



Schlussbericht

Verbund: 05M2020 - SOPRANN

Zuwendungsempfänger:	Technische Universität Chemnitz
Projektleitung:	Prof. Dr. habil. Oliver Ernst
E-Mail:	oernst@math.tu-chemnitz.de
Förderkennzeichen:	05M20OCA
Förderzeitraum:	01.04.2020 - 30.09.2023
Zuwendung:	550.622,19 €
Projektträger:	Projektträger DESY
Zusätzlicher Kontakt:	jan.blechschmidt@mathematik.tu-chemnitz.de
Zusätzlicher Name:	Jan Blechschmidt

Genutzte Großgeräte:	Labor	Gerät	Experiment
Diplomarbeiten:	0		
Dissertationen:	1		
Habilitationen:	0		
Referierte Publikationen:	9		
Andere Veröffentlichungen:	1		
Patente:	0		
Bachelorarbeiten:	0		
Masterarbeiten:	3		
Staatsexamen:	0		

Dieser Bericht wurde beim Projektträger über einen individuellen Online-Zugang vom Projektleiter eingereicht und am 28.06.2024 10:05 für eine Veröffentlichung freigegeben.

Schlussbericht

zum

Verbundprojekt 05M2020 - SOPRANN: Synthese optimaler Regelungen und adaptiver Neuronaler Netze für Mobilitätsanwendungen. (Förderkennzeichen 05M20OCA)

Teilprojekt 2: Optimierungs- und Analysemethoden für NN

Teilprojektleitung:

Prof. Dr. Oliver Ernst, Professur für Numerische Mathematik

Fakultät für Mathematik,

Technische Universität Chemnitz, 09107 Chemnitz

Email: `oliver.ernst@math.tu-chemnitz.de`.

Teilprojekt 3: Stabilität und Adaptation von NN-basierten Regelungen

Teilprojektleitung:

Prof. Dr.-Ing. Stefan Streif, Professur für Regelungstechnik und Systemdynamik

Fakultät für Elektrotechnik und Informationstechnik,

Technische Universität Chemnitz, 09107 Chemnitz

Email: `stefan.streif@etit.tu-chemnitz.de`.

1 Zielsetzung

Der Einfluss von Methoden des maschinellen Lernens in technischen Anwendungen und speziell beim autonomen Fahren nimmt seit Jahren rapide zu mit dem Versprechen, komplexe Situationen sicher beherrschen zu können. Insbesondere im Bereich der maschinellen Wahrnehmung sind Deep Neural Networks inzwischen quasi zum Standard geworden. Im Bereich der Regelung finden aktuell ebenfalls sehr viele Forschungsaktivitäten statt, wobei die Absicherung der Funktionalität von lernenden Reglern immer noch ein im Allgemeinen ungelöstes Problem darstellt.

Hier setzte das Verbundprojekt SOPRANN (Synthese optimaler Regelungen und adaptiver Neuronaler Netze für Mobilitätsanwendungen) an mit dem Gesamtziel, datenbasierte Modelle und künstliche Intelligenz (KI) zur sicheren und optimalen Regelung im Bereich der Mobilität einzusetzen. Eine Herausforderung dabei war der Umgang mit der Komplexität, der Multimodalität und der Menge der Messdaten, die mit Hilfe von Neuronalen Netzen (NN) synthetisiert werden sollten. Die Auswahl geeigneter Netzarchitekturen und die Entwicklung von effizienten Trainingsverfahren sowie die Echtzeit-Adaptation der NN basierend auf neuen, während der Laufzeit gewonnenen Daten bei gleichzeitiger Gewährleistung von Stabilität, Optimalität und Sicherheit des Regelkreises – bildeten den mathematischen Forschungskern des Projekts.

Projektziele im Rahmen des Teilprojekts 2 – *Optimierungs- und Analysemethoden für NN* – des Verbundprojekts SOPRANN, welches von der Arbeitsgruppe Numerische Mathematik der Fakultät für Mathematik der TU Chemnitz durchgeführt wurde, waren unter anderem die Untersuchung von Methoden der Unsicherheitsquantifizierung für NN sowie die Entwicklung von neuen Berechnungsmethoden zur Lösung der für Optimalsteuerung dynamischer Systeme entscheidenden Hamilton-Jacobi-Bellman (HJB) Gleichung.

Das Teilprojekt 3 – *Stabilität und Adaptation von NN-basierten Regelungen* – durchgeführt von der Professur Regelungstechnik und Systemdynamik, hatte das Ziel, lernende Regler für Optimalsteuerungsprobleme mit unendlichem Horizont mit gesicherter Stabilität des geschlossenen Regelkreises und unter Berücksichtigung von Sicherheitseinschränkungen zu entwerfen und zu untersuchen.

Diese Arbeiten erfolgten in regelmäßigem Austausch mit den Projektpartnern von der Universität der Bundeswehr München (Arbeitsgruppe Ingenieurmathematik der Fakultät für Luft- und Raumfahrttechnik) sowie vom Fraunhofer ITWM Kaiserslautern (Abteilung Dynamik, Lasten und Umgebungsdaten des Bereichs Mathematik für die Fahrzeugentwicklung).

2 Zahlenmäßiger Nachweis

Während der Projektlaufzeit vom 1.4.2020 bis zum 30.9.2023 entstanden an der TU Chemnitz in Teilprojekt 2 (Numerische Mathematik) und Teilprojekt 3 (Regelungstechnik und Systemdynamik) Personalkosten in Höhe von 432.415,74 Euro und Reisekosten in Höhe von 11.140,28 Euro. Die entstandenen Ausgaben belaufen sich auf 443.556,02 Euro zuzüglich einer Projektpauschale von 88.711,20 Euro.

3 Darstellung der Projektarbeiten

3.1 Organisatorisches

Ein Austausch war insbesondere im ersten Jahr des Projekts wegen der Corona-Pandemie nur online möglich. Dies erschwerte die Zusammenarbeit zu Beginn zwischen den akademischen Partnern aber insbesondere mit den assoziierten Industriepartnern. Dennoch konnte ein regelmäßiger Austausch über Videokonferenzen etabliert werden. Nach dem Ende der Pandemie wurde die Zusammenarbeit intensiviert und gegenseitige Forschungstreffen fanden statt in München, Chemnitz und Kaiserslautern.

Schwerpunkt der Treffen war die Vorstellung der erzielten Ergebnisse durch die akademischen Partner sowie der Erfahrungsaustausch mit den Industriepartnern. Die Zusammenarbeit der Verbundpartner wurde in einem Kooperationsvertrag geregelt.

Eine Web-Präsenz, welche auf die Förderung, die Ziele des Projekts und die Verbundpartner hinweist, wurde unter Beteiligung aller Verbundpartner erstellt¹. Darüber hinaus wurde an der TU Chemnitz ein Cloud-Verzeichnis zum Datenaustausch für die Verbundpartner eingerichtet.

Nach Ende der Projektlaufzeit fand ein Treffen mit der Arbeitsgruppe von Matthias Gerdt (Universität der Bundeswehr in München) sowie Fabian Duddeck (Professur Computational Mechanics, TU München) im Rahmen eines gemeinsamen Workshops zum Thema *Uncertainty Quantification* statt, um über gemeinsame weiterführende Themen und Forschungsschwerpunkte zu sprechen.

¹siehe <https://math4innovations.de> und <https://www.chemnitz-am.de/soprann>

3.2 Definition von Anwendungsszenarien und Problemspezifikation

Zu Beginn des Projekts wurden zusammen mit den universitären Projektpartnern eine Reihe von Anwendungsszenarien aus dem Bereich des automatisierten Fahrens definiert, spezifiziert und als Modellprobleme ausgewählt.

Das wichtigste Szenario beschreibt im Teilprojekt 2 das automatisierte Ausweichen von Hindernissen, insbesondere Notausweichmanöver für plötzlich auftretende Hindernisse mit geringer Time-To-Collision. Aufgrund der hierbei möglicherweise entstehenden komplexen Verhältnisse stellen solche Situationen für klassische gradienten-basierte Optimierungsmethoden oft große Probleme dar. Aus diesem Grund wurde es als geeignetes Modellszenario für das maschinelle Lernen identifiziert und vorrangig von den Arbeitsgruppen von Matthias Gerdt (Universität der Bundeswehr, München) sowie unserer Arbeitsgruppe bearbeitet. Als Fahrzeugmodell wurde ein gängiges kinematisches Modell verwendet, welches Längs- und Querbewegung hinreichend gut abbildet. Die Steuereingaben sind Lenkwinkel und Beschleunigung, wobei beide Größen Steuerbeschränkungen unterliegen. Zur Beschreibung der Hindernisse im verwendeten Bezugssystem wurden einfache kreisförmige Objekte verwendet. Die Beachtung von Zustands- und Steuerschranken wurde durch Strafterme sichergestellt.

Eine ebenfalls geplante Zusammenstellung von Anwendungsszenarien zusammen mit den Industriepartnern gestaltete sich als schwierig, da die erste Hälfte des Projekts von der Corona-Pandemie geprägt war und somit kein Austausch mit den Industriepartnern vor Ort stattfand. Diese waren zu den regelmäßigen Treffen eingeladen, jedoch folgten den dabei stattfindenden Gesprächen keine weitere Kooperation. Aus diesem Grund haben sich die akademischen Partner für Teilprojekt 2 schnell über mögliche, auch für die außeruniversitären Partner interessanten Themen geeinigt. Diese waren unter anderem Szenarien der Bahnverfolgung, für die grundsätzlich sehr gute und echtzeitfähige Algorithmen, etwa lineare modell-prediktive Regelung (MPC), existieren. Dadurch war ein Vergleich von entwickelten KI-Reglern im Bezug auf Geschwindigkeit sowie Genauigkeit möglich. Weitere Beispielprobleme finden sich in den öffentlich verfügbaren Berichten der anderen Projektpartner.

Die untersuchten Lösungsmethoden lassen sich in zwei Kategorien einordnen:

- Die erste Variante untersuchte die Approximation eines Feedback-Regelungsgesetz der Form $\mathcal{N}_\theta : \Omega \rightarrow U$. Hierbei bezeichnet die Abbildung $\mathcal{N}_\theta$ ein neuronales Netz mit Parametern $\theta \in \Theta$ (Gewichte und Bias). Dieser Regler bildet jeden Zustand – also aktuelle Position, Geschwindigkeit, Auslenkung und Gierwinkel des Fahrzeugs – auf die im gegebenen Zustand beste Steuerung $u \in U$ ab. Damit ergibt sich der geschlossene Regelkreis

$$\dot{x}(t) = f(x(t), \mathcal{N}_\theta(x(t))), \quad x(0) = x_0, \quad t \geq 0,$$

wobei die Differentialgleichung die kinematische Bewegung des Fahrzeugs modelliert (inklusive Steuer- und Zustandsschranken). Für das Training von NN-Reglern wurden dabei verschiedene Techniken entwickelt und erprobt.

- Variante zwei untersucht die Approximation der Wertfunktion v durch ein neuronales Netz und deren Berechnung über die Lösung der zugehörigen Optimalitätsbedingungen (HJB-Gleichungen). Die optimale Regelung ergibt sich in der Regel durch Auswertung der Funktion

$$u^*(x(t)) = \mathcal{G}(t, x(t), v(t, x(t)), \nabla v(t, x(t))),$$

wobei sich $\mathcal{G}$ aus der HJB-Gleichung ergibt.

Details zu beiden Ansätzen finden sich in Abschnitt 3.3. Die Anwendungen von Methoden der Unsicherheitsquantifizierung für neuronale Netze im Rahmen von SOPRANN werden in Abschnitt 3.4 erläutert.

Für die im Teilprojekt 3 untersuchten Methoden wurden drei Anwendungsbeispiele berücksichtigt. Als erstes Beispiel würde die sogenannte Ring-Road gewählt. Dies ist ein vereinfachtes geschlossenes Verkehrsnetzwerk, welches aus einer kreisförmigen einspurigen Straße besteht, auf der autonome und von Menschen gesteuerte Fahrzeuge fahren. Dieses Szenario ermöglicht die Untersuchung des Verkehrsflusses und des Einflusses von autonomen Fahrzeugen. Insbesondere lassen sich sogenannte Stop-and-Go-Wellen erzeugen und untersuchen, welche eine erhebliche Störung des Verkehrsflusses verursachen. Es wurde der Fall mit einem autonomen Fahrzeug (AV) betrachtet, dessen Beschleunigung direkt vorgegeben werden kann, und einer variablen Anzahl an menschlichen Fahrern, welche durch das Intelligent Driver Model (IDM), eine Abstraktion des menschlichen Fahrverhaltens, beschrieben werden.

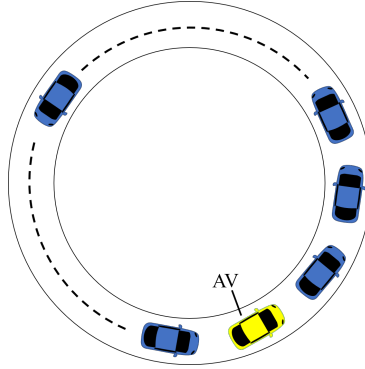


Abbildung 1: Ring-Road-Szenario

Als zweites Beispiel wurde aufgrund der Relevanz für das dynamische Verhalten eines Wagens die Traktionsregelung gewählt. Abbildung 2 zeigt das Schema des Modells. Die Variable m ist die Masse des Wagens, v die Geschwindigkeit, J die Radträgheit, τ_f ein kondensierter Reibungsterm, ω die Drehzahl, r_d der Radius des Rades und τ_d der Steuereingang. Das Schlupfverhältnis soll zu einem vorgegebenen Wert geregelt werden.

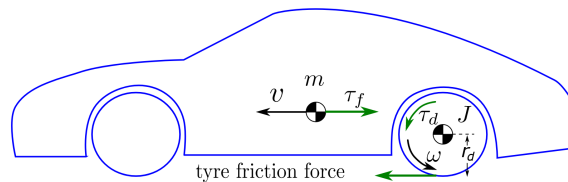


Abbildung 2: Traktionsmodell

Als drittes Beispiel wurde eine adaptive Geschwindigkeitsreglungsaufgabe (Adaptive Cruise Control). Das dynamische Verhalten des Wagens wird durch folgende Differentialgleichung beschrieben

$$\dot{v} = -\frac{1}{m} (f_0 + v f_1 + v^2 f_2) + \frac{1}{m} u,$$

wobei v die Geschwindigkeit, $f_{0,1,2}$ unbekannte Reibungskoeffizienten, m die Masse des Wagens und u der Steuereingang sind. Das Ziel ist es, dass der Wagen mit einer vorgegebenen Geschwindigkeit fährt.

Auf methodischer Seite wurden datenbasierte Optimalsteuerungsverfahren entwickelt und deren theoretischen Eigenschaften, wie Stabilität und Erfüllung der Systembeschränkungen, untersucht. Im MPC-Kontext wurden zwei Architekturen untersucht:

- Ein robuster adaptiver MPC-Ansatz mit online-gelerntem Modell. Dieser basiert auf der Minimierung einer Kostenfunktion, welche sich aus der Summe einer vorgegebenen Gütefunktion ergibt. Diese wird entlang der prädizierten Systemtrajektorie mithilfe der online-gelernten Modellparameter ausgewertet. Die Ergebnisse hierfür werden im Abschnitt 3.5 erläutert.
- Ein MPC-Ansatz mit gelernter Endkostenfunktion. Im Vergleich zum vorherigen Ansatz wird das Systemmodell als bekannt angenommen und der Fokus auf die Performanz des geschlossenen Kreises unter dem MPC-Regelgesetz gelegt. Insbesondere wird die Performanz untersucht, wenn die Endkostenfunktion basierend auf Reinforcement-Learning-Methoden bestimmt wird und die verwendeten iterativen Verfahren nicht exakt gelöst werden können. Die Erkenntnisse zu diesem Ansatz sind im Abschnitt 3.6 dargestellt.

Zuletzt wurde ein reiner Reinforcement-Learning-basierter Ansatz mit Sicherheitsgarantien für partiell unbekannte Systeme entworfen. Dieser basiert auf Policy Iteration und adaptiven Methoden, um die unbekannten Systemparameter zu bestimmen und die Performanz zu verbessern, ohne Offline-Training zu nutzen. Dieser Ansatz wird im Abschnitt 3.7 erläutert.

3.3 Neuronale Netze zur Lösung von Hamilton-Jacobi-Bellman-Gleichungen (TP 2)

Es wurden verschiedene neuartige Ansätze zur Lösung von Optimalsteuerproblemen für dynamische Systeme untersucht und implementiert. Dazu zählen Ansätze mit zusätzlichem stochastischen Rauschterm (HJB Lösungen mit verschwindender Viskosität), Ansätze, die neuronale Netze unter Beachtung von physikalischen Gesetzmäßigkeiten lernen (engl. PINNs), sowie Löser die auf der linearen bzw. nichtlinearen Feynman-Kac-Formel basieren. Diese wurden anhand von mathematisch gut beschreibbaren Problemen untersucht und anschließend auf Probleme der Fahrzeugdynamik und Hindernisvermeidung angewendet. Dabei zeigte sich, dass die Umgehung des Fluchs der Dimension, für den die auf den Feynman-Kac-Formeln basierten Methoden bekannt sind, auf die Optimalitätsbedingungen, die sich bspw. aus der Hindernisvermeidung ergeben, nicht ohne weiteres übertragbar sind. Für die Berechnung von KI-Reglern im Bereich des automatisierten Fahrens zeigte sich, dass die betrachteten Verfahren nur sehr eingeschränkt mit der komplexen Dynamik sowie den auftretenden Steuer- und Zustandsschranken zurechtkommen. Um physikalische Gesetzmäßigkeiten angereicherte neuronale Netze (PINNs) wurden in der Masterarbeit von Julius Peltri in Hinblick auf Fehlerabschätzungen nach [MISHRA, MOLINARO, 2021] weiter untersucht. Desweiteren wurde im Rahmen dieser Abschlussarbeit der Einsatz von Dünngitter-Quadratur durch Leja-Punkte, wie sie in Abb. 3 dargestellt sind, evaluiert. Dünngitter-Quadratur ist ein Standardverfahren der Numerik, um bekannte Ansätze auf

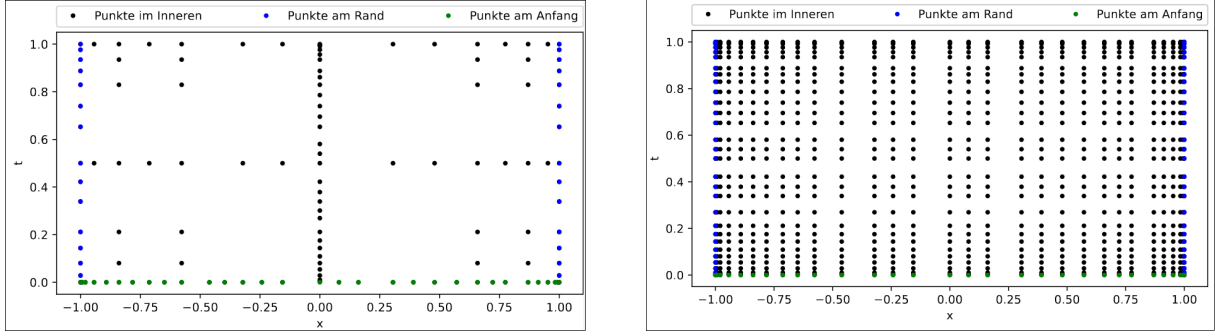


Abbildung 3: Untersuchte Quadraturpunkte: Leja Dünngitter (links); Leja Tensorproduktgitter (rechts) anstelle zufälliger Quadraturpunkte.

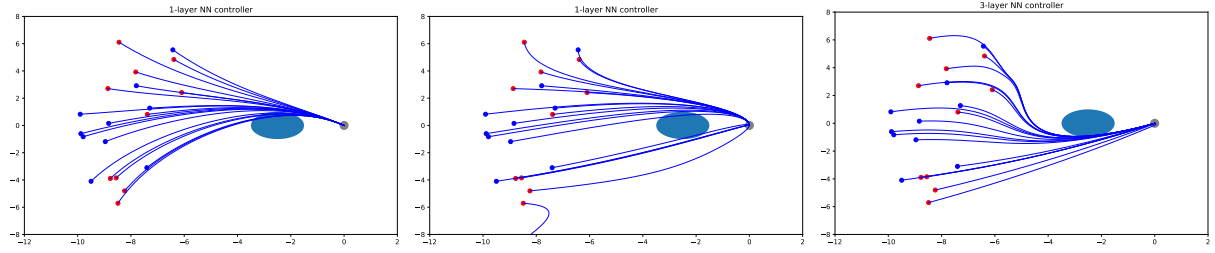


Abbildung 4: Optimale Trajektorien für ein einfaches Hindernisausweichproblem für verschiedene Netzwerkarchitekturen und Aktivierungsfunktionen. Trainiert wurden die Netzwerke ausschließlich mit Trajektorien, die an den blauen Punkten starten. Rote Punkte sind Startzustände der Testmenge.

hochdimensionale Problemstellungen anwendbar zu machen. Die Verwendung von Leja-Knoten anstelle zufällig bestimmter führten tatsächlich zu einer Verringerung des Quadraturfehlers. Der Effekt war jedoch nicht stark genug, um eine Anwendung von PINNs für deutlich höherdimensionale Probleme zu ermöglichen.

Nachdem sich gezeigt hat, dass die Methoden, deren Ziel die Lösung der HJB-Gleichung ist, für Anwendungen im Bereich des automatisierten Fahrens nur mit großer Mühe anwendbar sind, wurden im weiteren Verlauf Verfahren untersucht, die die Regelfunktion direkt durch ein neuronales Netz approximieren. Dieser Ansatz basiert auf [KUNISH, WALTER, 2021] und löst für eine kleine Anzahl an Startpositionen das Optimalsteuerproblem mit Hilfe des Pontryaginschen Maximumprinzips. Hierbei sorgt die Annahme eines glatten neuronalen Netzes für eine gute Generalisierbarkeit der gelernten Steuerung in einem großen Gebiet. Unter gewissen Annahmen konnten in [KUNISH, WALTER, 2021] für Probleme mit unendlichem Zeithorizont theoretisch fundierte Optimalitätsaussagen gezeigt werden, was den Ansatz mathematisch sehr vielversprechend machte. In Abb. 4 sind exemplarisch für drei verschiedene Netzwerkarchitekturen und Aktivierungsfunktionen Ausweichtrajektorien dargestellt.

Das Training der Netzwerke gestaltete sich dabei als herausfordernd. Gute Ergebnisse konnten in der Regel nur durch eine differenzierte Schrittweitensteuerung erzielt werden. Die für NN-Training üblichen stochastischen Gradientenverfahren, aber auch mathematisch fundierte exakte Gradientenverfahren mit Schrittweitensteuerung führten nicht immer zu plausiblen Ergebnissen. Gründe hierfür könnten sein, dass durch die hohe Anzahl an freien Parametern eine große Anzahl lokaler Minima entsteht. Desweiteren ist zu erkennen, dass die berechneten Trajektorien nicht in allen Fällen dazu in der Lage waren, dass Hindernis

beidseitig zu umfahren. Auch hier spielt die Konvergenz gegen lokale Optima eine Rolle. Ein möglicher Ausweg könnte durch die Kombination verschiedener KI-Regler gegeben sein, indem verschiedene lokal-optimale Trajektorien vorgeschlagen und verglichen werden. Aufgrund der schnellen Auswertbarkeit von NN-Reglern ist dies in Echtzeit möglich.

Parallel dazu wurde ein in [RUTHOTTO ET. AL, 2020] vorgestellter Ansatz, welcher die Lösung von HJB Gleichungen und Optimalsteuerproblemen in den Kontext der Lösung eines Mean-Field Games setzt, hinsichtlich seiner Anwendbarkeit auf Regelungsaufgaben aus dem Bereich des autonomen Fahrens untersucht und implementiert. Erste Versuche zeigen im Rahmen von einfachen Dynamiken vielversprechende Ergebnisse, eine Erweiterung auf realistische Fahrzeugdynamiken steht zum aktuellen Zeitpunkt noch aus.

3.4 Unsicherheitsquantifizierung für neuronale Netze und deren Anwendung (TP 2)

Ein trainiertes NN produziert für jede Belegung seiner Eingangsknoten eine Ausgabe ohne Information darüber, wie vertrauenswürdig dieses Ergebnis angesichts des Umfangs und der Qualität der Trainingsdaten sein kann. Im Extremfall können out-of-sample inputs, also Eingaben welche sich fundamental von den Trainingsdaten unterscheiden, zu bedeutungslosen Extrapolationen in den Resultaten des NN führen, ohne das dies der zugrundeliegenden Anwendung mitgeteilt wird.

In der Masterarbeit von Maximillian Kloppe mit dem Titel „Unsicherheit bei neuronalen Netzen“ wurde die Grundlagenarbeit [GAL, 2016] untersucht. Zu den wichtigsten Erkenntnissen zählt, dass Bayessche Neuronale Netze, also Netze, deren Gewichte durch Normalverteilungen mit jeweils eigenem trainierten Mittelwert und eigener trainierter Standardabweichung beschrieben werden, im Allgemeinen langsam und schwer zu optimieren sind. Es zeigt sich jedoch, dass diese Netze in einigen Fällen mathematisch äquivalent zu sogenannten Dropout-Netzen sind, welche sowohl zum Training als auch zur Prädiktion zufällig einen Teil der Neuronen ausschalten, wodurch eine Bandbreite von Vorhersagen entsteht. Diese ermöglichen Aussagen zur statistischen Sicherheit von Prädiktion und erlauben die Berechnung von Risikomaßen. Die in der Masterarbeit entwickelten Verfahren wurden an einfachen mathematischen Problemen aus dem Bereich der optimalen Regelung getestet. Leider stellte sich heraus, dass die im Rahmen des Projekts angewendeten Methoden durch die Hinzunahme von Dropout-Schichten wesentlich schwerer und instabiler zu trainieren waren. Insbesondere haben diese nicht zu einer gewünschten Stabilisierung und Verbesserung der Robustheit der berechneten Steuerungen gesorgt. Aus mathematischer Sicht sind weitere Untersuchungen notwendig, die nach Abschluss des SOPRANN-Projekts zusammen mit der Arbeitsgruppe von Matthias Gerds (Universität der Bundeswehr, München) betrachtet werden sollen.

3.5 Robuster Adaptiver MPC-Ansatz und dessen Performanzanalyse (TP 3)

Ein robuster adaptiver MPC-Ansatz ermöglicht die Regelung von unsicheren Systemen. Da in der Regel das dynamische Verhalten von Systemen nur zum Teil bekannt ist, liegen nur unsichere Modelle für den Reglerentwurf vor. Die Erfüllung der Systembeschränkungen erfordert die Berücksichtigung aller möglichen Systemtrajektorien, was zu

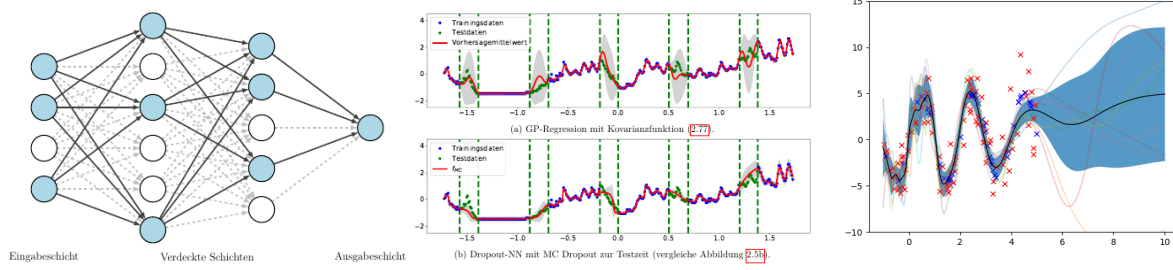


Abbildung 5: Neuronales Netz mit Dropout (links); Beispielanwendung von Dropout-NNs auf Regressionsaufgabe (Mitte); Illustration von Unsicherheitsbereichen durch Markov chain Monte Carlo-Sampling (rechts).

einem komplexen Optimierungsproblem führt. Zudem muss das Adaptationsverfahren des Systemmodells sorgfältig ausgewählt werden, damit die gewünschten Eigenschaften des geschlossenen Regelkreises beibehalten werden. Aus diesem Grund sind Performanzanalysen von robusten adaptiven MPC-Ansätzen rar. Da solche Analysen Aussagen zum Kosten-Nutzen-Verhältnis eines komplexen Optimierungsproblems liefern können, hat man trotz dieser Schwierigkeiten die Performanz eines robusten adaptiven MPC-Reglers untersucht. Es wurde ein lineares System betrachtet, dessen System- und Inputmatrix unbekannt sind

$$x_{k+1} = A(\theta^*) x_k + B(\theta^*) u_k,$$

wobei die Matrizen $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times n}$ affin abhängig sind vom unbekannten Parameter $\theta^* \in \mathbb{R}^p$. Der MPC-Ansatz wurde in [Lorenzen, Cannon, Allgöwer, 2019] vorgeschlagen. Es wird angenommen, dass der unbekannte Parameter in einer bekannten konvexen Menge $\Theta \subset \mathbb{R}^p$ liegt. Basierend auf Online-Messungen wird die Schätzung der Menge und des unbekannten Parameters verbessert. Dann wird zu jedem Zeitschritt folgendes Optimierungsproblem gelöst

$$\min_{\mathbf{d}_k} J_N(\mathbf{d}_k) = \sum_{i=0}^N l(x_{i|k}, u_{i|k}) + V_f(x_{N|k}) \quad \text{s.t.} \quad \mathbf{d}_k \in \mathbb{D}(x_k, \theta_k),$$

wobei θ_k die aktuelle Schätzung des unbekannten Parameters ist, $\mathbf{d}_k$ die Optimierungsvariablen darstellen und $\mathbb{D}$ die Gleichheits- und Ungleichungsbeschränkungen (Systembeschränkungen, Systemdynamik, etc.) zusammenfasst. Die Gütefunktion $l : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}_{\geq 0}$ und die Endkosten $V_f : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$ sind quadratisch und Letzteres entsteht aus der Lösung einer Matrix-Riccati-Gleichung, was die Stabilität des Regelgesetzes sicherstellt. Die Variablen $x_{i|k}$ und $u_{i|k}$ sind jeweils die prädizierte Systemtrajektorie und der prädizierte Steuereingang. Alle möglichen Trajektorien des Systems für alle θ in Θ_k , wobei Θ_k die aktuelle Schätzung der Unsicherheitsmenge ist, werden berücksichtigt, um sicherzustellen, dass die Beschränkungen erfüllt werden.

Die Performanz des geschlossenen Regelkreises J_{MPC}^∞ wurde im Vergleich zu den optimalen Kosten untersucht, d.h., die Lösung des eingeführten Optimierungsproblems, aber mit unendlichem Horizont und bekannter Systemdynamik. Es konnte gezeigt werden, dass im Worst-Case die Performanz quadratisch mit wachsendem Parameterfehler abnimmt, was sich in höheren Kosten im geschlossenen Kreis widerspiegelt. Andere Faktoren wie die Größe der Menge Θ können eine Verschlechterung der Performanz verursachen. Dieses Verhalten wurde in Simulation beobachtet, indem die Performanz für den Worst-Case

ausgerechnet wurde, siehe Abbildung 6. Da die Analyse nur für den vorgestellten MPC-Ansatz gilt, können in der Zukunft andere Adaptationsverfahren untersucht werden, um den Einfluss der Adaptation zu charakterisieren. Auch wenn lineare Systeme eine einfache Klasse von Systemen darstellen, lassen sich auch nichtlineare Systeme mit linearen Methoden untersuchen, z.B. durch eine geeignete Transformation (Koopman-Operator, zeitvariante Linearisierung, etc.).

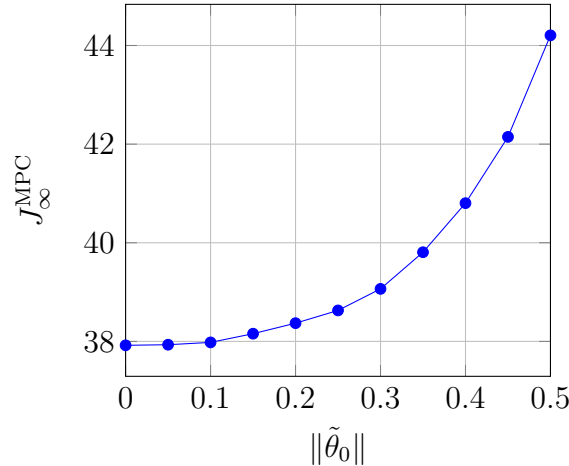


Abbildung 6: Worst-Case-Performanz für verschiedene $\|\tilde{\theta}_0\|$.

3.6 MPC-Ansatz mit gelernter Endkostenfunktion (TP 3)

Da es in der Praxis häufig eine Gütefunktion für eine Optimierungsaufgabe gibt (Kosten, Gewinn), kann ein Optimierungsproblem mit unendlichem Zeithorizont aufgestellt werden, um die kumulierte Summe der Gütefunktion für alle Zeiten zu minimieren, sofern sie endlich ist. Die Lösung dieses Problems kann für allgemeine nichtlineare Systeme nicht analytisch bestimmt werden, da dies die Lösung der Bellman-Gleichung erfordert, eine Fixpunktgleichung. Probleme wie der Curse of Dimensionality (Fluch der Dimensionalität) erschweren die Lösung der Gleichung. Jedoch kann ein MPC-Ansatz verwendet werden, um eine suboptimale Lösung des Problems zu finden. Darüber hinaus kann dieser MPC-Regler angepasst werden, um die Suboptimalität des Regelgesetzes zu reduzieren. Dafür bietet sich die Endkostenfunktion an, welche als Restkosten interpretiert werden kann, d. h. die kumulierte Summe der Gütefunktion für ein bestimmtes Regelgesetz. Um die Endkostenfunktion zu bestimmen, können Methoden des Reinforcement Learning verwendet werden, wie die Adaptive Dynamische Programmierung. Diese bezeichnet eine Familie iterativer Verfahren, um die Lösung der Bellman-Gleichung in einem beschränkten Gebiet zu bestimmen. Diese Verfahren können in zwei Gruppen unterteilt werden: Policy Iteration und Value Iteration Methoden. Im Projekt SOPRANN hat man beide Klassen untersucht.

Bei der Policy Iteration wird die Wertefunktion einer Anfangsstrategie bestimmt und sukzessiv verbessert. Diese erfordert eine stabilisierende Anfangsstrategie, weist aber den Vorteil auf, dass die generierten Strategien stabilisierend bleiben, und ist somit geeignet für Online-Anwendungen. Da die Iterationen in der Regel nicht exakt gelöst werden können, weil Funktionsapproximatoren verwendet werden, hat man für die Stabilitäts- und Performanzanalyse den dadurch entstandenen Fehler explizit berücksichtigt. Es wurde gezeigt,

dass ein ausreichend langer Prädiktionshorizont die Stabilität des geschlossenen Kreises sowie die rekursive Lösbarkeit sicherstellt. Diese hängt von vielen Faktoren ab, wie dem gewünschten Einzugsbereich (β) und dem Approximationsfehler (c). Die Ergebnisse der Prädiktionslänge für ein akademisches Beispiel sind in Abbildung 7a dargestellt.

Bei der Value Iteration wird eine Anfangsschätzung der optimalen Wertefunktion iterativ verbessert. Im Gegensatz zur Policy Iteration hat diese den Vorteil, dass keine stabilisierende Strategie erforderlich ist. Die Stabilität der generierten Strategien, welche aus den generierten Wertefunktionen hergeleitet werden können, ist hingegen unbekannt. Ähnlich wie bei der Policy Iteration konnte man zeigen, dass eine ausreichend lange Prädiktionslänge die Stabilität des geschlossenen Kreises und die rekursive Lösbarkeit gewährleistet. Diese hängt von dem Approximationsfehler, der Konvergenztoleranz und dem gewünschten Einzugsbereich ab. Auch wenn die Prädiktionslänge dieser Methode deutlich kürzer als für die Policy Iteration ist, ist diese immer noch sehr konservativ, da das System für deutlich kürzere Prädiktionslängen zum Ursprung konvergiert. Die Ergebnisse für ein akademisches Beispiel sind in Abbildung 7b dargestellt.

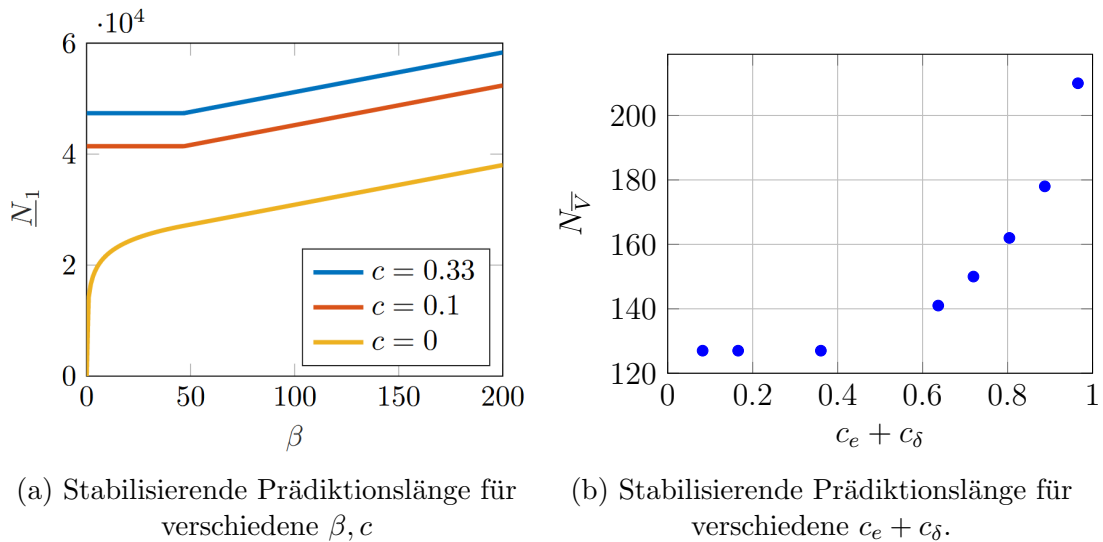


Abbildung 7: Stabilisierende Prädiktionslänge für MPC-Ansätze. (a) mit Policy Iteration und (b) mit Value Iteration

Die gewonnen Erkenntnisse zur stabilisierenden Prädiktionslänge konnten in Kooperation mit dem Partner Fraunhofer ITWM auf das Szenario der Ring-Road angewendet werden. Die geforderte Prädiktionslänge für Stabilität wurde für unterschiedliche Anzahlen an menschlichen Fahrern untersucht, wie in Abbildung 8 dargestellt. In diesem Fall ist das Ergebnis auch konservativ.

3.7 Reinforcement-Learning-Ansatz mit Stabilitätsgarantien (TP 3)

Ein Vorteil von Reinforcement-Learning-Methoden ist, dass diese kein Systemmodell benötigen um ein Regelgesetz zu lernen, nur Messungen der Ausgangsgrößen. Hingegen lassen sich größtenteils keine Aussagen zur Stabilität des geschlossenen Kreises treffen. In diesem Sinne können regelungstechnische Methoden verwendet werden, um die Stabilität zu gewährleisten. Im Teilprojekt 3 wurde das folgende System betrachtet

$$\dot{x} = f(x) + F(x)\theta + g(x)u$$

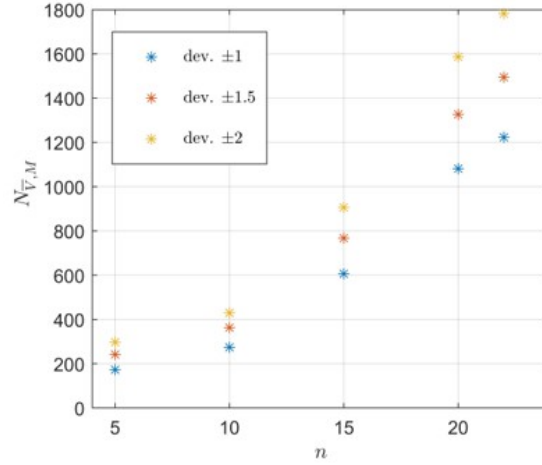


Abbildung 8: Stabilisierende Prädiktionslänge für das Ring-Road-Szenario

mit Zustand $x \in \mathbb{R}^n$, Steuereingang $u \in \mathbb{R}^m$, unbekannten Parametern $\theta \in \mathbb{R}^p$ und Funktionen $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times m}$ und $F : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times p}$. Als regelungstechnische Aufgabe soll das Integral über einen unendlichen Zeithorizont einer gegebenen Gütefunktion $r : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}_{\geq 0}$ minimiert werden. Eine adaptive Kontroll-Lyapunov-Funktion wird anschließend verwendet, um ein Adaptationsverfahren zu implementieren und die unbekannten Parameter anhand Online-Messungen zu bestimmen. Basierend auf Policy Iteration und der adaptiven Kontroll-Lyapunov-Funktion konnte ein Algorithmus entwickelt werden, welches das Regelgesetz online lernt und praktische Stabilität gewährleistet unter Berücksichtigung der Implementierung mit einem Sample-and-Hold-Verfahren.

Der Algorithmus wurde in Simulationen für eine Traktionsregelung und eine adaptive Geschwindigkeitsregelungsaufgabe getestet. Es wurde eine signifikante Verbesserung der Performanz beobachtet im Vergleich zu einem adaptiven Regler ohne Reinforcement-Learning-Anteil. In Abbildung 9 ist die Kostensenkung für die Traktionsregelung mit verschiedenen Anfangsbedingungen dargestellt. Es konnte eine Reduktion von bis zu 90% erreicht werden. Der weiße streifen ist die Menge, zu der das System konvergieren soll. Dementsprechend sind Performanzaussagen für Anfangsbedingungen in dieser Menge bedeutungslos.

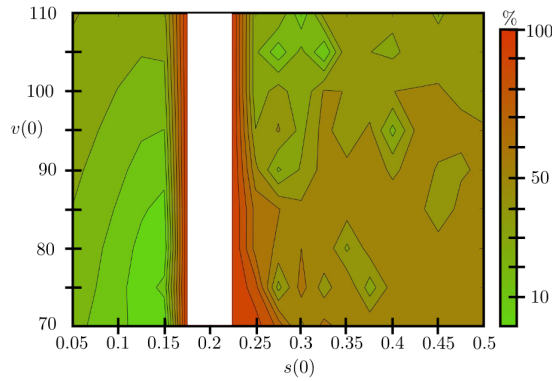


Abbildung 9: Kostensenkung für die Traktionsregelungsaufgabe

3.8 Zusammenfassung der Ergebnisse und Erfahrungen

Zusammenfassend lautet das wissenschaftlich-technische Ergebnis von Teilprojekt 2 des Verbundprojekts SOPRANN wie folgt:

- Vergleich von verschiedenen Ansätzen zur NN-basierten Lösung von partiellen Differentialgleichungen und Hamilton-Jacobi-Bellman Gleichungen.
- Implementierung und Erweiterung des semiglobalen Stabilisierungsansatzes für Probleme der Hindernisvermeidung für einfache und komplexe Dynamiken unter Beachtung von Steuer- und Zustandsschranken.
- Implementierung und Erweiterung des NN-basierten Mean-Field-Games Ansatzes für einfache Hindernisvermeidungsaufgaben.
- Implementierung von Imitation Learning durch klassische Regler sowie Reinforcement Learning Techniken zum Vortraining von generischen globalen NN-Reglern.
- Untersuchung von Imitation Learning durch klassische Regler zum Vortraining von generischen globalen NN-Reglern.
- Implementierung und Untersuchung von Dropout-Netzen in generischen NN-Reglern.
- Veröffentlichung eines Teils der Methoden und Algorithmen unter <https://github.com/janblechschmidt/PDEsByNNs>.

Im Verlauf des Projekts zeigte sich im Vergleich zu den anderen akademischen Partnern, dass das Problem-informierte Netz, welches in der Gruppe von Matthias Gerds entwickelt wurde, die besten Resultate erzielen konnte. Die von uns entwickelten generischen Feedforward-Netze sowie Residualnetze konnten trotz der Untersuchung einer Vielzahl von verschiedenen Netzwerkarchitekturen, Aktivierungsfunktionen sowie Optimierungsmethoden nicht die gleiche Performance erreichen. Unserer Ansicht nach könnte die Kombination aus einem solchen parametrischen Modell zusammen mit einem flexiblen neuronalen Netz zu weiteren Verbesserungen führen.

Für das Teilprojekt 3 lassen sich die wissenschaftlichen Ergebnisse wie folgt zusammenfassen:

- Sichere Einbettung eines online-adaptierten Modells in ein Optimalsteuerungsverfahren (MPC) und Quantifizierung der Suboptimalität des Regelgesetzes.
- Stabilitätszertifikat eines MPC-Ansatzes mit gelernter Endkostenfunktion für die Reinforcement-Learning-Methoden Policy Iteration und Value Iteration und Charakterisierung dessen Suboptimalität.
- Bestimmung einer stabilisierenden Prädiktionslänge für das Ring-Road-Szenario unter einem MPC-Regler in Kooperation mit dem Partner Fraunhofer ITWM. Evaluation des Verhaltens in Simulation und Vergleich mit anderen Steuerungsverfahren.
- Entwurf eines reinen Reinforcement-Learning-basierten Ansatzes für zum Teil unbekannte Systeme mit Stabilitätsgarantien.
- Signifikante Verbesserung der Performanz für Traktionsregelung- und adaptive Geschwindigkeitsregelungsaufgaben mit dem vorgeschlagenen Reinforcement-Learning-Ansatz im Vergleich zu anderen adaptiven Regelungsmethoden.
- Veröffentlichung der erzielten Ergebnisse auf internationalen Konferenzen (European Control Conference, International Federation of Automatic Control World Congress) und bei internationalen Journalen (European Journal of Control, IEEE Transactions of Automatic Control).

Die online Adaptation des Systemmodells stellt eine große Herausforderung für die Suboptimalitätsanalyse und die Einhaltung der Systembeschränkungen dar, da die tatsächliche Systemtrajektorie unbekannt ist. Besonders für nichtlineare Systeme kann die Berücksichtigung aller möglichen Systemtrajektorien zu einem sehr komplexen Optimierungsproblem führen. Zudem hat sich die resultierende stabilisierende Prädiktionslänge als sehr konservativ herausgestellt. Dies erschwert die Lösung des Optimierungsproblems, weil die Anzahl an Variablen exponentiell steigt. Da in Simulationen eine deutliche kürzere Länge ausreichend ist, wird vermutet, dass die aktuellen Ergebnisse verbessert werden können.

4 Verwertbarkeit des Ergebnisses und Planungen

Die bisherigen Ergebnisse des Projekts SOPRANN haben gezeigt, dass die Kombination von System- und Modellwissen mit Daten bzw. datengetriebenen Verfahren ein großes Potential hat. Im bisherigen Projektverlauf stand die Erforschung von mathematischen Methoden im Vordergrund. Eine weitere Erforschung von Methoden und deren Transfer in die Praxis ist sehr sinnvoll. Hierfür muss die technische Reife erhöht werden sowie weiterhin an methodischen Ansätzen, insbesondere im Bereich der robusten Steuerung, geforscht werden.

Im Rahmen des Teilprojekts 2 wurden KI-basierte Ausweichregler für das automatisierte Fahren konzipiert und in Simulationen getestet. Im Gesamtprojekt SOPRANN konnte durch die Kollegen der Universität der Bundeswehr der Nachweis der prinzipiellen Funktionstüchtigkeit im Rahmen einer Versuchsfahrt mit einem automatisierten Versuchsfahrzeug auf Basis eines Streetscooter auf dem Testgelände der Universität der Bundeswehr München erbracht werden. Insbesondere die Integration von KI-basierten Reglern in etablierte MPC-basierte Reglerarchitekturen erscheint sehr sinnvoll, um die Sicherheit im Echtzeitbetrieb durch redundante Reglerebenen zu erhöhen.

Auf methodischer Seite sollen zukünftig folgende Aspekte im Detail untersucht werden:

- Konzept des *Problem-informierten maschinellen Lernens (PINNs)*: Hierunter werden Lernverfahren verstanden, die ein Verhalten auf Basis einer physikalischen Gesetzmäßigkeit oder eines zu Grunde liegenden Optimalitätsprinzips erlernen, anstatt rein datenbasiert im Sinne eines überwachten Lernens zu arbeiten. Dieses Konzept lässt sich auf parametrische Optimierungsprobleme und HJB-Gleichungen der dynamischen Programmierung anwenden. Dieses Konzept ist vielversprechend, da mitunter Fehlerschätzer hergeleitet werden können, die eine quantitative Bewertung der KI-Lösungen ermöglichen.
- *Imitation Learning Regler* dienen als Werkzeug, um Echtzeitfähigkeit für bestehende, insbesondere optimierungsbasierte Regler zu erreichen und ggf. auch die Robustheit zu erhöhen. Geht man von einem stabilen Regler aus, so besteht die Hoffnung, auch für die imitierende Variante einen Stabilitätsnachweis erzielen zu können, siehe z.B. die Arbeit

Luis Rodrigues and Sidney Givigi (2022). “Analysis and Design of Quadratic Neural Networks for Regression, Classification, and Lyapunov Control of Dynamical Systems”. In: *arXiv preprint arXiv:2207.13120*

- Es sollen *hybride Regler* für Bahnverfolgungsaufgaben oder Bahnplanungsaufgaben konstruiert werden, die konventionelle Regler mit KI-basierten Reglern kombinieren.

Hierdurch wird eine Erhöhung der Effizienz und auch die Absicherung bei Fehlfunktionen angestrebt.

Hierbei ist eine weitere Zusammenarbeit mit der Arbeitsgruppe von Matthias Gerdts geplant. Die Universität der Bundeswehr bietet hier ideale Voraussetzungen, da die Einbettung in eine ROS-basierte Simulationsumgebung sowie die Erprobung im Versuchsfahrzeug wichtige Bestandteile weiterer Forschung in diesem Bereich sind.

Im Teilprojekt 3 wurden datenbasierte Optimalsteuerungsverfahren entwickelt und deren theoretischen Eigenschaften, wie Stabilität und Erfüllung der Systembeschränkungen, analysiert. Diese sollen wie folgt erweitert werden:

- Andere Modelladaptationsverfahren: es wurde nur ein Adaptationsverfahren untersucht. Um den Einfluss der Adaptation auf die Performanz zu charakterisieren, sollen andere Verfahren untersucht werden.
- Online-Anpassung der Endkostenfunktion: da für das Lernen der Endkostenfunktion die exakte Systemdynamik erforderlich ist, welche in der Regel nur zum Teil bekannt ist, soll ein Adaptationsverfahren entworfen werden, das mit Online-Messungen der Anpassung der Kostenfunktion ermöglicht und die Sicherheitsgarantien nicht verletzt.
- Berücksichtigung allgemeiner Gütefunktionen in einem ökonomischen MPC-Kontext mit lernenbasierten Endkostenfunktion. Eine Kooperation mit anderen Forschungseinrichtungen ist geplant.
- Die Ergebnisse zur Traktionsregelung sollen im Kontext des Gemeinschaftsvorhabens Smart Rail Connectivity Campus² auf Praxistauglichkeit für die Regelung von Zügen untersucht werden.
- Kürzere Prädiktionslänge: da die aktuelle resultierende stabilisierende Prädiktionslänge sehr lang ist, soll eine andere Stabilitätsanalyse durchgeführt werden. Der Fokus soll auf die Konvergenz der Wertefunktion des Optimierungsproblems des MPC-Reglers mit steigender Prädiktionslänge liegen.

Um die Praxistauglichkeit des MPC-Ansatzes mit stabilisierender Prädiktionslänge für die Ring-Road zu untersuchen, bietet sich eine Kooperation mit dem Fraunhofer ITWM an. Die Expertise des Institutes im Bereich autonome Systeme für Mobilitätsanwendungen dient als Basis zur weiteren Zusammenarbeit im Bereich sicherer autonomen Systeme für Mobilität.

5 Fortschritt auf dem Gebiet

Methoden der künstlichen Intelligenz und maschinelle Lernverfahren finden mit enormer Dynamik Einzug in nahezu alle Bereiche von Wissenschaft, Wirtschaft, Technik und Verwaltung und die Anzahl an Veröffentlichungen ist enorm. Daher ist es sehr schwierig, sämtliche Entwicklungen im Blick zu behalten. Zusammen mit den anderen akademischen Partnern wurden regelmäßig neue Entwicklungen in diesem Bereich diskutiert. Im Bereich der auf Reinforcement Learning basierten Regelung sind relevante Entwicklungen in den Bereichen Q-Learning, Imitation Learning sowie Counter Learning zu beobachten, etwa:

²<https://www.smart-rail-campus.de/>

- Das Imitation Learning zielt darauf ab, einen bestehenden, aber möglicherweise nicht echtzeitfähigen Regler durch Neuronale Netze zu imitieren. Im Fall von quadratischen Neuronalen Netzen können mitunter Stabilitätsnachweise unter Verwendung einer Lyapunov-Funktion geführt werden, siehe Luis Rodrigues and Sidney Givigi (2022). “Analysis and Design of Quadratic Neural Networks for Regression, Classification, and Lyapunov Control of Dynamical Systems”. In: *arXiv preprint arXiv:2207.13120*
- Ein weiteres interessantes Konzept ist durch das Erlernen des Reglers durch sogenannte DeepONets gegeben. Dabei wird das Optimierungsproblem von Anfang an im unendlichdimensionalen Setting der zugrundeliegenden Differentialgleichung formuliert und gelöst. Erste Anwendungen auf Optimalsteuerprobleme finden sich in Luke Bhan, Yuanyuan Shi, and Miroslav Krstic (2023). “Operator learning for nonlinear adaptive control”. In: *Learning for Dynamics and Control Conference*. PMLR, pp. 346–357

Im Bereich der mathematisch fundierten Lösung von Steuerproblemen mit neuronalen Netzen erscheinen uns zwei Arbeiten besonders wichtig, die wir im Verlauf des Projekts intensiv studiert haben und auf Fahrzeugdynamikprobleme mit Hindernisvermeidung angewendet haben.

- Ein Ansatz, der anstelle der HJB-Gleichung direkt das Feedback-Regelungsgesetz als neuronales Netz darstellt und für dieses Optimalitätsbedingungen in Form von KKT-Bedingungen herleiten. Insbesondere der Nachweis von Stabilitätseigenschaften von NN-Reglern ist bei diesem Ansatz hervorzuheben, siehe: Karl Kunisch and Daniel Walter (2021). “Semiglobal optimal feedback stabilization of autonomous systems via deep neural network approximation”. In: *ESAIM: Control, Optimisation and Calculus of Variations* 27, p. 16
- Ein Ansatz, der die Lösung der HJB-Gleichung in den Kontext eines Mean-Field-Games einbettet und durch die Berechnung einer Wahrscheinlichkeitsverteilung direkt Aussagen zur Stabilität und Robustheit der berechneten Lösung bietet, siehe: Lars Ruthotto, Stanley J Osher, Wuchen Li, Levon Nurbekyan, and Samy Wu Fung (2020). “A machine learning framework for solving high-dimensional mean field game and mean field control problems”. In: *Proceedings of the National Academy of Sciences* 117.17, pp. 9183–9193

Im Bereich der datenbasierten Optimalsteuerungsverfahren wurden folgende Arbeiten veröffentlicht:

- Ein robuster adaptiver MPC-Ansatz für nichtlineare Systeme, allerdings ohne Untersuchung der Performanz des geschlossenen Regelkreises András Sasfi, Melanie N. Zeilinger, and Johannes Köhler (2023). “Robust adaptive MPC using control contraction metrics”. In: *Automatica* 155, p. 111169. ISSN: 0005-1098
- Im Bereich der Reinforcement-Learning-basierten Ansätze wurde eine Methode vorgeschlagen, um die Parameter eines MPC-Reglers, wie die Gütefunktion und das Systemmodell, mit Reinforcement-Learning zu tunen: Sebastien Gros and Mario Zanon (2022). “Learning for MPC with stability & safety guarantees”. In: *Automatica* 146, p. 110598. ISSN: 0005-1098.

6 Veröffentlichungen

Publikationen:

- Jan Blechschmidt and Oliver G Ernst (2021). “Three ways to solve partial differential equations with neural networks—A review”. In: *GAMM-Mitteilungen* 44.2, e202100006. DOI: 10.1002/gamm.202100006
- Jan Blechschmidt (2022). “Numerical Methods for Stochastic Control Problems with Applications in Financial Mathematics”. PhD thesis. TU Chemnitz
- Francisco Moreno-Mora, Lukas Beckenbach, and Stefan Streif (2022). “Performance bounds of adaptive MPC with bounded parameter uncertainties”. In: *European Journal of Control* 68, p. 100688
- Lukas Beckenbach and Stefan Streif (2022). “Approximate infinite-horizon predictive control”. In: *2022 IEEE 61st Conference on Decision and Control (CDC)*. IEEE, pp. 3711–3717
- Francisco Moreno-Mora, Lukas Beckenbach, and Stefan Streif (2023). “Predictive Control with Learning-Based Terminal Costs Using Approximate Value Iteration”. In: *IFAC-PapersOnLine* 56.2. 22nd IFAC World Congress, pp. 3874–3879. ISSN: 2405-8963
- Urs Baumgart, Francisco Moreno-Mora, Lukas Beckenbach, Michael Burger, and Stefan Streif (2023). “Nonlinear Optimal Control of Traffic Flow with Stability Guarantees”. In: *IFAC-PapersOnLine* 56.2. 22nd IFAC World Congress, pp. 4965–4970. ISSN: 2405-8963
- Lukas Beckenbach, Pavel Osinenko, and Stefan Streif (2022). “A stabilizing reinforcement learning approach for sampled systems with partially unknown models”. In: *IEEE Transactions on Automatic Control*. In minor revision.

Masterarbeiten:

- Maximillian Kloppe (2021). “Unsicherheit bei neuronalen Netzen”
- Julius Peltri (2022). “Einsatz von Quadraturformeln basierend auf Leja-Punkten in Physics Informed Neural Networks”
- Salma Binsilm (2023). “Simulation-based development, optimization and comparison of methods for comfortable driving by means of adaptive cruise control”, in Kooperation mit FDTech GmbH

Vorträge:

- J. Blechschmidt, O. Ernst: Neuronale Netze: PDE-Löser und Unsicherheit; Konsortialtreffen SOPRANN, 2021
- J. Blechschmidt, O. Ernst: Optimalsteuerung mit neuronalen Netzen; Konsortialtreffen SOPRANN, 2022
- J. Blechschmidt, O. Ernst: Semiglobale Feedbackstabilisierung; Konsortialtreffen SOPRANN, 2022
- F. Moreno-Mora, L. Beckenbach, S. Streif: ADP-basierte modellprädiktive Regelung; Konsortialtreffen SOPRANN, 2022
- F. Moreno-Mora, L. Beckenbach, S. Streif: Adaptive Modellprädiktive Regelung und Stabilität des Ring-Roads unter MPC; Konsortialtreffen SOPRANN, 2022
- F. Moreno-Mora, L. Beckenbach, S. Streif: Performance bounds of adaptive MPC with bounded parameter uncertainties; European Control Conference; 2022
- J. Blechschmidt, O. Ernst: Neural Networks in Optimal Control; Workshop on Uncertainty Analysis, TU München, 2023
- J. Blechschmidt, O. Ernst: Neural Networks in Optimal Control; Research Seminar Numerical Mathematics, TU Chemnitz, 2024

Kurzbericht

zum

Verbundprojekt 05M2020 - SOPRANN: Synthese optimaler Regelungen und adaptiver Neuronaler Netze für Mobilitätsanwendungen. (Förderkennzeichen 05M20OCA)

Teilprojekt 2: Optimierungs- und Analysemethoden für NN

Teilprojektleitung:

Prof. Dr. Oliver Ernst, Professur für Numerische Mathematik

Fakultät für Mathematik,

Technische Universität Chemnitz, 09107 Chemnitz

Email: `oliver.ernst@math.tu-chemnitz.de`.

Teilprojekt 3: Stabilität und Adaptation von NN-basierten Regelungen

Teilprojektleitung:

Prof. Dr.-Ing. Stefan Streif, Professur für Regelungstechnik und Systemdynamik

Fakultät für Elektrotechnik und Informationstechnik,

Technische Universität Chemnitz, 09107 Chemnitz

Email: `stefan.streif@etit.tu-chemnitz.de`.

1 Aufgabenstellung

Der Einfluss von Methoden des maschinellen Lernens in technischen Anwendungen und speziell beim autonomen Fahren nimmt seit Jahren rapide zu mit dem Versprechen, komplexe Situationen sicher beherrschen zu können. Insbesondere im Bereich der maschinellen Wahrnehmung sind Deep Neural Networks inzwischen quasi zum Standard geworden. Im Bereich der Regelung finden aktuell ebenfalls sehr viele Forschungsaktivitäten statt, wobei die Absicherung der Funktionalität von lernenden Reglern immer noch ein im Allgemeinen ungelöstes Problem darstellt.

Hier setzte das Verbundprojekt SOPRANN (Synthese optimaler Regelungen und adaptiver Neuronaler Netze für Mobilitätsanwendungen) an mit dem Gesamtziel, datenbasierte Modelle und künstliche Intelligenz (KI) zur sicheren und optimalen Regelung im Bereich der Mobilität einzusetzen. Eine Herausforderung dabei war der Umgang mit der Komplexität, der Multimodalität und der Menge der Messdaten, die mit Hilfe von Neuronalen Netzen (NN) synthetisiert werden sollten. Die Auswahl geeigneter Netzarchitekturen und die Entwicklung von effizienten Trainingsverfahren sowie die Echtzeit-Adaptation der NN basierend auf neuen, während der Laufzeit gewonnenen Daten bei gleichzeitiger Gewährleistung von Stabilität, Optimalität und Sicherheit des Regelkreises – bildeten den mathematischen Forschungskern des Projekts.

Projektziele im Rahmen des Teilprojekts 2 – *Optimierungs- und Analysemethoden für NN* – des Verbundprojekts SOPRANN, welches von der Arbeitsgruppe Numerische Mathematik der Fakultät für Mathematik der TU Chemnitz durchgeführt wurde, waren unter anderem die Untersuchung von Methoden der Unsicherheitsquantifizierung für NN sowie die Entwicklung von neuen Berechnungsmethoden zur Lösung der für Optimalsteuerung dynamischer Systeme entscheidenden Hamilton-Jacobi-Bellman (HJB) Gleichung.

Im Teilprojekt 3 – *Stabilität und Adaptation von NN-basierten Regelungen*, durchgeführt von der Professur Regelungstechnik und Systemdynamik, war es Ziel, lernende Regler für Optimalsteuerungsprobleme mit unendlichem Horizont mit gesicherter Stabilität des geschlossenen Regelkreises unter Berücksichtigung von Sicherheitseinschränkungen zu entwerfen und zu untersuchen.

Die Arbeiten in beiden Teilprojekten erfolgten in regelmäßigem Austausch mit den Projektpartnern von der Universität der Bundeswehr München (Arbeitsgruppe Ingenieurmathematik der Fakultät für Luft- und Raumfahrttechnik) sowie vom Fraunhofer ITWM Kaiserslautern (Abteilung Dynamik, Lasten und Umgebungsdaten des Bereichs Mathematik für die Fahrzeugentwicklung).

2 Durchführung und Ergebnisse

Zu Beginn des Projekts wurden zusammen mit den anderen Verbundpartnern eine Reihe von Beispielproblemen festgelegt, die zum einen typischen Szenarien beim automatisierten Fahren möglichst nahe kommen, zum anderen jedoch mathematisch einfach beschreibbar sind. Diese reichen von einfachen Szenarien, bei denen ein Fahrzeug mit vorgegebener Dynamik ein bestimmtes Ziel unter Einhaltung von Zustands- und Steuerschranken erreichen soll bis hin zu Notfall-Ausweichmanövern, bei denen klassische gradientenbasierte Optimierungsverfahren oft an ihre Grenzen stoßen.

Im Teilprojekt 2 wurden für die Entwicklung von NN-basierten Verfahren zur Lösung der HJB Gleichung, welche eine Möglichkeit zur Lösung von optimalen Regelungsaufgaben darstellt, verschiedene neuartige Ansätze untersucht, darunter sogenannte *physikalisch informierte neuronale Netze (PINNs)* mit zufälligen Knoten sowie mittels Dünngitterkonstruktionen und Verfahren, die auf der linearen und nichtlinearen Feynman-Kac-Formel basieren. Deren grundsätzliche Eignung wurde zuerst an einer Reihe von mathematisch einfach beschreibbaren, jedoch keinesfalls einfach zu lösenden, Standardproblemen getestet und evaluiert. Diese schließen unter anderem Probleme, für die keine Lösung im klassischen Sinne existiert, sowie hochdimensionale Probleme mit ein. Anschließend fand eine schrittweise Anpassung der Ansätze auf die Aufgaben im Kontext von automatisiertem Fahren statt. Hierbei wurde schnell deutlich, dass insbesondere die exakte Abbildung der Fahrzeugdynamik sowie eine realitätsnahe Beachtung von Zustands- und Steuerbeschränkungen im Rahmen von NN-basierten Reglern sehr herausfordernd ist.

Aus diesem Grund wurden im weiteren Verlauf zwei neuartige Techniken, die Aufgaben der Optimalen Regelung unter Einsatz von KI-basierten Regelfunktionen, auf ihre Anwendbarkeit in diesem Kontext untersucht. Die erste Methode vermeidet dabei die Lösung der komplizierten HJB Gleichung und zielt auf das Erlernen eines NN-basierten Feedback-Reglers ab. Mit diesem Ansatz konnten erfolgreich Hindernisausweichprobleme gelöst werden. Eine zweite Methode basiert auf der Lösung von *mean-field games* und gliedert die Lösung der HJB Gleichung in einen größeren Kontext ein. Vorteil dieses Ansatzes ist ein im Vergleich zu den anderen Verfahren deutlich schnelleres Training. Auch hiermit konnten für einfache Dynamiken erfolgreich Hindernisausweichszenarios gelöst werden.

Methoden der Unsicherheitsquantifizierung unter Einsatz von neuronalen Netzen wurden parallel dazu für einfache mathematische Modelle untersucht und anschließend im Rahmen der entwickelten Verfahren untersucht und getestet. Beim Training der zugehörigen Netzwerke stellte sich jedoch heraus, dass das Training dieser *stochastischen* Netzwerke noch sensitiver ist als das Training der *deterministischen*. Die erhaltenen Ergebnisse konnten leider nicht die Hoffnung bestätigen, dadurch robustere Steuerungen für praktikable Anwendungen zu erhalten.

Zusammenfassend lässt sich sagen, dass alle Verfahren an akademischen Testbeispielen erfolgreich getestet werden konnten, jedoch klar wurde, dass nicht jeder untersuchte Ansatz Stand heute das Potential hat, in realistischen Anwendungen des automatisierten Fahrens zum Einsatz zu kommen. Fest steht, dass vor einer möglichen Anwendung auf realistische Mobilitätsszenarien weitergehende Entwicklungen notwendig sind.

Im Teilprojekt 3 wurden datenbasierte Optimalsteuerungsverfahren entwickelt und deren theoretischen Eigenschaften, wie Stabilität und Erfüllung der Systembeschränkungen untersucht. Im MPC-Kontext wurden zwei Architekturen untersucht.

Erstens, ein robuster adaptiver MPC-Ansatz mit online-gelerntem Modell. Dieser basiert auf der Minimierung einer Kostenfunktion. Diese ergibt sich aus der Summe einer vorgegebenen Gütefunktion, welche entlang der prädizierten Systemtrajektorie mithilfe der online-gelernten Modellparameter ausgewertet wird. Es konnte gezeigt werden, dass im Worst-Case die Performanz quadratisch abnimmt. Andere Faktoren, wie die Größe der Unsicherheitsmenge, können eine Verschlechterung der Performanz verursachen.

Zweitens, ein MPC-Ansatz mit gelernter Endkostenfunktion. Im Vergleich zum vorherigen Ansatz wird das dynamische Modell vom System als bekannt angenommen und der Fokus wird auf die Performanz des geschlossenen Kreises unter dem MPC-Regelgesetz gelegt. Insbesondere wird die Performanz untersucht, wenn die Endkostenfunktion basierend auf Reinforcement-Learning-Methoden, nämlich Policy Iteration und Value Iteration, bestimmt wird und die verwendeten iterativen Verfahren nicht exakt gelöst werden können. Um die Erfüllung der Sicherheitsbeschränkungen zu garantieren, wird eine stabilisierende Prädiktionslänge hergeleitet. In Simulationen hat sich diese als konservativ herausgestellt. Eine Kooperation mit anderen Forschungseinrichtungen für die Erweiterung dieser Ergebnisse in einem ökonomischen MPC-Kontext ist geplant.

Die gewonnen Erkenntnisse zur stabilisierenden Prädiktionslänge konnten in Kooperation mit dem Partner Fraunhofer ITWM auf das Szenario der Ring-Road angewendet werden. Die geforderte Prädiktionslänge für Stabilität würde für eine varierende Anzahl an menschlichen Fahrern untersucht. In diesem Fall ist das Ergebnis auch konservativ. Diese Resultate wurden im Rahmen einer gemeinsamen Publikation veröffentlicht und bilden die Basis für eine weitere Kooperation zwischen dem ITWM und der Professur Regelungstechnik und Systemdynamik der TU Chemnitz im Bereich sicherer autonomer Systeme für Mobilitätsanwendungen. Zudem werden die Erkenntnisse zum MPC-Ansatz mit Value Iteration gegenwärtig erweitert, um Online-Messungen zu integrieren und den Fall zu berücksichtigen, wenn die Systemdynamik nur zum Teil bekannt ist. Damit soll die Performanz des Regelgesetzes während des Betriebs verbessert werden.

Zuletzt wurde ein reiner Reinforcement-Learning-basierter Ansatz mit Sicherheitsgarantien für partiell unbekannte Systeme entworfen. Dieser basiert auf Policy Iteration und adaptiven Methoden, um die unbekannten Systemparameter zu bestimmen und die Performanz zu verbessern, ohne Offline-Training zu nutzen. Außerdem, wird die digitale Implementierung des Reglers berücksichtigt, um praktische Stabilität des geschlossenen Kreises zu gewährleisten. Der Ansatz wurde in Simulationen für Traktionsregelung und adaptive Geschwindigkeitsregelung (Adaptive Cruise Control) getestet. Es konnte eine signifikante Verbesserung der Performanz erreicht werden. Diese Ergebnisse spielen eine wichtige Rolle für die Arbeiten der Professur im Gemeinschaftsvorhaben Smart Rail Connectivity Campus¹, welches sich unter Anderem mit der Regelung von Zügen beschäftigt.

Die erzielten Ergebnisse wurden in insgesamt fünf Beiträgen auf internationalen Konferenzen und bei internationalen Journalen veröffentlicht.

¹<https://www.smart-rail-campus.de/>