

Approximation der zulässigen Parametermenge bei der Bounded-Error-Schätzung durch ein Ray-Shooting-Verfahren

Felix Wittich¹, Andreas Kroll²

^{1,2}Fachgebiet Mess- und Regelungstechnik, Universität Kassel
Mönchebergstr. 7, 34125 Kassel
E-Mail: {felix.wittich; andreas.kroll}@mrt.uni-kassel.de

1 Einführung

Sollen bei der datengetriebenen Modellbildung Unsicherheiten der zu schätzenden Parameter quantifiziert werden, wird typischerweise eine probabilistische Beschreibung der Unsicherheit herangezogen. Dabei wird eine bekannte Wahrscheinlichkeitsdichtefunktion (WDF) vorausgesetzt, bzw. es werden viele Daten für die Schätzung der WDF benötigt, was bei praktischen Anwendungsfällen meist nicht gegeben ist.

Ein alternativer Ansatz ist hier die mengenbasierte Bounded-Error-(BE-)Parameterschätzung [1]. Dabei wird die Annahme getroffen, dass der Prädiktionsfehler in einem Intervall mit garantierten Schranken liegt und gegeben der Annahmen wird die zulässige Parametermenge (Feasible Parameter Set, FPS) bestimmt [2]. Für einen Anwendungsfall der Randschichtprädiktion beim Hartdrehen mit Takagi-Sugeno-Multimodellen konnten BE-Verfahren in [3, 4] erfolgreich eingesetzt werden. Für einen Einsatz in der datengetriebenen Modellbildung mit einer großen Parameteranzahl n ergibt sich allerdings ein hoher Rechenaufwand für die exakte Bestimmung der zulässigen Parametermenge. Um diesem Problem bei der BE-Parameterschätzung zu begegnen, können Approximationsverfahren eingesetzt werden. Verbreitet werden dabei Hyperellipsoide als Kompromiss zwischen kompakter Beschreibung und Flexibilität

DOI: 10.58895/ksp/1000138532-6 erschienen in:

Proceedings - 31. Workshop Computational Intelligence : Berlin, 25. - 26. November 2021

DOI: 10.58895/ksp/1000138532 | <https://www.ksp.kit.edu/site/books/m/10.58895/ksp/1000138532/>

eingesetzt [5]. Allerdings kommt es hierbei zu starken Überschätzungen des FPS, da die Approximationen zwar in jedem Schritt optimal sind, global aber nicht. Somit kann nicht garantiert werden, dass die Fehlerschranken durch das approximierete Modell eingehalten werden.

Aus den oben genannten Problemen ist die Idee für ein neuartiges Verfahren zur approximativen Bestimmung der zulässigen Parametermenge entstanden. Dabei wird das aus der Computergrafik bekannte Ray-Tracing bzw. im n -Dimensionalen in der algorithmischen Geometrie als Ray-Shooting bekannte Verfahren der Strahlenverfolgung eingesetzt. Die Schnittpunkte des Ray-Shootings in konvexen Polytopen ergeben sich dabei als Lösung eines linearen Optimierungsproblems [6]. Durch die Strahlenverfolgung ergibt sich ein samplingbasierter Ansatz, der die Geometrie des konvexen Polytops des FPS ausnutzt. Das Ray-Shooting identifiziert garantiert eine Untermenge der wahren Parametermenge, wodurch die festgelegten Fehlerschranken immer eingehalten werden. In dieser Arbeit wird die Idee des Verfahrens konzeptionell vorgestellt und anhand einer Fallstudie demonstriert.

2 Methoden

2.1 Bounded-Error-Fehlerbeschreibung

In diesem Beitrag wird die Parameterschätzung aus einer mengentheoretischen Sichtweise betrachtet. Die Idee der Bounded-Error-Schätzung kann dabei wie folgt beschrieben werden. Es soll ein parametrisches Modell $\hat{y} = f(\boldsymbol{\theta}, \mathbf{x})$, mit dem Parametervektor $\boldsymbol{\theta} \in \mathbb{R}^n$ bestimmt werden, um den funktionalen Zusammenhang zwischen einer Ausgangsgröße $y \in \mathbb{R}$ und Eingangsgrößen $\mathbf{x} \in \mathbb{R}^{n_p}$ herzustellen. Das Modell soll dabei aus einem Datensatz $Z^N = \{\mathbf{x}(k), y(k)\}, k = 1, \dots, N$ gelernt werden. Das Ziel bei der BE-Schätzung ist dabei die Bestimmung einer zulässigen Untermenge des Parameterraums $\mathbb{S}_{\text{FPS}} \subseteq \mathbb{P}$, die zu einer Modellausgabe führt, die die Annahme einer spezifizierten Fehlerschrankenmenge \mathbb{E} erfüllt:

$$\mathbb{S}_{\text{FPS}} = \{\boldsymbol{\theta} \in \mathbb{R}^n \mid \mathbf{e}(\boldsymbol{\theta}) \in \mathbb{E}\} \quad (1)$$

Diese Menge wird als zulässige Parametermenge bezeichnet. Der Prädiktionsfehler wird als $e(k, \boldsymbol{\theta}) = y(k) - \hat{y}(k, \boldsymbol{\theta})$ definiert. Die Fehlerschranken werden als Intervalle beschrieben:

$$e(k, \boldsymbol{\theta}) \in [e_{\min}(k), e_{\max}(k)], k = 1, \dots, N \quad (2)$$

und können individuell für jeden Punkt festgelegt werden. Gewöhnlich werden dabei die selben symmetrischen Fehlerschranken $\delta = (e_{\max} - e_{\min})/2$ für jedes k angenommen. Somit folgt für (1):

$$\mathbb{S}_{\text{FPS}} = \{\boldsymbol{\theta} \in \mathbb{R}^n | y(k) - \delta \leq \hat{y}(k, \boldsymbol{\theta}) \leq y(k) + \delta \forall k\} \quad (3)$$

2.2 Bounded-Error-Schätzung für LiP-Modelle

Für Modelle, die linear-in-den-Parametern (LiP) sind, ist \mathbb{S}_{FPS} ein Polytop, wenn $N \geq d_{\text{ind}} > n$ unabhängige lineare Ungleichungen existieren. Ist ein LiP-Modell $\hat{y} = \boldsymbol{\varphi}^T \boldsymbol{\theta}$ mit dem Regressionsvektor $\boldsymbol{\varphi} \in \mathbb{R}^n$ und sind N Beobachtungen gegeben, schränkt jedes der N Intervalle

$$y(k) - \delta \leq \boldsymbol{\varphi}(k)^T \boldsymbol{\theta} \leq y(k) + \delta \quad (4)$$

das FPS durch zwei Ungleichungen im \mathbb{R}^n ein. Insgesamt existieren also $2N$ Ungleichungen. Ein Polytop \mathcal{P} kann durch die eingrenzenden Halbräume beschrieben werden:

$$\mathcal{P} = \{\boldsymbol{\theta} \in \mathbb{R}^{n_\theta} | \boldsymbol{\Phi} \boldsymbol{\theta} \leq \mathbf{Y}\} \quad (5)$$

mit

$$\mathbf{Y} = \begin{bmatrix} -y(1) + \delta \\ y(1) + \delta \\ \vdots \\ -y(N) + \delta \\ y(N) + \delta \end{bmatrix}, \boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\varphi}(1)^T \\ \boldsymbol{\varphi}(1)^T \\ \vdots \\ \boldsymbol{\varphi}(N)^T \\ \boldsymbol{\varphi}(N)^T \end{bmatrix}, \boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_{n_\theta} \end{bmatrix}, \quad (6)$$

somit folgt: $\boldsymbol{\Phi} \in \mathbb{R}^{2N \times n_\theta}$, $\boldsymbol{\theta} \in \mathbb{R}^{n_\theta}$, und $\mathbf{Y} \in \mathbb{R}^{2N}$.

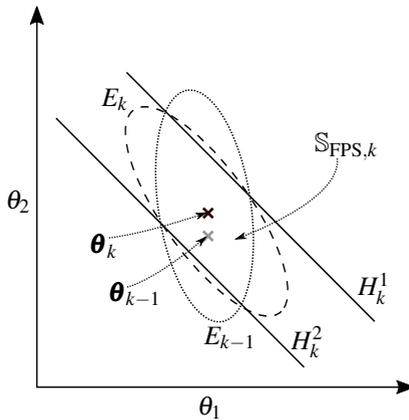


Bild 1: REOB-Verfahren (Dargestellt ist die k -te Iteration)

2.3 Rekursive Approximation des FPS mit Ellipsoiden

In der Literatur werden verschiedene Methoden zur approximativen Schätzung des FPS vorgeschlagen. Dazu wird das exakte FPS durch Geometrien wie Parallelotope, Zonotope oder Ellipsoide angenähert. Ein verbreiteter Ansatz ist dabei die rekursive Schätzung mit begrenzenden Ellipsoiden (Recursive Ellipsoidal Outer Bounding, REOB). Dabei wird eine äußere Approximation $\hat{\mathbb{S}}_{\text{FPS}} \supset \mathbb{S}_{\text{FPS}}$ durch die rekursive Berechnung der Schnittmengen von volumenminimalen Ellipsoiden E_k in der k -ten Iteration mit dem k -ten Paar der Halbebenen H_k^1 und H_k^2 berechnet, wie in Bild 1 illustriert. Eine ausführliche Beschreibung des Verfahrens ist in [5] zu finden. Neben den Verfahren zur äußeren Einschränkung des FPS existieren auch rekursive Verfahren zur inneren Einschränkung, die allerdings zu einer starken Unterschätzung tendieren [2].

2.4 Lokal-affine Multi-Modell

Als parametrischer Modellansatz werden im Folgenden lokal-affine Multimodelle betrachtet. Diese zeichnen sich durch eine hohe Modellflexibilität bei

kompakter Modellstruktur aus und können für die Identifikation nichtlinearer Systeme verwendet werden.

Dabei werden $c \in \mathbb{N}_+$ Teilmodelle $\hat{y}_j = f(\boldsymbol{\theta}_{j,\text{LM}}, \tilde{\boldsymbol{\varphi}}) : \mathbb{R}^{n_p} \rightarrow \mathbb{R}$, durch Fuzzy-Basisfunktionen $\phi_j(\mathbf{z}) : \mathbb{R}^{n_z} \rightarrow [0, 1]$ überlagert, welche von Schedulingvariablen $\mathbf{z} = [z_1 \dots z_{n_z}]^T \in \mathbb{R}^{n_z}$ abhängen:

$$\hat{y}(\mathbf{z}, \boldsymbol{\theta}, \tilde{\boldsymbol{\varphi}}) = \sum_{j=1}^c \phi_j(\boldsymbol{\theta}_{\text{MF}}, \mathbf{z}) \hat{y}_j(\boldsymbol{\theta}_{j,\text{LM}}, \tilde{\boldsymbol{\varphi}}). \quad (7)$$

Dabei werden affine Teilmodelle verwendet:

$$\hat{y}_j(\boldsymbol{\theta}_{j,\text{LM}}, \tilde{\boldsymbol{\varphi}}) = \sum_{r=0}^n \theta_{j,r,\text{LM}} \cdot \tilde{\varphi}_r = \boldsymbol{\theta}_{j,\text{LM}} \cdot \tilde{\boldsymbol{\varphi}}, \quad (8)$$

mit dem r -ten Element $\tilde{\varphi}_r$ des Regressionsvektors

$$\tilde{\boldsymbol{\varphi}} = [1 \ x_1 \ \dots \ x_{n_p}]^T \quad (9)$$

und dem r -ten Element $\theta_{j,r,\text{LM}}$ des entsprechenden lokalen Parametervektor $\boldsymbol{\theta}_{j,\text{LM}} \in \mathbb{R}^n$. Das Multimodell (7) kann dann wie folgt geschrieben werden:

$$\hat{y} = \tilde{\boldsymbol{\varphi}}^T \boldsymbol{\theta}_{\text{LM}} \quad (10)$$

mit dem erweiterten Regressionsvektor

$$\tilde{\boldsymbol{\varphi}} = [\phi_1 \ \phi_1 x_1 \ \dots \ \phi_1 x_{n_p} \ | \ \dots \ | \ \phi_c \ \phi_c x_1 \ \dots \ \phi_c x_{n_p}]^T \quad (11)$$

und dem Vektor der lokalen Modellparameter

$$\begin{aligned} \boldsymbol{\theta}_{\text{LM}}^T &= [a_{0,1} \ a_{1,1} \ \dots \ a_{n_p,1} \ | \ \dots \ | \ a_{0,c} \ a_{1,c} \ \dots \ a_{n_p,c}] \\ &= [\boldsymbol{\theta}_{1,\text{LM}}^T \ \dots \ \boldsymbol{\theta}_{c,\text{LM}}^T] \in \mathbb{R}^{n \times c}. \end{aligned} \quad (12)$$

Dabei sind

$$\phi_j(\mathbf{z}) = \frac{\mu_j(\mathbf{z})}{\sum_{m=1}^c \mu_m(\mathbf{z})}, \quad (13)$$

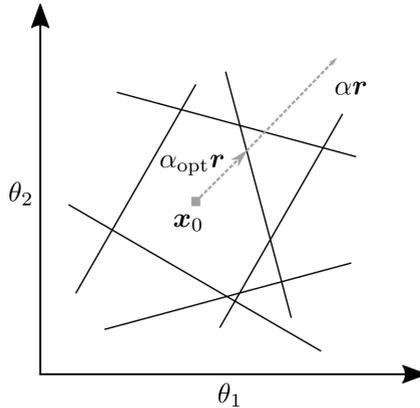


Bild 2: Ray-Shooting im konvexen Polytop

die Fuzzy-Basisfunktionen mit der Fuzzy-C-Means-Zugehörigkeitsfunktion

$$\mu_j(\mathbf{z}) = \left[\sum_{i=1}^c \left(\frac{\|\mathbf{z} - \mathbf{v}_j\|_2}{\|\mathbf{z} - \mathbf{v}_i\|_2} \right)^{\frac{2}{v-1}} \right]^{-1}. \quad (14)$$

Die Partitionierungsparameter $\mathbf{v}_j, \mathbf{v}_i \in \mathbb{R}^{n_z}$ werden im Parametervektor $\boldsymbol{\theta}_{\text{MF}} = [\mathbf{v}_1^T, \dots, \mathbf{v}_c^T]^T$ aggregiert. Im Folgenden wird für die Schedulingvariable $\mathbf{z} = \mathbf{x}$ angenommen.

2.5 Ray-Shooting in konvexen Polytopen

Das Prinzip des Ray-Shooting wird in Bild 2 dargestellt. Ein Strahl wird ausgehend von einem initialen Punkt $\mathbf{x}_0 \in \mathbb{R}_n$ im Inneren eines durch M Halbebenen $H_k, k = 1, \dots, M$ beschriebenen konvexen Polytops gesendet. Das Ziel ist es, den Schnittpunkt $\alpha_{\text{opt}} \mathbf{r} + \mathbf{x}_0$ zwischen dem Strahl und der ersten Halbebene, die getroffen wird, zu bestimmen. Die Richtung des Strahls wird dabei durch den Vektor $\mathbf{r} \in \mathbb{R}^n$ und die Länge des Strahls durch $\alpha \in \mathbb{R}^+$ festgelegt.

Der Strahl geht von einem initialen Punkt $\mathbf{x}_0 \in \mathbb{R}_n$ im inneren des Polytops aus. Um α_{opt} zu bestimmen wird hierzu das lineare Optimierungsproblem:

$$\begin{aligned} \alpha_{\text{opt}} &= \max \alpha \\ \text{u.d.N. } \mathbf{A}(\alpha \mathbf{r} + \mathbf{x}_0) &\leq \mathbf{b} \end{aligned} \quad (15)$$

gelöst, das sich in polynomialer Zeit berechnen lässt.

2.6 Ray-Shooting für die Approximation des FPS

Die Idee besteht darin, das Ray-Shooting in konvexen Polytopen für eine innere Approximation des FPS bei LiP-Modellen einzusetzen. Für die Bounded-Error-Schätzung wird das Ray-Shooting im Raum der Parameter $\boldsymbol{\theta} \in \mathbb{R}_n$ durchgeführt, um Randpunkte des durch (5) gegebenen Polytops zu finden. Als Suchgebiet wird die achsparallele begrenzende Box gewählt, also das (Hyper-)Rechteck, das durch $[\theta_{\min,1}, \theta_{\max,1}] \times [\theta_{\min,2}, \theta_{\max,2}] \times \cdots \times [\theta_{\min,n}, \theta_{\max,n}]$ gegeben ist. Dieses Rechteck lässt sich durch das Lösen von $2n$ linearen Programmen:

$$\theta_{\min,k} = \min \theta_k \quad (16)$$

$$\theta_{\max,k} = \max \theta_k \quad (17)$$

$$\text{u.d.N. } \mathbf{A}\boldsymbol{\theta} \leq \mathbf{b} \quad (18)$$

für $k = 1, \dots, n$ finden [1]. Es werden Parametervektoren zufällig gleichverteilt innerhalb der achsparallelen Box erzeugt und mit (4) auf Zulässigkeit überprüft. Die zulässigen Punkte werden dann als initiale Punkte $\boldsymbol{\theta}_0$ verwendet. Um eine gute Abdeckung im Parameterraum zu erreichen, müssen bei hoher Parameteranzahl n entsprechend viele Punkte initialisiert werden, da deren Dichte exponentiell mit zunehmendem n sinkt. Die Richtungsvektoren \mathbf{r} , die die Suchrichtungen vorgeben, werden ebenfalls zufällig gleichverteilt gewählt. Dazu werden diese als Vektoren vom Ursprung zur Oberfläche einer n -dimensionalen Einheitskugel festgelegt. Die mit dem Ray-Shooting aus Abschnitt 2.5 bestimmten Randpunkte bilden dann die konvexe Hülle der approximierten zulässigen Parametermenge $\text{FPS}_{\text{RS-BE}}$.

3 Fallstudie

Die RS-BE-Methode soll anhand einer akademischen Fallstudie demonstriert werden. Hierzu wurden $N = 50$ Trainingsdaten mit einer einfachen nichtlinearen Funktion $y = f(x) = x^2$ gleichverteilt im Intervall $[-1; 1]$ erzeugt. Der Ausgang $\tilde{y} = y + d$ wurde mit gleichverteiltem Rauschen $d \sim \mathcal{U}(-0,05; 0,05)$ beaufschlagt. Für die Modellierung wurden TS-Multi-Modelle mit $v = 1,4$ und $c = 4$ Teilmodellen herangezogen, wobei die Clusterzentren äquidistant im Intervall der Eingangsgröße verteilt wurden. Im Fallbeispiel wird die Unsicherheit durch die lokalen Modellparametern θ_{LM} ausgedrückt und die Partitionierungsparameter θ_{MF} werden als fixiert angenommen, wodurch ein LiP-Model resultiert. Als zulässige Fehlerschranke wurde $\delta = 0,15$ gewählt. Weitere Informationen zu Modellansatz und Identifikation sind in [3] zu finden. Die zulässige Parametermenge FPS_{RS-BE} wurde mit dem RS-BE-Verfahren aus Abschnitt 2.6 bestimmt. Es wurden 10^6 initiale Punkte θ_0 erzeugt und für jeden der 221 zulässigen Punkte wurden 100 zufällige Richtungsvektoren r erzeugt. Somit wurden insgesamt 22100 Randpunkte gefunden. Die Rechenzeit für das Sampling der zulässigen initialen Punkte betrug 1,8 s und für die Ray-Shooting-Prozedur 21,4 s. Außerdem wurde eine Approximation mit dem REOB-Verfahren aus Abschnitt 2.3 bestimmt. Die Rechenzeit beträgt hierbei 2,6 s.

In Bild 3 sind die geschätzten zulässigen Parametermengen für die einzelnen Teilmodelle dargestellt. Dabei ist die deutliche Überschätzung mit dem REOB-Verfahren erkennbar. In Bild 4 sind die Trainingsdaten mit Fehlerschranken und die Prädiktionsfehlerschranken für beide Ansätze dargestellt. Es ist zu erkennen, dass RS-BE die vorgegebenen Fehlerschranken einhält, während REOB die Schranken deutlich reißt. Alle Berechnungen wurden auf einer Workstation mit Intel i5-6500 3,2 GHz CPU und 16 GB RAM in MATLAB durchgeführt. Dabei wurde die Implementierung nicht auf Parallelisierbarkeit hin optimiert. Dies ist ein potentieller Ansatzpunkt, die Geschwindigkeit des Verfahrens zu erhöhen.

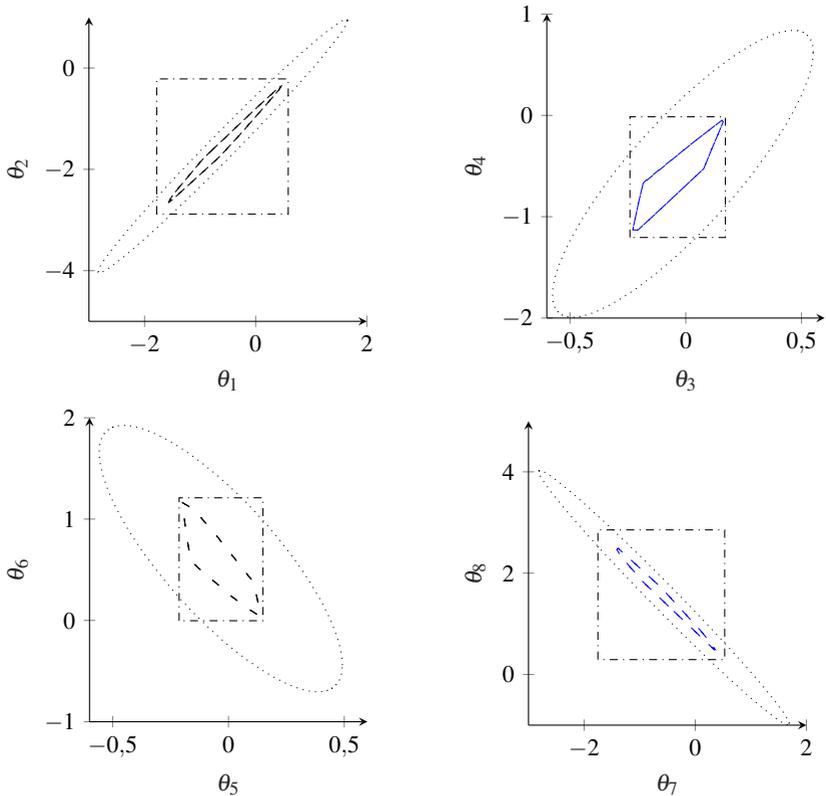


Bild 3: Projektionen für die Ergebnisse der Parameterschätzung mit REOB (gepunktet) und RS-BE (gestrichelt), sowie die achsparallelen begrenzenden Boxen (gestrichpunktet)

4 Zusammenfassung und Ausblick

In diesem Beitrag wurden die Idee und erste Ergebnisse für ein neuartiges Verfahren für die Bounded-Error-Schätzung bei LiP-Modellen vorgestellt. Es wurde gezeigt, dass eine Approximation der zulässigen Parametermenge gefunden wird, die eine Modellprädiktion innerhalb der vorgegebenen Fehler-schranken gewährleistet. Das Verfahren lässt sich über die Anzahl der initialen Punkte und der ausgesendeten Strahlen skalieren. Somit erlaubt das Verfahren

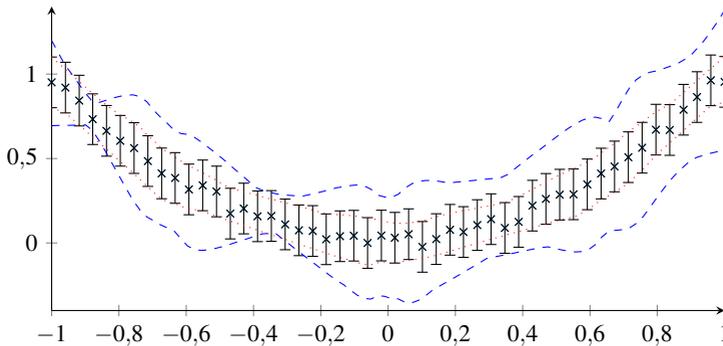


Bild 4: Trainingsdaten (x) mit Fehlerschranke (Fehlerbalken) und Prädictionsschranken für FPS_{REOB} (gestrichelt) und FPS_{RS-BE} (gepunktet)

durch Festlegen von Abbruchkriterien den Trade-Off zwischen Genauigkeit und Rechenzeit einzustellen.

Als samplingbasiertes Verfahren unterliegt es allerdings dem "Curse of Dimensionality", d.h. bei steigender Parameterdimension werden exponentiell mehr Samplingpunkte benötigt. Hierzu soll untersucht werden, wie ein effizienteres Sampling der initialen Punkte möglich ist. Zudem soll in Zukunft untersucht werden, inwiefern sich der Ansatz der Strahlenverfolgung auch für nicht-LiP BE-Schätzungen eignet. Dabei treten aber weitere Probleme wie nicht-konvexe und nicht-zusammenhängende Parametermengen auf. Zudem lässt sich das Ray-Shooting dann nicht mehr als Lösung eines linearen Programms (15) formulieren.

5 Danksagung

Dieser Beitrag wurde von der Deutschen Forschungsgemeinschaft DFG im Rahmen des Schwerpunktprogramms SPP 2086 (KR 3795/8-1) gefördert. Die Autoren danken der DFG für die finanzielle und technische Unterstützung.

Literatur

- [1] M. Milanese, J. Norton, H. Piet-Lahanier, É. Walter. „Bounding approaches to system identification“. New York: Springer. 2013.
- [2] E. Walter, H. Piet-Lahanier. „Estimation of parameter bounds from bounded-error data: a survey“. In: *Mathematics and Computers in Simulation* Bd. 32. S. 449–468. 1990.
- [3] F. Wittich, M. Kahl, A. Kroll. „Zur Schätzung zulässiger Parametermengen nichtlinearer Takagi-Sugeno-Multi-Modelle mit garantierten Fehlerschranken“. In: *Proc., 29. Workshop Computational Intelligence* (Mikut, R.; Reischl, M., Hg.), S. 247–254. KIT Scientific Publishing. 2019.
- [4] F. Wittich, L. Kistner, A. Kroll, C. Schott, T. Niendorf. „On data-driven nonlinear uncertainty modeling: Methods and application for control-oriented surface condition prediction in hard turning“. In: *tm - Technisches Messen*, Bd. 87, Nr. 11, S. 732–741. 2020.
- [5] G. Favier, L. Arruda. „Review and comparison of ellipsoidal bounding algorithms“. In: *Bounding approaches to system identification* S. 42-68. New York: Springer. 1996.
- [6] J. Matoušek, O. Schwarzkopf. „On ray shooting in convex polytopes“. In: *Discrete & Computational Geometry* Bd. 10. S. 215–232. 1993.