entwickelt und anschließend in realistischeren Szenarien oder in Situationen, in denen das AF eine Entscheidung trifft, die der Nutzer nicht erwartet und daher hinterfragt, getestet werden.

8 Zusammenfassung und Ausblick

Die Transparenz von Verhaltensentscheidungen automatisierter Fahrzeuge ist auf verschiedenen Ebenen bedeutsam. Einerseits ist es für Entwickler wichtig, die Funktionsweise der Algorithmen nachzuvollziehen, um z. B. das Fortpflanzen von Unsicherheiten entlang der Wirkkette verstehen zu können. Andererseits muss es aber auch möglich sein, die Planungsentscheidungen so erläutern zu können, dass auch Personen ohne detailliertes Wissen eine Vorstellung von den Fähigkeiten und Grenzen eines AF bekommen. Diese Arbeit hat beide Aspekte transparenter Verhaltensentscheidungen beleuchtet.

Noch ist unklar, ob zukünftige AF das menschliche Fahrverhalten exakt nachbilden. Allerdings werden die Nutzer eines AF dieses nach menschlichen Maßstäben bewerten. Zu den Kriterien werden vermutlich nicht nur ihre Sicherheit und ihr Fahrstil gehören, sondern auch ihre Fahrentscheidungen. Daher ist es kritisch zu betrachten, dass derzeit oft nur das Auftreten von Kollisionen mit anderen Verkehrsteilnehmern als Metrik für die Zuverlässigkeit von Planungsalgorithmen genutzt wird. Aus diesem Grund hat diese Arbeit ein neues Maß dafür eingeführt, wie stark ein Planungsergebnis schwankt: Es betrachtet die Planungsentscheidung als stochastische Größe und bezieht so die Auswirkungen der Messabweichung in die Beurteilung der Planungsqualität ein. Das Ergebnis ist eine Definition, die bestimmt, ob eine Planungsentscheidung zulässig ist oder nicht.

Falls eine Planungsentscheidung als nicht zulässig eingestuft wird, so untersucht die sampling-basierte SA nach Morris, welche Inputs des Planers einen großen Anteil an den Planungsschwankungen hatten. Morris (1991) stellt dafür die Behauptung auf, dass bereits eine kleine Stichprobengröße ausreicht, um einflussreiche Inputs identifizieren zu können. Genauere Untersuchungen zeigen jedoch, dass die resultierenden charakteristischen Größen für kleine Stichprobengrößen starken Schwankungen unterliegen und daher eine größere Stichprobe nötig sein kann. Daher wird das DSC entwickelt, das die Stichprobengröße dynamisch an die Komplexität einer Fahrsituation anpasst. Verschiedene

Experimente zeigen, dass dieses Vorgehen insbesondere in der Anwendung des automatisierten Fahrens die Rechenzeit deutlich reduziert, weil es häufig viele Objekte gibt, die keinen Einfluss auf die Fahrentscheidung haben und damit schon nach wenigen Stichprobenwerten als irrelevant eingestuft werden können. Zukünftige Arbeiten sollten sich damit beschäftigen, wie die Voraussetzung für einen sicheren Abbruch des Kriteriums weiter gelockert (vgl. Satz 4.1.1) oder die Auswertung für einzelne Inputs noch früher gestoppt werden kann. Dafür können Heuristiken, die den konkreten Anwendungsfall des automatisierten Fahrens berücksichtigen, genutzt werden, z. B. dass die Geschwindigkeit und Position eines Objekts nur dann untersucht werden müssen, falls es existiert oder existieren könnte.

Bei den Informationen, die eine SA als relevant einstuft, schließt sich die Frage an, ob die zugehörige Messunsicherheit bereits zu groß ist und der Planer nicht mehr in der Lage ist, zulässige Ergebnisse zu berechnen. Leider gibt es bisher nur wenige Arbeiten, die die Fortpflanzung von Unsicherheiten an der Schnittstelle zwischen Wahrnehmung und Planung untersuchen. Daher ist das oben definierte stochastische Optimierungsproblem einer der ersten Ansätze, der aus einem gegebenen Planungsalgorithmus Anforderungen an die Wahrnehmung ableitet. Dieses bietet also die Möglichkeit zu entscheiden, welche Sensorkonfiguration verbaut werden sollte oder ob Rechenzeit zwischen Prozessoren anders verteilt werden muss, damit Ergebnisse rechtzeitig zur Verfügung stehen.

Henning, Müller, Gies, Buchholz und Dietmayer (2022) stellen einen Ansatz für eine an die Situation angepasste Wahrnehmung vor, um den Energieverbrauch zu reduzieren. Sie argumentieren, dass nicht in allen Situationen die gesamte Umgebung verarbeitet werden muss, sondern es ausreichen kann, die Situationinterpretation auf bestimmte Bereiche zu beschränken: Beispielsweise muss im urbanen Kontext die Umgebung im näheren Umfeld gut beobachtet werden, während auf der Autobahn auch Informationen zu Objekten wichtig sein können, die weiter entfernt sind. Dafür unterteilen sie die Umgebung mithilfe von Expertenwissen in relevante und irrelevante Bereiche. Die hier vorgestellte SA ist eine Möglichkeit, dieses Vorgehen zu ersetzen und die Unterscheidung automatisiert vorzunehmen, sodass Fehleinschätzungen durch Experten vermieden werden.

Henning et al. (2022) schlagen außerdem vor, dass die Definition von relevanten Gebieten verwendet werden kann, um die Eigenwahrnehmung (engl.

self-awareness) des AF zu unterstützen: Sie definieren Mindestanforderungen an die Sensoren, die für einen einwandfreien Betrieb notwendig sind und bei einer Verletzung der Anforderungen ein angemessenes Notfallmanöver auslösen. Dieses Konzept kann zukünftig mithilfe des Optimierungsproblems erweitert werden, um die notwendigen Anforderungen direkt aus dem Planungsalgorithmus abzuleiten.

Insgesamt wird der Nutzen der entwickelten Methoden anhand von zwei Beispielen deutlich: Die Freifahrt (Beispiel 3.1.8) und das Überholmanöver (Beispiel 3.2.6) zeigen, dass die SA, das DSC und das Optimierungsproblem dazu beitragen, dass Planungsalgorithmen und Wahrnehmung besser aufeinander abgestimmt werden. Insgesamt haben diese Methoden jedoch derzeit noch den Nachteil, dass die Rechenzeiten zu groß für den Einsatz im Realverkehr sind. Dies ist vor allem auf die prototypische Implementierung in Python zurückzuführen, die nicht auf Laufzeiteffizienz oder Speicherplatzoptimalität ausgelegt ist. Es ist zu erwarten, dass bei einer Implementierung, die auf diese Aspekte mehr Wert legt, und durch die Wahl einer anderen Programmiersprache deutliche Laufzeiteinsparungen möglich sind (Pereira et al., 2017). Auch falls diese Einsparungen nicht ausreichen sollten, um die Ergebnisse im Realverkehr berechnen zu können, sind diese dennoch wertvoll: Einerseits können sie im Entwicklungsprozess für Analysen im Anschluss an Testfahrten genutzt werden. Andererseits können sie in der Validierung der Fahrfunktion genutzt werden, um Schwachstellen zu identifizieren und die Fehlerfortpflanzung zu untersuchen. Damit haben diese Methoden das Potential, die Transparenz für Entwickler während des Entwicklungsprozess deutlich zu steigern.

Der zweite Schwerpunkt der Arbeit lag darauf, die Transparenz auch für Nutzer ohne Fachwissen zugänglich zu machen. Dafür wurde zunächst mit einer Fokusgruppe untersucht, welche unerwarteten Verhaltensentscheidungen eines AF, die bei Testfahrten auftreten, auch für Entwickler weitere Erklärungen erfordern. Aus diesen Ergebnissen können Anforderungen an Aussagen abgeleitet werden, die das Verhalten eines AF für Nutzer ohne Fachwissen erläutern.

Mit der SA stellt die Arbeit eine technische Umsetzung vor, um Informationen für zukünftige HMI-Konzepte zur Verfügung zu stellen: Die große Menge an Informationen, die ein Planungsalgorithmus verarbeitet, wird auf eine kleine Menge relevanter Inputs reduziert. Die anschließende Weiterverarbeitung setzt diese in Bezug zur Umgebung und zu den geltenden Verkehrsregeln, sodass das Verhalten des AF auch ohne detailliertes Wissen gut nachvollziehbar ist. Auch

wenn die hier vorgestellten erläuternden Beschreibungen noch Schwachstellen aufweisen, z. B. die Länge der Aussagen, wurde durch die durchgeführte Online-Studie doch deutlich, dass sie das Potential haben, um die technische Grundlage für leicht verständliche Erklärungen des Fahrverhaltens zu liefern. Falls die oben angesprochenen Laufzeitverbesserungen für die SA nicht ausreichen, um die Aussagen während der Fahrt zu generieren, ist eine Kopplung mit einem Machine-Learning-Ansatz eine mögliche Erweiterung: Dafür wird eine gute Kategorisierung von möglichen Verkehrssituationen benötigt, die eine Situation z. B. anhand der Straßentopologie oder der räumlichen Beziehung des Egos zu anderen Verkehrsteilnehmer einer Gruppe zuordnet. In jeder dieser Gruppen wird die SA für verschiedene Situationen ausgewertet. Diese Daten werden als Trainingsdatensatz verwendet, um neue Situationen durch Überlagerung der gegebenen Gruppen zusammensetzen und anhand der SA-Ergebnisse der überlagerten Gruppen die wichtigen Informationen zu identifizieren. Durch das so entstehende neuronale Netz kann die im Realverkehr beobachtete Situation verschiedenen Gruppen zugeordnet, die entsprechenden SA-Ergebnisse abgerufen und darauf basierend schneller beschrieben werden, was das AF entschieden hat. Falls der Nutzer eine Beschreibung in Frage stellt und weitere Informationen verlangt, kann er die etwas langsamere Auswertung durch die SA selbstständig triggern und nach kurzer Wartezeit eine passendere erläuternde Beschreibung erhalten.

Die entstehenden Erläuterungen des Fahrverhaltens werden vielfältige Informationen beinhalten, z. B. zu anderen Verkehrsteilnehmern und der Umgebung. Daher sollten dafür auch neue Anzeigekonzepte entwickelt werden. Vor diesem Hintergrund ist es ratsam, dass die HMI-Entwicklung die genannten Informationen noch besser weiter verarbeitet und die Brücke zu gelungenen, nutzerzentrierten Anzeigekonzepten schlägt. Dabei sollte insbesondere darauf Wert gelegt werden, den Nutzer nicht mit zu vielen Informationen zu überfordern und ein gelungenes Gleichgewicht zwischen einfacher Darstellung und angemessen komplexer Erklärung zu schaffen.

A Anhang

A.1 Erwartungswert und Varianz für Beispiel 3.1.8.e

Mit der W-Dichte aus Beispiel 3.1.8.a für die Freifahrt gilt für den Erwartungswert der elementaren Effekte (3.16) für die ZV $\hat{V} \sim \mathcal{N}(0, \sigma_v^2)$ und $\Delta_v \in \{-\sigma_v, \sigma_v\}$ mit $\mathbb{P}(\Delta_v = -\sigma_v) = \mathbb{P}(\Delta_v = \sigma_v) = 0.5$

$$\begin{aligned}
 \mathbb{E}(d_{a,i,t}(\hat{V}_{t,j})) &= \frac{\bar{a}}{v_w^4} \mathbb{E} \left(\left(\frac{(v_t + \hat{V})^4}{\Delta_v} - \frac{(v_t + \hat{V} + \Delta_v)^4}{\Delta_v} \right) \right) \\
 &= \frac{\bar{a}}{v_w^4} \mathbb{E} \left(\frac{(v_t + \hat{V})^4}{\Delta_v} - \frac{(v_t + \hat{V})^4}{\Delta_v} - \frac{1}{\Delta_v} \sum_{l=1}^4 \binom{4}{l} (v + \hat{V})^{4-l} \Delta_v^l \right) \\
 &= -\frac{\bar{a}}{v_w^4} \mathbb{E} \left(\frac{1}{\Delta_v} \sum_{l=1}^4 \binom{4}{l} (v + \hat{V})^{4-l} \Delta_v^l \right) \\
 &= -\frac{\bar{a}}{v_w^4} \left(4\mathbb{E} \left((v + \hat{V})^3 \right) + 6\mathbb{E} \left((v + \hat{V})^2 \right) \mathbb{E}(\Delta_v) \right. \\
 &\quad \left. + \mathbb{E} \left((v + \hat{V}) \right) \mathbb{E}(\Delta_v^2) + \mathbb{E}(\Delta_v^3) \right),
 \end{aligned}$$

wobei die letzte Gleichung aus der stochastischen Unabhängigkeit der ZV \hat{V} , Δ_v folgt. Für die Momente der ZV Δ_v gilt $\mathbb{E}(\Delta_v^l) = 0.5((-\Delta_v)^l + \Delta_v^l)$ und daher gilt für $l \in \{1, 2, 3\}$

$$\mathbb{E}(\Delta_v) = 0, \quad \mathbb{E}(\Delta_v^2) = \sigma_v^2, \quad \mathbb{E}(\Delta_v^3) = 0.$$

Weiterhin gilt für die Momente der ZV \hat{V}

$$\mathbb{E}(\hat{V}) = 0, \quad \mathbb{E}(\hat{V}^2) = \sigma_v^2, \quad \mathbb{E}(\hat{V}^3) = 0.$$

Daher vereinfacht sich der Erwartungswert zu

$$\begin{aligned} \mathbb{E}(d_{a,i,t}(\hat{V}_{t,j})) &= -\frac{\bar{a}}{v_w^4} \left(4\mathbb{E}((v + \hat{V})^3) + \mathbb{E}((v + \hat{V}))\sigma^2 \right) \\ &= -\frac{\bar{a}}{v_w^4} \left(4(v_t^3 + 3v_t^2\mathbb{E}(\hat{V}) + 3v_t\mathbb{E}(\hat{V}^2) + \mathbb{E}(\hat{V}^3) + 4\sigma_v^2(v_t + E(\hat{V}))) \right) \\ &= -\frac{\bar{a}}{v_w^4} \left(4v_t^3 + 12v_t\sigma_v^2 + 4\sigma_v^2v_t \right) \\ &= -\frac{\bar{a}}{v_w^4} \left(4v_t^3 + 16v_t\sigma_v^2 \right). \end{aligned}$$

Für die Varianz der elementaren Effekte gilt

$$\begin{aligned} \text{Var}(d_{a,i,t}(\hat{V}_{t,j})) &= \mathbb{E} \left(\left(\frac{\bar{a}}{v_w^4} \frac{(v_t + \hat{V})^4}{\Delta_v} - \frac{\bar{a}}{v_w^4} \frac{(v_t + \hat{V} + \Delta_v)^4}{\Delta_v} - \mathbb{E}(d_{a,i,t}(\hat{V}_{t,j})) \right)^2 \right) \\ &= \frac{\bar{a}^2}{v_w^8} \mathbb{E} \left(\left(\frac{(v_t + \hat{V})^4}{\Delta_v} - \frac{(v_t + \hat{V} + \Delta_v)^4}{\Delta_v} + 4v_t^3 + 16v_t\sigma_v^2 \right)^2 \right) \\ &= \frac{\bar{a}^2}{v_w^8} \mathbb{E} \left(\Delta_v^6 + 8\Delta_v^5\hat{V} + 8\Delta_v^5v_t + 28\Delta_v^4\hat{V}^2 + 56\Delta_v^4\hat{V}v_t + 28\Delta_v^4v_t^2 \right. \\ &\quad - 32\Delta_v^3\sigma_v^2v_t + 56\Delta_v^3\hat{V}^3 + 168\Delta_v^3\hat{V}^2v_t + 168\Delta_v^3\hat{V}v_t^2 \\ &\quad + 48\Delta_v^3v_t^3 - 128\Delta_v^2\sigma_v^2\hat{V}v_t - 128\Delta_v^2\sigma_v^2v_t^2 + 68\Delta_v^2\hat{V}^4 \\ &\quad + 272\Delta_v^2\hat{V}^3v_t + 408\Delta_v^2\hat{V}^2v_t^2 + 240\Delta_v^2\hat{V}v_t^3 + 36\Delta_v^2v_t^4 \\ &\quad - 192\Delta_v\sigma_v^2\hat{V}^2v_t - 384\Delta_v\sigma_v^2\hat{V}v_t^2 - 192\Delta_v\sigma_v^2v_t^3 \\ &\quad + 48\Delta_v\hat{V}^5 + 240\Delta_v\hat{V}^4v_t + 480\Delta_v\hat{V}^3v_t^2 + 432\Delta_v\hat{V}^2v_t^3 \\ &\quad + 144\Delta_v\hat{V}v_t^4 + 256\sigma_v^4v_t^2 - 128\sigma_v^2\hat{V}^3v_t - 384\sigma_v^2\hat{V}^2v_t^2 \\ &\quad - 384\sigma_v^2\hat{V}v_t^3 + 16\hat{V}^6 + 96\hat{V}^5v_t + 240\hat{V}^4v_t^2 + 288\hat{V}^3v_t^3 \\ &\quad \left. + 144\hat{V}^2v_t^4 \right). \end{aligned}$$

Mit den oben angegebenen Momenten und den weiteren Momenten

$$\begin{aligned}\mathbb{E}\left(\Delta_v^4\right) &= \sigma_v^4, & \mathbb{E}\left(\Delta_v^5\right) &= 0, & \mathbb{E}\left(\Delta_v^6\right) &= \sigma_v^6, \\ \mathbb{E}\left(\hat{V}^4\right) &= 3\sigma_v^4, & \mathbb{E}\left(\hat{V}^5\right) &= 0, & \mathbb{E}\left(\hat{V}^6\right) &= 15\sigma_v^6\end{aligned}$$

gilt mit der Linearität des Erwartungswertes

$$\begin{aligned}\text{Var}(d_{a,i,t}(\hat{V}_{t,j})) &= \frac{\bar{a}^2}{v_w^8} \left(\sigma_v^6 + 28\sigma_v^4\sigma_v^2 + 28\sigma_v^6 - 128\sigma_v^4v_t^2 \right. \\ &\quad + 204\sigma_v^6 + 408\sigma_v^4v_t^2 + 36\sigma_v^2v_t^4 + 256\sigma_v^4v_t^2 \\ &\quad \left. - 384\sigma_v^4v_t^2 + 240\sigma_v^6 + 720\sigma_v^4v_t^2 + 144\sigma_v^2v_t^4 \right) \\ &= \frac{\bar{a}^2}{v_w^8} \left(473\sigma_v^6 + 900\sigma_v^4v_t^2 + 180\sigma_v^2v_t^4 \right).\end{aligned}$$

A.2 Sensitivitätsanalyse für Beispiel 3.2.6.c

Abb. A.1 zeigt die zeitabhängige SA des Überholmanöver-Beispiel 3.2.6. Die linke Spalte zeigt den Einfluss der verschiedenen Inputs auf die Beschleunigung $o = a$, die rechte Spalte auf den FSW $o = lc$.

A.3 Beweis: Abbruch des dynamischen Abbruchkriteriums

Im Folgenden wird bewiesen, dass das DSC abbricht:

Satz 4.1.1. *Für eine Funktion $f : \mathbb{R}^k \rightarrow \mathbb{R}^m$ bricht das DSC mit den Konstanten $\kappa_{act}, \kappa_{stop} > 0$ ab, falls $\sqrt{km}\kappa_{act} \leq \kappa_{stop}$ gilt.*

Um diesen Satz zu beweisen, verwenden wir das Lemma:

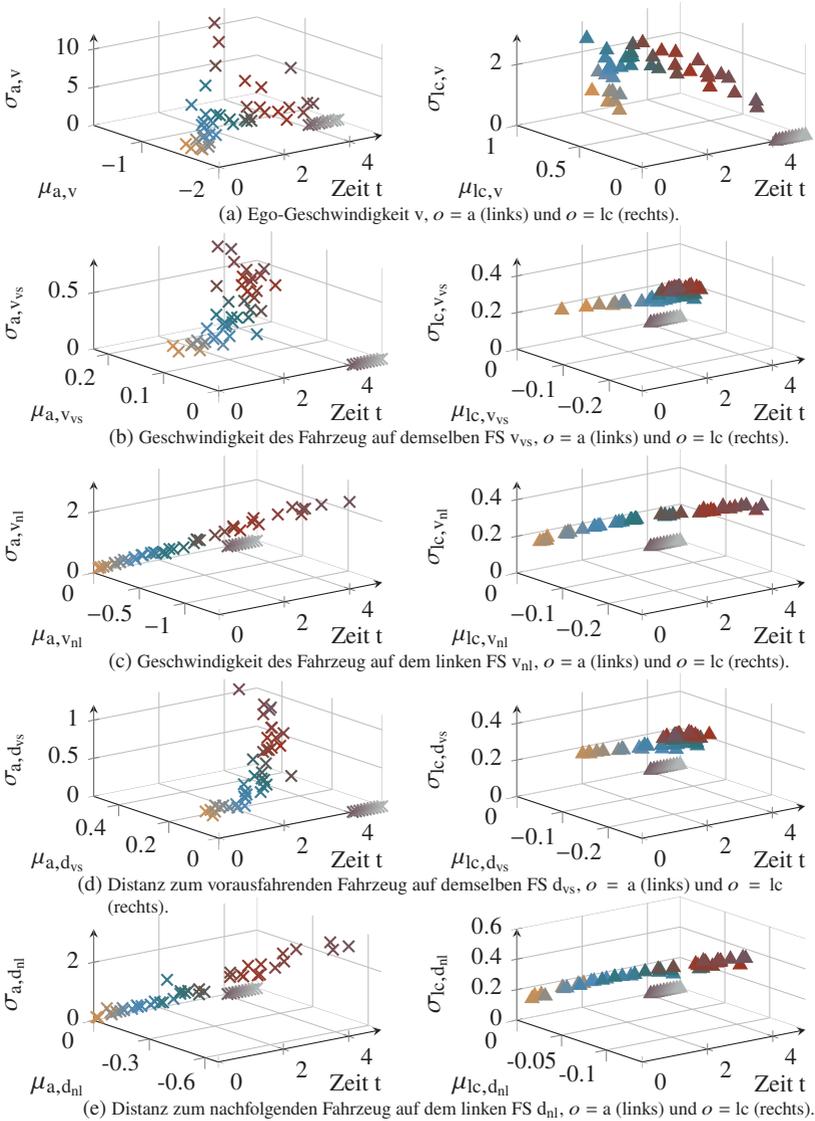


Abbildung A.1: SA für Beispiel 3.2.6 (Überholmanöver). Zeitlicher Verlauf in 0 s 5 s.

Lemma A.3.1. Die relativen Differenzen $\epsilon_{\mu,o,i,M}$, $\epsilon_{\mu^*,o,i,M}$, $\epsilon_{\sigma,o,i,M}$ der charakteristischen Größen $\mu_{o,i,M}$, $\mu_{o,i,M}^*$, $\sigma_{o,i,M}$ konvergieren schwach gegen 0 für $M \rightarrow \infty$, d.h. für alle $o = 1, \dots, m$, $i = 1, \dots, k$, $\delta > 0$ gilt

$$\lim_{M \rightarrow \infty} \mathbb{P}(\epsilon_{\mu,o,i,M} > \delta) = \lim_{M \rightarrow \infty} \mathbb{P}(\epsilon_{\mu^*,o,i,M} > \delta) = \lim_{M \rightarrow \infty} \mathbb{P}(\epsilon_{\sigma,o,i,M} > \delta) = 0.$$

Beweis (Lemma A.3.1). Für den Beweis nutzen wir zunächst das Schwache Gesetz der großen Zahlen (Khinchins Version) und die Definition von Cauchy-Folgen in Wahrscheinlichkeit. Aus Khinchins Gesetz der großen Zahlen (vgl. z. B. (Borovkov, 2013, Theorem 8.1.1)) folgen für $\delta > 0$ die schwachen Konvergenzen

$$\begin{aligned} \lim_{M \rightarrow \infty} \mathbb{P}(|\mu_{o,i,M} - \mathbb{E}(d_{o,i})| > \delta) &= 0 \\ \lim_{M \rightarrow \infty} \mathbb{P}(|\mu_{o,i,M}^* - \mathbb{E}(|d_{o,i}|)| > \delta) &= 0 \\ \lim_{M \rightarrow \infty} \mathbb{P}(|\sigma_{o,i,M} - \sqrt{\text{Var}(d_{o,i})}| > \delta) &= 0. \end{aligned}$$

Da eine Folge von Zufallsvariablen, die schwach konvergiert, auch eine schwache Cauchy-Folge ist (vgl. z. B. (Borovkov, 2013, Theorem 6.1.3)), sind die Folgen $\mu_{o,i,M}$, $\mu_{o,i,M}^*$, $\sigma_{o,i,M}$ schwache Cauchy-Folgen, d.h. für $\delta > 0$, $M, \tilde{M} \rightarrow \infty$ gilt

$$\lim_{M, \tilde{M} \rightarrow \infty} \mathbb{P}(|\mu_{o,i,\tilde{M}} - \mu_{o,i,M}| > \delta) = 0$$

und analoges für die Folgen $\mu_{o,i,M}^*$, $\sigma_{o,i,M}$. Für $\tilde{M} := M - l$ folgt also die Konvergenz der Zähler von $\nu_{\mu,M,o,i}$, $\nu_{\mu^*,M,o,i}$, $\nu_{\sigma,M,o,i}$. Da für zwei schwach konvergente Folgen $X_M \rightarrow X$, $Y_M \rightarrow Y$ auch das Produkt $X_M Y_M \rightarrow XY$ schwach konvergiert und dasselbe für eine Folge $Y_M := \frac{1}{Z_M} \rightarrow \frac{1}{Z} = Y$ gilt, folgt auch für die Matrizen $\nu_{\mu,M,M-l}$, $\nu_{\mu^*,M,M-l}$, $\nu_{\sigma,M,M-l}$ die komponentenweise Konvergenz gegen $\mathbf{0} \in \mathbb{R}^{m \times k}$. Damit konvergieren die relativen Differenzen $\epsilon_{\mu,o,i,M}$, $\epsilon_{\mu^*,o,i,M}$, $\epsilon_{\sigma,o,i,M}$ schwach gegen 0 für $M \rightarrow \infty$ und $o = 1, \dots, m$, $i = 1, \dots, k$. \square

Damit kann nun Satz 4.1.1 gezeigt werden:

Beweis Satz 4.1.1. Für jedes $M \in \mathbb{N}$ existiert eine Teilmenge $\mathcal{I}_M := \{1, \dots, k\} \setminus \mathcal{A}_M \subseteq \{1, \dots, k\}$ von inaktiven Inputs, sodass

$$\mathcal{I}_M := \{i = 1, \dots, k \mid \epsilon_{\mu, o, i, M_i} \leq \kappa_{\text{act}} \vee \epsilon_{\mu^*, o, i, M_i} \leq \kappa_{\text{act}} \\ \vee \epsilon_{\sigma, o, i, M_i} > \kappa_{\text{act}} \text{ für } M_i < M\}$$

mit $M_i := \min\{\tilde{M} \in \mathbb{N} \mid \epsilon_{\mu, o, i, M_i} \leq \kappa_{\text{act}} \vee \epsilon_{\mu^*, o, i, M_i} \leq \kappa_{\text{act}} \vee \epsilon_{\sigma, o, i, M_i} > \kappa_{\text{act}}\}$ gilt ($\mathcal{I}_M = \emptyset$ ist auch möglich). Aus der schwachen Konvergenz nach Lemma A.3.1 ist bekannt, dass für alle κ_{act} ein $\bar{M} := \max_{i=1, \dots, k} M_i$ existiert, sodass $\epsilon_{\mu, o, i, M} \leq \kappa_{\text{act}}$, $\epsilon_{\mu^*, o, i, M} \leq \kappa_{\text{act}}$ und $\epsilon_{\sigma, o, i, M} \leq \kappa_{\text{act}}$ für alle $M \geq \bar{M}$ gilt. Dann folgt für $M \geq \bar{M}$ und $\mathbf{M} \in \mathbb{N}^k$, $M_i < M$, falls $i \in \mathcal{I}_M$ und $M_i = M$ sonst mit der Definition der Frobenius-Norm

$$\begin{aligned} \|\mathbf{v}_{\mu^*, \mathbf{M}, \mathbf{M}-l}\|_{\text{F}}^2 &= \sum_{o, i} \frac{|\mu_{o, i, M_i-l}^* - \mu_{o, i, M_i}^*|^2}{|\mu_{o, i, M_i}^*|^2} \\ &= \sum_{i \in \mathcal{A}_M} \sum_o \frac{|\mu_{o, i, M-l}^* - \mu_{o, i, M}^*|^2}{|\mu_{o, i, M}^*|^2} \\ &\quad + \sum_{i \in \mathcal{I}_M} \sum_o \frac{|\mu_{o, i, M_i-l}^* - \mu_{o, i, M_i}^*|^2}{|\mu_{o, i, M_i}^*|^2} \\ &\leq \sum_{i \in \mathcal{A}_M} \sum_o \kappa_{\text{act}}^2 + \sum_{i \in \mathcal{I}_M} \sum_o \kappa_{\text{act}}^2 \\ &= km\kappa_{\text{act}}^2 \end{aligned}$$

(und analog für $\|\mathbf{v}_{\mu, M, M-l}\|_F^2$, $\|\mathbf{v}_{\sigma, M, M-l}\|_F^2$). Damit gilt

$$\begin{aligned} \epsilon_M &= \begin{pmatrix} \epsilon_{\mu, M} \\ \epsilon_{\mu^*, M} \\ \epsilon_{\sigma, M} \end{pmatrix} = \frac{1}{10} \sum_{l=1}^{10} \begin{pmatrix} \|\mathbf{v}_{\mu, M, M-l}\|_F \\ \|\mathbf{v}_{\mu^*, M, M-l}\|_F \\ \|\mathbf{v}_{\sigma, M, M-l}\|_F \end{pmatrix} \\ &= \frac{1}{10} \sum_{l=1}^{10} \sqrt{km} \kappa_{\text{act}} = \sqrt{km} \kappa_{\text{act}} \leq \kappa_{\text{stop}}. \end{aligned}$$

Damit bricht das DSC ab. □

A.4 Ergänzende Abbildungen zum dynamischen Abbruchkriterium

Abb. A.2 und A.3 zeigen den Verlauf der charakteristischen Größen μ_{i, M_i} , σ_{i, M_i} in Abhängigkeit von κ_{stop} . Die Berechnungen für Abb. A.2 wurden wie in Satz 4.1.1 vorausgesetzt mit $\kappa_{\text{act}} = 1/\sqrt{km} \kappa_{\text{stop}}$ durchgeführt, für Abb. A.3 wurde $\kappa_{\text{act}} = 0.5 \kappa_{\text{stop}}$ gewählt. In beiden Fällen werden die Konfidenzintervalle mit sich verringernden κ_{stop} schmaler. Außerdem liegt auch hier die Referenz (gestrichelte Linien), die gemäß Campolongo und Saltelli (1997) berechnet wurde, außerhalb der 95 %-Konfidenzintervalle. Dies deutet daraufhin, dass das DSC die Berechnung der charakteristischen Größen angemessen wählt.

A.5 Monotonie der Nebenbedingung für Beispiel 3.1.8.g

Es wird gezeigt, dass die stochastische Nebenbedingung (5.3.2) des Optimierungsproblems (5.3) streng monoton steigend ist, vgl. Beispiel 3.1.8.g. Dafür

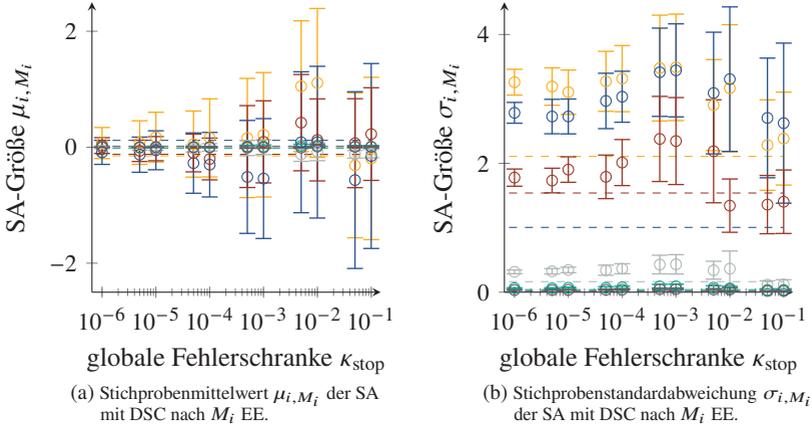


Abbildung A.2: Vergleich SA mit und ohne DSC (vgl. Satz 4.1.1) für die g -Funktion (4.5) mit $k = 6$ Inputs. $M_{\min} = 10$, $\kappa_{\text{act}} = 1/\sqrt{km}\kappa_{\text{stop}}$. $\circ a_1 = 0$; $\circ a_2 = 0.2$; $\circ a_3 = 0.9$; $\circ a_4 = 9$; $\circ a_5 = 50$; $\circ a_6 = 99$; Fehlerbalken zeigen 95% Konfidenzintervalle, gestrichelte Linien geben Ergebnisse analog zu Campolongo und Saltelli (1997) an.

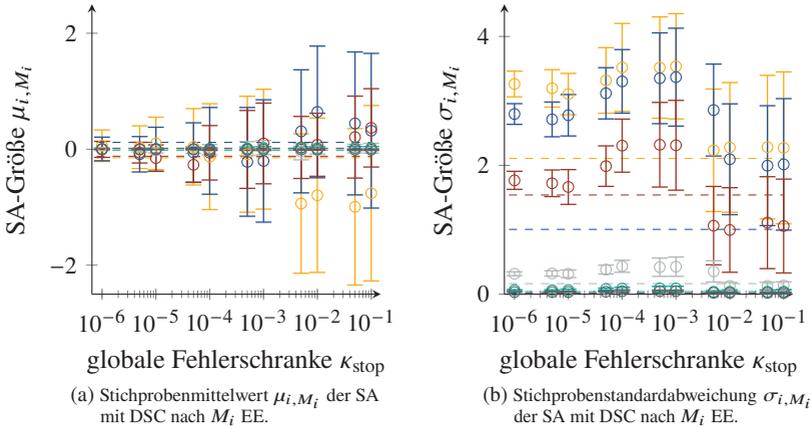


Abbildung A.3: Vergleich SA mit und ohne DSC für die g -Funktion (4.5) mit $k = 6$ Inputs. $M_{\min} = 10$, $\kappa_{\text{act}} = 0.5\kappa_{\text{stop}}$. $\circ a_1 = 0$; $\circ a_2 = 0.2$; $\circ a_3 = 0.9$; $\circ a_4 = 9$; $\circ a_5 = 50$; $\circ a_6 = 99$; Fehlerbalken zeigen 95% Konfidenzintervalle, gestrichelte Linien geben Ergebnisse analog zu Campolongo und Saltelli (1997) an.

gilt analog zu Beispiel 3.1.8.b mit $\delta = 4$ und $E(\sigma_v) := \mathbb{E}(f_a(v_t + \hat{V})) = \bar{a} - \frac{\bar{a}}{\sqrt{w}^4} (v_t^4 + 6v_t^2\sigma_v^2 + 3\sigma_v^4)$

$$\begin{aligned}
& \frac{d}{d\sigma_v} \mathbb{P}_{f_a(v_t + \hat{V})} \left(|f_a(v_t + \hat{V}) - \mathbb{E}(f_a(v_t + \hat{V}))| > \bar{b}_a \right) \\
&= \frac{d}{d\sigma_v} \left(1 - \Phi_{f_a}(E(\sigma_v) + \bar{b}_a) + \Phi_{f_a}(E(\sigma_v) - \bar{b}_a) \right) \\
&= \frac{d}{d\sigma_v} \frac{1}{2} \operatorname{erf} \left(\underbrace{\frac{\sqrt[4]{1 - \frac{\min(E(\sigma_v) + \bar{b}_a, \bar{a})}{\bar{a}}} v_w - v_t}{\sqrt{2}\sigma_v}}_{=: A_+(\sigma_v)} \right) \\
&\quad - \frac{d}{d\sigma_v} \frac{1}{2} \operatorname{erf} \left(\underbrace{\frac{\sqrt[4]{1 - \frac{\min(E(\sigma_v) - \bar{b}_a, \bar{a})}{\bar{a}}} v_w - v_t}{\sqrt{2}\sigma_v}}_{=: A_-(\sigma_v)} \right) \\
&= \frac{1}{\sqrt{\pi}} \underbrace{\exp(A_+^{-2}(\sigma_v))}_{>0} \frac{dA_+(\sigma_v)}{d\sigma_v} - \frac{1}{\sqrt{\pi}} \underbrace{\exp(A_-^{-2}(\sigma_v))}_{>0} \frac{dA_-(\sigma_v)}{d\sigma_v}.
\end{aligned}$$

Es wird also gezeigt, dass $\frac{dA_+(\sigma_v)}{d\sigma_v} \geq 0$ und $\frac{dA_-(\sigma_v)}{d\sigma_v} < 0$ gilt. Damit gilt nun für den Zähler von $\frac{dA_+(\sigma_v)}{d\sigma_v}$ für $\bar{a} \leq E(\sigma_v) + \bar{b}_a$

$$\frac{d}{d\sigma_v} \left(\sqrt[4]{1 - \frac{\bar{a}}{\bar{a}}} v_w - v_t \right) = 0.$$

Daher gilt

$$\frac{dA_+(\sigma_v)}{d\sigma_v} = -\frac{\sqrt[4]{1 - \frac{\bar{a}}{\bar{a}}} v_w - v_t}{2\sigma_v^2} = \frac{v_t}{2\sigma_v^2} > 0.$$

Für $\bar{a}_a > E(\sigma_v) + \bar{b}_a$ gilt für den Zähler von $A_+(\sigma_v)$:

$$\begin{aligned}
 \frac{d}{d\sigma_v} \left(\sqrt[4]{1 - \frac{E(\sigma_v) + \bar{b}_a}{\bar{a}}} v_w - v_t \right) &= \frac{d}{d\sigma_v} \left(1 - \frac{E(\sigma_v) + \bar{b}_a}{\bar{a}} \right)^{1/4} v_w \\
 &= \frac{1}{4} v_w \left(1 - \frac{E(\sigma_v) + \bar{b}_a}{\bar{a}} \right)^{-3/4} \frac{d}{d\sigma_v} \left(1 - \frac{E(\sigma_v) + \bar{b}_a}{\bar{a}} \right) \\
 &= \frac{1}{4} \frac{1}{v_w^3} \left(1 - \frac{E(\sigma_v) + \bar{b}_a}{\bar{a}} \right)^{-3/4} \left(12v_t^2 \sigma_v + 12\sigma_v^3 \right) \\
 &= 3 \frac{v_t^2 \sigma_v + \sigma_v^3}{v_w^3} \left(1 - \frac{E(\sigma_v) + \bar{b}_a}{\bar{a}} \right)^{-3/4}
 \end{aligned}$$

und somit für die Ableitung

$$\begin{aligned}
 \frac{dA_+(\sigma_v)}{d\sigma_v} &= \frac{1}{2\sigma_v^2} \left(3\sqrt{2}\sigma_v \frac{v_t^2 \sigma_v + \sigma_v^3}{v_w^3} \left(1 - \frac{E(\sigma_v) + \bar{b}_a}{\bar{a}} \right)^{-3/4} - \right. \\
 &\quad \left. \sqrt{2} \left(\sqrt[4]{1 - \frac{E(\sigma_v) + \bar{b}_a}{\bar{a}}} v_w - v_t \right) \right) \\
 &= \frac{1}{2\sigma_v^2} \left(3\sqrt{2}\sigma_v \frac{v_t^2 \sigma_v + \sigma_v^3}{v_w^3} \left(\frac{v_t^4 + 6v_t^2 \sigma_v^2 + 3\sigma_v^4}{v_d^4} - \frac{\bar{b}_a}{\bar{a}} \right)^{-3/4} - \right. \\
 &\quad \left. \sqrt{2} \left(\sqrt[4]{\frac{v_t^4 + 6v_t^2 \sigma_v^2 + 3\sigma_v^4}{v_d^4} - \frac{\bar{b}_a}{\bar{a}}} v_w - v_t \right) \right) \\
 &= \frac{1}{\sqrt{2}\sigma_v^2} \left(\frac{v_t^4 + 6v_t^2 \sigma_v^2 + 3\sigma_v^3}{v_w^4} - \frac{\bar{b}_a}{\bar{a}} \right)^{-3/4} \\
 &\quad \cdot \left(\frac{3v_t^2 \sigma_v^2 + 3\sigma_v^4}{v_w^3} - \frac{v_t^4 + 6v_t^2 \sigma_v^2 + 3\sigma_v^4 - \bar{b}_a v_w^4 / \bar{a}}{v_w^3} \right) + \frac{v_t}{\sqrt{2}\sigma_v^2} \\
 &= \frac{1}{\sqrt{2}\sigma_v^2} \left(v_t^4 + 6v_t^2 \sigma_v^2 + 3\sigma_v^3 - \frac{v_w^4 \bar{b}_a}{\bar{a}} \right)^{-3/4} \left(v_t^4 - 3v_t^2 \sigma_v^2 + \frac{v_w^4 \bar{b}_a}{\bar{a}} \right) + \frac{v_t}{\sqrt{2}\sigma_v^2}
 \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\sqrt{2}\sigma_v^2} \underbrace{\left(v_t^4 + 6v_t^2\sigma_v^2 + 3\sigma_v^4 - \frac{v_w^4\bar{b}_a}{\bar{a}} \right)^{-3/4}}_{\geq 0 \text{ wegen } \bar{a} > E(\sigma_v) + \bar{b}_a} \underbrace{\left(v_t^4 - 3v_t^2\sigma_v^2 + \frac{v_w^4\bar{b}_a}{\bar{a}} \right)}_{\geq 0 \text{ für } \sigma_v^2 \leq \frac{4}{3}v_t^2 + \frac{\bar{b}_a v_w^4}{3v_t^2 \bar{a}}} + \frac{v_t}{\sqrt{2}\sigma_v^2} \\
&\hspace{20em} (A.1)
\end{aligned}$$

Für $\sigma_v^2 \geq \frac{4}{3}v_t^2 + \frac{\bar{b}_a v_w^4}{3v_t^2 \bar{a}}$ gilt:

$$\begin{aligned}
3\sigma_v^4 &\geq 3 \left(\frac{4}{3}v_t^2 + \frac{\bar{b}_a v_w^4}{3\bar{a}v_t^2} \right)^2 = 3 \frac{4^2}{3^2} v_t^4 + \frac{8}{3} \frac{\bar{b}_a v_w^4}{\bar{a}} + \frac{3\bar{b}_a^2 v_w^8}{9\bar{a}^2 v_t^4} \\
&= \frac{16}{3} v_t^4 + \frac{8}{3} \frac{\bar{b}_a}{\bar{a}} v_w^4 + \frac{\bar{b}_a^2 v_w^8}{3\bar{a}^2 v_t^4}.
\end{aligned}$$

Damit folgt für den ersten Faktor von (A.1):

$$\begin{aligned}
&\left(v_t^4 + 6v_t^2\sigma_v^2 + 3\sigma_v^4 - \frac{v_w^4\bar{b}_a}{\bar{a}} \right)^{-3/4} \\
&\leq \left(v_t^4 + 6v_t^2\sigma_v^2 + \frac{16}{3}v_t^4 + \frac{8}{3}\frac{\bar{b}_a}{\bar{a}}v_w^4 + \frac{\bar{b}_a^2 v_w^8}{3\bar{a}^2 v_t^4} - \frac{\bar{b}_a v_w^4}{\bar{a}} \right)^{-3/4} \\
&= \left(v_t^4 + 6v_t^2\sigma_v^2 + \frac{16}{3}v_t^4 + \frac{5}{3}\frac{\bar{b}_a}{\bar{a}}v_w^4 + \frac{\bar{b}_a^2 v_w^8}{3\bar{a}^2 v_t^4} \right)^{-3/4}.
\end{aligned}$$

Einsetzen in (A.1) ergibt:

$$\begin{aligned}
&\frac{1}{\sqrt{2}\sigma_v^2} \left(v_t^4 + 6v_t^2\sigma_v^2 + 3\sigma_v^4 - \frac{v_w^4\bar{b}_a}{\bar{a}} \right)^{-3/4} \left(v_t^4 - 3v_t^2\sigma_v^2 + \frac{v_w^4\bar{b}_a}{\bar{a}} \right) + \frac{v_t}{\sqrt{2}\sigma_v^2} \\
&\geq \frac{1}{\sqrt{2}\sigma_v^2} \frac{v_t^4 - 3v_t^2\sigma_v^2 + \frac{\bar{b}_a v_w^4}{\bar{a}}}{\left(v_t^4 + 6v_t^2\sigma_v^2 + \frac{16}{3}v_t^4 + \frac{5}{3}\frac{\bar{b}_a}{\bar{a}}v_w^4 + \frac{\bar{b}_a^2 v_w^8}{3\bar{a}^2 v_t^4} \right)^{3/4}} + \frac{v_t}{\sqrt{2}\sigma_v^2} \\
&\geq \frac{1}{\sqrt{2}\sigma_v^2} \frac{v_t^4 + \frac{\bar{b}_a}{\bar{a}}v_w^4}{\left(v_t^4 + 6v_t^2\sigma_v^2 + \frac{16}{3}v_t^4 + \frac{5}{3}\frac{\bar{b}_a}{\bar{a}}v_w^4 + \frac{\bar{b}_a^2 v_w^8}{3\bar{a}^2 v_t^4} \right)^{3/4}} \dots
\end{aligned}$$

$$\begin{aligned}
 & - \frac{1}{\sqrt{2}\sigma_v^2} \frac{1}{\frac{1}{3} \frac{v_t^2}{\sigma_v^2} + 2 + \frac{16}{9} \frac{v_t^2}{\sigma_v^2} + \frac{5}{9} \frac{\bar{b}_a}{\bar{a}} \frac{v_w^4}{v_t^2 \sigma_v^2} + \frac{\bar{b}_a^2 v_w^8}{9 \bar{a}^2 v_t^6 \sigma_v^2}} + \frac{v_t}{\sqrt{2}\sigma_v^2} \\
 \geq & \frac{1}{\sqrt{2}\sigma_v^2} \frac{v_t^4 + \frac{\bar{b}_a}{\bar{a}} v_w^4}{\left(v_t^4 + 6v_t^2\sigma_v^2 + \frac{16}{3}v_t^4 + \frac{5}{3} \frac{\bar{b}_a}{\bar{a}} v_w^4 + \frac{\bar{b}_a^2 v_w^8}{3 \bar{a}^2 v_t^6}\right)^{3/4}} - \underbrace{\frac{1}{2\sqrt{2}\sigma_v^2} + \frac{v_t}{\sqrt{2}\sigma_v^2}}_{\geq 0} \geq 0,
 \end{aligned}$$

wobei die zweite Ungleichung aus $f(x_1 + x_2) = (x_1 + x_2)^{-3/4} \geq 1/(x_1^{3/4} + x_2^{3/4})$ folgt. Für den zweiten Term $\frac{dA_-(\sigma_v)}{d\sigma_v}$ gilt wegen $E(\sigma_v) \leq \bar{a} \leq \bar{a} + \bar{b}_a$ immer auch $E(\sigma_v) - \bar{b}_a \leq \bar{a}$, sodass gilt:

$$\begin{aligned}
 \frac{dA_-(\sigma_v)}{d\sigma_v} &= \frac{d}{d\sigma_v} \frac{\sqrt[4]{1 - \frac{E(\sigma_v) - \bar{b}_a}{\bar{a}} v_w} - v_t}{\sqrt{2}\sigma_v} \\
 &= \frac{1}{2\sigma_v^2} \left(3\sqrt{2} \frac{v_t^2 \sigma_v^2 + \sigma_v^4}{v_w^3} \left(1 - \frac{E(\sigma_v) - \bar{b}_a}{\bar{a}} \right)^{-3/4} \right. \\
 &\quad \left. - \sqrt{2} \left(\left(1 - \frac{E(\sigma_v) - \bar{b}_a}{\bar{a}} \right)^{1/4} v_w - v_t \right) \right) \\
 &= \frac{1}{2\sigma_v^2} \left(3\sqrt{2} \frac{v_t^2 \sigma_v^2 + \sigma_v^4}{v_w^3} \left(\frac{v_t^4 + 6v_t^2 \sigma_v^2 + 3\sigma_v^4}{v_w^4} + \frac{\bar{b}_a}{\bar{a}} \right)^{-3/4} \right. \\
 &\quad \left. - \sqrt{2} \left(\left(\frac{v_t^4 + 6v_t^2 \sigma_v^2 + 3\sigma_v^4}{v_w^4} + \frac{\bar{b}_a}{\bar{a}} \right)^{-1/4} v_w - v_t \right) \right).
 \end{aligned}$$

Aus dem Optimierungsproblem wissen wir, dass $\sigma_v \in (0, 2]$ gilt. Daher können die beiden Terme durch ihr Supremum bzw. Infimum auf diesem Intervall abgeschätzt werden. Damit ergibt sich

$$\begin{aligned} \frac{d A_-(\sigma_v)}{d \sigma_v} &\leq \frac{1}{2\sigma_v^2} \sup_{\sigma_v \in (0,2]} 3\sqrt{2} \frac{v_t^2 \sigma_v^2 + \sigma_v^4}{v_w^3} \left(\frac{v_t^4 + 6v_t^2 \sigma_v^2 + 3\sigma_v^4}{v_w^4} + \frac{\bar{b}_a}{\bar{a}} \right)^{-3/4} \\ &\quad - \frac{1}{2\sigma_v^2} \inf_{\sigma_v \in (0,2]} \sqrt{2} \left(\left(\frac{v_t^4 + 6v_t^2 \sigma_v^2 + 3\sigma_v^4}{v_w^4} + \frac{\bar{b}_a}{\bar{a}} \right)^{1/4} v_w - v_t \right) \\ &= \frac{1}{2\sigma_v^2} (1.2837 - 2.01) < 0, \end{aligned}$$

wobei das Supremum des ersten Terms für $\sigma_v = 2 \text{ m/s}$ und das Infimum des zweiten Terms für $\sigma_v \searrow 0 \text{ m/s}$ angenommen wird. Damit gilt für die Ableitung der stochastischen Nebenbedingung:

$$\begin{aligned} &\frac{d}{d \sigma_v} \mathbb{P}_{f_a(v_t + \hat{v})} \left(|f_a(v_t + \hat{v}) - \mathbb{E}(f_a(v_t + \hat{v}))| > \bar{b}_a \right) \\ &= \frac{1}{\sqrt{\pi}} \underbrace{\exp(A_+(\sigma_v))}_{>0} \underbrace{\frac{d A_+(\sigma_v)}{d \sigma_v}}_{\geq 0} - \frac{1}{\sqrt{\pi}} \underbrace{\exp(A_-(\sigma_v))}_{>0} \underbrace{\frac{d A_-(\sigma_v)}{d \sigma_v}}_{<0} > 0 \end{aligned}$$

und diese ist damit streng monoton steigend.

A.6 Optimierungsproblem für Beispiel 5.3.1

Für $k = m = 2$ wird das kontinuierliche Optimierungsproblem

$$\begin{aligned} \max_{\substack{\sigma_1 \in (0,1], \\ \sigma_2 \in (0,0.5]}} J(\sigma) &:= \max_{\substack{\sigma_1 \in (0,1], \\ \sigma_2 \in (0,0.5]}} 1/\sqrt{2} (\sigma_1 + \sigma_2) \\ \text{u. d. N. } \mathbb{P}_{X_i} (|x_i - \mathbb{E}(X_i)| > \bar{b}_i) &\leq \bar{p}_i \quad \forall i = 1, 2 \\ X_i &\sim \mathcal{N}(0, \sigma_i^2) \quad \forall i = 1, 2 \end{aligned}$$

definiert mit der oberen Schranke $\bar{\mathbf{b}} = (3, 0.329)^\top$ und den Wahrscheinlichkeiten $\bar{p}_1 = \bar{p}_2 = 0.1$. Da die ZV X_1, X_2 zentriert sind, gilt für $i = 1, 2$

$$\mathbb{P}_{X_i} (|x_i - \mathbb{E}(X_i)| > \bar{b}_i) = \mathbb{P} (|x_i| > \bar{b}_i) = \mathbb{P} (x_i < -\bar{b}_i) + \mathbb{P} (x_i > \bar{b}_i).$$

Mit $\bar{\mathbf{b}} = (3, 0.329)^\top \approx 1.64(1.83, 0.2)^\top$, dem 95 %-Quantil $z_{0.95} = 1.64$ der Normalverteilung und dem zulässigen Suchraum $(0, 1] \times (0, 0.2]$ folgt für die Standardabweichungen $\sigma = (1, 0.2)^\top$ die Wahrscheinlichkeit $\mathbb{P} (x_i < -\bar{b}_i) + \mathbb{P} (x_i > \bar{b}_i) \leq 0.05 + 0.05 = 0.1$, sodass die Standardabweichungen $\sigma^* = (1, 0.2)^\top$ das analytische Optimum des kontinuierlichen Problems sind. Mit der Chebychev-Ungleichung (3.7) kann aus der Abschätzung der stochastischen Nebenbedingung $\mathbb{P}_{X_i} (|x_i - \mathbb{E}(X_i)| > \bar{b}_i) \leq \text{var}(X_i)/\bar{b}_i^2 \stackrel{!}{=} 0.1$ die optimale Standardabweichung σ^* approximiert werden durch $\sigma_C^* = \sqrt{0.1}\bar{\mathbf{b}} = (0.9487, 0.1040)^\top$. Das zugehörige diskrete Optimierungsproblem für eine Stichprobe $\{\mathbf{x}_j\}_{j=1}^M$ der ZV \mathbf{X} ist mit $\lambda_{S,i}$ aus Gleichung (3.12) definiert durch

$$\begin{aligned} \max_{\substack{\sigma_1 \in (0,1], \\ \sigma_2 \in (0,0.5]}} J(\sigma) &:= \max_{\substack{\sigma_1 \in (0,1], \\ \sigma_2 \in (0,0.5]}} 1/\sqrt{2} (\sigma_1 + \sigma_2) \\ \text{u. d. N. } \frac{1}{M+1} g_{M+1} \left(\frac{M\lambda_{S,i}^2}{M-1+\lambda_{S,i}^2} \right) &\leq 0.1 \quad \forall i = 1, 2 \\ x_{j,i} &\sim \mathcal{N}(0, \sigma_i^2) \quad \forall j = 1, \dots, M, i = 1, 2. \end{aligned}$$

Dieses Optimierungsproblem kann mit dem im Folgenden vorgestellten MCS-Algorithmus 5.1 gelöst werden. Für $s_{\max} = 15$ und $\bar{l} = 8$ bricht dieser mit dem Ergebnis $\sigma_S^* = (0.89, 0.07)^\top$ ab. Zwischenergebnisse des Lösungsprozesses sind in Abb. 5.3 bis 5.5 dargestellt.

A.7 Optimierungsergebnisse für Ego-Position p, Beispiel 3.2.6.d

Abb. A.4 zeigt die Optimierungsergebnisse für die Position des Ego-Fahrzeugs während des Überholmanövers in Beispiel 3.2.6. Außer im Zeitraum [1.7 s, 2.0 s] sind die Optimierungsergebnisse größer als 40 m. Daher hat die Position keinen großen Einfluss auf die Berechnung der Beschleunigung. Dies stimmt mit der

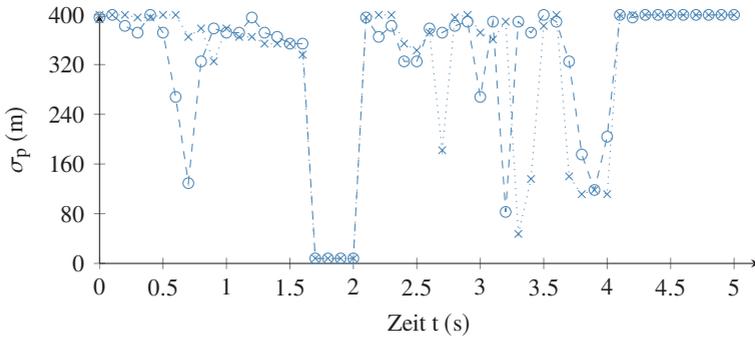


Abbildung A.4: Optimierungsergebnisse für Beispiel 3.2.6.d (Überholmanöver) für die Optimierung der Position p des Ego-Fahrzeugs. \times $\sigma_{S,p}^*$ Berechnung mit SA ohne DSC; \circ $\sigma_{S,p}^*$ Berechnung mit SA mit DSC.

Gleichung des IDM überein (vgl. Beispiel 3.1.4), da die Distanzen zu anderen Verkehrsteilnehmern unabhängig von der Ego-Position veranschlagt werden.

A.8 Maschinenlesbare Verkehrsregeln für die Textgenerierung

Tabelle A.1 gibt einen Überblick über die *if-else*-Regeln, die das AF in der Simulation einhalten musste. Wenn die Bedingung verletzt ist, dann muss es seine Entscheidungen an die entsprechende Regel anpassen.

Tabelle A.1: Verkehrsregeln, die das AF in der Simulation befolgen muss. Wenn die *if-else*-Bedingung erfüllt ist, dann verletzt das Ego diese Regel.

Regel	<i>if-else</i> -Bedingung
Überholverbot auf dem rechten Fahrstreifen	$v > v_{vl} > v_{krit} = 60 \text{ km/h}$
Geschwindigkeitsbegrenzung	$v > v_{max}$

Sicherheitsabstand zum vorausfahrenden Fahrzeug auf dem linken Fahrstreifen	$\rho v + 0.5 a_{\max} \rho^2 + 0.5 (v + \rho a_{\max})^2 / b_{\min} - 0.5 v_{vl}^2 / b_{\max} > d_{vl}$ (Shalev-Shwartz et al., 2017)
Sicherheitsabstand zum vorausfahrenden Fahrzeug auf demselben Fahrstreifen	$\rho v + 0.5 a_{\max} \rho^2 + 0.5 (v + \rho a_{\max})^2 / b_{\min} - 0.5 v_{slv}^2 / b_{\max} > d_{slv}$ (Shalev-Shwartz et al., 2017)
Sicherheitsabstand zum vorausfahrenden Fahrzeug auf dem rechten Fahrstreifen	$\rho v + 0.5 a_{\max} \rho^2 + 0.5 (v + \rho a_{\max})^2 / b_{\min} - 0.5 v_{rlv}^2 / b_{\max} > d_{rlv}$ (Shalev-Shwartz et al., 2017)
fahre möglichst schnell	$v < v_{\max}$
Rechtsfahrgebot	$v > v_{\text{urban}} \wedge$ existiert ein rechter Fahrstreifen?

A.9 Grundlagen, Vor- und Nachteile der erläuternden Aussagen

Tabelle A.2 gibt eine Übersicht über die Eigenschaften der Wichtige-Inputs-Aussage sowie der schlussfolgernden und regelbasierten Aussagen.

Tabelle A.2: Überblick über Aussagentypen: Literatur, Charakteristik und Beispiel für das Überholmanöver, vgl. Abb. 7.2 (angelehnt an Henze, Rufer et al. (2022)).

	Wichtige-Inputs-Aussage	Schlussfolgernde Aussage	Regelbasierte Aussage
Literatur	Binns et al. (2018), Meck und Precht (2021)	Koo et al. (2015), Lim et al. (2009), Meck und Precht (2021)	Koo et al. (2015), Lim et al. (2009), Meck und Precht (2021)
Informationen	Fahrentscheidung relevante Inputs aus der SA	Fahrentscheidung relevante Inputs und ihre Auswirkungen	Fahrentscheidung

		Verkehrsregeln, die verletzt werden	Verkehrsregeln, die verletzt werden
Unsicherheiten	nein	optional	optional
Vorteile	+ Umweltinformationen + keine Interpretation + hohe HMI Flexibilität	+ detaillierte Umweltinformationen + relevante Verkehrsregeln + nur abgesicherte Argumentation + kein zusätzliches Wissen notwendig	+ relevante Verkehrsregeln + nur abgesicherte Argumentation + kein zusätzliches Wissen notwendig + kurze Sätze
Nachteile	- wenige Details, keine Verkehrsregeln - benötigt zusätzliches Hintergrundwissen	- lange Sätze - (zu) viele Details	- hohes Abstraktionslevel - keine Details zu verarbeiteten Informationen

Literaturverzeichnis

- Ajanovic, Z., Lacevic, B., Shyrokau, B., Stolz, M. & Horn, M. (2018). Search-Based Optimal Motion Planning for Automated Driving. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (S. 4523–4530). doi: 10.1109/IROS.2018.8593813
- Allen, M. D. (2014). Telephone focus groups: Strengths, challenges, and strategies for success. *Qualitative Social Work: Research and Practice*, 13 (4), 571–583. doi: 10.1177/1473325013499060
- Amditis, A., Andreone, L., Pagle, K., Markkula, G., Derigibus, E., Romero Rue, M., Bellotti, F., Engelsberg, A., Brouwer, R., Peters, B. et al. (2010). Towards the Automotive HMI of the Future: Overview of the AIDE-Integrated Project Results. *IEEE Transactions on Intelligent Transportation Systems*, 11 (3), 567–578. doi: 10.1109/TITS.2010.2048751
- Audet, C., Custódio, A. L. & Dennis, J. E., Jr. (2008). Mesh Adaptive Direct Search Algorithms for Constrained Optimization. *SIAM Journal on Optimization*, 18 (4), 1501–1503. doi: 10.1137/060671267
- Ayoub, J., Zhou, F., Bao, S. & Yang, X. J. (2019). From Manual Driving to Automated Driving: A Review of 10 Years of AutoUI. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (S. 70–90). Utrecht, Niederlande: Association for Computing Machinery. doi: 10.1145/3342197.3344529
- Becker, J. (2022). *The automotive industry is driving toward software-defined vehicles – What if cars could be as easy to program as smartphones are today?* Aachen, Deutschland. (Vortrag beim 2022 IEEE Intelligent Vehicles Symposium (IV))
- Beggiato, M., Hartwicht, F., Schleinitz, K., Krems, J., Othersen, I. & Petermann-Stock, I. (2015). What would drivers like to know during automated driving? Information needs at different levels of automation. In *7. Tagung Fahrerassistenz*. München, Deutschland. doi: 10.13140/RG.2.1.2462.6007
- Bey, H., Dierkes, F., Bayerl, S., Lange, A., Faßbender, D. & Thielecke, J. (2019). Optimization-Based Tactical Behavior Planning for Autonomous

- Freeway Driving in Favor of the Traffic Flow. In *2019 IEEE Intelligent Vehicles Symposium (IV)* (S. 1033-1040). Paris, Frankreich. doi: 10.1109/IVS.2019.8813787
- Bey, H., Tratz, M., Sackmann, M., Lange, A. & Thielecke, J. (2020). Tutorial on Sampling-based POMDP-planning for Automated Driving. In K. Berns, M. Helfert & O. Gusikhin (Hrsg.), *VEHITS 2020 - Proceedings of the 6th International Conference on Vehicle Technology and Intelligent Transport Systems* (S. 312–321). Prag, Tschechien: SciTePress. doi: 10.5220/0009344703120321
- Bhattacharya, S., Kumar, V. & Likhachev, M. (2010). Search-based Path Planning with Homotopy Class Constraints. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Bd. 24, S. 1230-1237). Atlanta, USA.
- Binns, R., Van Kleek, M., Veale, M., Lyngs, U., Zhao, J. & Shadbolt, N. (2018). 'It's Reducing a Human Being to a Percentage': Perceptions of Justice in Algorithmic Decisions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (S. 1–14). New York, USA: Association for Computing Machinery. doi: 10.1145/3173574.3173951
- Boeije, H. (2002). A Purposeful Approach to the Constant Comparative Method in the Analysis of Qualitative Interviews. *Quality and Quantity*, 36 (4), 391–409. doi: 10.1023/A:1020909529486
- Borgo, R., Cashmore, M. & Magazzeni, D. (2018). *Towards Providing Explanations for AI Planner Decisions*. Zugriff am 24.01.2020 auf <http://arxiv.org/abs/1810.06338> (arXiv:1810.06338)
- Borovkov, A. A. (2013). *Probability Theory*. London, Vereinigtes Königreich: Springer. doi: 10.1007/978-1-4471-5201-9
- Brislin, R. W. (1970). Back-Translation for Cross-Cultural Research. *Journal of Cross-Cultural Psychology*, 1 (3), 185–216. doi: 10.1177/135910457000100301
- Brooks, R. A. (1986). A robust layered control system for a mobile robot. *IEEE Journal on Robotics and Automation*, 2 (1), 14–23. doi: 10.1109/JRA.1986.1087032
- Bundesministerium für Verkehr und digitale Infrastruktur. (2017). *Ethik-Kommission Automatisiertes und Vernetztes Fahren*. Zugriff am 15.08.2022 auf <https://www.bmvi.de/SharedDocs/DE/Publikationen/DG/bericht-der-ethik-kommission.pdf>

- Campolongo, F., Cariboni, J. & Saltelli, A. (2007). An effective screening design for sensitivity analysis of large models. *Environmental Modelling & Software*, 22 (10), 1509–1518. doi: 10.1016/j.envsoft.2006.10.004
- Campolongo, F. & Saltelli, A. (1997). Sensitivity analysis of an environmental model: an application of different analysis methods. *Reliability Engineering & System Safety*, 57 (1), 49–69. doi: 10.1016/S0951-8320(97)00021-5
- Campolongo, F., Saltelli, A. & Cariboni, J. (2011). From screening to quantitative sensitivity analysis. A unified approach. *Computer Physics Communications*, 182 (4), 978–988. doi: 10.1016/j.cpc.2010.12.039
- Campolongo, F., Tarantola, S. & Saltelli, A. (1999). Tackling quantitatively large dimensionality problems. *Computer Physics Communications*, 117 (1), 75–85. doi: 10.1016/S0010-4655(98)00165-9
- de Groot, O., Brito, B., Ferranti, L., Gavrila, D. & Alonso-Mora, J. (2021). Scenario-Based Trajectory Optimization in Uncertain Dynamic Environments. *IEEE Robotics and Automation Letters*, 6 (3), 5389–5396. doi: 10.1109/LRA.2021.3074866
- Doran, D., Schulz, S. & Besold, T. R. (2017). *What Does Explainable AI Really Mean? A New Conceptualization of Perspectives*. Zugriff am 24.02.2020 auf <http://arxiv.org/pdf/1710.00794v1> (arXiv:1710.00794)
- Dudenredaktion. (2022). *erklären*. Zugriff am 18.08.2022 auf <https://www.duden.de/rechtschreibung/erklaeren>
- Eder, M., Skibinski, S., Mickler, F. & Ulbrich, M. (2022). A Rigorous Optimization Approach for the Generic Derivation of Secondary Information from Digital Maps for Urban Scenarios. In *2022 25th IEEE International Conference on Intelligent Transportation Systems (ITSC)*. Macau, China. (akzeptiert)
- Esterle, K., Gressenbuch, L. & Knoll, A. (2020). Formalizing Traffic Rules for Machine Interpretability. In *2020 IEEE 3rd Connected and Automated Vehicles Symposium (CAVS)* (S. 1–7). Victoria, Canada. doi: 10.1109/CAVS51000.2020.9334599
- Europäische Kommission. (2021). *Vorschlag für eine Verordnung des europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Gesetz über künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union*. Zugriff am 18.08.2022 auf <https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A52021PC0206>

- Faßbender, D., Heinrich, B. C. & Wünsche, H.-J. (2016). Motion planning for autonomous vehicles in highly constrained urban environments. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (S. 4708-4713). Daejeon, Südkorea. doi: 10.1109/IROS.2016.7759692
- Feinauer, S., Schuller, L., Groh, I., Huestegge, L. & Petzoldt, T. (2022). The potential of gamification for user education in partial and conditional driving automation: A driving simulator study. *Transportation Research Part F: Traffic Psychology and Behaviour*, 90, 252-268. doi: 10.1016/j.trf.2022.08.009
- Fox, M., Long, D. & Magazzeni, D. (2017). Explainable Planning. In *Proceedings of IJCAI-17 Workshop on Explainable AI (XAI)* (S. 24–31). Melbourne, Australien.
- Fraikin, N., Funk, K., Frey, M. & Gauterin, F. (2019). Dimensionality reduction and identification of valid parameter bounds for the efficient calibration of automated driving functions. *Automotive and Engine Technology*, 4, 75–91. doi: 10.1007/s41104-019-00043-z
- Fraikin, N., Funk, K., Frey, M. & Gauterin, F. (2020). Efficient Simulation Based Calibration of Automated Driving Functions Based on Sensitivity Based Optimization. *IEEE Open Journal of Intelligent Transportation Systems*, 1, 63–79. doi: 10.1109/OJITS.2020.3001801
- Frandi, E. & Papini, A. (2014). Coordinate Search Algorithms in Multilevel Optimization. *Optimization Methods and Software*, 29 (5), 1020–1041.
- Gatt, A. & Krahmer, E. (2018). Survey of the State of the Art in Natural Language Generation: Core tasks, applications and evaluation. *Journal of Artificial Intelligence Research (JAIR)*, 61, 65–170. doi: 10.1613/jair.5477
- Ge, Q. & Menendez, M. (2017). Extending Morris method for qualitative global sensitivity analysis of models with dependent inputs. *Reliability Engineering & System Safety*, 162, 28–39. doi: 10.1016/j.res.2017.01.010
- Goldman, C. V. & Bustin, R. (2022). Trusting Explainable Autonomous Driving: Simulated Studies. In *2022 IEEE Intelligent Vehicles Symposium (IV)* (S. 1255–1260). Aachen, Deutschland. doi: 10.1109/IV51971.2022.9827312
- Google Maps, GeoBasis-DE/BKG, GeoContent & Maxar Technologies. (2022). *Kreisverkehr Gaimersheim*. Gaimersheim, Deutschland. Zugriff am 15.02.2022 auf <https://goo.gl/maps/8uKAtUmSnQXQ3NhX7>

- Hartwich, F., Schmidt, C., Gräfung, D. & Krems, J. F. (2020). In the Passenger Seat: Differences in the Perception of Human vs. Automated Vehicle Control and Resulting HMI Demands of Users. In H. Krömker (Hrsg.), *HCI in Mobility, Transport, and Automotive Systems. Automated Driving and In-Vehicle Experience Design* (Bd. 12212, S. 31–45). Cham, Deutschland: Springer International Publishing. doi: 10.1007/978-3-030-50523-3_3
- Henning, M., Müller, J. C., Gies, F., Buchholz, M. & Dietmayer, K. (2022). Situation-Aware Environment Perception Using a Multi-Layer Attention Map. *IEEE Transactions on Intelligent Vehicles*. doi: 10.1109/TIV.2022.3164236
- Henze, F., Faßbender, D. & Stiller, C. (2020). Sensitivity Analysis of a Planning Algorithm Considering Uncertainties. In *2020 IEEE Intelligent Vehicles Symposium (IV)* (S. 1128–1134). Las Vegas, USA. doi: 10.1109/IV47402.2020.9304575
- Henze, F., Faßbender, D. & Stiller, C. (2022). How Can Automated Vehicles Explain Their Driving Decisions? Generating Clarifying Summaries Automatically. In *2022 IEEE Intelligent Vehicles Symposium (IV)* (S. 935–942). Aachen, Deutschland. doi: 10.1109/IV51971.2022.9827197
- Henze, F., Faßbender, D. & Stiller, C. (2021). Identifying Admissible Uncertainty Bounds for the Input of Planning Algorithms. *IEEE Transactions on Intelligent Vehicles*. doi: 10.1109/TIV.2021.3119352
- Henze, F., Rußer, M., Faßbender, D., Stiller, C. & Schneider, S.-A. (2022). Dynamic Sampling Strategy for Morris' Method of Elementary Effects. In *10th International Conference on Sensitivity Analysis of Model Output* (S. 14–18). Tallahassee, USA.
- Henze, F., Stasinski, N. M., Faßbender, D. & Stiller, C. (2021). Developers' Information Needs during Test Drives with Automated Vehicles in Real Traffic: A Focus Group Study. In *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '21 Adjunct)* (S. 81–85). Leeds, Vereinigtes Königreich: Association for Computing Machinery. doi: 10.1145/3473682.3480269
- Hoel, C.-J., Driggs-Campbell, K., Wolff, K., Laine, L. & Kochenderfer, M. J. (2020). Combining Planning and Deep Reinforcement Learning in Tactical Decision Making for Autonomous Driving. *IEEE Transactions on Intelligent Vehicles*, 5 (2), 294–305. doi: 10.1109/TIV.2019.2955905
- Hoffmann, R. R., Mueller, S. T., Klein, G. & Litman, J. (2019). *Metrics for Explainable AI: Challenges and Prospects*. Zugriff am 17.08.2021 auf <https://arxiv.org/abs/1812.04608> (arXiv:1812.04608)

- Hubmann, C., Becker, M., Althoff, D., Lenz, D. & Stiller, C. (2017). Decision making for autonomous driving considering interaction and uncertain prediction of surrounding vehicles. In *2017 IEEE Intelligent Vehicles Symposium (IV)* (S. 1671–1678). Redondo Beach, USA. doi: 10.1109/IVS.2017.7995949
- Hubmann, C., Schulz, J., Becker, M., Althoff, D. & Stiller, C. (2018). Automated Driving in Uncertain Environments: Planning With Interaction and Uncertain Maneuver Prediction. *IEEE Transactions on Intelligent Vehicles*, 3 (1), 5–17. doi: 10.1109/TIV.2017.2788208
- Huyer, W. & Neumaier, A. (1999). Global Optimization by Multilevel Coordinate Search. *Journal of Global Optimization*, 14, 331–355. doi: 10.1023/A:1008382309369
- International Organization for Standardization. (1994). *Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions (ISO 5725-1:1994)*. Zugriff am 11.08.2022 auf <https://www.iso.org/obp/ui/#iso:std:iso:5725:-1:ed-1:v1:en>
- Iooss, B. & Lemaitre, P. (2015). A review on global sensitivity analysis methods. In C. Meloni & G. Dellino (Hrsg.), *Uncertainty management in Simulation-Optimization of Complex Systems: Algorithms and Applications* (S. 101–122). Boston, USA: Springer. doi: 10.1007/978-1-4899-7547-8_5
- Jung, M. F., Sirkin, D., Gür, T. M. & Steinert, M. (2015). Displayed Uncertainty Improves Driving Experience and Behavior: The Case of Range Anxiety in an Electric Car. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (S. 2201–2210). New York, USA: Association for Computing Machinery. doi: 10.1145/2702123.2702479
- Kesting, A., Treiber, M. & Helbing, D. (2007). General Lane-Changing Model MOBIL for Car-Following Models. *Transportation Research Record: Journal of the Transportation Research Board*, 1999 (1), 86–94. doi: 10.3141/1999-10
- Kim, J., Rohrbach, A., Darrell, T., Canny, J. & Akata, Z. (2018). Textual Explanations for Self-Driving Vehicles. In V. Ferrari, M. Hebert, C. Sminchisescu & Y. Weiss (Hrsg.), *Computer Vision – ECCV 2018* (Bd. 11206, S. 577–593). Cham, Deutschland: Springer International Publishing. doi: 10.1007/978-3-030-01216-8_35
- Klenke, A. (2013). *Wahrscheinlichkeitstheorie* (3. Aufl.). Berlin, Heidelberg, Deutschland: Springer-Verlag. doi: 10.1007/978-3-642-36018-3

- Kolonko, M. (2008). *Stochastische Simulation: Grundlagen, Algorithmen und Anwendungen*. Wiesbaden, Deutschland: Vieweg + Teubner. doi: 10.1007/978-3-8348-9290-4
- Koo, J., Kwac, J., Ju, W., Steinert, M., Leifer, L. & Nass, C. (2015). Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. *International Journal on Interactive Design and Manufacturing (IJDeM)*, 9 (4), 269–275. doi: 10.1007/s12008-014-0227-2
- Korpan, R. & Epstein, S. L. (2018). Toward Natural Explanations for a Robot's Navigation Plans. In *Notes from the Explainable Robotic Systems Workshop, Human-Robot Interaction HRI'18* (S. 1-3). Chicago, USA.
- Kortenkamp, D. & Simmons, R. (2008). Robotic Systems Architectures and Programming. In B. Siciliano & O. Khatib (Hrsg.), *Springer Handbook of Robotics* (S. 187–206). Berlin, Heidelberg, Deutschland: Springer. doi: 10.1007/978-3-540-30301-5_9
- Krueger, R. A. & Casey, M. A. (2014). *Focus Groups. A practical Guide for Applied Research* (5. Aufl.). Thousand Oaks, USA: Sage Publications Inc.
- Krämer, S., Stiller, C. & Bouzouraa, M. E. (2018). LiDAR-Based Object Tracking and Shape Estimation Using Polylines and Free-Space Information. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (S. 4515–4522). Madrid, Spanien. doi: 10.1109/IROS.2018.8593385
- Kucherenko, S., Rodriguez-Fernandez, M., Pantelides, C. & Shah, N. (2009). Monte Carlo evaluation of derivative-based global sensitivity measures. *Reliability Engineering & System Safety*, 94 (7), 1135–1148. doi: 10.1016/j.res.2008.05.006
- Kyriakidis, M., de Winter, J. C. F., Stanton, N., Bellet, T., van Arem, B., Brookhuis, K., Martens, M. H., Bengler, K., Andersson, J., Merat, N. et al. (2019). A human factors perspective on automated driving. *Theoretical Issues in Ergonomics Science*, 20 (3), 223–249. doi: 10.1080/1463922X.2017.1293187
- Lejeune, M. A., Margoty, F. & Oliveira, A. D. (2018). Chance-Constrained Programming with Decision-Dependent Uncertainty. *Proceedings of the Workshop New Directions in Stochastic Optimisation*, 31 (38), 1–44. doi: 10.2139/ssrn.3201121
- Li, P., Arellano-Garcia, H. & Wozny, G. (2008). Chance constrained programming approach to process optimization under uncertainty.

- Computers & Chemical Engineering*, 32 (1), 25–45. doi: 10.1016/j.compchemeng.2007.05.009
- Lim, B. Y., Dey, A. K. & Avrahami, D. (2009). Why and Why Not Explanations Improve the Intelligibility of Context-Aware Intelligent Systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (S. 2119–2128). New York, USA: Association for Computing Machinery. doi: 10.1145/1518701.1519023
- Maierhofer, S., Rettinger, A.-K., Mayer, E. C. & Althoff, M. (2020). Formalization of Interstate Traffic Rules in Temporal Logic. In *2020 IEEE Intelligent Vehicles Symposium (IV)* (S. 752–759). Las Vegas, USA. doi: 10.1109/IV47402.2020.9304549
- Marti, E., de Miguel, M. A., Garcia, F. & Perez, J. (2019). A Review of Sensor Technologies for Perception in Automated Driving. *IEEE Intelligent Transportation Systems Magazine*, 11 (4), 94–108. doi: 10.1109/MITS.2019.2907630
- Meck, A.-M. & Precht, L. (2021). How to Design the Perfect Prompt: A Linguistic Approach to Prompt Design in Automotive Voice Assistants – An Exploratory Study. In *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '21 Adjunct)* (S. 237–246). New York, NY, USA: Association for Computing Machinery. doi: 10.1145/3409118.3475144
- Menberg, K., Heo, Y., Augenbroe, G. & Choudhary, R. (2016). New Extension of Morris Method for Sensitivity Analysis of Building Energy Models. In *Proceedings of BSO Conference 2016: Third Conference of IBPSA England* (Bd. 3, S. 89–96). Newcastle, Vereinigtes Königreich: IBPSA England.
- Mishra, L. (2016). Focus Group Discussion in Qualitative Research. *Techno-Learn: An International Journal of Educational Technology*, 6 (1), 1-5. doi: 10.5958/2249-5223.2016.00001.2
- Montavon, G., Bach, S., Binder, A., Samek, W. & Müller, K.-R. (2017). Explaining Nonlinear Classification Decisions with Deep Taylor Decomposition. *Pattern Recognition*, 65 (C), 211–222. doi: 10.1016/j.patcog.2016.11.008
- Morris, M. D. (1991). Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics*, 33 (2), 161–174. doi: 10.2307/1269043
- Nahata, R., Omeiza, D., Howard, R. & Kunze, L. (2021). Assessing and Explaining Collision Risk in Dynamic Environments for Autonomous Driving Safety. In *2021 IEEE International Intelligent Transportation*

- Systems Conference (ITSC)* (S. 223–230). Indianapolis, USA. doi: 10.1109/ITSC48978.2021.9564966
- Nolte, M., Ernst, S., Richelmann, J. & Maurer, M. (2018). Representing the Unknown - Impact of Uncertainty on the Interaction between Decision Making and Trajectory Generation. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)* (S. 2412–2418). doi: 10.1109/ITSC.2018.8569490
- OLG Hamm. (2012). *Urteil vom 17.07.2012 - 9 U 200/11*. Zugriff am 02.02.2022 auf <https://openjur.de/u/654453.html>
- Omeiza, D., Kollnig, K., Web, H., Jirotko, M. & Kunze, L. (2021). Why Not Explain? Effects of Explanations on Human Perceptions of Autonomous Driving. In *2021 IEEE International Conference on Advanced Robotics and Its Social Impacts (ARSO)* (S. 194–199). Tokoname, Japan. doi: 10.1109/ARSO51874.2021.9542835
- Omeiza, D., Webb, H., Jirotko, M. & Kunze, L. (2021). Towards Accountability: Providing Intelligible Explanations in Autonomous Driving. In *2021 IEEE Intelligent Vehicles Symposium (IV)* (S. 231–237). Nagoya, Japan. doi: 10.1109/IV48863.2021.9575917
- Omeiza, D., Webb, H., Jirotko, M. & Kunze, L. (2022). Explanations in Autonomous Driving: A Survey. *IEEE Transactions on Intelligent Transportation Systems*, 23 (8), 10142–10162. doi: 10.1109/TITS.2021.3122865
- On-Road Automated Driving (ORAD) Committee. (2021). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*. Zugriff am 07.09.2022 auf https://www.sae.org/content/j3016_202104 doi: 10.4271/J3016_202104
- Ossig, J. & Cramer, S. (2020). Tactical Decisions for Lane Changes or Lane Following? Development of a Study Design for Automated Driving. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (S. 23–26). New York, USA: Association for Computing Machinery. doi: 10.1145/3409251.3411714
- Ossig, J., Cramer, S., Eckl, A. & Bengler, K. (2022). Tactical Decisions for Lane Changes or Lane Following: Assessment of Automated Driving Styles under Real-World Conditions. *IEEE Transactions on Intelligent Vehicles*, 1–10. doi: 10.1109/TIV.2022.3161570
- Paden, B., Čáp, M., Yong, S. Z., Yershov, D. & Frazzoli, E. (2016). A Survey of Motion Planning and Control Techniques for Self-Driving Urban

- Vehicles. *IEEE Transactions on Intelligent Vehicles*, 1 (1), 33–55. doi: 10.1109/TIV.2016.2578706
- Pereira, R., Couto, M., Ribeiro, F., Rua, R., Cunha, J., Fernandes, J. P. & Saraiva, J. (2017). Energy Efficiency across Programming Languages: How Do Energy, Time, and Memory Relate? In *Proceedings of the 10th ACM SIGPLAN International Conference on Software Language Engineering* (S. 256–267). Vancouver, Kanada: Association for Computing Machinery. doi: 10.1145/3136014.3136031
- Perera, R. & Nand, P. (2017). Recent Advances in Natural Language Generation: A Survey and Classification of the Empirical Literature. *Computing and Informatics*, 36 (1), 1–32. doi: 10.4149/cai_2017_1_1
- Pew Research Center. (2017). *Automation in Everyday Life*. (durchgeführt von Smith, A. & Anderson, M.)
- Philipp, R., Qian, H., Hartjen, L., Schuldt, F. & Howar, F. (2021). Simulation-Based Elicitation of Accuracy Requirements for the Environmental Perception of Autonomous Vehicles. In T. Margaria & B. Steffen (Hrsg.), *Leveraging Applications of Formal Methods, Verification and Validation* (Bd. 13036, S. 129–145). Cham, Deutschland: Springer International Publishing. doi: 10.1007/978-3-030-89159-6_9
- Piazzoni, A., Cherian, J., Slavik, M. & Dauwels, J. (2021). Modeling Perception Errors towards Robust Decision Making in Autonomous Vehicles. In *Proceedings of the 29th International Joint Conference on Artificial Intelligence* (S. 3494–3500). Yokohama, Japan. doi: 10.5555/3491440.3491923
- Pošík, P., Huyer, W. & Pál, L. (2012). A Comparison of Global Search Algorithms for Continuous Black Box Optimization. *Evolutionary Computation*, 20 (4), 509–541. doi: 10.1162/EVCO_a_00084
- Reid, T. G., Houts, S. E., Cammarata, R., Mills, G., Agarwal, S., Vora, A. & Pandey, G. (2019). Localization Requirements for Autonomous Vehicles. *SAE International Journal of Connected and Automated Vehicles*, 2 (3), 173–190. doi: 10.4271/12-02-03-0012
- Reinhardt, R., Hoffmann, A. & Gerlach, T. (2013). *Nichtlineare Optimierung*. Berlin, Heidelberg, Deutschland: Springer Spektrum. doi: 10.1007/978-3-8274-2949-0
- Rizaldi, A., Keinholtz, J., Huber, M., Feldle, J., Immler, F., Althoff, M., Hilgendorf, E. & Nipkow, T. (2017). Formalising and Monitoring Traffic Rules for Autonomous Vehicles in Isabelle/HOL. In N. Polikarpova & S. Schneider (Hrsg.), *Integrated Formal Methods* (Bd. 10510, S.

- 50–66). Cham, Deutschland: Springer International Publishing. doi: 10.1007/978-3-319-66845-1_40
- Rohde, J., Stellet, J. E., Mielenz, H. & Zöllner, J. M. (2015). Model-Based Derivation of Perception Accuracy Requirements for Vehicle Localization in Urban Environments. In *2015 IEEE 18th International Conference on Intelligent Transportation Systems (ITSC)* (S. 712–718). Kanarische Inseln, Spanien. doi: 10.1109/ITSC.2015.121
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1 (5), 206–215. doi: 10.1038/s42256-019-0048-x
- Rußer, M. (2021). *Screening-Strategien für die Sensitivitätsanalyse von automatisierten Fahrfunktionen* (Masterarbeit). Hochschule für angewandte Wissenschaften Kempten, Fakultät Elektrotechnik, Kempten, Deutschland.
- Rutjens, R. J. L., Band, L. R., Jones, M. D. & Owen, M. R. (2022). Elementary Effects for Models with Dimensional Inputs of Arbitrary Type and Range: Scaling and Trajectory Generation. *SSRN Electronic Journal*. doi: 10.2139/ssrn.4102825
- Ryan, K. E., Gandha, T., Culbertson, M. J. & Carlson, C. (2014). Focus Group Evidence: Implications for Design and Analysis. *American Journal of Evaluation*, 35 (3), 328–345. doi: 10.1177/1098214013508300
- Sackmann, M., Bey, H., Hofmann, U. & Thielecke, J. (2020). Prediction Error Reduction of Neural Networks for Car-Following Using Multi-Step Training. In *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)* (S. 1-7). Rhodos, Griechenland. doi: 10.1109/ITSC45102.2020.9294646
- Sackmann, M., Leemann, T., Bey, H., Hofmann, U. & Thielecke, J. (2021). Multi-Step Training for Predicting Roundabout Traffic Situations. *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, 1982-1989. doi: 10.1109/ITSC48978.2021.9564547
- Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M. & Tarantola, S. (2010). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181 (2), 259–270. doi: 10.1016/j.cpc.2009.09.018
- Samek, W. & Müller, K.-R. (2019). Towards Explainable Artificial Intelligence. In W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen & K.-R. Müller (Hrsg.), *Explainable AI: Interpreting, Explaining and Visualizing*

- Deep Learning* (S. 5–22). Cham, Deutschland: Springer International Publishing. doi: 10.1007/978-3-030-28954-6_1
- Saw, J. G., Yang, M. C. K. & Chin Mo, T. (1984). Chebyshev Inequality with Estimated Mean and Variance. *The American Statistician*, 38 (2), 130–132. doi: <https://doi.org/10.2307/2683249>
- Schmidt, L. M., Kontes, G., Plinge, A. & Mutschler, C. (2021). Can You Trust Your Autonomous Car? Interpretable and Verifiably Safe Reinforcement Learning. In *2021 IEEE Intelligent Vehicles Symposium (IV)* (S. 171–178). Nagoya, Japan. doi: 10.1109/IV48863.2021.9575328
- Sedol, L. (2016a). *Google's AI beats world Go champion in first of five matches*. Zugriff am 15.08.2022 auf <https://www.bbc.com/news/technology-35761246>
- Sedol, L. (2016b). *Google's AI Wins Pivotal Second Game in Match With Go Grandmaster*. Zugriff am 15.08.2022 auf <https://www.wired.com/2016/03/googles-ai-wins-pivotal-game-two-match-go-grandmaster/>
- Shalev-Shwartz, S., Shammah, S. & Shashua, A. (2017). *On a Formal Model of Safe and Scalable Self-driving Cars*. Zugriff am 20.08.2019 auf <https://arxiv.org/abs/1708.06374> (arXiv:1708.06374)
- Shariff, A., Bonnefon, J.-F. & Rahwan, I. (2021). How safe is safe enough? Psychological mechanisms underlying extreme safety demands for self-driving cars. *Transportation Research Part C: Emerging Technologies*, 126, 1–12. doi: 10.1016/j.trc.2021.103069
- Sin, G. & Gernaey, K. V. (2009). Improving the Morris method for sensitivity analysis by scaling the elementary effects. In J. Jeżowski & J. Thullie (Hrsg.), *19th European Symposium on Computer Aided Process Engineering* (Bd. 26, S. 925-930). Krakau, Polen: Elsevier. doi: 10.1016/S1570-7946(09)70154-3
- Stoter, A., Cornelisse, E. & Dalmolen, S. (2011). Context-aware prioritization of information: an architecture for real-time in-vehicle information management. In *Proceedings of the 8th European Congress and Exhibition on Intelligent Transport Systems and Services (ITS)* (S. 1–10). Lyon, Frankreich.
- StVO. (2013). *Straßenverkehrs-Ordnung vom 6. März 2013 (BGBl. I S. 367), die zuletzt durch Artikel 13 des Gesetzes vom 12. Juli 2021 (BGBl. I S. 3091) geändert worden ist*.
- Stöckle, C., Utschick, W., Herrmann, S. & Dirndorfer, T. (2018). Robust Design of an Automatic Emergency Braking System Considering Sensor Measurement Errors. In *2018 21st International Conference on Intelligent*

- Transportation Systems (ITSC)* (S. 2018–2023). Maui, USA. doi: 10.1109/ITSC.2018.8569495
- Stöckle, C., Utschick, W., Herrmann, S. & Dirndorfer, T. (2019). Robust Function and Sensor Design Considering Sensor Measurement Errors Applied to Automatic Emergency Braking. In *2019 IEEE Intelligent Vehicles Symposium (IV)* (S. 2284–2290). Paris, Frankreich. doi: 10.1109/IVS.2019.8814142
- Sweet, C. (2001). Designing and conducting virtual focus groups. *Qualitative Market Research: An International Journal*, 4 (3), 130–135. doi: 10.1108/13522750110393035
- Tas, Ö. S., Hauser, F. & Stiller, C. (2018). Decision-Time Postponing Motion Planning for Combinatorial Uncertain Maneuvering. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)* (S. 2419–2425). Maui, USA. doi: 10.1109/ITSC.2018.8569580
- Tchébychef, P. L. & Khanikof, N. (1867). The mean values (Des valeurs moyennes; Traduction du russe). *Journal de Mathématiques Pures et Appliquées*, 2 (12), 177–184.
- Touzani, S. & Busby, D. (2014). Screening Method Using the Derivative-based Global Sensitivity Indices with Application to Reservoir Simulator. *Oil & Gas Science and Technology – Revue d’IFP Energies nouvelles*, 69 (4), 619–632. doi: 10.2516/ogst/2013195
- Treiber, M., Hennecke, A. & Helbing, D. (2000). Congested Traffic States in Empirical Observations and Microscopic Simulations. *Physical Review E*, 62 (2), 1805–1824. doi: 10.1103/PhysRevE.62.1805
- Tsang, S., Royse, C. & Terkawi, A. (2017). Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. *Saudi Journal of Anaesthesia*, 11 (5), 80–89. doi: 10.4103/sja.SJA_203_17
- Turney, L. & Pocknee, C. (2005). Virtual Focus Groups: New Frontiers in Research. *International Journal of Qualitative Methods*, 4 (2), 32–43. doi: 10.1177/160940690500400203
- van der Bles, A. M., van der Linden, S., Freeman, A. L. J. & Spiegelhalter, D. J. (2020). The effects of communicating uncertainty on public trust in facts and numbers. *Proceedings of the National Academy of Sciences*, 117 (14), 7672–7683. doi: 10.1073/pnas.1913678117
- Vogel, P. & Althoff, M. (2020). Rekonstruktion von durch vollautomatisierte Fahrzeuge verursachten Verkehrsunfällen. *InTeR – Zeitschrift zum Innovations- und Technikrecht*, 8 (2), 89–94.

- Vogl, C., Sackmann, M., Kürzinger, L. & Hofmann, U. (2020). Frenet Coordinate Based Driving Maneuver Prediction at Roundabouts Using LSTM Networks. In *Computer Science in Cars Symposium* (S. 1–9). Feldkirchen, Deutschland: Association for Computing Machinery. doi: 10.1145/3385958.3430475
- Wald, P., Haentjes, J., Albert, M., Cramer, S. & Bengler, K. (2021). Active Vehicle Motion as Feedback during Different Levels of Automation. In *2021 24th IEEE International Intelligent Transportation Systems Conference (ITSC)* (S. 1713–1720). Indianapolis, USA. doi: 10.1109/ITSC48978.2021.9564982
- Wald, P., Henreich, N., Albert, M., Ossig, J. & Bengler, K. (2022). Different feedback strategies: Evaluation of active vehicle motions in a multi-level system. In *13th International Conference on Applied Human Factors and Ergonomics (AHFE 2022)* (S. 367–375). New York, USA. doi: 10.54941/ahfe1002468
- Wang, N., Pynadath, D. V. & Hill, S. G. (2016). The Impact of POMDP-Generated Explanations on Trust and Performance in Human-Robot Teams. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems* (S. 997–1005). Singapur, Singapur: International Foundation for Autonomous Agents and Multiagent Systems. doi: 10.5555/2936924.2937071
- Xu, W., Pan, J., Wei, J. & Dolan, J. M. (2014). Motion Planning under Uncertainty for On-Road Autonomous Driving. In *2014 IEEE International Conference of Robotics and Automation (ICRA)* (S. 2507–2512). Hong Kong, China. doi: 10.1109/ICRA.2014.6907209
- Xu, Y., Yang, X., Gong, L., Lin, H.-C., Wu, T.-Y., Li, Y. & Vasconcelos, N. (2020). Explainable Object-Induced Action Decision for Autonomous Vehicles. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (S. 9520–9529). Seattle, USA. doi: 10.1109/CVPR42600.2020.00954
- You, C., Lu, J., Filev, D. & Tsiotras, P. (2019). Advanced planning for autonomous vehicles using reinforcement learning and deep inverse reinforcement learning. *Robotics and Autonomous Systems*, 114, 1–18. doi: 10.1016/j.robot.2019.01.003
- Zhou, Y. & Danks, D. (2020). Different „Intelligibility“ for Different Folks. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (S. 194–199). New York, USA: Association for Computing Machinery. doi: 10.1145/3375627.3375810