Systemidentifikation eines Einmassenschwingers mit spärlicher linearer Regression

Yaxiong Ren[1], Christian Adams[1], Tobias Melz[1,2]
[1] *Technische Universität Darmstadt, Fachbereich Maschinenbau, Fachgebiet Systemzuverlässigkeit, Adaptronik und Maschinenakustik SAM, Magdalenenstr. 4, 64289 Darmstadt*
E-Mail: yaxiong.ren@sam.tu-darmstadt.de
[2] *Fraunhofer-Institut für Betriebsfestigkeit und Systemzuverlässigkeit LBF, Bartningstraße 47, 64289 Darmstadt*

Einleitung

Systemidentifikation als eine Methodik der Modellbildung bedeutet Entwicklung oder Optimierung von Modellen mithilfe von Eingangs- und Ausgangssignalen aus Messungen [1]. Innerhalb dieser Gruppe zeigt *Spärliche Identifikation Nichtlinearer Dynamischer Systeme* (SINDy) [2], die auf Differentialpolynomen und spärlicher linearer Regression beruht, ein hohes Potenzial zur Identifikation von Schwingungssystemen [3]. In dieser Arbeit wird ein Systemidentifikationsverfahren auf Basis von SINDy vorgestellt. Die originale spärliche Regression von SINDy ist die *Kleinste-Quadrate-Methode mit festem Threshold* (KQFT). Diese Methode wird durch eine Alternative ersetzt, die *Kleinste-Quadrate-post-LASSO* (KQPL). LASSO steht für Least Absolute Shrinkage and Selection Operator. Der Vergleich beider Methoden erfolgt durch numerische Untersuchungen anhand eines Einmassenschwingers im Rahmen von Monte-Carlo-Simulationen (MCSs). Den numerisch generierten Trainingsdaten wird Rauschen unterschiedlicher Intensität hinzugefügt, um reale Messsignale nachzubilden. Die SINDy-Modelle werden für Polynomordnungen von eins bis fünf identifiziert und im Hinblick auf Spärlichkeit, Konvergenz, Eigenfrequenz und Bestimmtheitsmaß ausgewertet.

Identifikation eines Einmassenschwingers mit SINDy

Das Systemidentifikationsverfahren SINDy basiert auf Differentialpolynomen und spärlicher Regression. Angenommen, die Bewegungsgleichung eines zu identifizierenden Systems sei

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t)) = (\boldsymbol{\theta}(\boldsymbol{x}^{\mathrm{T}}(t))\boldsymbol{\Xi})^{\mathrm{T}} \tag{1}$$

mit dem Zustandsvektor $\boldsymbol{x} \in \mathbb{R}^{n_x}$ und einer polynomischen Modellstruktur mit zeitunabhängigen Parametern in der Koeffizientenmatrix $\boldsymbol{\Xi} \in \mathbb{R}^{n_\xi \times n_x}$. Die i-te Spalte der Matrix $\boldsymbol{\Xi}$, also $\boldsymbol{\xi}_i \in \mathbb{R}^{n_\xi}$, entspricht der i-ten Dimension des zeitlich abgeleiteten Zustandsvektors $\dot{\boldsymbol{x}}(t)$. Der n_ξ-dimensionale Zeilenvektor $\boldsymbol{\theta}(\boldsymbol{x}^{\mathrm{T}}(t))$ umfasst die Konstante 1 und alle Potenzen des Vektors $\boldsymbol{x}(t)$ bis zu einer bestimmten Ordnung, die auch Kandidatenfunktionen genannt werden, beispielsweise $\boldsymbol{\theta}([x_1, x_2])|_{Ord=2} = [1, x_1, x_2, x_1^2, x_1x_2, x_2^2]$. Diese zeitlichen Signale von $\dot{\boldsymbol{x}}$ und $\boldsymbol{\theta}(\boldsymbol{x}^{\mathrm{T}})$ aus Messungen und Berechnungen können in den Matrizen

$$\dot{\boldsymbol{X}} = \begin{bmatrix} \dot{\boldsymbol{x}}^{\mathrm{T}}(t_1) \\ \dot{\boldsymbol{x}}^{\mathrm{T}}(t_2) \\ \vdots \\ \dot{\boldsymbol{x}}^{\mathrm{T}}(t_{n_t}) \end{bmatrix}, \boldsymbol{\Theta}(\boldsymbol{X}) = \begin{bmatrix} \boldsymbol{\theta}(\boldsymbol{x}^{\mathrm{T}}(t_1)) \\ \boldsymbol{\theta}(\boldsymbol{x}^{\mathrm{T}}(t_2)) \\ \vdots \\ \boldsymbol{\theta}(\boldsymbol{x}^{\mathrm{T}}(t_{n_t})) \end{bmatrix} \tag{2}$$

zusammengefasst werden. Somit wird die Gleichung (1) zu einer Regressionsaufgabe

$$\dot{\boldsymbol{X}} = \boldsymbol{\Theta}(\boldsymbol{X})\boldsymbol{\Xi}. \tag{3}$$

Das Modell wird durch die Bestimmung der Parameter erstellt. Weil bei den meisten physikalischen Systemen nur einige Terme von den Kandidatenfunktionen relevant sind, ist die Koeffizientenmatrix $\boldsymbol{\Xi}$ spärlich. Die Bewegungsgleichung eines Einmassenschwingers kann mit der Gleichung

$$\begin{bmatrix} \dot{q} \\ \ddot{q} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{d}{m} \end{bmatrix} \begin{bmatrix} q \\ \dot{q} \end{bmatrix} \tag{4}$$

im Zustandsraum dargestellt werden. Mit den Kandidatenfunktionen höherer Ordnungen ist die kleinste-Quadrate-(KQ-)Methode für eine beliebige i-te Dimension

$$\hat{\boldsymbol{\xi}}_i = \underset{\boldsymbol{\xi}_i \in \mathbb{R}^{n_\xi}}{\operatorname{argmin}} \{\|\dot{\boldsymbol{x}}_i - \boldsymbol{\Theta}\boldsymbol{\xi}_i\|_2^2\} \tag{5}$$

$$= (\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\Theta})^{-1}\boldsymbol{\Theta}^{\mathrm{T}}\dot{\boldsymbol{x}}_i \tag{6}$$

nicht geeignet. Da ein Parameter mit dem wahren Wert Null wegen des real vorhandenen Messrauschens als ein Nicht-Null-Wert abgeschätzt wird, tritt Overfitting (Überanpassung) auf. Daher wird spärliche Regression eingesetzt.

Eine Möglichkeit der spärlichen Regression ist die KQFT-Methode. Spärlichkeit wird durch iterative Nullsetzung der abgeschätzten Koeffizienten unterhalb eines festen Thresholds λ_{KQFT} erreicht. Unter der Annahme, dass das Messrauschen in den Signalen niedrig ist, sind die fälschlich als Nicht-Null-Wert geschätzten Koeffizienten im Vergleich mit den tatsächlichen Nicht-Null-Koeffizienten absolut gesehen kleiner. Somit lässt sich ein Threshold einsetzen. Der Ablauf von KQFT ist in der Abbildung 1 dargestellt.

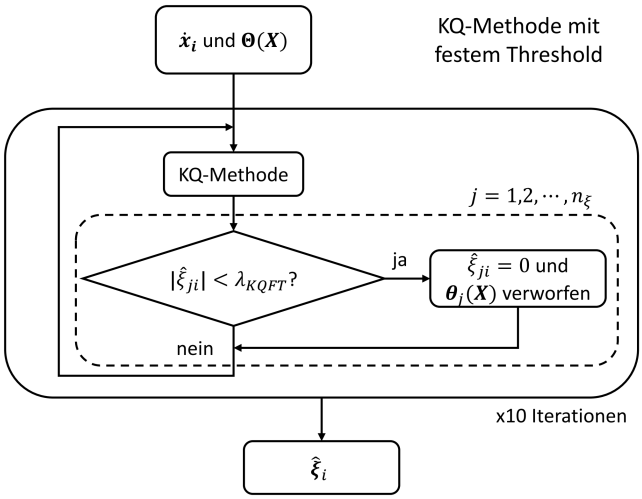

Abbildung 1: Ablauf der KQFT Methode

Ein passender Threshold wird nicht gefunden, wenn ein Null-Wert-Koeffizient größer als ein Nicht-Null-Koeffizient abgeschätzt wird. Diese Situation tritt häufig auf, wenn die Koeffizienten verschiedene Größenordnungen haben.

KQ-post-LASSO

LASSO ist ein Optimierungsverfahren zur Abschätzung der Koeffizienten. Es automatisiert den Aufbau von Modellen mit Spärlichkeit. Allerdings ist die Schrumpfung der abgeschätzten Koeffizienten nicht wünschenswert, d. h. die Absolutwerte der Koeffizienten werden kleiner als die wahren Werte abgeschätzt. Um die Einschränkung zu vermeiden, kann die KQ-Methode nach LASSO mit ausgewählten Kandidatenfunktionen durchgeführt werden. Diese Methode wird KQ-post-LASSO (KQPL, engl. LS post-LASSO) genannt [4]. Der Ablauf von KQPL ist in der Abbildung 2 dargestellt.

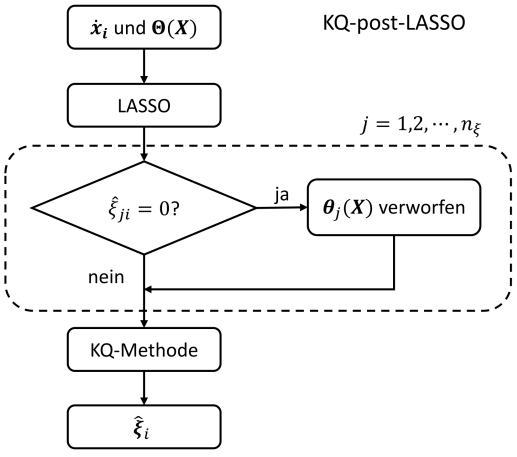

Abbildung 2: Ablauf der KQPL-Methode

Numerische Untersuchung

In diesem Abschnitt wird SINDy in den numerischen Untersuchungen implementiert und mit MCSs bewertet. Der Vorteil der numerischen Untersuchung liegt darin, dass die realen Werte der identifizierten Koeffizienten bekannt sind. Damit können die Eigenschaften des Identifikationsverfahrens bewertet werden.

Das zu untersuchende Schwingungssystem ist ein freischwingender Einmassenschwinger mit der Bewegungsgleichung (4). Mit der Anfangsbedingung $\boldsymbol{x}(0)$ lassen sich die zeitlichen Signalen von Verschiebung $q(t)$, Schnelle $\dot{q}(t)$ und Beschleunigung $\ddot{q}(t)$ als eine Lösung eines Anfangswertproblems berechnen. Um die realen Messsignale nachzubilden, wird Rauschen $S(t)$ in $q(t), \dot{q}(t)$ und $\ddot{q}(t)$ hinzugefügt. Das stochastische Rauschen $S(t)$, allgemein für ein beliebiges Signal $Signal(t)$, wird als die Normalverteilung

$$S(t) \sim \mathcal{N}(0, (\epsilon_R \max(Signal(t)))^2) \tag{7}$$

definiert. Der Rauschfaktor ϵ_R bezeichnet die Intensität des künstlichen Rauschens. Unter Berücksichtigung der Unsicherheit werden 10.000 Wiederholungen im Rahmen der MCSs als Trainingsdaten generiert. Innerhalb der MCSs sind die Parameter Masse m, Steifigkeit k und Dämpfung d in der Gleichung (4) normalverteilt. Die mit Rauschen generierten Signale $\boldsymbol{X}$ und $\dot{\boldsymbol{X}}$ dienen als Trainingssätze zur Identifikation. Außerdem werden die Testsätze mit unterschiedlichen Anfangsbedingungen erzeugt, um die Ergebnisse umfangreicher auszuwerten. Die Testsätze enthalten kein Messrauschen, weil die Testsätze nur als Referenz verwendet werden.

Die Systemidentifikation aus den Trainingssätzen erfolgt mit zwei Verfahren: SINDy mit KQFT und SINDy mit KQPL. Zum Test der Spärlichkeit eines Modells werden auch die Polynome höherer Ordnung als Kandidatenfunktionen angenommen. Die Polynomordnung Ord der Modellstruktur steigt von eins auf fünf. Anschließend werden die zeitlichen Signalen mit numerischer Integration aus den polynomischen Modellen berechnet. In der Tabelle 1 sind die realen Werte der eingesetzten Parameter aufgelistet. Dabei stammen die Verteilungen der Parametern m, d und k aus Erfahrungswerten [5]. Die Rauschfaktoren sind äquivalent zu den Signal-Rausch-Verhältnissen (engl.: signal to noise ratio, SNR) $\{\infty; 40{,}1; 34{,}1; 30{,}5; 28{,}0\}$ in dB. Der Threshold von KQFT λ_{KQFT} liegt unterhalb des kleinsten Mittelwertes „1“ von den realen Parametern mit einem zweckmäßigen Abstand um 0,5. Im Vergleich zur Anfangsbedingung $\boldsymbol{x}(0)$ haben die Anfangsbedingungen $\boldsymbol{x}_1^*(0)$ und $\boldsymbol{x}_2^*(0)$ ungefähr die doppelte bzw. die 20-fache Amplitude, wie in der Abbildung 3 gezeigt wird.

Parameter	Wert
m in kg	$\mathcal{N}(5; 0{,}05^2)$
d in kg/s	$\mathcal{N}(5; 0{,}5^2)$
k in kg/s^2	$\mathcal{N}(2000; 66{,}7^2)$
ϵ_R	$\{0; 0{,}0025; 0{,}005; 0{,}0075; 0{,}01\}$
$\boldsymbol{x}(0)$	$[0; 1]$
$\boldsymbol{x}_1^*(0)$	$[0{,}1; 0]$
$\boldsymbol{x}_2^*(0)$	$[1; 0]$

Tabelle 1: Parameter des Einmassenschwingers für die numerische Untersuchung

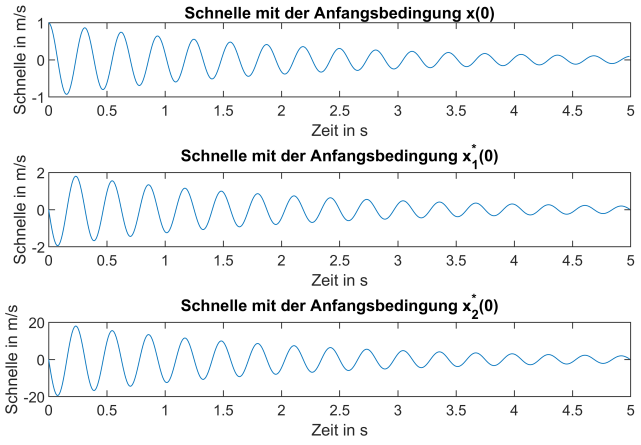

Abbildung 3: Schwingschnelle mit den Anfangsbedingung $\boldsymbol{x}(0)$ (oben), $\boldsymbol{x}_1^*(0)$ (Mitte) und $\boldsymbol{x}_2^*(0)$ (unten)

Auswertungskriterien

Zur Auswertung eines identifizierten Modells werden die Koeffizientenmatrix $\boldsymbol{\Xi}$ und das zeitliche Signal der Beschleunigung $\ddot{\boldsymbol{q}}(t)$ verwendet. Einerseits besitzt die zweite Zeile der Gleichung (4) alle maßgeblichen Koeffizienten, andererseits enthält die Beschleunigung mehr dynamische Information im Vergleich mit der Schnelle oder Verschiebung [6].

In dieser Untersuchung bleibt die wahre Anzahl der Nicht-Null-Parameter (ANNP) in der zweiten Zeile der Koeffizientenmatrix $\boldsymbol{\Xi}$ gleich zwei. Mit dem Vergleich zwischen den identifizierten ANNPs und dem idealen Wert zwei kann die Spärlichkeit der Matrix $\boldsymbol{\Xi}$ ausgewertet werden.

Bei der MCS treten häufig physikalisch unrealistische Modelle auf, deren numerische Integration mit einer bestimmten Anfangsbedingung nicht konvergiert. Das Verhältnis der nicht konvergierten Integrationen zur Gesamtzahl der Integrationen (VNKI) wird bei jeder MCS berechnet und als ein Auswertungskriterien verwendet.

Außerdem werden die identifizierte Eigenfrequenz und das Bestimmtheitsmaß R^2 zur Auswertung verwendet.

Ergebnisse

Die ANNPs bei fünf Polynomordnungen und fünf Rauschfaktoren (insgesamt 25 MCSs) weisen nicht symmetrische Verteilungen auf. In der Abbildung 4 sind die Verteilungen von ANNPs durch Median und 0,25- bzw. 0,75-Quantile in der Form von Fehlerbalken visualisiert.

Bei den Simulationen ohne Rauschen und/oder mit der Polynomordnung $Ord = 1$ fallen die ANNPs beider Abschätzverfahren streuungsfrei mit dem wahren Wert zwei zusammen. In den Fällen mit Rauschen steigen die ANNPs bei KQFT mit größerer Polynomordnung, wenn $Ord \geq 2$, wie in Abbildung 4 auf der linken Seite gezeigt. Je stärker das Rauschen ist, desto näher liegen die ANNPs an der Anzahl der Kandidatenfunktionen. Demgegenüber stimmen bei KQPL die Mediane der ANNPs aller MCSs mit dem wahren Wert überein. Die Abweichungen der 0,75-Quantile treten mit $Ord \geq 3$ und $\epsilon_R \geq$ 0,0075 auf und betragen eins. Im Sinne der Spärlichkeit zeigt die Methode KQPL einen beträchtlichen Vorteil gegenüber KQFT. Außerdem ist die Tendenz beider Methoden zu erkennen, dass sich die Gefahr des Overfittings mit größerer Polynomordnung und stärkerem Rauschen erhöht.

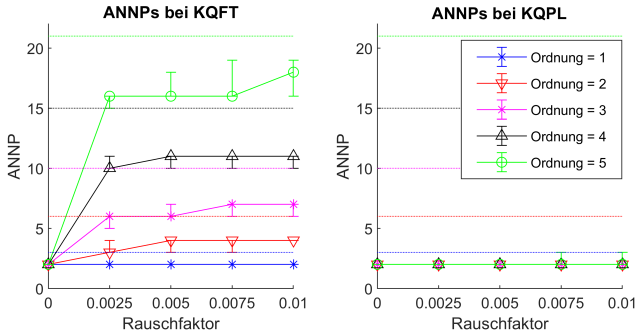

Abbildung 4: Mediane und Quantile der ANNPs bei KQFT und KQPL

Die Abbildung 5 zeigt vier Gruppen von MCSs zur Auswertung des VNKIs, die mithilfe der zwei Anfangsbedingungen $\boldsymbol{x}_1^*(0)$ und $\boldsymbol{x}_2^*(0)$ sowie mit den beiden Identifikationsverfahren KQFT und KQPL entstanden sind.

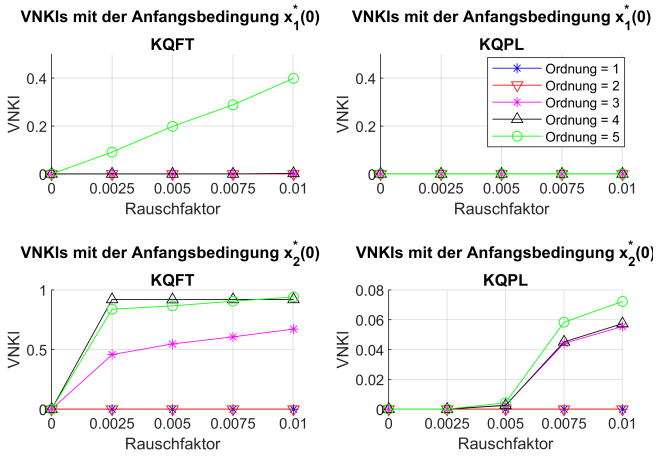

Abbildung 5: VNKI bei KQPT und KQPL unter den Anfangsbedingungen $\boldsymbol{x}_1^*(0)$ und $\boldsymbol{x}_2^*(0)$

Bei der Anfangsbedingung $\boldsymbol{x}_1^*(0)$ konvergiert die Integration in den MCSs mit KQFT zufriedenstellend bis Ordnung 4. Das maximale VNKI beträgt 0,24% bei $\epsilon_R = 0{,}01$ und $Ord = 4$. Die wesentliche Verschlechterung des VNKIs tritt bei $Ord = 5$ auf, das proportional zum Rauschfaktor bis auf 40% steigt. Demgegenüber bleiben die VNKIs jeder MCS bei KQPL null.

Die zweite Anfangsbedingung $\boldsymbol{x}_2^*(0)$ ist deutlich kritischer als die erste. Die VNKIs beider Abschätzverfahren bleiben nur bis $Ord = 2$ komplett null. Ab $Ord = 3$ steigen die VNKIs bei KQFT schnell mit höherem Rauschen. Insbesondere liegen die VNKIs bei $Ord = 4$ und 5 ab $\epsilon_R = 0{,}0025$ in der Nähe von eins. Dagegen hat KQPL den Vorteil, dass die VNKIs nur langsam auf maximal 7,2% steigen.

Um die beiden Abschätzverfahren einfach vergleichen zu können, werden im Folgenden die Simulationen beider Methoden mit $Ord = 3$ und $\epsilon_R = 0{,}005$ für die Auswertung der Eigenfrequenz und des Bestimmtheitsmaßes betrachtet.

Die Eigenfrequenzen beider Methoden weisen eine ähnliche Verteilung für die Anfangsbedingung $\boldsymbol{x}_1^*(0)$ auf, wie in der Abbildung 6 gezeigt. Die maximale Abweichung entspricht der Auflösung des Frequenzvektors, also 0,01 Hz. Bei $\boldsymbol{x}_2^*(0)$ ist die Verteilung der Eigenfrequenzen bei Verwendung von KQFT flacher und breiter als bei Verwendung von KQPL. Außerdem werden die Maxima im Histogramm von KQFT um 0,01 Hz unterschätzt. Das ist verständlich im Kontext mit der Konvergenz in der Abbildung 5 ($\boldsymbol{x}_2^*(0)$-KQFT), da das VNKI ca. 50% beträgt. Die nicht konvergierten Signale bringen einen zusätzlichen Fehler in die Auswertung der Eigenfrequenz ein.

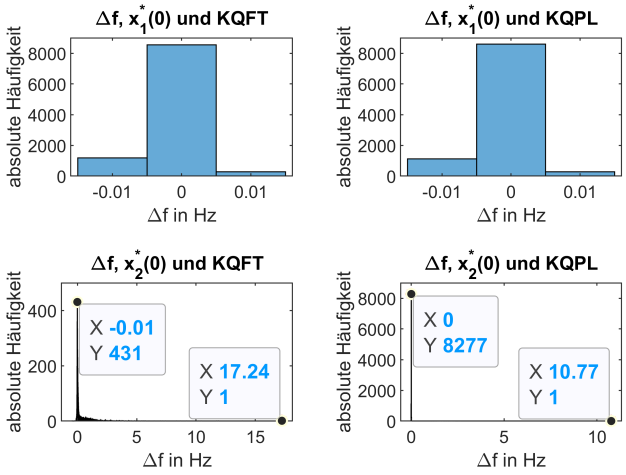

Abbildung 6: Abweichung der Eigenfrequenz bei KQFT und KQPL für die Anfangsbedingungen $\boldsymbol{x}_1^*(0)$ und $\boldsymbol{x}_2^*(0)$, $Ord = 3$, $\epsilon_R = 0{,}005$

Das Bestimmtheitsmaß R^2 wird nur für den Trainingssatz mit der Anfangsbedingung $\boldsymbol{x}(0)$ betrachtet. In der Abbildung 7 ist zu erkennen, dass KQPL einen höheren Medianwert und ein kleineres Intervall zwischen den Quantilen im Vergleich mit KQFT hat. D. h. die durch KQPL abgeschätzten Modelle stimmen besser mit den Trainingsdaten überein.

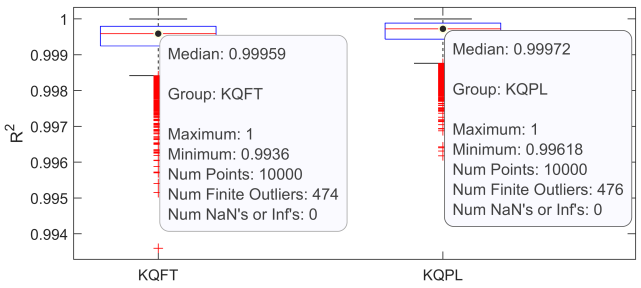

Abbildung 7: Boxplots des Bestimmtheitsmaßes bei KQFT und KQPL

Zusammenfassung und Ausblick

Die Methoden KQFT und KQPL beruhen auf der KQ-Methode und zeigen daher die gleiche Tendenz bei unterschiedlicher Rauschstärke und Polynomordnungen. Je intensiver das Rauschen ist, desto größer ist die Abweichung beider Methoden hinsichtlich der Spärlichkeit. Aufgrund zu viel identifizierter Koeffizienten, steigt die Gefahr des Overfittings mit größerem Rauschen an. Mit steigender Ordnung des Polynommodells werden mehr Terme identifiziert. Einerseits wird das Modell durch die zusätzlichen Termen ergänzt. Andererseits ist die Tendenz zum Overfitting stärker. Die Wechselwirkung von Rauschen und Polynomordnung führt häufig zu unbrauchbaren Modellen, insbesondere wenn beide hoch sind, so dass die numerische Integration nicht konvergiert.

Im Vergleich zu der Methode KQFT ist KQPL robuster gegen Overfitting und liefert bei den Untersuchungen spärlichere Modelle. Damit ist das Modell bei KQPL allgemeingültiger als bei KQFT. In den Auswertungen mittels ANNP, VNKI, identifizierter Eigenfrequenz und R^2 zeigt die Methode KQPL deutlich bessere Eigenschaften, insbesondere bei hohen Ordnungen $Ord \geq 3$. Deswegen ist die Methode KQPL eine gute Alternative zur KQFT.

In weiteren Schritten soll das Identifikationsverfahren für diskrete Systeme mit mehreren Freiheitsgraden sowie für kontinuierliche Systeme auf Basis von SINDy-KQPL entwickelt werden. Die Nichtlinearität soll auch berücksichtigt werden. Experimentelle Untersuchungen sind zur Validierung der Ergebnisse geplant.

Literatur

[1] Kerschen, G., Worden, K., Vakakis, A. F., Golinval, J.-C.: Past, present and future of nonlinear system identification in structural dynamics. Mechanical Systems and Signal Processing. 20 (3), S. 505–592, 2006.

[2] Brunton, S. L., Proctor, J. L., Kutz, J. N.: Discovering governing equations from data by sparse identification of nonlinear dynamical systems. In: Proceedings of the National Academy of Sciences of the United States of America 113 (15), S. 3932–3937, 2016.

[3] Adams, C., Bös, J., Melz, T.: Evidenzbasierte Modellierung von Schwingungssystemen. In: DAGA – Deutsche Jahrestagung für Akustik, Deutschland, 18.-21. März 2019, S. 449–452, 2019.

[4] Belloni, A., Chernozhukov, V.: Least squares after model selection in high-dimensional sparse models. Bernoulli 19 (2), S. 521–547, 2013.

[5] Platz, R., Enss, G. C.: Comparison of uncertainty in passive and active vibration isolation. Model Validation and Uncertainty Quantification, Volume 3, S. 15–25, 2015.

[6] Storm, R., Hanselka, H.: Kompendium Maschinenakustik Teil A: „Maschinenakustik – Grundlagen“ Band 1, S. 147–149, 2008.