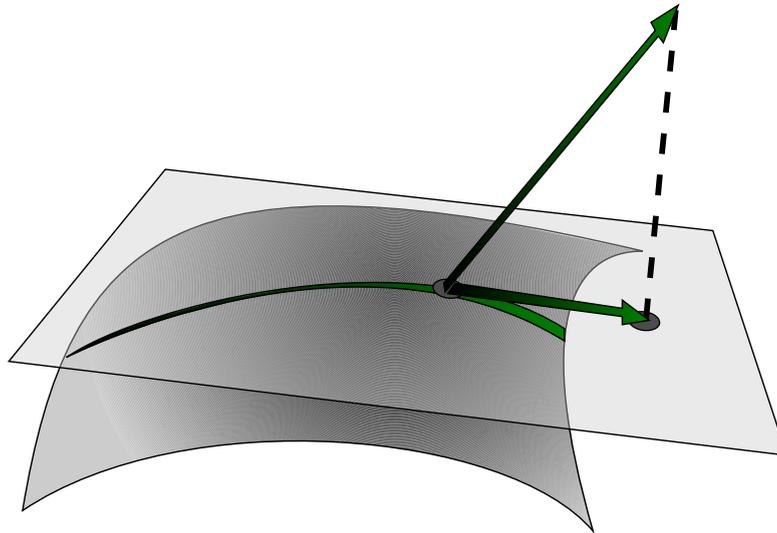


Lösung von hochdimensionalen Anfangswertproblemen durch hierarchische Tensor-Approximation



Zur Erlangung des akademisches Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

von der Fakultät für Mathematik des
Karlsruher Instituts für Technologie (KIT)
genehmigte

DISSERTATION

von

Andreas Thomas Arnold
(DIPL.-TECH. MATH. UNIV.)
aus Augsburg

Tag der mündlichen Prüfung: 17. Juli 2013

Referent: Prof. Dr. Tobias Jahnke
Korreferent: Prof. Dr. Christian Lubich

Inhaltsverzeichnis

1	Einleitung	1
2	Niedrigrangapproximation der Lösung von Matrix-Differentialgleichungen	5
3	Ein Überblick über Tensor-Formate	17
3.1	Kanonisches Format	19
3.2	Tucker-Format	20
3.3	Hierarchisches Tucker-Format	22
4	Zeitabhängige Approximation im hierarchischen Tucker-Format	33
5	Matrix-Darstellung von Tensoren	45
6	Differentialgeometrie von Tensoren im hierarchischen Tucker-Format	57
6.1	Interpretation als Mannigfaltigkeit	57
6.2	Krümmung der Mannigfaltigkeit	64
7	A posteriori Fehleranalyse der Niedrigrangapproximation	83
8	Zeitabhängige Approximation der Black-Scholes-Gleichung im \mathcal{H}-Tucker-Format	95
8.1	Die Black-Scholes-Gleichung	95
8.2	Formulierung des diskreten Anfangswertproblems	98
8.3	Numerische Lösung	102
	Literaturverzeichnis	115

1. Einleitung

Diese Arbeit befasst sich mit der zeitabhängigen Approximation von hochdimensionalen Anfangswertproblemen der Form

$$\dot{Y}_{ex}(t) = F(Y_{ex}(t)) \quad , \quad Y_{ex}(0) = Y_0 \in \mathbb{R}^{N_1 \times \dots \times N_d} \quad , \quad t \geq 0, \quad (1.1)$$

mit einer Funktion $F : \mathbb{R}^{N_1 \times \dots \times N_d} \rightarrow \mathbb{R}^{N_1 \times \dots \times N_d}$. Geht das Anfangswertproblem aus der Diskretisierung einer partiellen Differentialgleichung hervor, so sind die Werte N_1, \dots, N_d ein Maß für die Feinheit der Raumdiskretisierung. Bekannte hochdimensionale partielle Differentialgleichungen sind beispielsweise die Black-Scholes-Gleichung aus der Finanzmathematik, die Fokker-Planck-Gleichung aus der Wahrscheinlichkeitstheorie oder die Schrödingergleichung aus der Quantenmechanik. Setzen wir die Existenz einer Lösung des Anfangswertproblems voraus, so ist $Y_{ex}(t)$ zu jedem Zeitpunkt ein d -dimensionaler Tensor mit insgesamt $\prod_{\omega=1}^d N_\omega$ Einträgen. Für $N_1 = \dots = N_d = N$ besitzt der Tensor also insgesamt N^d Einträge. Obgleich viele Verfahren zur numerischen Lösung eines Anfangswertproblems bekannt sind, scheitert deren Anwendung in hohen Dimensionen aufgrund der enormen Anzahl an Freiheitsgraden. Selbst wenn es möglich wäre, den Tensor trotz seiner vielen Einträge im Computer-Speicher abzulegen, so würde mit der Dimension auch der numerische Aufwand für die Zeitintegration exponentiell steigen. Richard Bellman bezeichnet in [Bel61] das exponentielle Ansteigen an Freiheitsgraden mit der Dimension als „Fluch der Dimensionen“.

Da es schon für relativ kleine Werte von $d \in \mathbb{N}$ nicht möglich ist, ein Standardverfahren auf das Problem (1.1) anzuwenden, besteht ein möglicher Ausweg darin, **vor** dem numerischen Lösen das ursprüngliche Anfangswertproblem durch ein neues Anfangswertproblem zu approximieren, das weniger Unbekannte enthält und daher mit klassischen ODE-Verfahren gelöst werden kann. Um ein reduziertes Problem zu erhalten, nimmt man an, dass die

Lösung des Ausgangsproblems durch Funktionen einer bestimmten Form hinreichend gut approximiert werden kann. Diese Ansatzfunktionen sind dabei so gewählt, dass sie nur von einer gewissen Anzahl an Parametern abhängen und weitaus weniger Freiheitsgrade aufweisen als beliebige Tensoren. Mit Hilfe des Variationsprinzips von Dirac-Frenkel [Lub08] können dann Bewegungsgleichungen für die Parameter der Ansatzfunktionen hergeleitet werden. Nach dem Lösen der Bewegungsgleichungen ermöglichen die Ansatzfunktionen die Darstellung eines zeitabhängigen Tensors anhand der zeitabhängigen Parameter. Mit diesem Ansatz erhalten wir aus einem Anfangswertproblem immenser Größe ein nichtlineares Problem moderater Größe. Liegt ein lineares Anfangswertproblem vor, so wird die Linearität nicht an das neue Problem vererbt. Während die Lösung von (1.1) ein zeitabhängiger Tensor mit einer beträchtlichen Anzahl an Einträgen ist, besteht die Lösung des neuen Problems aus einer mäßigen Anzahl an zeitabhängigen Parametern.

Die Frage ist nun, wie die Ansatzfunktionen in einer sinnvollen Weise gewählt werden können. Vielversprechende Kandidaten sind Tensorprodukte. Sie besitzen die Eigenschaft, einen hochdimensionalen Tensor aus dem Produkt von niederdimensionalen Tensoren zu erzeugen. Nehmen wir als Beispiel zwei Vektoren $u \in \mathbb{R}^m$ und $v \in \mathbb{R}^n$. Insgesamt besitzen die beiden Vektoren $m + n$ Freiheitsgrade. Betrachten wir nun aber die Matrix $M = u v^T$, so besitzt die Matrix $m \cdot n$ Einträge und wird dabei vollständig durch die beiden Vektoren beschrieben. Das Prinzip, eine Matrix aus dem Produkt zweier Vektoren zu erzeugen, lässt sich auf höhere Dimensionen erweitern, indem man beispielsweise das Tensorprodukt von d Vektoren verwendet, um einen d -dimensionalen Tensor zu erzeugen. Leider reicht ein einziges solches Tensorprodukt für eine zufriedenstellende Approximation in den meisten Fällen noch nicht aus. Aus diesem Grund betrachtet man in der Regel Linearkombinationen von solchen Tensorprodukten. Dabei kann man verschiedene Ansätze verfolgen, was auf verschiedene sogenannte Tensor-Formate führt. Das hierarchische Tucker-Format [HK09] ist eines dieser Formate. Es zeichnet sich dadurch aus, dass ein hochdimensionaler Tensor in einer rekursiven Weise durch niederdimensionale Objekte dargestellt wird. Die Struktur des Formats führt dazu, dass die Anzahl an Freiheitsgraden formal nur **linear** mit der Dimension wächst [Gra10], so dass die numerische Lösung von weitaus höherdimensionalen Anfangswertproblemen möglich ist.

Ausgehend von dem Anfangswertproblem (1.1) findet die numerische Lösung in zwei Schritten statt: Als Erstes stellt man anhand des Anfangswertproblems (1.1) die Bewegungsgleichungen für die Parameter eines Tensors im hierarchischen Tucker-Format auf. Dieser Approximationsschritt ist als eine Reduktion des Modells zu verstehen. In einem zweiten Schritt sind diese Bewegungsgleichungen mit einem numerischen Integrator zu lösen. Abbildung 1.1 illustriert die Approximationsschritte dieser Vorgehensweise.

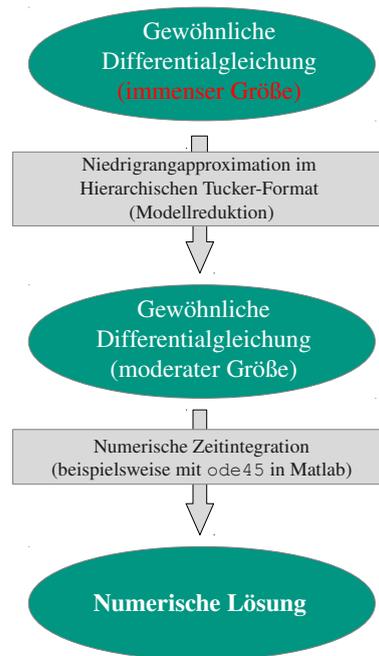


Abbildung 1.1: Approximationsschritte

Diese Arbeit ist der Analyse der Modellreduktion, also des ersten Approximationsschritts, gewidmet. Dabei orientiert sich unser Vorgehen an der Arbeit [KL07] von Koch und Lubich, wo mit Hilfe des Variationsprinzips eine zeitabhängige Approximation der Lösung einer Matrix-Differentialgleichung bestimmt wird. Die dort entwickelten Methoden werden in [KL10] von den gleichen Autoren auf die Approximation von hochdimensionalen Anfangswertproblemen im Tucker-Format [KB09] erweitert. In der vorliegenden Dissertation übertragen wir die Techniken von Koch und Lubich aus [KL07] und [KL10] auf das hierarchische Tucker-Format. Dabei leiten wir nicht nur die Bewegungsgleichungen der Darstellung her, sondern entwickeln nach den Ideen von Koch und Lubich in [KL07] beziehungsweise [KL10] eine *a posteriori* Fehleranalyse der Modellreduktion.

Die vorliegende Dissertation ist in folgender Weise gegliedert: In Kapitel 2 erläutern wir die Vorgehensweise von Koch und Lubich in [KL07]. Die Ideen und Konzepte dieses Kapitels bilden die Grundlage für die Herleitung einer zeitabhängigen Approximation im hierarchischen Tucker-Format in Kapitel 4. In Kapitel 3 stellen wir die drei wohl bekanntesten Tensor-Formate vor: Das kanonische Format, das Tucker-Format und das hierarchische Tucker-Format. In Kapitel 4 präsentieren wir ein Hauptresultat dieser Arbeit: Die Formeln zur zeitabhängigen Approximation eines Anfangswertproblems im hierarchischen Tucker-Format. Zu diesem Zweck bestimmen wir mit Hilfe des Variationsprinzips die Bewegungsgleichungen für die Parameter der Darstellung. Das Lösen dieser Bewegungsgleichungen führt zu einem zeitabhängigen Tensor, welcher die Lösung des hochdimensionalen Anfangswertproblems im hierarchischen Tucker-Format approximiert. In Kapitel 5 führen wir einen Formalismus ein, welcher uns erlaubt, einen Tensor als eine Matrix

oder als einen Vektor darzustellen. Dieser Formalismus hilft uns dabei, Rekursionen in den Formeln aus Kapitel 4 aufzulösen. In Abschnitt 6.1 aus Kapitel 6 greifen wir die Resultate von Uschmajew und Vandereycken in [UV13] auf und zeigen, dass es sich bei der Menge an Tensoren im hierarchischen Tucker-Format um eine Untermannigfaltigkeit des $\mathbb{R}^{N_1 \times \dots \times N_d}$ handelt. Mit dem Formalismus aus Kapitel 5 sind wir dann in Abschnitt 6.2 in der Lage, eine Abschätzung über die Krümmung der Mannigfaltigkeit herzuleiten. Nachdem wir in Kapitel 4 eine Möglichkeit vorstellen, ein hochdimensionales Anfangswertproblem in der Menge der Tensoren im hierarchischen Tucker-Format zu approximieren, untersuchen wir in Kapitel 7 die Genauigkeit dieser Approximation im Rahmen einer *a posteriori* Fehleranalyse. In den Beweisen dieses Kapitels wird die Abschätzung über die Krümmung der Mannigfaltigkeit die wesentliche Ingredienz sein. In Kapitel 8 betrachten wir ein numerisches Beispiel: Wir benutzen das hierarchische Tucker-Format zur Approximation der Lösung einer vierdimensionalen Black-Scholes-Gleichung und untersuchen das Approximationsverhalten.

Wir weisen darauf hin, dass zentrale Resultate der vorliegenden Arbeit in ähnlicher Form auch in dem kürzlich erschienenen Artikel [LRSV13] von Lubich, Rohwedder, Schneider und Vandereycken zu finden sind. Die Bewegungsgleichungen der Niedrigrangapproximation werden dort mit etwas anderen Argumenten hergeleitet, doch sowohl die Abschätzungen der Krümmung der Mannigfaltigkeit als auch die *a posteriori* Fehlerabschätzung werden mit einer ähnlichen Strategie bewiesen. Wir möchten jedoch betonen, dass die Resultate der vorliegenden Dissertation zeitgleich und völlig unabhängig von der Arbeit [LRSV13] erzielt wurden und bereits in dem Preprint [AJ12] dokumentiert sind.

2. Niedrigrangapproximation der Lösung von Matrix-Differentialgleichungen

In diesem Kapitel stellen wir ein Verfahren zur dynamischen Approximation von Matrix-Differentialgleichungen vor. Die Resultate dieses Kapitels beruhen auf der Arbeit [KL07] von Koch und Lubich. Zwar sind Matrizen lediglich zweidimensionale Objekte, jedoch dienen die hier vorgestellten Konzepte als Grundlage für das hierarchische Vorgehen in Kapitel 4.

Als Modellproblem betrachten wir für eine Funktion $F : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{m \times n}$ das matrixwertige Anfangswertproblem

$$\dot{Y}_{ex}(t) = F(Y_{ex}(t)) \quad , \quad Y_{ex}(0) = Y_0 \in \mathbb{R}^{m \times n} \quad , \quad t \geq 0. \quad (2.1)$$

Wir nehmen im Folgenden stets an, dass das Anfangswertproblem (2.1) eine eindeutige Lösung besitzt. Das ist beispielsweise dann erfüllt, wenn die Funktion $F : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{m \times n}$ einer Lipschitz-Bedingung genügt, siehe [Wal00, Kapitel 2 §6]. Obgleich eine Lösung $Y_{ex}(t)$ zu jedem Zeitpunkt stets eine Matrix wäre, kann selbst in diesem Fall die Anzahl an Freiheitsgraden sehr groß werden. Resultiert beispielsweise das zweidimensionale Anfangswertproblem aus der Diskretisierung einer partiellen Differentialgleichung auf einem feinen Gitter, so erhalten wir große Werte für m und n .

Um die Anzahl an Freiheitsgraden zu reduzieren, suchen wir eine zeitabhängige Approximation von $Y_{ex}(t)$, welche in der Menge

$$\mathcal{M}_r = \left\{ Y \in \mathbb{R}^{m \times n} : \text{rang}(Y) = r \right\}$$

verläuft. Ein solches Vorgehen ist sinnvoll, weil die Lösung eines matrixwertigen Anfangswertproblems oft einen niedrigen „essentiellen“ Rang besitzt.

Die folgenden Beispiel illustriert, wie eine Matrix, welche die Graustufenwerte der Pixel einer Fotografie enthalten, für verschiedene Werte von r durch ein Element aus der Menge \mathcal{M}_r approximiert wird.

(a) Bestapproximation in \mathcal{M}_{10} (b) Bestapproximation in \mathcal{M}_{20} (c) Bestapproximation in \mathcal{M}_{50} 

(d) Originalbild

Abbildung 2.1: Bestapproximation einer Graustufenmatrix in \mathcal{M}_r

Die Grafiken in Abbildung 2.1 zeigen die Bestapproximationen jener Graustufenmatrix in den Mengen \mathcal{M}_{10} , \mathcal{M}_{20} und \mathcal{M}_{50} . Dabei verstehen wir unter der Bestapproximation einer Matrix $M \in \mathbb{R}^{m \times n}$ in der Menge \mathcal{M}_r eine Matrix $X \in \mathcal{M}_r$ mit

$$\|X - M\|_F = \min_{\hat{X} \in \mathcal{M}_r} \|\hat{X} - M\|_F.$$

Das Originalbild¹ aus Abbildung 2.1 besteht aus 512×512 Bildpunkten.

¹University of Southern California. Signal and Image Processing Institute: Lena (Testbild).
Verfügbar online: <http://sipi.usc.edu/database/database.php?volume=misc&image=12>

Ehe wir auf das Variationsprinzip eingehen, welches uns eine zeitabhängige Approximation der exakten Lösung $Y_{ex}(t)$ in der Menge \mathcal{M}_r liefert, möchten wir noch einige Eigenschaften der Menge \mathcal{M}_r anführen. Zum Einen handelt es sich bei der Menge \mathcal{M}_r um eine $(m-r)(n-r)$ -dimensionale Untermannigfaltigkeit des $\mathbb{R}^{m \times n}$, siehe [Lee03, Example 8.14]. Darüber hinaus bietet die Menge \mathcal{M}_r den Vorteil, dass jede Matrix $Y \in \mathcal{M}_r$ eine Niedrigrangzerlegung der Form

$$Y = U S V^T \quad (2.2)$$

besitzt mit $U \in \mathbb{R}^{m \times r}$ und $V \in \mathbb{R}^{n \times r}$ als Matrizen mit orthonormalen Spalten und $S \in \text{GL}(r, \mathbb{R})$, also

$$U^T U = I, \quad V^T V = I \quad \text{und} \quad \text{rang}(S) = r. \quad (2.3)$$

Die Existenz einer solchen Zerlegung garantiert beispielsweise die Singulärwertzerlegung, siehe [DH08, Satz 5.15]. Wendet man die Singulärwertzerlegung an, um Matrizen U, S und V zu erhalten, welche (2.2) erfüllen, so ist in diesem speziellen Fall die Matrix S sogar eine Diagonalmatrix. Eine Zerlegung von der Form (2.2) ist nicht eindeutig: Seien $M_u, M_v \in O(r)$ beliebige orthogonale Matrizen, so wäre

$$Y = \tilde{U} \tilde{S} \tilde{V}^T \quad \text{mit} \quad \tilde{U} = U M_u^T, \quad \tilde{S} = M_u S M_v^T \quad \text{und} \quad \tilde{V} = V M_v^T$$

ebenfalls eine Zerlegung mit den gewünschten Eigenschaften. Der große Vorteil in der Darstellung der Matrix Y durch die Matrizen U, S und V wird dann deutlich, wenn wir den Speicheraufwand vergleichen: Setzen wir $m = n$ voraus, so müssen für eine Matrix Y insgesamt m^2 Zahlen gespeichert werden. Betrachten wir hingegen den Speicheraufwand für eine Darstellung durch die Matrizen U, S und V , so liegt dieser bei $2mr + r^2$ Einträgen. Wir erreichen also eine Situation, in welcher der Speicheraufwand nur noch linear in m zu Buche schlägt. Der Nachteil dieser Darstellung liegt darin, dass zur Rekonstruktion eines Eintrages der Matrix Y aus den Matrizen U, S und V insgesamt $2r^2 + r - 1$ Rechenoperationen nötig sind.

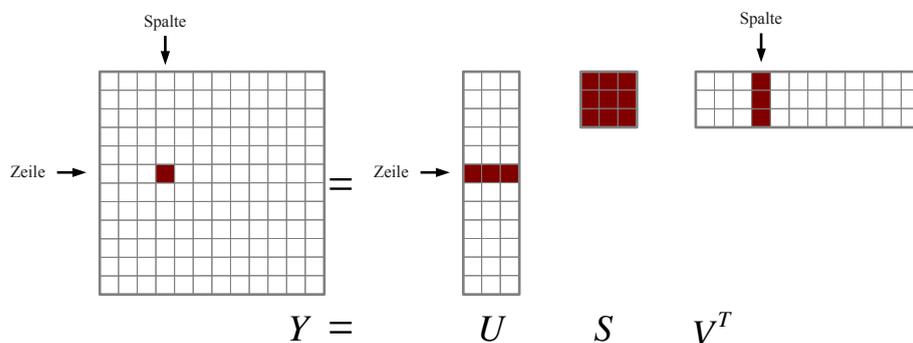


Abbildung 2.2: Rekonstruktion eines Eintrages aus der Matrix Y

Abbildung 2.2 zeigt schematisch, welche Einträge aus den Matrizen U, S und V verrechnet werden müssen, um einen Eintrag in der Matrix Y zu rekonstruieren.

Es macht also nur dann Sinn, eine Matrix Y durch eine Zerlegung von der Form (2.2) darzustellen, wenn der Rang der Matrix Y moderat ist, also

$$r \ll m \quad \text{und} \quad r \ll n$$

erfüllt ist; dann ist der Speicheraufwand erheblich geringer und die Rekonstruktion eines Eintrages der Matrix Y mit mäßigem Rechenaufwand verbunden. Die Motivation, eine zeitabhängige Approximation des Anfangswertproblems (2.1) in der Menge \mathcal{M}_r zu suchen, besteht also in erster Linie darin, die Anzahl der Freiheitsgrade zu reduzieren. Ein naives Vorgehen bestünde darin, das Anfangswertproblem (2.1) zu lösen und zu jedem Zeitschritt anhand der Singulärwertzerlegung eine Bestapproximation $X(t)$ zu bestimmen, siehe [EY36] und [GHS87]. Solch ein Vorgehen würde aber bedeuten, ein matrixwertiges Anfangswertproblem immenser Größe zu lösen und hätte nicht zuletzt aufgrund der Singulärwertzerlegung erhebliche numerische Kosten zur Folge. Für $m = n$ sind zur Berechnung der Singulärwertzerlegung $\mathcal{O}(\frac{4}{3}n^3)$ Rechenoperationen nötig.

Ein alternatives Vorgehen liefert das Variationsprinzip von Dirac-Frenkel. Diese Formulierung geht zurück auf die Mathematiker Paul Adrien Maurice Dirac und Yakov Illich Frenkel, siehe [Dir30] und [Fre34], sowie [Lub08] für weitere Informationen. Das Ziel des Variationsprinzips von Dirac-Frenkel besteht darin, ein neues Anfangswertproblem herzuleiten, dessen Lösung auf einer Mannigfaltigkeit verläuft und die Lösung des eigentlichen Problems approximiert. Wir erklären den Mechanismus des Variationsprinzips am Beispiel der Mannigfaltigkeit \mathcal{M}_r und des Anfangswertproblems (2.1).

Angenommen, der Anfangswert Y_0 ist ein Element der Mannigfaltigkeit \mathcal{M}_r und wir befinden uns zu einem Zeitpunkt t in einem Punkt $Y(t)$ auf der Mannigfaltigkeit. Betrachten wir nun unser Anfangswertproblem (2.1), so wäre die Zeitableitung in diesem Punkt $F(Y(t))$. Anstelle dieser Zeitableitung verwenden wir jedoch die Projektion davon in den Tangentialraum und erhalten auf diese Weise das neue Anfangswertproblem

$$\dot{Y}(t) = P(Y(t))F(Y(t)), \quad Y(0) = Y_0 \in \mathcal{M}_r, \quad t \geq 0 \quad (2.4)$$

mit $P(Y(t))$ als den orthogonalen Projektor in den Tangentialraum $\mathcal{T}_Y\mathcal{M}_r$. Dadurch, dass die Zeitableitung zu jedem Zeitpunkt ein Element des Tangentialraumes ist, bewegt sich die zeitabhängige Approximation nur in Richtungen, in denen sie die Mannigfaltigkeit \mathcal{M}_r nicht verlässt, siehe [HLW06, Theorem 5.2]. Wir nehmen im Folgenden an, dass das Anfangswertproblem (2.4) stets eine Lösung besitzt.

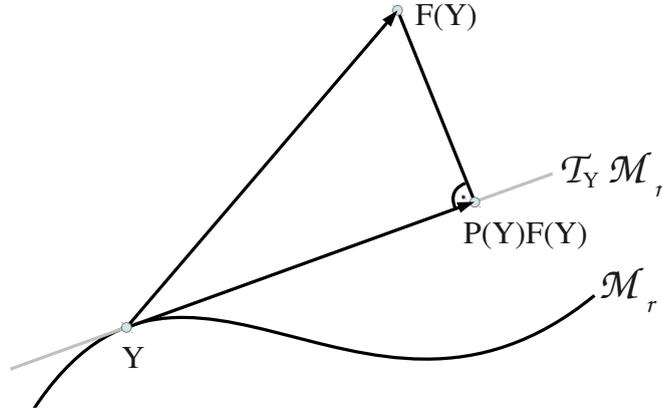


Abbildung 2.3: Projektion in den Tangentialraum $\mathcal{T}_Y \mathcal{M}_r$

Abbildung 2.3 illustriert das Vorgehen des Variationsprinzips, wobei wir aus Gründen der Übersichtlichkeit in der Abbildung auf das Zeitargument verzichtet haben. Eine zu (2.4) äquivalente Formulierung ist das Minimierungsproblem

$$\min_{\dot{Y}(t) \in \mathcal{T}_Y \mathcal{M}_r} \|\dot{Y}(t) - F(Y(t))\|_F$$

oder die Variationsformulierung

$$\langle \dot{Y}(t) - F(Y(t)), \delta Y \rangle = 0 \quad \text{für alle } \delta Y \in \mathcal{T}_Y \mathcal{M}_r, \quad (2.5)$$

wobei wir als Skalarprodukt das Frobenius-Skalarprodukt

$$\langle A, B \rangle := \text{tr}(A^T B) \quad \text{für } A, B \in \mathbb{R}^{m \times n}$$

wählen. Dadurch, dass die Approximation $Y(t)$ in der Menge \mathcal{M}_r verläuft, existieren zeitabhängige Matrizen $U(t)$, $S(t)$ und $V(t)$ mit den Eigenschaften (2.3), welche die Gleichung

$$Y(t) = U(t) S(t) V^T(t)$$

erfüllen. Zur dynamischen Approximation der Lösung von Matrix-Differentialgleichungen ist es das Ziel dieses Kapitels, ähnlich wie in [KL07] Bewegungsgleichungen für die Matrizen $U(t)$, $S(t)$ und $V(t)$ herzuleiten. Diese Bewegungsgleichungen begünstigen die Zeitintegration in der Hinsicht, dass zur Berechnung der Zeitableitung lediglich ein Gleichungssystem mit $r(r + m + n)$ Unbekannten gelöst werden muss. Ehe wir in Theorem 2 die Bewegungsgleichungen für die zeitabhängigen Matrizen $U(t)$, $S(t)$ und $V(t)$ vorstellen, ist es nötig, den Tangentialraum der Mannigfaltigkeit zu bestimmen. Anschließend nutzen wir die Variationsformulierung (2.5), um Bewegungsgleichungen für die Matrizen $U(t)$, $S(t)$ und $V(t)$ herzuleiten. Für die Definition des Tangentialraumes greifen wir auf die Arbeit von Koch und Lubich in [KL07, Abschnitt 2.1] zurück.

Lemma 1. Sei $\mathcal{M}_r \ni Y = USV^T$ eine Zerlegung wie in (2.2). Dann ist δY genau dann ein Element des Tangentialraumes $\mathcal{T}_Y \mathcal{M}_r$, wenn Matrizen $\delta U \in \mathbb{R}^{m \times r}$, $\delta S \in \mathbb{R}^{r \times r}$ und $\delta V \in \mathbb{R}^{n \times r}$ existieren, welche die **Gauge-Bedingungen**

$$U^T \delta U = 0 \quad \text{und} \quad V^T \delta V = 0 \quad (2.6)$$

erfüllen, so dass δY die Darstellung

$$\delta Y = \delta U S V^T + U \delta S V^T + U S \delta V^T \quad (2.7)$$

besitzt. Darüber hinaus sind die Matrizen δU , δS und δV durch (2.7) und die Gauge-Bedingungen (2.6) eindeutig bestimmt und erfüllen mit

$$P_U = U U^T, \quad P_V = V V^T \quad P_U^\perp = (I - U U^T), \quad P_V^\perp = (I - V V^T)$$

die Gleichungen

$$\begin{aligned} \delta S &= U^T \delta Y V, \\ \delta U &= P_U^\perp \delta Y S^{-1}, \\ \text{und } \delta V &= P_V^\perp \delta Y^T U S^{-T}. \end{aligned}$$

Beweis. Die Aussage des Lemmas wird in [KL07, Abschnitt 2.1] bewiesen. ■

Mit Lemma 1 haben wir eine formale Beschreibung für die Elemente des Tangentialraumes in einem Punkt der Mannigfaltigkeit gefunden. Eine interessante Eigenschaft des Tangentialraumes besteht darin, dass für jedes Element $\delta Y \in \mathcal{T}_Y \mathcal{M}_r$ eindeutige Matrizen δU , δS und δV existieren, welche den Gauge-Bedingungen genügen und die Gleichung

$$\delta Y = \delta U S V^T + U \delta S V^T + U S \delta V^T$$

erfüllen. Mit der Charakterisierung des Tangentialraumes sind wir in der Lage, die Bewegungsgleichungen der Niedrigrangdarstellung zu bestimmen.

Theorem 2. Für $t \in [0, T]$ sei $Y(t)$ die Lösung des Anfangswertproblems

$$\dot{Y}(t) = P(Y(t))F(Y(t)), \quad Y(0) = Y_0 \in \mathcal{M}_r$$

mit $Y_0 = U_0 S_0 V_0^T$ wie in (2.2) und $\text{rang}(Y(t)) = r$. Dann gilt

$$\dot{Y}(t) = \dot{U}(t) S(t) V^T(t) + U(t) \dot{S}(t) V^T(t) + U(t) S(t) \dot{V}^T(t),$$

wobei die zeitabhängigen Matrizen $U(t)$, $S(t)$ und $V(t)$ für alle $t \in [0, T]$ die Bewegungsgleichungen

$$\begin{aligned}\dot{S}(t) &= U^T(t) F(Y(t)) V(t), \\ \dot{U}(t) &= P_{U(t)}^\perp F(Y(t)) V(t) S^{-1}(t), \\ \text{und } \dot{V}(t) &= P_{V(t)}^\perp F^T(Y(t)) U(t) S^{-T}(t)\end{aligned}$$

erfüllen ($U(0) = U_0$, $S(0) = S_0$ und $V(0) = V_0$ ist eine mögliche Wahl für Anfangswerte).

In den Bewegungsgleichungen von Theorem 2 erfordert die Berechnung der Zeitableitung die Invertierung der zeitabhängigen Matrix $S(t)$, beziehungsweise das Lösen eines linearen Gleichungssystems.

Ehe wir Theorem 2 beweisen, zeigen wir, dass die Gleichungen des Theorems sowohl die Gauge-Bedingungen, als auch die Orthogonalität der Matrizen $U(t)$ und $V(t)$ erhalten.

Lemma 3. *Seien $U(t)$, $S(t)$ und $V(t)$ die Lösungen der Bewegungsgleichungen aus Theorem 2. Dann gelten einerseits die Gauge-Bedingungen*

$$U^T(t) \dot{U}(t) = 0 \quad \text{und} \quad V^T(t) \dot{V}(t) = 0$$

und andererseits die Orthogonalitätsbedingungen

$$U^T(t) U(t) = I \quad \text{und} \quad V^T(t) V(t) = I$$

für alle $t \in [0, T]$.

Beweis von Lemma 3. Wir zeigen zunächst, dass die Orthogonalität erhalten bleibt. Die Zeitableitung für den Term $U^T(t)U(t)$ erfüllt die Gleichung

$$\begin{aligned}\frac{d}{dt} [U^T(t) U(t)] &= \dot{U}^T(t) U(t) + U^T(t) \dot{U}(t) \\ &= [P_{U(t)}^\perp F(Y(t)) V(t) S^{-1}(t)]^T U(t) + U^T(t) P_{U(t)}^\perp F(Y(t)) V(t) S^{-1}(t) \\ &= S^{-T}(t) V^T(t) F^T(Y(t)) P_{U(t)}^\perp U(t) + [P_{U(t)}^\perp U(t)]^T F(Y(t)) V(t) S^{-1}(t).\end{aligned}$$

Mit $H(t) := S^{-T}(t) V^T(t) F^T(Y(t)) P_{U(t)}^\perp U(t)$ folgt

$$\frac{d}{dt} [U^T(t) U(t)] = \begin{cases} H(t) + H^T(t) & \text{für } U^T(t) U(t) \neq I \\ 0 & \text{für } U^T(t) U(t) = I. \end{cases} \quad (2.8)$$

Das bedeutet, wenn die Gleichung

$$U^T(0)U(0) = I$$

erfüllt ist, so gilt wegen (2.8) für alle $t \in [0, T]$ auch

$$U^T(t)U(t) = I.$$

Die Gauge-Bedingung folgt unmittelbar aus der Orthogonalität

$$U^T(t) \dot{U}(t) = [P_{U(t)}^\perp U(t)]^T F(Y(t)) V(t) S^{-1}(t) = 0.$$

Analog erhalten wir $V^T(t)V(t) = I$ und $V^T(t)\dot{V}(t) = 0$ für alle $t \in [0, T]$. ■

Beweis von Theorem 2. Dadurch, dass die Gleichungen des Theorems die Gauge-Bedingungen erhalten, folgt mit Lemma 1

$$\dot{Y}(t) \in \mathcal{T}_Y \mathcal{M}_r.$$

Wir beweisen die Aussage des Theorems, indem wir nachweisen, dass die Gleichung

$$\langle \dot{Y}(t) - F(Y(t)), \delta Y \rangle = 0 \quad (2.9)$$

mit $\dot{Y}(t)$ wie aus Theorem 2 für alle $\delta Y \in \mathcal{T}_Y \mathcal{M}_r$ erfüllt ist. In den folgenden Gleichungen nutzen wir frequentiert die Zyklizität der Spur. Das bedeutet, dass die Spur für beliebige Matrizen $A \in \mathbb{R}^{m \times n}$ und $B \in \mathbb{R}^{n \times m}$ die Gleichung

$$\text{tr}(AB) = \text{tr}(BA)$$

erfüllt. Wir beginnen mit der Variation

$$\delta Y = \delta U S(t) V^T(t) \quad (2.10)$$

mit einem beliebigen $\delta U \in \mathbb{R}^{m \times r}$, das die Gauge-Bedingung $U^T(t)\delta U = 0$ erfüllt. Dann gilt

$$\begin{aligned} \langle \dot{Y}(t), \delta Y \rangle &= \langle \dot{U}(t) S(t) V^T(t) + U(t) \dot{S}(t) V^T(t) + U(t) S(t) \dot{V}^T(t), \delta U S(t) V^T(t) \rangle \\ &= \langle P_{U(t)}^\perp F(Y(t)) V(t) S^{-1}(t) S(t) V^T(t), \delta U S(t) V^T(t) \rangle \\ &= \langle P_{U(t)}^\perp F(Y(t)) V(t) V^T(t), \delta U S(t) V^T(t) \rangle \\ &= \text{tr} \left(V(t) V^T(t) F^T(Y(t)) P_{U(t)}^\perp \delta U S(t) V^T(t) \right) \\ &= \text{tr} \left(F^T(Y(t)) \delta U S(t) V^T(t) \right) \\ &= \langle F(Y(t)), \delta U S(t) V^T(t) \rangle \\ &= \langle F(Y(t)), \delta Y \rangle. \end{aligned}$$

Demnach gilt die Variationsformulierung (2.9) für Variationen (2.10).
Betrachten wir nun die Variation

$$\delta Y = U(t) S(t) \delta V^T \quad (2.11)$$

mit einem beliebigen $\delta V \in \mathbb{R}^{n \times r}$, welches ebenfalls der Gauge-Bedingung $V^T(t)\delta V = 0$ genügt, so gilt

$$\begin{aligned} \langle \dot{Y}(t), \delta Y \rangle &= \langle \dot{U}(t) S(t) V^T(t) + U(t) \dot{S}(t) V^T(t) + U(t) S(t) \dot{V}^T(t), U(t) S(t) \delta V^T \rangle \\ &= \langle U(t) S(t) \dot{V}^T(t), U(t) S(t) \delta V^T \rangle \\ &= \langle U(t) S(t) [P_{V(t)}^\perp F^T(Y(t)) U(t) S^{-T}(t)]^T, U(t) S(t) \delta V^T \rangle \\ &= \langle U(t) U^T(t) F(Y(t)) P_{V(t)}^\perp, U(t) S(t) \delta V^T \rangle \\ &= \text{tr} \left(P_{V(t)}^\perp F^T(Y(t)) U(t) U^T(t) U(t) S(t) \delta V^T \right) \\ &= \text{tr} \left(F^T(Y(t)) U(t) S(t) \delta V^T \right) \\ &= \langle F(Y(t)), U(t) S(t) \delta V^T \rangle \\ &= \langle F(Y(t)), \delta Y \rangle. \end{aligned}$$

Damit besitzt die Variationsformulierung (2.9) auch für Variationen (2.11) Gültigkeit.
Zuletzt betrachten wir die Variation

$$\delta Y = U(t) \delta S V^T(t) \quad (2.12)$$

mit einem beliebigen $\delta S \in \mathbb{R}^{r \times r}$. Dann gilt

$$\begin{aligned} \langle \dot{Y}(t), \delta Y \rangle &= \langle \dot{U}(t) S(t) V^T(t) + U(t) \dot{S}(t) V^T(t) + U(t) S(t) \dot{V}^T(t), U(t) \delta S V^T(t) \rangle \\ &= \langle U(t) \dot{S}(t) V^T(t), U(t) \delta S V^T(t) \rangle \\ &= \text{tr} \left(V(t) \dot{S}^T(t) U^T(t) U(t) \delta S V^T(t) \right) \\ &= \text{tr} \left([U^T(t) F(Y(t)) V(t)]^T \delta S \right) \\ &= \text{tr} \left(F^T(Y(t)) U(t) \delta S V^T(t) \right) \\ &= \langle F(Y(t)), U(t) \delta S V^T(t) \rangle \\ &= \langle F(Y(t)), \delta Y \rangle. \end{aligned}$$

Die Variationsformulierung (2.9) ist also auch für Variationen (2.12) gültig.

Betrachten wir nun die Superposition der Variationen (2.10), (2.11) und (2.12), so spannen diese den gesamten Tangentialraum auf und wir erhalten

$$\langle \dot{Y}(t) - F(Y(t)), \delta Y \rangle = 0 \quad \text{für alle } \delta Y \in \mathcal{T}_Y \mathcal{M}_r.$$

■

Unser Vorgehen in dem Beweis von Theorem 2 unterscheidet sich von dem Vorgehen von Koch und Lubich in [KL07, Beweis von Proposition 2.1]. Wir haben uns für den Alternativweg entschieden, weil wir dieselben Techniken in Kapitel 4 anwenden.

Hinsichtlich der Güte der Approximation sei darauf hingewiesen, dass die zeitabhängige Approximation $Y(t)$ aus Theorem 2 nicht zwangsläufig einer Bestapproximation gleicht. In [KL07, Theorem 6.1] stellen Koch und Lubich eine *a posteriori* Analyse des Fehlers der Niedrigrangapproximation vor.

Das Variationsprinzip offeriert nicht nur die Möglichkeit, die Lösung eines Anfangswertproblems im Niedrigrangformat anzunähern, sondern auch eine zeitabhängige Matrix $M(t) \in \mathbb{R}^{m \times n}$ auf der Mannigfaltigkeit \mathcal{M}_r zu approximieren. Zu diesem Zweck benötigen wir eine Formel für die Projektion in den Tangentialraum von \mathcal{M}_r , siehe [KL07, Proposition 2.1].

Korollar 4. Sei $B \in \mathbb{R}^{m \times n}$ und $\mathcal{M}_r \ni Y = U S V^T$ wie in (2.2). Ferner seien

$$\begin{aligned} \delta S &= U^T B V, \\ \delta U &= P_U^\perp B V S^{-1}, \\ \text{und } \delta V &= P_V^\perp B^T U S^{-T}. \end{aligned}$$

Dann erfüllt die Projektion von B in den Tangentialraum $\mathcal{T}_Y \mathcal{M}_r$ die Gleichung

$$P(Y)B = \delta U S V^T + U \delta S V^T + U S (\delta V)^T.$$

Beweis. Betrachten wir den Beweis von Theorem 2 und vertauschen den Term $\dot{Y}(t)$ mit $P(Y)B$ und $F(Y(t))$ mit B , so folgt, dass

$$P(Y)B = \delta U S V^T + U \delta S V^T + U S (\delta V)^T$$

die Variationsformulierung

$$\langle B - P(Y)B, \delta Y \rangle = 0$$

für alle Elemente des Tangentialraumes $\delta Y \in \mathcal{T}_Y \mathcal{M}_r$ erfüllt. ■

Setzen wir voraus, dass $M(t)$ eine zeitabhängige Matrix ist, die sich glatt in der Zeit bewegt, so können wir unter Verwendung der Zeitableitung der Matrix, sowie der Projektion in den Tangentialraum durch das Anfangswertproblem

$$\dot{Y}(t) = P(Y(t)) \dot{M}(t), \quad Y(0) = Y_0 \in \mathcal{M}_r, \quad t \geq 0 \quad (2.13)$$

eine Approximation $Y(t) \approx M(t)$ konstruieren, welche sich auf der Mannigfaltigkeit \mathcal{M}_r bewegt.

Koch und Lubich untersuchen in [KL07] die Approximation einer zeitabhängigen Matrix durch die Lösung des Anfangswertproblems (2.13) eingehend. Darüber hinaus illustrieren Nonnenmacher und Lubich in [NL08] anhand von numerischen Experimenten die zeitabhängige Approximation der Lösung von matrixwertigen Anfangswertproblemen beziehungsweise einer zeitabhängigen Matrix.

3.

Ein Überblick über Tensor-Formate

Ehe wir auf verschiedene Tensor-Formate eingehen, stellen wir vorab die dafür notwendigen Definitionen und Notationen vor. Ein Tensor $T \in \mathbb{R}^{N_1 \times \dots \times N_d}$ ist ein d -dimensionales Zahlenfeld mit insgesamt $\prod_{\omega=1}^d N_\omega$ Einträgen. Eine äquivalente Darstellung eines Tensors ist die Interpretation des Zahlenfeldes als multivariate Funktion. Mit den Indextmengen

$$I_\omega = \{1, \dots, N_\omega\}, \quad \omega \in \{1, \dots, d\},$$

interpretieren wir T als eine multivariate Funktion

$$T : I_1 \times \dots \times I_d \rightarrow \mathbb{R}, \quad T = T(i_1, \dots, i_d).$$

Eine Matrix, also ein zweidimensionales Zahlenfeld, kann beispielsweise als eine bivariate Funktion betrachtet werden; die Funktionsargumente sind die Zeile und die Spalte der Matrix und die Auswertung der Funktion entspricht dem Eintrag, welcher die Matrix in der betreffenden Zeilen- und Spaltenposition annimmt. Diese Darstellung ist vorteilhaft bei der Konstruktion von hochdimensionalen Tensoren aus dem Produkt von niederdimensionalen Tensoren. Als nächstes wollen wir die Multiplikation zweier Tensoren definieren. Seien $\mu = \{\mu_1, \dots, \mu_m\} \subseteq \{1, \dots, d\}, \nu = \{\nu_1, \dots, \nu_n\} \subseteq \{1, \dots, d\}$ und

$$\begin{aligned} U : I_\mu &\rightarrow \mathbb{R}, & I_\mu &= I_{\mu_1} \times \dots \times I_{\mu_m}, \\ V : I_\nu &\rightarrow \mathbb{R}, & I_\nu &= I_{\nu_1} \times \dots \times I_{\nu_n}. \end{aligned}$$

Da sich die Tensoren womöglich in unterschiedliche Richtungen erstrecken, definieren wir Hilfstensoren \tilde{U} und \tilde{V} , welche in den gemeinsamen Richtungen $\kappa = \{\kappa_1, \dots, \kappa_k\} = \mu \cup \nu$ verlaufen, als

$$\tilde{U} : I_\kappa \rightarrow \mathbb{R}, \quad \tilde{U}(i_{\kappa_1}, \dots, i_{\kappa_k}) = U(i_{\mu_1}, \dots, i_{\mu_m})$$

beziehungsweise

$$\tilde{V} : I_\kappa \rightarrow \mathbb{R}, \quad \tilde{V}(i_{\kappa_1}, \dots, i_{\kappa_k}) = V(i_{\nu_1}, \dots, i_{\nu_n}).$$

Damit definieren wir das Produkt von U und V als

$$(U \cdot V) : I_\kappa \rightarrow \mathbb{R}, \quad (U \cdot V)(i_{\kappa_1}, \dots, i_{\kappa_k}) = \tilde{U}(i_{\kappa_1}, \dots, i_{\kappa_k}) \tilde{V}(i_{\kappa_1}, \dots, i_{\kappa_k}).$$

Die Multiplikation zweier Tensoren kann also als die punktweise Multiplikation ihrer Hilfstensoren interpretiert werden. Für Tensoren $A = A(i_2, i_4, i_5)$ und $B = B(i_1, i_4, i_5)$ lautet das Produkt beispielsweise

$$(A \cdot B)(i_1, i_2, i_4, i_5) = A(i_2, i_4, i_5) B(i_1, i_4, i_5).$$

Insbesondere in Kapitel 4 wird eine weitere Abbildung von Bedeutung sein. Mit den Indexmengen

$$\begin{aligned} \iota &= \{\iota_1, \dots, \iota_j\} = \mu \cap \nu, \\ \lambda &= \{\lambda_1, \dots, \lambda_l\} = (\mu \cup \nu) \setminus (\mu \cap \nu), \\ \text{also } \kappa &= \{\iota_1, \dots, \iota_j, \lambda_1, \dots, \lambda_l\}, \end{aligned}$$

definieren wir die Abbildung

$$\begin{aligned} \langle \cdot, \cdot \rangle : \mathbb{R}^{N_{\mu_1} \times \dots \times N_{\mu_m}} \times \mathbb{R}^{N_{\nu_1} \times \dots \times N_{\nu_n}} &\rightarrow \mathbb{R}^{N_{\lambda_1} \times \dots \times N_{\lambda_l}} \quad (3.1) \\ \langle U, V \rangle(i_{\lambda_1}, \dots, i_{\lambda_l}) &= \sum_{i_{\iota_1}}^{N_{\iota_1}} \dots \sum_{i_{\iota_j}}^{N_{\iota_j}} (U \cdot V)(i_{\iota_1}, \dots, i_{\iota_j}, i_{\lambda_1}, \dots, i_{\lambda_l}). \end{aligned}$$

Eine anschauliche Interpretation dieser Abbildung besteht darin, dass in den gemeinsamen Richtungen aufsummiert wird. Wenden wir beispielsweise Abbildung (3.1) auf zwei Tensoren $A = A(i_2, i_4, i_5)$ und $B = B(i_1, i_4, i_5)$ an, so erhalten wir

$$\langle A, B \rangle(i_1, i_2) = \sum_{i_4=1}^{N_4} \sum_{i_5=1}^{N_5} (A \cdot B)(i_1, i_2, i_4, i_5) = \sum_{i_4=1}^{N_4} \sum_{i_5=1}^{N_5} A(i_2, i_4, i_5) \cdot B(i_1, i_4, i_5).$$

Erstrecken sich die Tensoren U und V in (3.1) in identische Richtungen, also $\lambda = \emptyset$, so legen wir per Konvention \mathbb{R} als Zielmenge von (3.1) fest. In diesem speziellen Fall stimmt die Abbildung mit dem Frobenius-Skalarprodukt überein; die Einträge der zwei Tensoren werden punktweise miteinander multipliziert und anschließend aufsummiert.

Setzen sich zwei Tensoren A und B aus dem Produkt von niederdimensionalen Tensoren zusammen, zum Beispiel

$$A(i_1, i_2, i_3) = A^1(i_1)A^2(i_2)A^3(i_3) \text{ und } B(i_1, i_2, i_3) = B^1(i_1)B^2(i_2)B^3(i_3),$$

so vereinfacht sich Abbildung (3.1) hinsichtlich der Auswertung auf die folgende Weise

$$\langle A, B \rangle = \langle A^1, B^1 \rangle \langle A^2, B^2 \rangle \langle A^3, B^3 \rangle. \quad (3.2)$$

Während auf der linken Seite von (3.2) die Abbildung $\langle \cdot, \cdot \rangle$ dreidimensionale Tensoren als Argumente besitzt, erscheint die Abbildung auf der rechten Seite lediglich mit eindimensionalen Argumenten. Vergleichen wir den numerischen Aufwand zur Auswertung, so beträgt dieser auf der linken Seite $2N_1N_2N_3 - 1$ Rechenoperationen und auf der rechten Seite $2(N_1 + N_2 + N_3) - 1$. Für $N_1 \approx N_2 \approx N_3$ wird der numerische Mehraufwand zur Auswertung der linken Seite von (3.2) deutlich.

Mit den zuvor vorgestellten Formalismen sind wir in der Lage, die wohl bekanntesten drei Tensor-Formate vorzustellen.

3.1 Kanonisches Format

Das kanonische Format ist eines der ersten Tensor-Formate und geht zurück auf den Mathematiker Frank Lauren Hitchcock in [Hit27a] und [Hit27b].

Definition 1 (Kanonisches Format). *Sei $r \in \mathbb{N}$ eine ganze Zahl. Ein Tensor $Y \in \mathbb{R}^{N_1 \times \dots \times N_d}$ ist genau dann im kanonischen Format, wenn für alle $\omega \in \{1, \dots, d\}$ und $j \in \{1, \dots, r\}$ univariate Funktionen*

$$U_j^\omega : I_\omega \rightarrow \mathbb{R} \quad \left(U_j^\omega \in \mathbb{R}^{N_\omega} \right)$$

existieren, so dass Y der CP-Zerlegung

$$Y(i_1, \dots, i_d) = \sum_{j=1}^r \prod_{\omega=1}^d U_j^\omega(i_\omega)$$

genügt. Dabei wird die Variable r als Darstellungsrang des kanonischen Formats bezeichnet. Die Menge aller Tensoren im kanonischen Format mit Darstellungsrang r bezeichnen wir als $\text{CP}(r)$.

Der Begriff CP steht für Candecomp/Parafac und geht sowohl auf die Arbeit von Carroll und Chang in [CC70], als auch auf die Arbeit von Harshman in [Har70] zurück, siehe [KB09, Abschnitt 3].

Das kanonische Format bietet den großen Vorteil, dass es für moderate Werte von r möglich ist, einen hochdimensionalen Tensor mit sehr geringem Speicheraufwand darzustellen; zur Rekonstruktion eines Tensors im kanonischen Format genügt es, lediglich die Funktionen beziehungsweise Vektoren U_i^ω abzuspeichern. Setzen wir $N_1 = \dots = N_d = N$ voraus, so ist es auf diese Weise möglich, einen Tensor, welcher aus N^d Einträgen besteht, aus lediglich rdN Unbekannten zu rekonstruieren. Unglücklicherweise ist das kanonische Format für viele numerischen Anwendungen ungeeignet. Setzen wir voraus, dass ein Tensor

zu einer vorgegebenen Güte mindestens eine Approximation in der Menge $\text{CP}(r)$ besitzt, so ist kein stabiles Verfahren bekannt, welches die univariaten Basisfunktionen U_j^ω einer der Approximationen bestimmt [Gra11, Conclusion 11]. Silva und Lim zeigen in [dSL08, Proposition 4.1 (a)], dass die Menge $\text{CP}(r)$ für $d > 2$ nicht abgeschlossen ist. Dies führt dazu, dass ein Tensor $A \in \mathbb{R}^{N_1 \times \dots \times N_d}$ nicht notwendigerweise eine Bestapproximation in der Menge $\text{CP}(r)$ besitzt, das heißt

$$\|A - \tilde{A}\|_F \neq \inf \left\{ \|A - T\|_F : T \in \text{CP}(r) \right\}$$

für alle $\tilde{A} \in \text{CP}(r)$, siehe [dSL08, Proposition 4.1 (d)].

Nähere Informationen zum kanonischen Format sind in [Hac12, Kapitel 7 und 9], [HK09, Abschnitt 3.1] und [KB09, Abschnitt 3] zu finden.

3.2 Tucker-Format

Im Gegensatz zum kanonischen Format weist das Tucker-Format nicht die zuvor beschriebenen Defizite auf und eignet sich insbesondere zur Approximation von zeitabhängigen Problemen.

Definition 2 (Tucker-Format). *Sei $(r_\omega)_{\omega \in \{1, \dots, d\}}$ eine Familie positiver ganzer Zahlen. Ein Tensor $Y \in \mathbb{R}^{N_1 \times \dots \times N_d}$ ist genau dann im Tucker-Format, wenn ein Core-Tensor $A \in \mathbb{R}^{r_1 \times \dots \times r_d}$ und für alle $\omega \in \{1, \dots, d\}, j \in \{1, \dots, r_\omega\}$ univariate Basisfunktionen*

$$U_j^\omega : I_\omega \rightarrow \mathbb{R} \quad \left(U_j^\omega \in \mathbb{R}^{N_\omega} \right)$$

existieren, so dass Y die **Tucker-Zerlegung**

$$Y(i_1, \dots, i_d) = \sum_{j_1=1}^{r_1} \dots \sum_{j_d=1}^{r_d} A(j_1, \dots, j_d) U_{j_1}^1(i_1) \cdot \dots \cdot U_{j_d}^d(i_d)$$

besitzt und die folgenden Eigenschaften erfüllt sind:

(T1) Die univariaten Basisfunktionen U_j^ω sind orthonormal, das heißt

$$\langle U_i^\omega, U_j^\omega \rangle = \delta_{i,j}$$

für alle $\omega \in \{1, \dots, d\}$ und $i, j \in \{1, \dots, r_\omega\}$.

(T2) Die Folge

$$\langle U_1^\omega, Y \rangle, \dots, \langle U_{r_\omega}^\omega, Y \rangle$$

ist für alle $\omega \in \{1, \dots, d\}$ linear unabhängig.

Die Familie $(r_\omega)_{\omega \in \{1, \dots, d\}}$ wird **Tucker-Rang** genannt. Die Menge aller Tensoren im Tucker-Format mit Tucker-Rang $(r_\omega)_{\omega \in \{1, \dots, d\}}$ bezeichnen wir mit $\text{Tucker}(r_1, \dots, r_d)$.

Eigenschaft (T1) bedeutet keine Einschränkung an die Menge Tucker (r_1, \dots, r_d) . Würde eine Darstellung vorliegen, in der Eigenschaft (T1) nicht erfüllt ist, so könnte man mittels der QR-Zerlegung eine Darstellung mit orthonormalen Basisfunktionen konstruieren. Eigenschaft (T2) ist genau dann erfüllt, wenn die Darstellung minimalen Rang besitzt; das bedeutet, dass für einen Tensor im Tucker-Format $Y \in \text{Tucker}(r_1, \dots, r_d)$ keine Richtung $\omega^* \in \{1, \dots, d\}$ und keine Familie positiver Zahlen $(r_\omega^*)_{\omega \in \{1, \dots, d\}}$ existieren mit

$$r_\omega^* \leq r_\omega \text{ für alle } \omega \in \{1, \dots, d\}$$

und $r_{\omega^*}^* < r_{\omega^*}$,

so dass $Y \in \text{Tucker}(r_1^*, \dots, r_d^*)$. Betrachten wir den Speicheraufwand der Darstellung und setzen $r_1 = \dots = r_d = r$ und $N = N_1 = \dots = N_d = N$ voraus, so müssen für den Core-Tensor und den Basisfunktionen insgesamt $r^d + rdN$ Unbekannte abgespeichert werden. Vergleichen wir das mit der Anzahl an Einträgen eines hochdimensionalen Tensors, so wächst der Speicheraufwand eines Tensors im Tucker-Format exponentiell mit der Anzahl an Dimensionen, jedoch zu einer niedrigeren Basis. Das Tucker-Format erweist sich also dann vorteilhaft, wenn ein Tensor moderaten Tucker-Rang besitzt, also

$$\prod_{\omega=1}^d r_\omega \ll \prod_{\omega=1}^d N_\omega$$

erfüllt ist. In [DLDMV00] stellen Lathauwer, Moor und Vandewalle die Singulärwertzerlegung höherer Ordnung vor. Das ist eine Zerlegung, welche die bekannte Singulärwertzerlegung auf Tensoren verallgemeinert. Liegt ein Tensor im Tucker-Format vor, so lassen sich unter Verwendung der Singulärwertzerlegung höherer Ordnung die Basisfunktionen, sowie der Core-Tensor der Darstellung ermitteln. Darüber hinaus lässt sich mit Hilfe der Singulärwertzerlegung höherer Ordnung für einen Tensoren $\hat{Y} \in \text{Tucker}(r'_1, \dots, r'_d)$ mit

$$r'_\omega \geq r_\omega \text{ für alle } \omega \in \{1, \dots, d\}$$

und $r'_{\omega^*} > r_{\omega^*}$ für ein $\omega^* \in \{1, \dots, d\}$

eine Approximation in der Menge Tucker (r_1, \dots, r_d) bestimmen. Lathauwer, Moor und Vandewalle geben in diesem Fall in [DLDMV00, Property 10] eine Abschätzung über die Güte der Approximation an.

Weitere Informationen, welche das Format betreffen, sind in [Hac12, Kapitel 8 und 10], [HK09, Abschnitt 3.2] und [KB09, Abschnitt 4] angeführt. Wir weisen darauf hin, dass das Tucker-Format in der Literatur häufig ohne der Bedingung (T2) zu finden ist. Wir setzen diese Annahme jedoch voraus, weil die Menge Tucker (r_1, \dots, r_d) auf diese Weise auf eine Teilmenge eingeschränkt wird, welche eine Untermannigfaltigkeit des $\mathbb{R}^{N_1 \times \dots \times N_d}$ ist

[KL10, Abschnitt 2.2 und 2.3]. Analog zum Vorgehen in Kapitel 2 ist es dadurch möglich, Anfangswertprobleme von der Gestalt

$$\dot{Y}_{ex}(t) = F(Y_{ex}(t)) \quad , \quad Y_{ex}(0) = Y_0 \in \mathbb{R}^{N_1 \times \dots \times N_d} \quad , \quad t \geq 0.$$

mit $F : \mathbb{R}^{N_1 \times \dots \times N_d} \rightarrow \mathbb{R}^{N_1 \times \dots \times N_d}$ mittels des Variationsprinzips von Dirac-Frenkel in der Menge Tucker (r_1, \dots, r_d) zu approximieren. In [KL10, Theorem 2.1] leiten Koch und Lubich die daraus resultierenden Bewegungsgleichungen für die univariaten Basisfunktionen, sowie für den Core-Tensor her. Darüber hinaus entwickeln sie eine Analyse für den Fehler der Approximation. In [JH08] nutzen Jahnke und Huisinga diese Vorgehensweise zur Approximation der Lösung der Chemischen Mastergleichung.

3.3 Hierarchisches Tucker-Format

In Abschnitt 3.1 wurde mit dem kanonischen Format eine Darstellung eingeführt, welche formal kein exponentielles Ansteigen in der Anzahl an Unbekannten besitzt, jedoch Instabilitäten in vielen numerischen Anwendungen aufweist. Mit dem Tucker-Format hingegen wurde in Abschnitt 3.2 eine Darstellung eingeführt, welche sich insbesondere zur zeitabhängigen Approximation von höherdimensionalen Anfangswertproblemen eignet, jedoch aufgrund des Core-Tensors mit einem exponentiellen Ansteigen an Unbekannten mit der Anzahl an Dimensionen inhärent ist. Das Ziel des hierarchischen Tucker-Formats besteht darin, die Schwächen der zuvor vorgestellten Formate zu überwinden. Ehe wir das hierarchische Tucker-Format definieren, ist es nötig, den Begriff des Dimensionsbaumes einzuführen.

Definition 3 (Dimensionsbaum). *Ein Dimensionsbaum ist ein Binärbaum, welcher eine rekursive Zerlegung der Menge $\{1, \dots, d\}$ darstellt. Jeder Knoten $\mathcal{N} \in T$ besteht aus einer Auswahl an Richtungen $\mathcal{N} \subseteq \{1, \dots, d\}$. Die Wurzel des Baumes $\mathcal{N}_0 = \{1, \dots, d\}$ enthält stets alle Richtungen. Mit Ausnahme der Wurzel \mathcal{N}_0 besitzt jeder Knoten \mathcal{N} einen Mutterknoten $\tilde{\mathcal{N}} \supset \mathcal{N}$. Die Kindknoten \mathcal{N}_1 und \mathcal{N}_2 eines Knotens \mathcal{N} werden auch als Sukzessoren bezeichnet, $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$, und erfüllen die Relation*

$$\mathcal{N} = \mathcal{N}_1 \cup \mathcal{N}_2 \quad \text{mit} \quad \mathcal{N}_1 \cap \mathcal{N}_2 = \emptyset.$$

Jeder Knoten ist entweder ein innerer Knoten oder ein Blatt. Ein Knoten \mathcal{N} liegt im Inneren des Baumes, wenn

$$\mathcal{N} \in \mathcal{I}(T) := \{\tilde{\mathcal{N}} \in T : \tilde{\mathcal{N}} \text{ hat Sukzessoren}\}.$$

Die Menge

$$\mathcal{L}(T) := \{\tilde{\mathcal{N}} \in T : \tilde{\mathcal{N}} \text{ hat keine Sukzessoren}\}$$

bezeichnet die Blätter des Baumes.

Abbildung 8.3 illustriert wie ein Dimensionsbaum einen dreidimensionalen Raum aufspaltet. Mutter- und Kindknoten sind darin durch eine Kante verbunden.

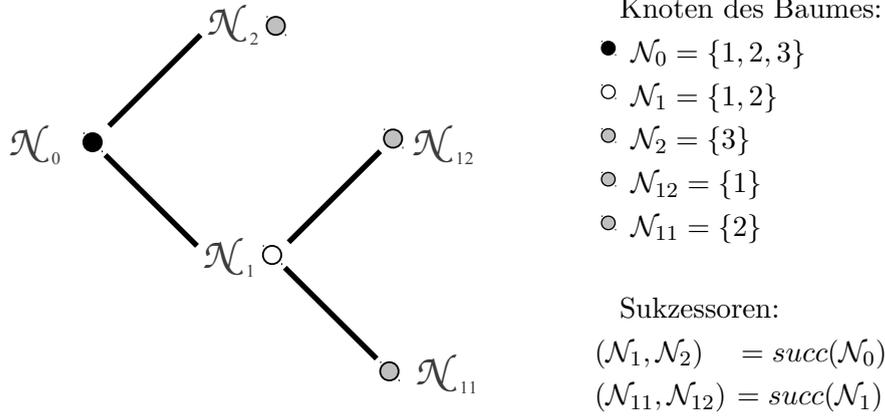


Abbildung 3.1: Dimensionsbaum T

Die Wurzel \mathcal{N}_0 besteht aus der Menge $\{1, 2, 3\}$ und enthält somit alle Richtungen eines dreidimensionalen Raumes. Diese Menge wird aufgespaltet zu den Knoten \mathcal{N}_1 und \mathcal{N}_2 , wobei \mathcal{N}_1 erneut ein innerer Knoten ist. Die Blätter des Baumes sind in Abbildung 8.3 durch das Symbol ◉ dargestellt und enthalten nur noch ein einziges Element, welches eine einzelne Richtung repräsentiert.

Der Abstand eines Knotens zu der Wurzel des Dimensionsbaumes wird auch als das **Level** des Knotens bezeichnet. In unserem Beispiel ist das Level der Knoten \mathcal{N}_1 und \mathcal{N}_2 gleich 1, während die Knoten \mathcal{N}_{11} und \mathcal{N}_{12} auf Level 2 sind.

Mit dem Dimensionsbaum sind wir in der Lage, das hierarchische Tucker-Format zu definieren. Zu diesem Zweck haben wir Definition 3.1 in [Gra10] angepasst.

Definition 4 (Hierarchisches Tucker-Format). *Sei T ein Dimensionsbaum und $(r_{\mathcal{N}})_{\mathcal{N} \in T}$ eine Familie positiver ganzer Zahlen mit $r_{\mathcal{N}_0} = 1$. Ein Tensor $Y \in \mathbb{R}^{N_1 \times \dots \times N_d}$ ist genau dann im hierarchischen Tucker-Format, wenn **Transfertensoren***

$$\mathcal{B}^{\mathcal{N}} \in \mathbb{R}^{r_{\mathcal{N}} \times r_{\mathcal{N}_1} \times r_{\mathcal{N}_2}} \text{ für alle } \mathcal{N} \in \mathcal{I}(T) \text{ mit } (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$$

und univariate Funktionen

$$U_i^\ell : I_\omega \rightarrow \mathbb{R} \text{ für alle } \ell = \{\omega\} \in \mathcal{L}(T), i = 1, \dots, r_\ell$$

existieren, so dass Y die Zerlegung

$$Y = \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{1,j,l}^{\mathcal{N}_0} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}, \quad (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N}_0)$$

besitzt, wobei die sogenannten **\mathcal{N} -frames**

$$U^{\mathcal{N}} : I_{\mathcal{N}} \rightarrow \mathbb{R} \text{ für alle } \mathcal{N} \in T \setminus \{\mathcal{N}_0\}, i = 1, \dots, r_{\mathcal{N}}$$

die Rekursion

$$U_i^{\mathcal{N}} = \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i,j,l}^{\mathcal{N}} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \text{ für alle } \mathcal{N} \in \mathcal{I}(T), (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$$

erfüllen. Die Familie $(r_{\mathcal{N}})_{\mathcal{N} \in T}$ wird als **hierarchischer Darstellungsrang** bezeichnet. Die Menge aller Tensoren im hierarchischen Tucker-Format mit hierarchischem Darstellungsrang $(r_{\mathcal{N}})_{\mathcal{N} \in T}$ bezeichnen wir mit $\mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$.

Die Transfertensoren sind stets dreidimensionale Objekte; deshalb interpretieren wir diese Tensoren nicht als multivariate Funktionen. Wie aus der Definition deutlich wird, sind zur Rekonstruktion eines Tensors im hierarchischen Tucker-Format $Y \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ lediglich die Transfertensoren und die \mathcal{N} -frames an den Blättern erforderlich. Sind die \mathcal{N} -frames der Sukzessoren eines Knotens bekannt, so kann mit dem entsprechenden Transfertensor das \mathcal{N} -frame des Mutterknotens rekonstruiert werden. Auf diese Weise können die \mathcal{N} -frames sukzessive wiederhergestellt werden.

Greifen wir das vorhergehende Beispiel eines Dimensionsbaumes auf, so könnte ein dreidimensionaler Tensor durch die Transfertensoren an den Knoten \mathcal{N}_0 und \mathcal{N}_1 und den eindimensionalen \mathcal{N} -frames an den Blättern $\mathcal{N}_{11}, \mathcal{N}_{12}$ und \mathcal{N}_2 rekonstruiert werden.

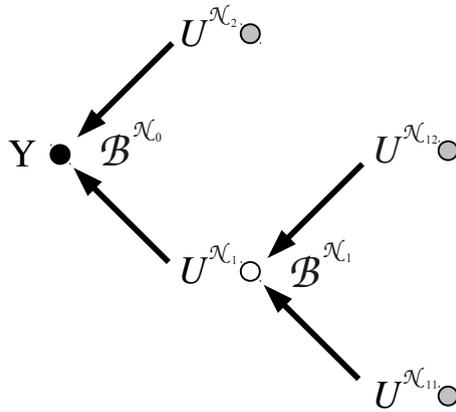


Abbildung 3.2: Rekonstruktionsschema eines Tensors im hierarchischen Tucker-Format

Die Wiederherstellung aus dem Tupel $\left((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^{\ell})_{\ell \in \mathcal{L}(T)} \right)$ ist auch ohne das Aufstellen der \mathcal{N} -frames im Inneren des Dimensionsbaumes möglich.

Lemma 5 (Rekonstruktionsformel). *Sei $Y \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ ein Tensor im hierarchischen Tucker-Format, welcher durch das Tupel $\left((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^{\ell})_{\ell \in \mathcal{L}(T)} \right)$ dargestellt wird. Dann besitzt die Rekonstruktion*

$$Y = \Psi \left(\left((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^{\ell})_{\ell \in \mathcal{L}(T)} \right) \right)$$

mit

$$\Psi : \left(\prod_{\mathcal{N} \in \mathcal{I}(T), (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})} \mathbb{R}^{r_{\mathcal{N}} \times r_{\mathcal{N}_1} \times r_{\mathcal{N}_2}} \right) \times \left(\prod_{\{j\} \in \mathcal{L}(T)} \mathbb{R}^{N_j \times r_{\{j\}}} \right) \rightarrow \mathbb{R}^{N_1 \times \dots \times N_d}$$

die Darstellung

$$\begin{aligned} & \Psi \left(\left((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^\ell)_{\ell \in \mathcal{L}(T)} \right) \right) \\ &= \sum_{J \leq (r_{\mathcal{N}})_{\mathcal{N} \in T}} \left(\prod_{\mathcal{N} \in \mathcal{I}(T), (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})} \mathcal{B}_{J_{\mathcal{N}}, J_{\mathcal{N}_1}, J_{\mathcal{N}_2}}^{\mathcal{N}} \right) \prod_{\ell \in \mathcal{L}(T)} U_{J_\ell}^\ell. \end{aligned} \quad (3.3)$$

In Gleichung (3.3) stellt J einen Multi-Index dar, dessen Anzahl an Einträgen mit der Anzahl an Knoten im Dimensionsbaum übereinstimmt. Die Schreibweise $J \leq (r_{\mathcal{N}})_{\mathcal{N} \in T}$ bedeutet, dass der Multi-Index für jeden Knoten $\mathcal{N} \in T$ die Werte $1, \dots, r_{\mathcal{N}}$ annehmen kann; die Summe in Gleichung (3.3) reicht also über insgesamt $\prod_{\mathcal{N} \in T} r_{\mathcal{N}}$ Summanden. Die Bezeichnung $J_{\mathcal{N}}$ gibt für einen Knoten $\mathcal{N} \in T$ die Auswertung des Multi-Index J an jener Indexposition an, welche mit dem Knoten \mathcal{N} assoziiert ist.

Beweis. Durch das Auflösen der Rekursion treten \mathcal{N} -frames lediglich an den Blättern des Dimensionsbaumes auf und wir erhalten insgesamt so viel Summen, wie der Dimensionsbaum Knoten besitzt. Anstelle eines formalen Beweises wollen wir den Sachverhalt an dem Dimensionsbaum verdeutlichen, welcher uns zuvor bereits als Beispiel gedient hat, siehe Abbildung 8.3 und 3.2. Dort gilt

$$\begin{aligned} Y &= \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}^{\mathcal{N}_0}_{1,j,l} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \\ &= \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}^{\mathcal{N}_0}_{1,j,l} \left(\sum_{j'=1}^{r_{\mathcal{N}_{11}}} \sum_{l'=1}^{r_{\mathcal{N}_{12}}} \mathcal{B}_{j,j',l'}^{\mathcal{N}_1} U_{j'}^{\mathcal{N}_{11}} \cdot U_{l'}^{\mathcal{N}_{12}} \right) \cdot U_l^{\mathcal{N}_2} \\ &= \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \sum_{j'=1}^{r_{\mathcal{N}_{11}}} \sum_{l'=1}^{r_{\mathcal{N}_{12}}} \left(\mathcal{B}_{1,j,l}^{\mathcal{N}_0} \mathcal{B}_{j,j',l'}^{\mathcal{N}_1} \right) U_{j'}^{\mathcal{N}_{11}} \cdot U_{l'}^{\mathcal{N}_{12}} \cdot U_l^{\mathcal{N}_2}. \end{aligned}$$

■

Die Rekonstruktionsformel verdeutlicht, dass ein Tensor im hierarchischen Tucker-Format vollständig durch das Tupel $\left((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^\ell)_{\ell \in \mathcal{L}(T)} \right)$ beschrieben ist. Liegen andererseits nicht die Transfertensoren und die \mathcal{N} -frames an den Blättern des Dimensionsbaumes vor, sondern der gesamte Tensor, also die insgesamt $\prod_{\omega=1}^d N_\omega$ Einträge des Zahlenfeldes, so existieren Algorithmen, um daraus ein Tupel $\left((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^\ell)_{\ell \in \mathcal{L}(T)} \right)$ zu berechnen, welches den Tensor gemäß Definition 4 rekonstruiert. Einen dieser Algorithmen stellen wir in Kapitel 5 vor, weitere Algorithmen zu diesem Zweck werden in [Gra10] angeführt.

Lemma 6 (Speicheraufwand). *Sei $Y \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ ein Tensor im hierarchischen Tucker-Format, so genügen*

$$\sum_{\substack{\mathcal{N} \in \mathcal{I}(T) \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} r_{\mathcal{N}} r_{\mathcal{N}_1} r_{\mathcal{N}_2} + \sum_{\{j\} \in \mathcal{L}(T)} r_{\{j\}} N_j$$

Freiheitsgrade, um den Tensor darzustellen.

Beweis. Ein Tensor im hierarchischen Tucker-Format kann durch seine Transfertensoren, sowie den \mathcal{N} -frames an den Blättern des Dimensionsbaumes rekonstruiert werden, siehe Rekonstruktionsformel. Die erste Summe in Lemma 6 zählt die Freiheitsgrade für die Transfertensoren, während die zweite Summe die Freiheitsgrade der \mathcal{N} -frames an den Blättern des Dimensionsbaumes wiedergibt, siehe auch [Gra10, Lemma 3.7]. ■

Der große Vorteil, den das hierarchische Tucker-Format bietet, ist der, dass die Anzahl an Freiheitsgraden nicht länger exponentiell mit der Dimension wächst. Wird ein Tensor im hierarchischen Tucker-Format dazu verwendet, eine hochdimensionale Funktion an diskreten Gitterpunkten darzustellen, so kann die Raumdiskretisierung erheblich feiner gewählt werden. Darüber hinaus ist bei gleicher Diskretisierung die numerische Lösung von weitaus höherdimensionalen Problemen möglich.

Ein Spezialfall des hierarchischen Tucker-Formats ist das sogenannte Tensor-Train-Format. Hierbei weist der Dimensionsbaum eine spezielle Struktur auf. Eine Verzweigung findet nur entlang eines Astes des Baumes statt.

Definition 5 (Tensor-Train-Format). *Ein Tensor $Y \in \mathbb{R}^{N_1 \times \dots \times N_d}$ ist genau dann im Tensor-Train-Format, wenn $Y \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ und für alle $\mathcal{N} \in \mathcal{I}(T)$ mit $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$ die Eigenschaft*

$$\mathcal{N}_1 \in \mathcal{L}(T) \quad \text{oder} \quad \mathcal{N}_2 \in \mathcal{L}(T)$$

erfüllt ist.

Abbildung 3.3 illustriert einen möglichen Dimensionsbaum eines fünfdimensionalen Tensors im Tensor-Train-Format.

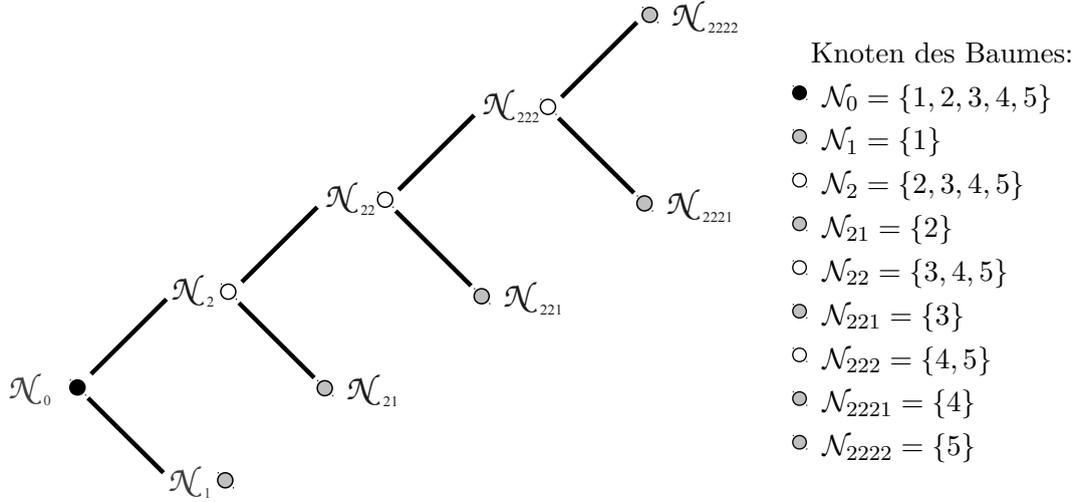


Abbildung 3.3: Dimensionsbaum (Tensor-Train-Format)

Dadurch, dass das Tensor-Train-Format nur ein Spezialfall des hierarchischen Tucker-Formats ist, gehen wir im Rahmen dieser Arbeit nicht näher darauf ein. Stattdessen verweisen wir auf Literatur, welche speziell das Tensor-Train-Format betrachtet: Oseledets gibt in [Ose11] einen Überblick über das Tensor-Train-Format. Darüber hinaus stellt er in [Ose11] einen Algorithmus vor, welcher einen beliebigen Tensor im Tensor-Train-Format approximiert und zeigt wie allgemeine Tensor-Operationen an Effizienz gewinnen können, wenn sie auf Tensoren im Tensor-Train-Format ausgeführt werden. In [OT10] stellen Oseledets und Tyrtysnikov einen Algorithmus vor, welcher einen hochdimensionalen Tensor im Tensor-Train-Format interpoliert. Des Weiteren zeigen sie eine Methode zur Approximation von hochdimensionalen Integralen mit Hilfe des Tensor-Train-Formats. In [HRS12] zeigen Holtz, Rohwedder und Schneider, dass die Menge an Tensoren im Tensor-Train-Format eine Untermannigfaltigkeit des $\mathbb{R}^{N_1 \times \dots \times N_d}$ ist. Sie bestimmen den Tangentialraum der Mannigfaltigkeit und weisen auf numerische Anwendungen hin.

Lemma 7 (Orthonormalität der Transfertensoren). *Sei $Y \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ ein Tensor im hierarchischen Tucker-Format, welcher durch das Tupel $((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^{\ell})_{\ell \in \mathcal{L}(T)})$ dargestellt wird. Dann sind die folgenden Aussagen äquivalent:*

(a) *Die \mathcal{N} -frames $(U^{\mathcal{N}})_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}}$ sind orthonormal bezüglich des Frobenius-Skalarprodukts, also*

$$\langle U_i^{\mathcal{N}}, U_j^{\mathcal{N}} \rangle = \delta_{i,j} \text{ für alle } \mathcal{N} \in T \setminus \{\mathcal{N}_0\}, i, j \in \{1, \dots, r_{\mathcal{N}}\}.$$

(b) Die Gleichung

$$\sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i_1,j,l}^{\mathcal{N}} \mathcal{B}_{i_2,j,l}^{\mathcal{N}} = \delta_{i_1,i_2} \quad (3.4)$$

gilt für alle $\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\}$, $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$, $i_1, i_2 \in \{1, \dots, r_{\mathcal{N}}\}$. Des Weiteren sind die ℓ -frames $(U^\ell)_{\ell \in \mathcal{L}(T)}$ orthonormal bezüglich des Frobenius-Skalarprodukts, also

$$\langle U_i^\ell, U_j^\ell \rangle = \delta_{i,j} \text{ für alle } \ell \in \mathcal{L}(T), i, j \in \{1, \dots, r_\ell\}.$$

Beweis. Sei (a) erfüllt, $\mathcal{N} \in \mathcal{I}(T)$ ein innerer Knoten, $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$ und $i_1, i_2 \in \{1, \dots, r_{\mathcal{N}}\}$. Dann folgt aus Definition 4

$$\begin{aligned} \langle U_{i_1}^{\mathcal{N}}, U_{i_2}^{\mathcal{N}} \rangle &= \sum_{j_1=1}^{r_{\mathcal{N}_1}} \sum_{l_1=1}^{r_{\mathcal{N}_2}} \sum_{j_2=1}^{r_{\mathcal{N}_1}} \sum_{l_2=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i_1,j_1,l_1}^{\mathcal{N}} \mathcal{B}_{i_2,j_2,l_2}^{\mathcal{N}} \langle U_{j_1}^{\mathcal{N}_1} \cdot U_{l_1}^{\mathcal{N}_2}, U_{j_2}^{\mathcal{N}_1} \cdot U_{l_2}^{\mathcal{N}_2} \rangle \\ &= \sum_{j_1=1}^{r_{\mathcal{N}_1}} \sum_{l_1=1}^{r_{\mathcal{N}_2}} \sum_{j_2=1}^{r_{\mathcal{N}_1}} \sum_{l_2=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i_1,j_1,l_1}^{\mathcal{N}} \mathcal{B}_{i_2,j_2,l_2}^{\mathcal{N}} \langle U_{j_1}^{\mathcal{N}_1}, U_{j_2}^{\mathcal{N}_1} \rangle \langle U_{l_1}^{\mathcal{N}_2}, U_{l_2}^{\mathcal{N}_2} \rangle \\ &= \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i_1,j,l}^{\mathcal{N}} \mathcal{B}_{i_2,j,l}^{\mathcal{N}}, \end{aligned} \quad (3.5)$$

woraufhin (b) folgt.

Sei nun (b) erfüllt, $\mathcal{N} \in \mathcal{I}(T)$ ein innerer Knoten mit $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$, so dass $\mathcal{N}_1, \mathcal{N}_2 \in \mathcal{L}(T)$. Dann folgt mit Gleichung (3.5)

$$\langle U_{i_1}^{\mathcal{N}}, U_{i_2}^{\mathcal{N}} \rangle = \delta_{i_1,i_2}.$$

Dieselbe Argumentation zeigt wiederum die Orthonormalität des Mutterknotens von \mathcal{N} . Mittels Induktion über die Level des Dimensionsbaumes folgt (a). \blacksquare

Für die meisten numerischen Anwendungen ist das hierarchische Tucker-Format gemäß Definition 4 ungeeignet. Betrachten wir beispielsweise einen Algorithmus, der zu vorgegebenen \mathcal{N} -frames die Transfertensoren der Darstellung bestimmt, so besäße dieser Algorithmus eine schlechte Kondition. Wären zwei der \mathcal{N} -frames eines Knotens nahezu identisch, so würde unter Umständen eine kleine Störung zu einer großen Änderung im Transfertensor des Mutterknotens führen.

Um diese Instabilität des Formats zu vermeiden, setzen wir für den Rest dieser Arbeit die folgende Annahme voraus.

Annahme 1 (Orthonormalität der \mathcal{N} -frames). *Sei $Y \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ ein Tensor im hierarchischen Tucker-Format, welcher durch das Tupel $\left((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^\ell)_{\ell \in \mathcal{L}(T)} \right)$ dargestellt wird, so setzen wir fortan die Orthonormalität der \mathcal{N} -frames voraus, also*

$$\langle U_i^{\mathcal{N}}, U_j^{\mathcal{N}} \rangle = \delta_{i,j} \text{ für alle } \mathcal{N} \in T \setminus \{\mathcal{N}_0\}, i, j \in \{1, \dots, r_{\mathcal{N}}\}.$$

Dadurch, dass zu jedem Tensor im hierarchischen Tucker-Format ein Tupel $\left((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^\ell)_{\ell \in \mathcal{L}(T)} \right)$ gefunden werden kann, welches Annahme 1 erfüllt, bedeutet diese Annahme keine Einschränkung der Menge \mathcal{H} -Tucker $((r_{\mathcal{N}})_{\mathcal{N} \in T})$.

Lemma 8. *Ist $Y \in \mathcal{H}$ -Tucker $((r_{\mathcal{N}})_{\mathcal{N} \in T})$ ein Tensor im hierarchischen Tucker-Format, welcher durch das Tupel $\left((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^\ell)_{\ell \in \mathcal{L}(T)} \right)$ dargestellt wird und $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$, so gilt*

$$Y = \sum_{i=1}^{r_{\mathcal{N}}} \langle Y, U_i^{\mathcal{N}} \rangle \cdot U_i^{\mathcal{N}}. \quad (3.6)$$

Beweis. Für $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$, $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$ und $i \in \{1, \dots, r_{\mathcal{N}}\}$, $j \in \{1, \dots, r_{\mathcal{N}_1}\}$ gilt

$$\begin{aligned} \langle U_i^{\mathcal{N}}, U_j^{\mathcal{N}_1} \rangle &= \left\langle \sum_{j'=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i,j',l}^{\mathcal{N}} U_{j'}^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}, U_j^{\mathcal{N}_1} \right\rangle \\ &= \sum_{j'=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i,j',l}^{\mathcal{N}} U_l^{\mathcal{N}_2} \underbrace{\langle U_{j'}^{\mathcal{N}_1}, U_j^{\mathcal{N}_1} \rangle}_{\text{Annahme 1 } \delta_{j',j}} \\ &= \sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i,j,l}^{\mathcal{N}} U_l^{\mathcal{N}_2}, \end{aligned}$$

woraufhin wir für $U_i^{\mathcal{N}}$ die Darstellung

$$\begin{aligned} U_i^{\mathcal{N}} &= \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i,j,l}^{\mathcal{N}} U_j^{\mathcal{N}_1} U_l^{\mathcal{N}_2} \\ &= \sum_{j=1}^{r_{\mathcal{N}_1}} U_j^{\mathcal{N}_1} \left(\sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i,j,l}^{\mathcal{N}} U_l^{\mathcal{N}_2} \right) \\ &= \sum_{j=1}^{r_{\mathcal{N}_1}} U_j^{\mathcal{N}_1} \langle U_i^{\mathcal{N}}, U_j^{\mathcal{N}_1} \rangle \end{aligned} \quad (3.7)$$

erhalten. Wir zeigen nun (3.6) mittels Induktion über die Level des Dimensionsbaumes. Für $\mathcal{N} \in \{\mathcal{N}_1^0, \mathcal{N}_2^0\}$ mit $(\mathcal{N}_1^0, \mathcal{N}_2^0) = \text{succ}(\mathcal{N}_0)$ ist (3.6) erfüllt (Induktionsanfang). Nehmen wir nun an, (3.6) gilt für einen beliebigen Knoten $\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\}$ mit $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$. Dann gilt

$$\begin{aligned} \sum_{j=1}^{r_{\mathcal{N}_1}} \langle Y, U_j^{\mathcal{N}_1} \rangle U_j^{\mathcal{N}_1} &= \sum_{j=1}^{r_{\mathcal{N}_1}} \left\langle \sum_{i=1}^{r_{\mathcal{N}}} \langle Y, U_i^{\mathcal{N}} \rangle U_i^{\mathcal{N}}, U_j^{\mathcal{N}_1} \right\rangle U_j^{\mathcal{N}_1} \\ &= \sum_{i=1}^{r_{\mathcal{N}}} \langle Y, U_i^{\mathcal{N}} \rangle \underbrace{\sum_{j=1}^{r_{\mathcal{N}_1}} \langle U_i^{\mathcal{N}}, U_j^{\mathcal{N}_1} \rangle U_j^{\mathcal{N}_1}}_{\stackrel{(3.7)}{=} U_i^{\mathcal{N}}} \\ &= Y. \end{aligned}$$

Analog lässt sich für \mathcal{N}_2 die Gleichung

$$\sum_{l=1}^{r_{\mathcal{N}_2}} \langle Y, U_j^{\mathcal{N}_2} \rangle U_l^{\mathcal{N}_2} = Y$$

zeigen, wodurch der Induktionsschritt vollzogen wird und die Behauptung folgt. \blacksquare

Lemma 9. *Sei $Y \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ ein Tensor im hierarchischen Tucker-Format, welcher durch das Tupel $\left((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^\ell)_{\ell \in \mathcal{L}(T)} \right)$ dargestellt wird. Dann sind die folgenden Aussagen äquivalent:*

(a) *Die Funktionen*

$$\langle Y, U_1^{\mathcal{N}} \rangle, \dots, \langle Y, U_{r_{\mathcal{N}}}^{\mathcal{N}} \rangle$$

sind linear unabhängig für alle $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$.

(b) *Y besitzt minimalen hierarchischen Darstellungsrang, das heißt, es existiert keine andere Familie positiver ganzer Zahlen $(r_{\mathcal{N}}^*)_{\mathcal{N} \in T}$ und keinen Knoten \mathcal{N}^* mit*

$$r_{\mathcal{N}}^* \leq r_{\mathcal{N}} \text{ für alle } \mathcal{N} \in T \setminus \{\mathcal{N}_0\} \text{ und } r_{\mathcal{N}^*}^* < r_{\mathcal{N}^*}, \quad (3.8)$$

so dass $Y \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}}^)_{\mathcal{N} \in T})$.*

Beweis. An dieser Stelle zeigen wir lediglich (a) \Rightarrow (b). Die Rückrichtung des Beweises wird in Kapitel 5 auf Seite 56 nachgetragen, da die Formalismen in diesem Kapitel den Beweisschritt erheblich verkürzen.

Angenommen (a) ist erfüllt und es gäbe ein Tupel $\left((\tilde{\mathcal{B}}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (\tilde{U}^\ell)_{\ell \in \mathcal{L}(T)} \right)$, durch welches Y dargestellt wird, das den hierarchischen Darstellungsrang $(r_{\mathcal{N}}^*)_{\mathcal{N} \in T}$ aufweist und die Eigenschaften (3.8) erfüllt. Mit Lemma 8 erhalten wir für Y die beiden Darstellungen

$$\sum_{i^*=1}^{r_{\mathcal{N}^*}^*} \tilde{U}_{i^*}^{\mathcal{N}^*} \langle Y, \tilde{U}_{i^*}^{\mathcal{N}^*} \rangle = Y = \sum_{i=1}^{r_{\mathcal{N}^*}} U_i^{\mathcal{N}^*} \langle Y, U_i^{\mathcal{N}^*} \rangle.$$

Setzen wir nun die eine Darstellung in die andere ein, so erhalten wir

$$Y = \sum_{i=1}^{r_{\mathcal{N}^*}} U_i^{\mathcal{N}^*} \sum_{i^*=1}^{r_{\mathcal{N}^*}^*} \langle U_i^{\mathcal{N}^*}, \tilde{U}_{i^*}^{\mathcal{N}^*} \rangle \langle Y, \tilde{U}_{i^*}^{\mathcal{N}^*} \rangle.$$

Daraus folgt

$$\langle Y, U_i^{\mathcal{N}^*} \rangle = \sum_{i^*=1}^{r_{\mathcal{N}^*}^*} \langle U_i^{\mathcal{N}^*}, \tilde{U}_{i^*}^{\mathcal{N}^*} \rangle \langle Y, \tilde{U}_{i^*}^{\mathcal{N}^*} \rangle.$$

Die Funktionen $\langle Y, U_1^{\mathcal{N}^*} \rangle, \dots, \langle Y, U_{r_{\mathcal{N}^*}}^{\mathcal{N}^*} \rangle$ wären demnach linear abhängig. Dies ist ein Widerspruch zu Aussage (a). \blacksquare

Lemma 9 bringt einen Makel in der Darstellung von Tensoren im hierarchischen Tucker-Format zum Vorschein: In Übereinstimmung mit Definition 4 und Annahme 1 kann die Darstellung mehr \mathcal{N} -frames besitzen, als zur Rekonstruktion benötigt werden. Um diese Redundanz in der Darstellung zu vermeiden, setzen wir eine weitere Annahme voraus.

Annahme 2. Sei Y ein Tensor im hierarchischen Tucker-Format. Fortan setzen wir die lineare Unabhängigkeit der Funktionen

$$\langle Y, U_1^{\mathcal{N}} \rangle, \dots, \langle Y, U_{r_{\mathcal{N}}}^{\mathcal{N}} \rangle$$

für alle $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ voraus.

Im Gegensatz zu Annahme 1 bedeutet Annahme 2 eine Einschränkung der Menge \mathcal{H} -Tucker $((r_{\mathcal{N}})_{\mathcal{N} \in T})$. So liegt beispielsweise ein Tensor, dessen Einträge allesamt gleich 0 sind, nun nicht mehr in der Menge. Um nicht mehr Bezeichnungen einzuführen, als nötig sind, sehen wir davon ab, der eingeschränkten Menge einen neuen Namen zu verleihen. Es sei darauf hingewiesen, dass wir fortan mit \mathcal{H} -Tucker $((r_{\mathcal{N}})_{\mathcal{N} \in T})$ die Menge der Tensoren bezeichnen, welche Definition 4 und die Annahmen 1 und 2 erfüllen. Die Anpassung des Formats bewirkt darüber hinaus, dass es sich bei der Menge \mathcal{H} -Tucker $((r_{\mathcal{N}})_{\mathcal{N} \in T})$ nun um eine Mannigfaltigkeit handelt, was von essentieller Bedeutung zur Herleitung der Bewegungsgleichungen in Kapitel 4 sein wird. Aufgrund der Anpassungen des Formats bezeichnen wir den hierarchischen Darstellungsrang fortan als **hierarchischer Rang** und die \mathcal{N} -frames als **Basisfunktionen**.

In Kapitel 4 werden zwei weitere Definitionen eine wichtige Rolle spielen.

Definition 6 (Single-hole Operator). Sei $Y \in \mathcal{H}$ -Tucker $((r_{\mathcal{N}})_{\mathcal{N} \in T})$ ein Tensor im hierarchischen Tucker-Format, dargestellt durch das Tupel $\left((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^{\ell})_{\ell \in \mathcal{L}(T)} \right)$. Für $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ und $i \in \{1, \dots, r_{\mathcal{N}}\}$ definieren wir den Single-hole Operator $\mathcal{S}(Y, \mathcal{N}, i)$ als

$$\mathcal{S}(Y, \mathcal{N}, i) := \langle Y, U_i^{\mathcal{N}} \rangle.$$

Übertragen wir den Formalismus des Single-hole Operators auf Lemma 8, so erhalten wir für ein $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ die Gleichung

$$Y = \sum_{i=1}^{r_{\mathcal{N}}} \mathcal{S}(Y, \mathcal{N}, i) \cdot U_i^{\mathcal{N}}.$$

Definition 7 (Orthogonalprojektion). Für $Y \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T}), \mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ und $X \in \mathbb{R}^{N_1 \times \dots \times N_d}$ definieren wir die Orthogonalprojektion von X in den Raum, welcher durch die Basisfunktionen $U_1^{\mathcal{N}}, \dots, U_{r_{\mathcal{N}}}^{\mathcal{N}}$ aufgespannt wird, als

$$\mathcal{P}_{\mathcal{N}}X := \sum_{i=1}^{r_{\mathcal{N}}} \langle X, U_i^{\mathcal{N}} \rangle \cdot U_i^{\mathcal{N}}.$$

Die Projektion auf das orthogonale Komplement ist definiert durch $\mathcal{P}_{\mathcal{N}}^{\perp}X := X - \mathcal{P}_{\mathcal{N}}X$.

4.

Niedrigrangapproximation der Lösung von hochdimensionalen Differentialgleichungen im hierarchischen Tucker-Format

Dieses Kapitel befasst sich mit der Approximation der Lösung von hochdimensionalen Anfangswertproblemen von der Gestalt

$$\dot{Y}_{ex}(t) = F(Y_{ex}(t)) \quad , \quad Y(0) = Y_0 \in \mathbb{R}^{N_1 \times \dots \times N_d} \quad , \quad t \geq 0. \quad (4.1)$$

Dabei sei $F : \mathbb{R}^{N_1 \times \dots \times N_d} \rightarrow \mathbb{R}^{N_1 \times \dots \times N_d}$ eine beliebige, womöglich nichtlineare, Funktion, welche Tensoren auf Tensoren abbildet. Im Folgenden setzen wir die Existenz und Eindeutigkeit einer Lösung $Y_{ex}(t)$ auf einem kompakten Zeitintervall $[0, T]$ voraus. Die Lösung $Y_{ex}(t)$ ist zu jedem Zeitpunkt des Intervalls ein d -dimensionaler Tensor mit $\prod_{\omega=1}^d N_\omega$ Einträgen. Resultiert die gewöhnliche Differentialgleichung aus der Diskretisierung einer partiellen Differentialgleichung, beispielsweise mit finiten Differenzen, so sind die Werte N_μ ein Maß für die Feinheit dieser Diskretisierung. Für $N_1 = \dots = N_d$ wächst die Anzahl an Freiheitsgraden exponentiell mit der Dimension, wodurch Standardverfahren der Numerik bereits für geringe Werte von d scheitern. Richard Bellman bezeichnet das rapide Ansteigen an Freiheitsgraden durch das Hinzufügen weiterer Dimensionen in [Bel61] als den „Fluch der Dimensionen“. Um diesen „Fluch“ zu umgehen, suchen wir eine Approximation $Y(t) \approx Y_{ex}(t)$ der exakten Lösung, welche auf der Mannigfaltigkeit

$$\mathcal{M} := \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$$

verläuft. Dieser Ansatz wurde bereits in der Quantenchemie dazu verwendet, die Lösung einer hochdimensionalen Schrödinger-Gleichung zu approximieren, siehe [WT03], [Man08], [VM11] und [Lub08]. Unser Vorgehen orientiert sich an der Arbeit von Koch und Lubich: In [KL07] betrachten Koch und Lubich die Niedrigrangapproximation der Lösung von matrixwertigen Differentialgleichungen, siehe Kapitel 2. Des Weiteren beschäftigen sie sich in [KL10] mit der Niedrigrangapproximation der Lösung eines hochdimensionalen Anfangswertproblems in der Menge Tucker (r_1, \dots, r_d) und leiten Bewegungsgleichungen für den Core-Tensor und die Basisfunktionen der Darstellung her.

Analog zum Vorgehen in Kapitel 2 wenden wir das Variationsprinzip von Dirac-Frenkel an: Ausgehend von einem Punkt $Y(t)$, welcher auf der Mannigfaltigkeit liegt, projizieren wir die rechte Seite der Differentialgleichung $F(Y(t))$ in den Tangentialraum der Mannigfaltigkeit und erhalten auf diese Weise eine neue Zeitableitung, die uns ein neues Anfangswertproblem definiert. Mit $P(Y(t))$ als den orthogonalen Projektor, welcher einen Tensor in den Tangentialraum $\mathcal{T}_Y \mathcal{M}$ projiziert, lautet das neue Anfangswertproblem

$$\dot{Y}(t) = P(Y(t)) F(Y(t)) \quad , \quad Y(0) = Y_0 \in \mathcal{M} \quad , \quad t \in [0, T]. \quad (4.2)$$

Abbildung 4.1 illustriert das Vorgehen des Variationsprinzips, wobei wir aus Gründen der Übersichtlichkeit auf das Zeitargument verzichtet haben.

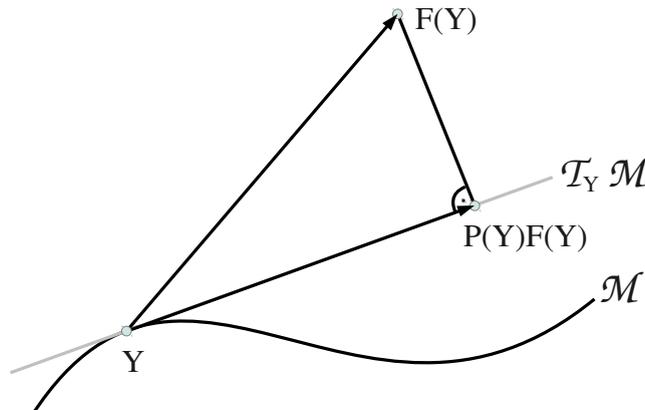


Abbildung 4.1: Orthogonalprojektion in den Tangentialraum

Ein zu (4.2) äquivalentes Problem ist das Minimierungsproblem

$$\min_{\dot{Y}(t) \in \mathcal{T}_Y \mathcal{M}} \|\dot{Y}(t) - F(Y(t))\|_F$$

oder die Variationsformulierung

$$\langle F(Y(t)) - Y(t), \delta Y \rangle = 0 \quad \text{für alle } \delta Y \in \mathcal{T}_Y \mathcal{M}. \quad (4.3)$$

Anhand der Variationsformulierung leiten wir in Theorem 11 Bewegungsgleichungen für die Transfertensoren und den Basisfunktionen an den Blättern des Dimensionsbaumes

her. Die Lösung der Bewegungsgleichungen, also die zeitabhängigen Transfertensoren $\mathcal{B}^{\mathcal{N}}(t)$, $\mathcal{N} \in \mathcal{I}(T)$ und Basisfunktionen $U^\ell(t)$, $\ell \in \mathcal{L}(T)$, ermöglicht uns zu jedem Zeitpunkt die Approximation $Y(t)$ zu rekonstruieren. Während die Zeitableitung in (4.1) ein Tensor mit insgesamt $\prod_{\omega=1}^d N_\omega$ Einträgen ist, führt die Variationsformulierung zu einem System, dessen Zeitableitung aus lediglich

$$\sum_{\substack{\mathcal{N} \in \mathcal{I}(T) \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} r_{\mathcal{N}} r_{\mathcal{N}_1} r_{\mathcal{N}_2} + \sum_{\{j\} \in \mathcal{L}(T)} r_{\{j\}} N_j$$

Freiheitsgraden besteht. Das neue Anfangswertproblem weist formal kein exponentielles Ansteigen an Freiheitsgraden bezüglich der Dimension mehr auf.

Ehe wir in Theorem 11 auf Seite 36 die Bewegungsgleichungen vorstellen, nehmen wir ein Resultat aus Kapitel 6 vorweg und geben an, aus welchen Elementen der Tangentialraum der Mannigfaltigkeit besteht.

Lemma 10 (Tangentialraum). *Sei $Y \in \mathcal{H}$ -Tucker $((r_{\mathcal{N}})_{\mathcal{N} \in T})$ ein Tensor im hierarchischen Tucker-Format, welcher durch das Tupel $((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^\ell)_{\ell \in \mathcal{L}(T)})$ dargestellt wird. Dann ist δY genau dann ein Element des Tangentialraumes $\mathcal{T}_Y \mathcal{M}$, wenn ein Tupel*

$$\left((\delta \mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (\delta U^\ell)_{\ell \in \mathcal{L}(T)} \right) \in \left(\prod_{\substack{\mathcal{N} \in \mathcal{I}(T), (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} \mathbb{R}^{r_{\mathcal{N}} \times r_{\mathcal{N}_1} \times r_{\mathcal{N}_2}} \right) \times \left(\prod_{\{\omega\} \in \mathcal{L}(T)} \mathbb{R}^{N_\omega \times r_{\{\omega\}}} \right)$$

existiert, so dass

$$\begin{aligned} \delta Y &= \sum_{j=1}^{r_{\mathcal{N}_1^0}} \sum_{l=1}^{r_{\mathcal{N}_2^0}} \delta \mathcal{B}_{1,j,l}^{\mathcal{N}_0} U_j^{\mathcal{N}_1^0} \cdot U_l^{\mathcal{N}_2^0} \\ &+ \sum_{\substack{\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\} \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} \sum_{i=1}^{r_{\mathcal{N}}} \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} \mathcal{S}(Y, \mathcal{N}, i) \cdot U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \\ &+ \sum_{\ell \in \mathcal{L}(T)} \sum_{i=1}^{r_\ell} \delta U_i^\ell \cdot \mathcal{S}(Y, \ell, i), \end{aligned}$$

mit $(\mathcal{N}_1^0, \mathcal{N}_2^0) = \text{succ}(\mathcal{N}_0)$.

Beweis. Der Beweis von Lemma 10 wird in Kapitel 6 auf Seite 63 nachgetragen. ■

Mit Lemma 10 sind wir in der Lage, das Hauptresultat dieses Kapitels vorstellen, die Bewegungsgleichungen der Transfertensoren und der Basisfunktionen an den Blättern des Dimensionsbaumes. Wir weisen darauf hin, dass die Bewegungsgleichungen mit anderen

Techniken von Lubich, Rohwedder, Schneider und Vandereycken in [LRSV13, Abschnitt 3.2] und von Uschmajew und Vandereycken in [UV13, Abschnitt 6.2] hergeleitet werden.

Theorem 11 (Bewegungsgleichungen). *Für $t \in [0, T]$ sei $Y(t) \in \mathcal{M}$ die Lösung des Anfangswertproblems*

$$\dot{Y}(t) = P(Y(t))F(Y(t)) \quad , \quad Y(0) = Y_0 \in \mathcal{M} \quad (4.4)$$

mit $\left((\mathcal{B}_0^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U_0^\ell)_{\ell \in \mathcal{L}(T)} \right)$ als ein Tupel, durch das Y_0 dargestellt werde. Des Weiteren definieren wir für $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ die zeitabhängige Lastmatrix

$$M^{\mathcal{N}}(t) = \left(m_{i,j}^{\mathcal{N}}(t) \right)_{i,j=1}^{r_{\mathcal{N}}} \in \mathbb{R}^{r_{\mathcal{N}} \times r_{\mathcal{N}}}$$

anhand der Einträge

$$m_{i,j}^{\mathcal{N}}(t) = \left\langle \mathcal{S}(Y(t), \mathcal{N}, i), \mathcal{S}(Y(t), \mathcal{N}, j) \right\rangle$$

und setzen auf dem gesamten Zeitintervall die Regularität der Matrix voraus. Darüber hinaus sei $W^{\mathcal{N}}(t) = (M^{\mathcal{N}}(t))^{-1}$ die Inverse mit Einträgen $w_{i,j}^{\mathcal{N}}(t)$.

Dann erfüllt die Lösung des Anfangswertproblems (4.4) mit $(\mathcal{N}_1^0, \mathcal{N}_2^0) = \text{succ}(\mathcal{N}_0)$ die Gleichung

$$\begin{aligned} \dot{Y}(t) &= \sum_{j=1}^{r_{\mathcal{N}_1^0}} \sum_{l=1}^{r_{\mathcal{N}_2^0}} \dot{\mathcal{B}}_{1,j,l}^{\mathcal{N}_0}(t) U_j^{\mathcal{N}_1^0}(t) \cdot U_l^{\mathcal{N}_2^0}(t) \\ &+ \sum_{j=1}^{r_{\mathcal{N}_1^0}} \sum_{l=1}^{r_{\mathcal{N}_2^0}} \mathcal{B}_{1,j,l}^{\mathcal{N}_0}(t) \dot{U}_j^{\mathcal{N}_1^0}(t) \cdot U_l^{\mathcal{N}_2^0}(t) + \sum_{j=1}^{r_{\mathcal{N}_1^0}} \sum_{l=1}^{r_{\mathcal{N}_2^0}} \mathcal{B}_{1,j,l}^{\mathcal{N}_0}(t) U_j^{\mathcal{N}_1^0}(t) \cdot \dot{U}_l^{\mathcal{N}_2^0}(t). \end{aligned} \quad (4.5)$$

Die Basisfunktionen im Inneren des Dimensionsbaumes erfüllen für $\mathcal{N} \in \mathcal{I}(T)$ mit $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$ die Rekursion

$$\begin{aligned} \dot{U}_i^{\mathcal{N}}(t) &= \sum_{j=1}^{r_{s_1}} \sum_{l=1}^{r_{s_2}} \dot{\mathcal{B}}_{i,j,l}^{\mathcal{N}}(t) U_j^{\mathcal{N}_1}(t) \cdot U_l^{\mathcal{N}_2}(t) \\ &+ \sum_{j=1}^{r_{s_1}} \sum_{l=1}^{r_{s_2}} \mathcal{B}_{i,j,l}^{\mathcal{N}}(t) \dot{U}_j^{\mathcal{N}_1}(t) \cdot U_l^{\mathcal{N}_2}(t) + \sum_{j=1}^{r_{s_1}} \sum_{l=1}^{r_{s_2}} \mathcal{B}_{i,j,l}^{\mathcal{N}}(t) U_j^{\mathcal{N}_1}(t) \cdot \dot{U}_l^{\mathcal{N}_2}(t). \end{aligned} \quad (4.6)$$

An den Blättern des Dimensionsbaumes genügen die Basisfunktionen der Gleichung

$$\dot{U}_i^\ell(t) = \sum_{i'=1}^{r_\ell} w_{i',i}^\ell(t) \mathcal{P}_\ell^\perp \left\langle F(Y(t)), \mathcal{S}(Y(t), \ell, i') \right\rangle, \quad \ell \in \mathcal{L}(T). \quad (4.7)$$

Die Ableitungen der Transfertensoren sind durch

$$\dot{\mathcal{B}}_{1,j,l}^{\mathcal{N}_0}(t) = \left\langle F(Y(t)), U_j^{\mathcal{N}_1^0}(t) \cdot U_l^{\mathcal{N}_2^0}(t) \right\rangle$$

$$\text{und } \dot{\mathcal{B}}_{i,j,l}^{\mathcal{N}}(t) = \sum_{i'=1}^{r_{\mathcal{N}}} w_{i,i'}^{\mathcal{N}}(t) \left\langle F(Y(t)), \mathcal{S}(Y(t), \mathcal{N}, i') \cdot \mathcal{P}_{\mathcal{N}}^{\perp}(U_j^{\mathcal{N}_1}(t) \cdot U_l^{\mathcal{N}_2}(t)) \right\rangle,$$

für $\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\}$ und $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$

beschrieben. Zusammen mit dem Tupel $\left((\mathcal{B}_0^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U_0^{\ell})_{\ell \in \mathcal{L}(T)} \right)$ definieren die Differentialgleichungen für die Transfertensoren und den Basisfunktionen an den Blättern des Dimensionsbaumes ein Anfangswertproblem.

Es sei darauf hingewiesen, dass die Zeitableitung in Gleichung (4.7) lediglich an Blättern des Dimensionsbaumes gegeben ist. Für Basisfunktionen im Inneren des Dimensionsbaumes muss die Zeitableitung sukzessive mit Formel (4.6) aus den Transfertensoren und den Basisfunktionen aus dem darunterliegenden Level rekonstruiert werden. Ehe wir uns dem Beweis der Bewegungsgleichungen widmen, zeigen wir in Lemma 12, dass die Lösung der Bewegungsgleichungen die Orthonormalitätsbedingungen, welche wir in Annahme 1 gefordert haben, erhalten.

Lemma 12 (Gauge-Bedingung). Sei $\left((\mathcal{B}^{\mathcal{N}}(t))_{\mathcal{N} \in \mathcal{I}(T)}, (U^{\ell}(t))_{\ell \in \mathcal{L}(T)} \right)$ die Lösung der Bewegungsgleichungen aus Theorem 11. Dann erfüllen für alle $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ die Zeitableitungen $\dot{U}_i^{\mathcal{N}}(t)$ die **Gauge-Bedingung**

$$\left\langle \dot{U}_i^{\mathcal{N}}(t), U_j^{\mathcal{N}}(t) \right\rangle = 0.$$

Darüber hinaus wird durch die Gauge-Bedingung für alle Zeiten $t \geq 0$ die Orthonormalität der Basisfunktionen erhalten, also

$$\left\langle U_i^{\mathcal{N}}(t), U_j^{\mathcal{N}}(t) \right\rangle = \delta_{i,j}. \quad (4.8)$$

Beweis von Lemma 12. Einerseits gelten für alle $\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\}$, $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$, $i_1, i_2 \in \{1, \dots, r_{\mathcal{N}}\}$ die Gleichungen

$$\begin{aligned} & \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \dot{\mathcal{B}}_{i_1,j,l}^{\mathcal{N}}(t) \mathcal{B}_{i_2,j,l}^{\mathcal{N}}(t) \\ &= \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i_2,j,l}^{\mathcal{N}}(t) \sum_{i=1}^{r_{\mathcal{N}}} w_{i_1,i}^{\mathcal{N}} \left\langle F(Y), \mathcal{S}(Y, \mathcal{N}, i) \cdot \mathcal{P}_{\mathcal{N}}^{\perp}(U_j^{\mathcal{N}_1}(t) \cdot U_l^{\mathcal{N}_2}(t)) \right\rangle \\ &= \sum_{i=1}^{r_{\mathcal{N}}} w_{i_1,i}^{\mathcal{N}} \left\langle F(Y), \mathcal{S}(Y, \mathcal{N}, i) \cdot \mathcal{P}_{\mathcal{N}}^{\perp} \left(\underbrace{\sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i_2,j,l}^{\mathcal{N}}(t) U_j^{\mathcal{N}_1}(t) \cdot U_l^{\mathcal{N}_2}(t)}_{=U_{i_2}^{\mathcal{N}}} \right) \right\rangle \\ &= 0. \end{aligned} \quad (4.9)$$

Andererseits erhalten wir für ein Blatt $\ell \in \mathcal{L}(T)$, $i_1, i_2 \in \{1, \dots, r_\ell\}$ die Gleichungen

$$\begin{aligned} \langle \dot{U}_{i_1}^\ell(t), U_{i_2}^\ell(t) \rangle &= \left\langle \sum_{i=1}^{r_\ell} w_{i,i_1}^\ell \mathcal{P}_\ell^\perp \langle F(Y), \mathcal{S}(Y, \ell, i) \rangle, U_{i_2}^\ell(t) \right\rangle \\ &= \left\langle \sum_{i=1}^{r_\ell} w_{i,i_1}^\ell \langle F(Y), \mathcal{S}(Y, \ell, i) \rangle, \mathcal{P}_\ell^\perp U_{i_2}^\ell(t) \right\rangle \\ &= 0. \end{aligned}$$

Nehmen wir nun an, dass die Gauge-Bedingung (4.8) für alle Knoten eines Levels erfüllt ist und $\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\}$ ein innerer Knoten des nächstkleineren Levels ist, so gilt

$$\begin{aligned} &\langle \dot{U}_{i_1}^{\mathcal{N}}(t), U_{i_2}^{\mathcal{N}}(t) \rangle \\ &= \sum_{j_1=1}^{r_{\mathcal{N}_1}} \sum_{l_1=1}^{r_{\mathcal{N}_2}} \sum_{j_2=1}^{r_{\mathcal{N}_1}} \sum_{l_2=1}^{r_{\mathcal{N}_2}} \dot{\mathcal{B}}_{i_1, j_1, l_1}^{\mathcal{N}}(t) \mathcal{B}_{i_2, j_2, l_2}^{\mathcal{N}}(t) \langle U_{j_1}^{\mathcal{N}_1}(t) \cdot U_{l_1}^{\mathcal{N}_2}(t), U_{j_2}^{\mathcal{N}_1}(t) \cdot U_{l_2}^{\mathcal{N}_2}(t) \rangle \\ &\quad + \sum_{j_1=1}^{r_{\mathcal{N}_1}} \sum_{l_1=1}^{r_{\mathcal{N}_2}} \sum_{j_2=1}^{r_{\mathcal{N}_1}} \sum_{l_2=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i_1, j_1, l_1}^{\mathcal{N}}(t) \dot{\mathcal{B}}_{i_2, j_2, l_2}^{\mathcal{N}}(t) \langle \dot{U}_{j_1}^{\mathcal{N}_1}(t) \cdot U_{l_1}^{\mathcal{N}_2}(t), U_{j_2}^{\mathcal{N}_1}(t) \cdot U_{l_2}^{\mathcal{N}_2}(t) \rangle \\ &\quad + \sum_{j_1=1}^{r_{\mathcal{N}_1}} \sum_{l_1=1}^{r_{\mathcal{N}_2}} \sum_{j_2=1}^{r_{\mathcal{N}_1}} \sum_{l_2=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i_1, j_1, l_1}^{\mathcal{N}}(t) \mathcal{B}_{i_2, j_2, l_2}^{\mathcal{N}}(t) \langle U_{j_1}^{\mathcal{N}_1}(t) \cdot \dot{U}_{l_1}^{\mathcal{N}_2}(t), U_{j_2}^{\mathcal{N}_1}(t) \cdot U_{l_2}^{\mathcal{N}_2}(t) \rangle \\ &= \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \dot{\mathcal{B}}_{i_1, j, l}^{\mathcal{N}}(t) \mathcal{B}_{i_2, j, l}^{\mathcal{N}}(t) \\ &\stackrel{(4.9)}{=} 0 \end{aligned}$$

für alle $i_1, i_2 \in \{1, \dots, r_{\mathcal{N}}\}$. Mittels Induktion über die Level erhalten wir Bedingung (4.9) für alle Knoten $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$. Mit

$$\begin{aligned} \langle U_i^{\mathcal{N}}(t), U_j^{\mathcal{N}}(t) \rangle &= \langle U_i^{\mathcal{N}}(0), U_j^{\mathcal{N}}(0) \rangle + \int_0^t \langle \dot{U}_i^{\mathcal{N}}(s), U_j^{\mathcal{N}}(s) \rangle + \langle U_i^{\mathcal{N}}(s), \dot{U}_j^{\mathcal{N}}(s) \rangle ds \\ &= \delta_{i,j} \end{aligned}$$

folgt die Erhaltung der Orthonormalität der Basisfunktionen. ■

Mit den Gauge-Bedingungen sind wir in der Lage Theorem 11 zu beweisen.

Beweis von Theorem 11. In diesem Beweis verzichten wir auf die Herleitung der Bewegungsgleichungen aus Theorem 11; wir verifizieren jedoch die Richtigkeit, indem wir zeigen, dass mit $\dot{Y}(t)$ wie aus Theorem 11 die Variationsformulierung

$$\langle F(Y(t)) - \dot{Y}(t), \delta Y \rangle = 0$$

für alle Elemente des Tangentialraumes $\delta Y \in \mathcal{T}_Y \mathcal{M}$ Gültigkeit besitzt. Das Vorgehen ist ähnlich wie im Beweis von Theorem 2 aus Kapitel 2, jedoch aufgrund der komplexen

Struktur der Mannigfaltigkeit $\mathcal{M} = \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ sehr viel technischer. Um die Lesbarkeit zu erleichtern, verzichten wir deshalb in diesem Beweis auf das Zeitargument, weisen jedoch darauf hin, dass sämtliche Transfertensoren und Basisfunktionen, sowie die Lastmatrizen zeitabhängige Objekte sind.

Verwenden wir für Y die Darstellung aus Lemma 8 auf Seite 29, so folgt

$$\dot{Y} = \sum_{i=1}^{r_{\mathcal{N}}} \dot{\mathcal{S}}(Y, \mathcal{N}, i) \cdot U_i^{\mathcal{N}} + \sum_{i=1}^{r_{\mathcal{N}}} \mathcal{S}(Y, \mathcal{N}, i) \cdot \dot{U}_i^{\mathcal{N}}. \quad (4.10)$$

Zusammen mit der Gauge-Bedingung, sowie der Orthonormalität der Basisfunktionen erhalten wir für $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$, $i \in \{1, \dots, r_{\mathcal{N}}\}$ die Gleichungen

$$\langle P(Y)F(Y), U_i^{\mathcal{N}} \rangle = \langle \dot{Y}, U_i^{\mathcal{N}} \rangle = \dot{\mathcal{S}}(Y, \mathcal{N}, i). \quad (4.11)$$

Darüber hinaus sind die Lastmatrizen $M^{\mathcal{N}}$, sowie deren Inverse symmetrisch, denn es gilt

$$m_{i,j}^{\mathcal{N}} = \langle \mathcal{S}(Y, \mathcal{N}, i), \mathcal{S}(Y, \mathcal{N}, j) \rangle = m_{j,i}^{\mathcal{N}}$$

für alle $j \in \{1, \dots, r_{\mathcal{N}}\}$. Damit folgt

$$\sum_{k=1}^{r_{\mathcal{N}}} m_{k,i}^{\mathcal{N}} w_{k,j}^{\mathcal{N}} = \delta_{i,j}. \quad (4.12)$$

Zusammen mit den Gleichungen (4.10), (4.11) und (4.12) können wir in den folgenden drei Beweisabschnitten zeigen, dass die Variationsformulierung für verschiedene Variationen $\delta Y \in \mathcal{T}_Y \mathcal{M}$ Gültigkeit besitzt:

- (i) Wir starten für beliebige $\delta \mathcal{B}^{\mathcal{N}_0} \in \mathbb{R}^{r_{\mathcal{N}_1} \times r_{\mathcal{N}_2}}$, $j \in \{1, \dots, r_{\mathcal{N}_1}\}$, $l \in \{1, \dots, r_{\mathcal{N}_2}\}$ mit der Variation

$$\delta Y = \delta \mathcal{B}_{1,j,l}^{\mathcal{N}_0} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \in \mathcal{T}_Y \mathcal{M}. \quad (4.13)$$

Mit (4.5) erhalten wir die Gleichungen

$$\begin{aligned} \langle \dot{Y}, \delta Y \rangle &= \left\langle \sum_{j'=1}^{r_{\mathcal{N}_1}} \sum_{l'=1}^{r_{\mathcal{N}_2}} \dot{\mathcal{B}}_{1,j',l'}^{\mathcal{N}_0} U_{j'}^{\mathcal{N}_1} \cdot U_{l'}^{\mathcal{N}_2}, \delta \mathcal{B}_{1,j,l}^{\mathcal{N}_0} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \right\rangle \\ &+ \left\langle \sum_{j'=1}^{r_{\mathcal{N}_1}} \sum_{l'=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{1,j',l'}^{\mathcal{N}_0} \dot{U}_{j'}^{\mathcal{N}_1} \cdot U_{l'}^{\mathcal{N}_2}, \delta \mathcal{B}_{1,j,l}^{\mathcal{N}_0} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \right\rangle \\ &+ \left\langle \sum_{j'=1}^{r_{\mathcal{N}_1}} \sum_{l'=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{1,j',l'}^{\mathcal{N}_0} U_{j'}^{\mathcal{N}_1} \cdot \dot{U}_{l'}^{\mathcal{N}_2}, \delta \mathcal{B}_{1,j,l}^{\mathcal{N}_0} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \right\rangle \\ &= \delta \mathcal{B}_{1,j,l}^{\mathcal{N}_0} \dot{\mathcal{B}}_{1,j,l}^{\mathcal{N}_0} \\ &= \delta \mathcal{B}_{1,j,l}^{\mathcal{N}_0} \langle P(Y)F(Y), U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \rangle \\ &= \langle F(Y), \delta \mathcal{B}_{1,j,l}^{\mathcal{N}_0} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \rangle \\ &= \langle F(Y), \delta Y \rangle. \end{aligned}$$

Daraus folgt, dass die Variationsformulierung für alle δY wie aus (4.13) erfüllt ist.

- (ii) Als nächsten Schritt betrachten wir für beliebige $\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\}$, $\delta \mathcal{B}^{\mathcal{N}} \in \mathbb{R}^{r_{\mathcal{N}} \times r_{\mathcal{N}_1} \times r_{\mathcal{N}_2}}$ und $i \in \{1, \dots, r_{\mathcal{N}}\}$, $j \in \{1, \dots, r_{\mathcal{N}_1}\}$, $l \in \{1, \dots, r_{\mathcal{N}_2}\}$ mit $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$ die Variation

$$\delta Y = \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} \mathcal{S}(Y, \mathcal{N}, i) \cdot U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}. \quad (4.14)$$

Mit (4.10) und (4.11) gelten die Gleichungen

$$\begin{aligned} \langle \dot{Y}, \delta Y \rangle &= \left\langle \sum_{i'=1}^{r_{\mathcal{N}}} \dot{\mathcal{S}}(Y, \mathcal{N}, i') \cdot U_{i'}^{\mathcal{N}}, \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} \mathcal{S}(Y, \mathcal{N}, i) \cdot U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \right\rangle \\ &\quad + \left\langle \sum_{i'=1}^{r_{\mathcal{N}}} \mathcal{S}(Y, \mathcal{N}, i') \cdot \dot{U}_{i'}^{\mathcal{N}}, \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} \mathcal{S}(Y, \mathcal{N}, i) \cdot U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \right\rangle \\ &= \sum_{i'=1}^{r_{\mathcal{N}}} \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} \left\langle \dot{\mathcal{S}}(Y, \mathcal{N}, i'), \mathcal{S}(Y, \mathcal{N}, i) \right\rangle \left\langle U_{i'}^{\mathcal{N}}, U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \right\rangle \\ &\quad + \sum_{i'=1}^{r_{\mathcal{N}}} \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} \left\langle \mathcal{S}(Y, \mathcal{N}, i'), \mathcal{S}(Y, \mathcal{N}, i) \right\rangle \left\langle \dot{U}_{i'}^{\mathcal{N}}, U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \right\rangle \\ &= \sum_{i'=1}^{r_{\mathcal{N}}} \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} \left\langle P(Y)F(Y), \mathcal{S}(Y, \mathcal{N}, i) \cdot U_{i'}^{\mathcal{N}} \right\rangle \mathcal{B}_{i',j,l}^{\mathcal{N}} \\ &\quad + \sum_{i'=1}^{r_{\mathcal{N}}} \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} m_{i',i}^{\mathcal{N}} \dot{\mathcal{B}}_{i',j,l}^{\mathcal{N}} \\ &= \sum_{i'=1}^{r_{\mathcal{N}}} \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} \left\langle P(Y)F(Y), \mathcal{S}(Y, \mathcal{N}, i) \cdot U_{i'}^{\mathcal{N}} \right\rangle \mathcal{B}_{i',j,l}^{\mathcal{N}} \\ &\quad + \sum_{i''=1}^{r_{\mathcal{N}}} \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} \left\langle F(Y), \mathcal{S}(Y, \mathcal{N}, i'') \cdot \mathcal{P}_{\mathcal{N}}^{\perp}(U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}) \right\rangle \underbrace{\left(\sum_{i'=1}^{r_{\mathcal{N}}} m_{i',i}^{\mathcal{N}} w_{i',i''}^{\mathcal{N}} \right)}_{\stackrel{(4.12)}{=} \delta_{i,i''}}. \end{aligned}$$

Dadurch, dass $\mathcal{S}(Y, \mathcal{N}, i) \cdot U_{i'}^{\mathcal{N}} \in \mathcal{T}_Y \mathcal{M}$ ein Element des Tangentialraumes ist, folgt

$$\begin{aligned} \left\langle P(Y)F(Y), \mathcal{S}(Y, \mathcal{N}, i) \cdot U_{i'}^{\mathcal{N}} \right\rangle &= \left\langle F(Y), P(Y) \mathcal{S}(Y, \mathcal{N}, i) \cdot U_{i'}^{\mathcal{N}} \right\rangle \\ &= \left\langle F(Y), \mathcal{S}(Y, \mathcal{N}, i) \cdot U_{i'}^{\mathcal{N}} \right\rangle \end{aligned}$$

und wir erhalten

$$\begin{aligned} \langle \dot{Y}, \delta Y \rangle &= \sum_{i'=1}^{r_{\mathcal{N}}} \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} \left\langle F(Y), \mathcal{S}(Y, \mathcal{N}, i) \cdot U_{i'}^{\mathcal{N}} \right\rangle \mathcal{B}_{i',j,l}^{\mathcal{N}} \\ &\quad + \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} \left\langle F(Y), \mathcal{S}(Y, \mathcal{N}, i) \cdot \mathcal{P}_{\mathcal{N}}^{\perp}(U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}) \right\rangle. \end{aligned} \quad (4.15)$$

Zugleich gilt

$$\begin{aligned}
\langle F(Y), \delta Y \rangle &= \left\langle F(Y), \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} \mathcal{S}(Y, \mathcal{N}, i) \cdot U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \right\rangle \\
&= \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} \left\langle F(Y), \mathcal{S}(Y, \mathcal{N}, i) \cdot \mathcal{P}_{\mathcal{N}}(U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}) \right\rangle \\
&\quad + \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} \left\langle F(Y), \mathcal{S}(Y, \mathcal{N}, i) \cdot \mathcal{P}_{\mathcal{N}}^{\perp}(U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}) \right\rangle \\
&= \sum_{i'=1}^{r_{\mathcal{N}}} \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} \left\langle F(Y), \mathcal{S}(Y, \mathcal{N}, i) \cdot U_{i'}^{\mathcal{N}} \right\rangle \mathcal{B}_{i',j,l}^{\mathcal{N}} \\
&\quad + \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} \left\langle F(Y), \mathcal{S}(Y, \mathcal{N}, i) \cdot \mathcal{P}_{\mathcal{N}}^{\perp}(U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}) \right\rangle.
\end{aligned} \tag{4.16}$$

Offensichtlich stimmen (4.15) und (4.16) überein und es folgt, dass die Variationsformulierung für alle Variationen δY wie in (4.14) Gültigkeit besitzt.

(iii) Zuletzt betrachten wir für beliebige $\ell \in \mathcal{L}(T)$, $i \in \{1, \dots, r_{\ell}\}$ und $\delta U_i^{\ell} : I_{\ell} \rightarrow \mathbb{R}$ die Variation

$$\delta Y = \mathcal{S}(Y, \ell, i) \cdot \delta U_i^{\ell}. \tag{4.17}$$

Mit (4.10) und (4.11) erhalten wir die Gleichungen

$$\begin{aligned}
\langle \dot{Y}, \delta Y \rangle &= \left\langle \sum_{i'=1}^{r_{\ell}} \dot{\mathcal{S}}(Y, \ell, i') \cdot U_{i'}^{\ell}, \mathcal{S}(Y, \ell, i) \cdot \delta U_i^{\ell} \right\rangle \\
&\quad + \left\langle \sum_{i'=1}^{r_{\ell}} \mathcal{S}(Y, \ell, i') \cdot \dot{U}_{i'}^{\ell}, \mathcal{S}(Y, \ell, i) \cdot \delta U_i^{\ell} \right\rangle \\
&= \sum_{i'=1}^{r_{\ell}} \left\langle \dot{\mathcal{S}}(Y, \ell, i'), \mathcal{S}(Y, \ell, i) \right\rangle \left\langle U_{i'}^{\ell}, \delta U_i^{\ell} \right\rangle \\
&\quad + \sum_{i'=1}^{r_{\ell}} \left\langle \mathcal{S}(Y, \ell, i'), \mathcal{S}(Y, \ell, i) \right\rangle \left\langle \dot{U}_{i'}^{\ell}, \delta U_i^{\ell} \right\rangle \\
&= \sum_{i'=1}^{r_{\ell}} \left\langle P(Y)F(Y), \mathcal{S}(Y, \ell, i) \cdot U_{i'}^{\ell} \right\rangle \left\langle U_{i'}^{\ell}, \delta U_i^{\ell} \right\rangle \\
&\quad + \sum_{i'=1}^{r_{\ell}} m_{i',i}^{\ell} \left\langle \sum_{i''=1}^{r_{\ell}} w_{i',i''}^{\ell} \mathcal{P}_{\ell}^{\perp} \langle F(Y), \mathcal{S}(Y, \ell, i'') \rangle, \delta U_i^{\ell} \right\rangle \\
&= \sum_{i'=1}^{r_{\ell}} \left\langle P(Y)F(Y), \mathcal{S}(Y, \ell, i) \cdot U_{i'}^{\ell} \right\rangle \left\langle U_{i'}^{\ell}, \delta U_i^{\ell} \right\rangle \\
&\quad + \sum_{i''=1}^{r_{\ell}} \left\langle \mathcal{P}_{\ell}^{\perp} F(Y), \mathcal{S}(Y, \ell, i'') \cdot \delta U_i^{\ell} \right\rangle \underbrace{\left(\sum_{i'=1}^{r_{\ell}} m_{i',i}^{\ell} w_{i',i''}^{\ell} \right)}_{\stackrel{(4.12)}{=} \delta_{i,i''}}.
\end{aligned}$$

Da $\mathcal{S}(Y, \ell, i) \cdot U_{i'}^\ell \in \mathcal{T}_Y \mathcal{M}$ ein Element des Tangentialraumes ist, folgt

$$\begin{aligned} \langle \dot{Y}, \delta Y \rangle &= \sum_{i'=1}^{r_\ell} \langle F(Y), \mathcal{S}(Y, \ell, i) \cdot U_{i'}^\ell \rangle \langle U_{i'}^\ell, \delta U_i^\ell \rangle \\ &\quad + \langle F(Y), \mathcal{S}(Y, \ell, i) \cdot \mathcal{P}_\ell^\perp \delta U_i^\ell \rangle. \end{aligned} \quad (4.18)$$

Andererseits gilt

$$\begin{aligned} \langle F(Y), \delta Y \rangle &= \langle F(Y), \mathcal{S}(Y, \ell, i) \cdot \delta U_i^\ell \rangle \\ &= \langle F(Y), \mathcal{S}(Y, \ell, i) \cdot \mathcal{P}_\ell \delta U_i^\ell \rangle \\ &\quad + \langle F(Y), \mathcal{S}(Y, \ell, i) \cdot \mathcal{P}_\ell^\perp \delta U_i^\ell \rangle \\ &= \sum_{i'=1}^{r_\ell} \langle F(Y), \mathcal{S}(Y, \ell, i) \cdot U_{i'}^\ell \rangle \langle U_{i'}^\ell, \delta U_i^\ell \rangle \\ &\quad + \langle F(Y), \mathcal{S}(Y, \ell, i) \cdot \mathcal{P}_\ell^\perp \delta U_i^\ell \rangle. \end{aligned} \quad (4.19)$$

Aus der Gleichheit von (4.18) und (4.19) folgt wiederum die Gültigkeit der Variationsformulierung für alle Variationen δY wie in (4.17).

Betrachten wir nun die Superposition der Variationen (4.13), (4.14) and (4.17), so spannt diese den gesamten Tangentialraum auf und es folgt

$$\langle F(Y) - \dot{Y}, \delta Y \rangle = 0 \text{ für alle } \delta Y \in \mathcal{T}_Y \mathcal{M}. \quad \blacksquare$$

Analog zu dem Vorgehen in Kapitel 2 bietet das Variationsprinzip die Möglichkeit, einen zeitabhängigen Tensor $A(t) \in \mathbb{R}^{N_1 \times \dots \times N_d}$ in der Mannigfaltigkeit $\mathcal{M} = \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ zu approximieren. Zu diesem Zweck benötigen wir eine Formel für die Projektion in den Tangentialraum.

Korollar 13. *Sei $B \in \mathbb{R}^{N_1 \times \dots \times N_d}$ ein beliebiger Tensor und $Y \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ ein Tensor im hierarchischen Tucker-Format, welcher durch das Tupel $((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^\ell)_{\ell \in \mathcal{L}(T)})$ dargestellt wird. Für $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ sei*

$$M^{\mathcal{N}} = \left(m_{i,j}^{\mathcal{N}} \right)_{i,j=1}^{r_{\mathcal{N}}} \in \mathbb{R}^{r_{\mathcal{N}} \times r_{\mathcal{N}}} \quad (4.20)$$

eine symmetrische Lastmatrix mit Einträgen

$$m_{i,j}^{\mathcal{N}} = \langle \mathcal{S}(Y, \mathcal{N}, i), \mathcal{S}(Y, \mathcal{N}, j) \rangle$$

und $W^{\mathcal{N}} = (M^{\mathcal{N}})^{-1}$ dessen Inverse. Dann erfüllt für $(\mathcal{N}_1^0, \mathcal{N}_2^0) = \text{succ}(\mathcal{N}_0)$ die Projektion von B in den Tangentialraum $\mathcal{T}_Y \mathcal{M}$ die Gleichung

$$\begin{aligned} P(Y)B &= \sum_{j=1}^{r_{\mathcal{N}_1^0}} \sum_{l=1}^{r_{\mathcal{N}_2^0}} \delta \mathcal{B}_{1,j,l}^{\mathcal{N}_0} U_j^{\mathcal{N}_1^0} \cdot U_l^{\mathcal{N}_2^0} \\ &+ \sum_{j=1}^{r_{\mathcal{N}_1^0}} \sum_{l=1}^{r_{\mathcal{N}_2^0}} \mathcal{B}_{1,j,l}^{\mathcal{N}_0} \delta U_j^{\mathcal{N}_1^0} \cdot U_l^{\mathcal{N}_2^0} + \sum_{j=1}^{r_{\mathcal{N}_1^0}} \sum_{l=1}^{r_{\mathcal{N}_2^0}} \mathcal{B}_{1,j,l}^{\mathcal{N}_0} U_j^{\mathcal{N}_1^0} \cdot \delta U_l^{\mathcal{N}_2^0} \end{aligned}$$

mit $\delta \mathcal{B}_{1,j,l}^{\mathcal{N}_0} = \langle B, U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \rangle$.

Die Funktionen $\delta U_i^{\mathcal{N}} : I_{\mathcal{N}} \rightarrow \mathbb{R}$ sind für $\mathcal{N} \in \mathcal{I}(T)$, $i \in \{1, \dots, r_{\mathcal{N}}\}$ und $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$ durch die Rekursion

$$\begin{aligned} \delta U_i^{\mathcal{N}} &= \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \\ &+ \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i,j,l}^{\mathcal{N}} \delta U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} + \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i,j,l}^{\mathcal{N}} U_j^{\mathcal{N}_1} \cdot \delta U_l^{\mathcal{N}_2} \end{aligned}$$

beschrieben, wobei die Einträge der Tensoren $\delta \mathcal{B}^{\mathcal{N}} \in \mathbb{R}^{r_{\mathcal{N}} \times r_{\mathcal{N}_1} \times r_{\mathcal{N}_2}}$ für $j \in \{1, \dots, r_{\mathcal{N}_1}\}$ und $l \in \{1, \dots, r_{\mathcal{N}_2}\}$ die Werte

$$\delta \mathcal{B}_{i,j,l}^{\mathcal{N}} = \sum_{i'=1}^{r_{\mathcal{N}}} w_{i,i'}^{\mathcal{N}} \langle B, \mathcal{S}(Y, \mathcal{N}, i') \cdot \mathcal{P}_{\mathcal{N}}^{\perp}(U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}) \rangle$$

annehmen. Die Funktionen $\delta U_i^{\ell} : I_{\mathcal{N}} \rightarrow \mathbb{R}$ erfüllen für $\ell \in \mathcal{L}(T)$ und $i \in \{1, \dots, r_{\ell}\}$ die Gleichung

$$\delta U_i^{\ell} = \sum_{i'=1}^{r_{\ell}} w_{i,i'}^{\ell} \mathcal{P}_{\ell}^{\perp} \langle B, \mathcal{S}(Y, \ell, i') \rangle.$$

Beweis. Vertauschen wir im Beweis von Theorem 11 den Term \dot{Y} mit $P(Y)B$ und $F(Y)$ mit B , so folgt mit den Formeln aus Korollar 13, dass die Variationsformulierung

$$\langle B - P(Y)B, \delta Y \rangle = 0$$

für alle $\delta Y \in \mathcal{T}_Y \mathcal{M}$ erfüllt ist. ■

Setzen wir voraus, dass sich $A(t)$ glatt in der Zeit bewegt und das Anfangswertproblem

$$\dot{Y}(t) = P(Y(t))\dot{A}(t), \quad Y(0) = Y_0 \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T}), \quad t \geq 0 \quad (4.21)$$

eine Lösung besitzt, so stellt $Y(t)$ eine Approximation des zeitabhängigen Tensors $A(t)$ dar, welche auf der Mannigfaltigkeit $\mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ verläuft. Falls der Tensor $A(0)$ kein Element der Mannigfaltigkeit ist, liefert Algorithmus 1 aus Kapitel 5 eine Möglichkeit, eine

Approximation in der Menge \mathcal{H} -Tucker $((r_{\mathcal{N}})_{\mathcal{N} \in \mathcal{T}})$ zu berechnen, welche sich als Startwert für das Anfangswertproblem (4.21) eignet.

5.

Matrix-Darstellung von Tensoren

Für die folgenden Kapitel wird eine weitere Darstellung von Tensoren eine bedeutende Rolle spielen. Ein Tensor kann nicht nur als eine Funktion interpretiert werden, sondern auch als eine Matrix, die man dadurch gewinnt, dass man alle Einträge des hochdimensionalen Zahlenfeldes in einer bestimmten Weise anordnet; die Einträge ändern sich nicht, nur die Anordnung. Diese Darstellung erlaubt es uns, eine äquivalente Definition für Tensoren im hierarchischen Tucker-Format zu formulieren; eine Definition, welche keine Rekursion mehr beinhaltet. Wir stellen zudem einen Algorithmus vor, welcher aus einem beliebigen Tensor im hierarchischen Tucker-Format ein Tupel $\left((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^\ell)_{\ell \in \mathcal{L}(T)} \right)$ bestimmt, durch welches dieser Tensor gemäß Definition 4 dargestellt werden kann und zugleich die Annahmen 1 und 2 erfüllt. Nachdem wir anschließend die Ergebnisse der letzten Kapitel in diese Notation übertragen haben, benutzen wir die Matrix-Darstellung, um für die Projektion eines Tensors in den Tangentialraum eine kompakte Darstellung zu finden; eine Darstellung, welche im Gegensatz zu Korollar 13 in Kapitel 4 keine Rekursion benötigt. Diese Darstellung dient als Grundlage für die Resultate in Kapitel 6. Zuletzt tragen wir in diesem Kapitel den noch ausstehenden Rest des Beweises von Lemma 9 nach.

Für $\mathcal{N} \in T$ definieren wir den **Komplementknoten** \mathcal{N}_C als

$$\mathcal{N}_C := \{1, \dots, d\} \setminus \mathcal{N}.$$

Um einen Tensor als Matrix oder auch als einen Vektor zu interpretieren, betrachten wir zwei Formalismen: Die **Matrifizierung** und die **Vektorisierung**. Für einen Knoten $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ verstehen wir unter der Matrifizierung eines Tensors $Y \in \mathbb{R}^{N_1 \times \dots \times N_d}$ die Matrix

$$Y^{(\mathcal{N}, \mathcal{N}_C)} \in \mathbb{R}^{\pi(\mathcal{N}) \times \pi(\mathcal{N}_C)}, \quad \pi(\mathcal{N}) = \prod_{\mu \in \mathcal{N}} N_\mu, \quad \pi(\mathcal{N}_C) = \prod_{\mu \in \mathcal{N}_C} N_\mu,$$

welche daraus resultiert, dass alle Richtungen, welche im Knoten \mathcal{N} enthalten sind, in den Zeilen der Matrix angeordnet werden und alle Richtungen in \mathcal{N}_C in den Spalten der Matrix. Welche Abbildung für die Anordnung verwendet wird, ist unerheblich; wichtig ist, dass die Anordnung in den Zeilen und in den Spalten auf dieselbe Weise geschieht, also

$$Y^{(\mathcal{N}, \mathcal{N}_C)} = \left(Y^{(\mathcal{N}_C, \mathcal{N})} \right)^T$$

erfüllt. Eine mögliche Wahl für eine Anordnung der Einträge eines Tensors in einer Matrix liefert beispielsweise der Befehl `reshape` in MATLAB. Für die Umkehrabbildung, also die Abbildung, welche aus einer Matrix einen Tensor erzeugt, verwenden wir die Notation

$$Y = \left(Y^{(\mathcal{N}, \mathcal{N}_C)} \right)_{(\mathcal{N}, \mathcal{N}_C)}.$$

Die Vektorisierung ist ein Spezialfall der Matrifizierung und ordnet alle Einträge eines Tensors in einem Vektor an. Wir verwenden für die Vektorisierung die Notation

$$Y^{(\mathcal{N}_0, \{\})} \in \mathbb{R}^{N_1 \cdots N_d}$$

und bezeichnen die inverse Abbildung als

$$Y = \left(Y^{(\mathcal{N}_0, \{\})} \right)_{(\mathcal{N}_0, \{\})}.$$

Es sei darauf hingewiesen, dass es für die Frobeniusnorm unerheblich ist, ob das Argument ein Tensor ist, dessen Matrifizierung, oder Vektorisierung, es gilt

$$\|Y\|_F = \left\| Y^{(\mathcal{N}, \mathcal{N}_C)} \right\|_F = \left\| Y^{(\mathcal{N}_0, \{\})} \right\|_F.$$

In Lemma 8 aus Abschnitt 3.3 haben wir mit der Formel

$$Y = \sum_{i=1}^{r_{\mathcal{N}}} \mathcal{S}(Y, \mathcal{N}, i) \cdot U_i^{\mathcal{N}}$$

für einen Knoten $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ eine Möglichkeit gefunden, einen Tensor im hierarchischen Tucker-Format durch seine Basisfunktionen, sowie den Single-hole Operator darzustellen. Diese Art der Darstellung wollen wir in Matrix-Schreibweise übertragen. Zu diesem Zweck definieren wir für $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ die orthonormale **Basismatrix**

$$\mathbb{U}^{\mathcal{N}} := \left((U_1^{\mathcal{N}})^{(\mathcal{N}, \{\})} | \dots | (U_{r_{\mathcal{N}}}^{\mathcal{N}})^{(\mathcal{N}, \{\})} \right) \in \mathbb{R}^{\pi(\mathcal{N}) \times r_{\mathcal{N}}},$$

sowie die **Single-Hole-Matrix**

$$\mathbb{S}^{\mathcal{N}} := \left((\mathcal{S}(Y, \mathcal{N}, 1))^{(\mathcal{N}, \{\})} | \dots | (\mathcal{S}(Y, \mathcal{N}, r_{\mathcal{N}}))^{(\mathcal{N}, \{\})} \right) \in \mathbb{R}^{\pi(\mathcal{N}_C) \times r_{\mathcal{N}}}.$$

Der Begriff orthonormale Basismatrix soll dabei nicht andeuten, dass es sich um eine quadratische Matrix handelt, sondern lediglich, dass die Spalten der Matrix orthonormal zueinander sind. Die Orthonormalität der Basismatrizen, also $(\mathbb{U}^{\mathcal{N}})^T \mathbb{U}^{\mathcal{N}} = I$, folgt unmittelbar aus Annahme 1. Übertragen wir nun Lemma 8 in Matrix-Schreibweise, so gilt für jeden Knoten $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ die Gleichung

$$Y^{(\mathcal{N}, \mathcal{N}_C)} = \mathbb{U}^{\mathcal{N}} \left(\mathbb{S}^{\mathcal{N}} \right)^T \quad \text{beziehungsweise} \quad \mathbb{S}^{\mathcal{N}} = \left(Y^{(\mathcal{N}, \mathcal{N}_C)} \right)^T \mathbb{U}^{\mathcal{N}}. \quad (5.1)$$

Im Folgenden bezeichnen wir den i -ten Singulärwert (der Größe nach absteigend) einer Matrix $M \in \mathbb{R}^{m \times n}$ mit $\sigma_i(M)$. Setzen wir ferner voraus, dass die Matrix vollen Spaltenrang besitzt, so bezeichnen wir mit $M^+ = (M^T M)^{-1} M^T$ die Pseudoinverse der Matrix. Aus der Orthonormalität der Basismatrizen und Gleichung (5.1) folgt, dass die Singulärwerte der Single-Hole-Matrix $\mathbb{S}^{\mathcal{N}}$ mit den Singulärwerten der Matrifizierung $Y^{(\mathcal{N}, \mathcal{N}_C)}$ übereinstimmen, also

$$\sigma_i(Y^{(\mathcal{N}, \mathcal{N}_C)}) = \sigma_i(\mathbb{S}^{\mathcal{N}})$$

für alle $i \in \{1, \dots, r_{\mathcal{N}}\}$. Mit den Basismatrizen definieren wir in Analogie zu Definition 7 die Projektion

$$P^{\mathcal{N}} := \mathbb{U}^{\mathcal{N}} (\mathbb{U}^{\mathcal{N}})^T \quad \text{beziehungsweise} \quad (P^{\mathcal{N}})^{\perp} := I - \mathbb{U}^{\mathcal{N}} (\mathbb{U}^{\mathcal{N}})^T.$$

Darüber hinaus definieren wir für $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$ den binären Matrix-Operator

$$\begin{aligned} \mathbb{U}^{\mathcal{N}_1} \odot \mathbb{U}^{\mathcal{N}_2} &\rightarrow \mathbb{R}^{\pi(\mathcal{N}) \times (r_{\mathcal{N}_1} r_{\mathcal{N}_2})} \\ \mathbb{U}^{\mathcal{N}_1} \odot \mathbb{U}^{\mathcal{N}_2} &= (v_{1,1} | v_{1,2} | \dots | v_{1,r_{\mathcal{N}_2}} | v_{2,1} | \dots | v_{2,r_{\mathcal{N}_2}} | \dots | \dots | v_{r_{\mathcal{N}_1}, r_{\mathcal{N}_2}}) \\ \text{mit } v_{j,l} &= \left(U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \right)^{(\mathcal{N}, \{\})}. \end{aligned}$$

Mit den vorgestellten Formalismen sind wir in der Lage, das erste Resultat dieses Abschnitts vorzustellen: Eine äquivalente Definition der Menge der Tensoren im hierarchischen Tucker-Format.

Lemma 14. *Es gilt*

$$\mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T}) = \left\{ X \in \mathbb{R}^{N_1 \times \dots \times N_d} : \text{rang} \left(X^{(\mathcal{N}, \mathcal{N}_C)} \right) = r_{\mathcal{N}} \text{ für alle } \mathcal{N} \in T \setminus \{\mathcal{N}_0\} \right\}.$$

Eine äquivalente Definition lässt sich ebenfalls für die Menge der Tensoren im Tucker-Format finden.

$$\begin{aligned} \text{Tucker}(r_1, \dots, r_d) & \\ &= \left\{ X \in \mathbb{R}^{N_1 \times \dots \times N_d} : \text{rang} \left(X^{(\{\omega\}, \omega_C)} \right) = r_{\omega} \text{ für } \omega \in \{1, \dots, d\}, \omega_C = \{1, \dots, d\} \setminus \{\omega\} \right\} \end{aligned} \quad (5.2)$$

Die Menge \mathcal{H} -Tucker $((r_{\mathcal{N}})_{\mathcal{N} \in T})$ ist demnach eine Teilmenge von Tucker (r_1, \dots, r_d) , wenn die Blätter des Dimensionsbaumes einelementig sind und der hierarchische Rang an den Blättern mit dem Tucker-Rang übereinstimmt. Da das Augenmerk dieser Arbeit auf Tensoren im hierarchischen Tucker-Format liegt, beweisen wir lediglich die Äquivalenz der Mengen aus Lemma 14 und weisen darauf hin, dass sich Gleichung (5.2) mit ähnlichen Argumenten zeigen lässt.

Beweis von Lemma 14. Wir definieren die Menge

$$\widehat{\mathcal{M}} := \left\{ X \in \mathbb{R}^{N_1 \times \dots \times N_d} : \text{rang}\left(X^{(\mathcal{N}, \mathcal{N}^c)}\right) = r_{\mathcal{N}} \text{ für alle } \mathcal{N} \in T \setminus \{\mathcal{N}_0\} \right\}$$

und zeigen zunächst \mathcal{H} -Tucker $((r_{\mathcal{N}})_{\mathcal{N} \in T}) \subseteq \widehat{\mathcal{M}}$. Sei $Y \in \mathcal{H}$ -Tucker $((r_{\mathcal{N}})_{\mathcal{N} \in T})$ ein Tensor im hierarchischen Tucker-Format, welcher durch das Tupel $\left(\left(\mathcal{B}^{\mathcal{N}}\right)_{\mathcal{N} \in \mathcal{I}(T)}, \left(U^{\ell}\right)_{\ell \in \mathcal{L}(T)}\right)$ dargestellt wird. Nach Annahme 2 ist die Sequenz

$$\langle Y, U_1^{\mathcal{N}} \rangle, \dots, \langle Y, U_{r_{\mathcal{N}}}^{\mathcal{N}} \rangle$$

linear unabhängig für alle $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$. Es folgt, dass die Spalten der Single-Hole-Matrix $\mathbb{S}^{\mathcal{N}}$ linear unabhängig sind, also

$$\text{rang}\left(\mathbb{S}^{\mathcal{N}}\right) = \text{rang}\left(Y^{(\mathcal{N}, \mathcal{N}^c)}\right) = r_{\mathcal{N}} \text{ für alle } \mathcal{N} \in T \setminus \{\mathcal{N}_0\}$$

und damit $Y \in \widehat{\mathcal{M}}$.

Sei nun $\widehat{Y} \in \widehat{\mathcal{M}}$. Um zu zeigen, dass \widehat{Y} ein Tensor im hierarchischen Tucker-Format ist, benötigen wir ein Tupel $\left(\left(\mathcal{B}^{\mathcal{N}}\right)_{\mathcal{N} \in \mathcal{I}(T)}, \left(U^{\ell}\right)_{\ell \in \mathcal{L}(T)}\right)$, durch das \widehat{Y} gemäß Definition 4 dargestellt wird und zugleich die Annahmen 1 und 2 erfüllt. Zu diesem Zweck wenden wir für jeden Knoten $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ die Singulärwertzerlegung auf die Matrifizierung $\widehat{Y}^{(\mathcal{N}, \mathcal{N}^c)}$ an und erhalten auf diese Weise die Links-Singulärvektoren $u_1^{\mathcal{N}}, \dots, u_{r_{\mathcal{N}}}^{\mathcal{N}}$, die Rechts-Singulärvektoren $v_1^{\mathcal{N}}, \dots, v_{r_{\mathcal{N}}}^{\mathcal{N}}$, sowie positive Skalare $d_1^{\mathcal{N}}, \dots, d_{r_{\mathcal{N}}}^{\mathcal{N}}$, so dass

$$\widehat{Y}^{(\mathcal{N}, \mathcal{N}^c)} = \sum_{i=1}^{r_{\mathcal{N}}} d_i^{\mathcal{N}} u_i^{\mathcal{N}} (v_i^{\mathcal{N}})^T.$$

Anhand der Singulärvektoren definieren wir für $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ und $i \in \{1, \dots, r_{\mathcal{N}}\}$ die orthonormalen Funktionen

$$U_i^{\mathcal{N}} := (u_i^{\mathcal{N}})_{(\mathcal{N}, \{\})} \quad \text{und} \quad V_i^{\mathcal{N}} := (v_i^{\mathcal{N}})_{(\mathcal{N}, \{\})},$$

woraufhin die Gleichung

$$\widehat{Y} = \sum_{i=1}^{r_{\mathcal{N}}} d_i^{\mathcal{N}} U_i^{\mathcal{N}} V_i^{\mathcal{N}}$$

folgt. Wir zeigen nun, dass es sich bei den Funktionen $U_1^{\mathcal{N}}, \dots, U_{r_{\mathcal{N}}}^{\mathcal{N}}$ für $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ um die Basisfunktionen der Darstellung handelt. Aufgrund der Orthonormalität ist Annahme 1, welche wir von den Basisfunktionen fordern, bereits erfüllt und wir erhalten zudem

$$\widehat{Y} = \sum_{i=1}^{r_{\mathcal{N}}} \langle U_i^{\mathcal{N}}, \widehat{Y} \rangle \cdot U_i^{\mathcal{N}} \quad \text{und} \quad U_i^{\mathcal{N}} = \frac{1}{d_i^{\mathcal{N}}} \langle V_i^{\mathcal{N}}, \widehat{Y} \rangle. \quad (5.3)$$

Mit $\langle U_i^{\mathcal{N}}, \widehat{Y} \rangle = d_i^{\mathcal{N}} V_i^{\mathcal{N}}$ für $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ und $i \in \{1, \dots, r_{\mathcal{N}}\}$ folgt die lineare Unabhängigkeit der Funktionen

$$\langle U_1^{\mathcal{N}}, \widehat{Y} \rangle, \dots, \langle U_{r_{\mathcal{N}}}^{\mathcal{N}}, \widehat{Y} \rangle,$$

wodurch Annahme 2 ebenfalls erfüllt ist. Falls nun

$$\begin{aligned} & \text{span}(\{U_i^{\mathcal{N}} : i \in \{1, \dots, r_{\mathcal{N}}\}\}) \\ & \subseteq \text{span}(\{U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} : j \in \{1, \dots, r_{\mathcal{N}_1}\}, l \in \{1, \dots, r_{\mathcal{N}_2}\}\}) \end{aligned} \quad (5.4)$$

für alle $\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\}$ mit $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$ gilt, erhalten wir mit

$$\mathcal{B}_{1,j,l}^{\mathcal{N}_0} := d_j^{\mathcal{N}_0} \delta_{j,l} \quad \text{und} \quad \mathcal{B}_{i,j,l}^{\mathcal{N}} := \langle U_i^{\mathcal{N}}, U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \rangle$$

ein Tupel $((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^{\ell})_{\ell \in \mathcal{L}(T)})$, durch welches der Tensor \widehat{Y} gemäß Definition 4 dargestellt wird und zugleich die Annahmen 1 und 2 erfüllt.

Sei $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ ein beliebiger Knoten und $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$. Nach (5.3) besitzt \widehat{Y} die Darstellungen

$$\sum_{j=1}^{r_{\mathcal{N}_1}} \langle U_j^{\mathcal{N}_1}, \widehat{Y} \rangle \cdot U_j^{\mathcal{N}_1} = \widehat{Y} = \sum_{l=1}^{r_{\mathcal{N}_2}} \langle U_l^{\mathcal{N}_2}, \widehat{Y} \rangle \cdot U_l^{\mathcal{N}_2}.$$

Setzen wir die eine Darstellung in die andere ein, so erhalten wir

$$\begin{aligned} \widehat{Y} &= \sum_{j=1}^{r_{\mathcal{N}_1}} \langle \widehat{Y}, U_j^{\mathcal{N}_1} \rangle \cdot U_j^{\mathcal{N}_1} \\ &= \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \langle \langle \widehat{Y}, U_l^{\mathcal{N}_2} \rangle \cdot U_l^{\mathcal{N}_2}, U_j^{\mathcal{N}_1} \rangle \cdot U_j^{\mathcal{N}_1} \\ &= \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \langle \langle \widehat{Y}, U_l^{\mathcal{N}_2} \rangle, U_j^{\mathcal{N}_1} \rangle \cdot U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \\ &= \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \langle \widehat{Y}, U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \rangle \cdot U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}. \end{aligned} \quad (5.5)$$

Das Einsetzen von (5.5) in (5.3) liefert

$$\begin{aligned}
U_i^{\mathcal{N}} &= \frac{1}{d_i} \langle \widehat{Y}, V_i^{\mathcal{N}} \rangle \\
&= \frac{1}{d_i} \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \langle \langle \widehat{Y}, U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \rangle \cdot U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}, V_i^{\mathcal{N}} \rangle \\
&= \frac{1}{d_i} \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \langle \langle \widehat{Y}, U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \rangle, V_i^{\mathcal{N}} \rangle \cdot U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \\
&= \frac{1}{d_i} \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \langle \widehat{Y}, U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \cdot V_i^{\mathcal{N}} \rangle \cdot U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}
\end{aligned}$$

und es folgt (5.4) und damit $\widehat{Y} \in \mathcal{H}$ -Tucker $((r_{\mathcal{N}})_{\mathcal{N} \in T})$. ■

Im Beweis von Lemma 14 haben wir für ein

$$\widehat{Y} \in \left\{ X \in \mathbb{R}^{N_1 \times \dots \times N_d} : \text{rang}(X^{(\mathcal{N}, \mathcal{N}^c)}) = r_{\mathcal{N}} \text{ für alle } \mathcal{N} \in T \setminus \{\mathcal{N}_0\} \right\}$$

ein Tupel $((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^{\ell})_{\ell \in \mathcal{L}(T)})$ konstruiert, welches den Tensor gemäß Definition 4 darstellt und die Annahmen 1 und 2 erfüllt. Analog zu dem Vorgehen im Beweis bestimmt Algorithmus 1 für einen Tensor im hierarchischen Tucker-Format ein Tupel $((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^{\ell})_{\ell \in \mathcal{L}(T)})$. Dabei ist Algorithmus 1 eine Verkettung der Algorithmen 1 und 3 aus [Gra10].

Algorithmus 1 : Finde Darstellung im hierarchischen Tucker-Format

Eingabe : Beliebiger Tensor $Y_{ex} \in \mathbb{R}^{N_1 \times \dots \times N_d}$ mit hierarchischem Rang $(r'_N)_{N \in T}$ und $r'_N \geq r_N$ für alle $N \in T$

Ausgabe : Tupel $\left((\mathcal{B}^N)_{N \in \mathcal{I}(T)}, (U^\ell)_{\ell \in \mathcal{L}(T)} \right)$ mit den Eigenschaften
 $Y_{ex} \approx \Psi \left((\mathcal{B}^N)_{N \in \mathcal{I}(T)}, (U^\ell)_{\ell \in \mathcal{L}(T)} \right) \in \mathcal{H}$ -Tucker $((r_N)_{N \in T})$
 und $Y_{ex} = \Psi \left((\mathcal{B}^N)_{N \in \mathcal{I}(T)}, (U^\ell)_{\ell \in \mathcal{L}(T)} \right)$ für $Y_{ex} \in \mathcal{H}$ -Tucker $((r_N)_{N \in T})$

for Jedes Blatt $\ell \in \mathcal{L}(T)$ **do**

Berechne die Singulärwertzerlegung von $Y_{ex}^{(\ell, \ell_C)}$ und definiere anhand der Links-Singulärvektoren $u_1^\ell, \dots, u_{r_\ell}^\ell$ für alle $i \in \{1, \dots, r_\ell\}$ die Funktionen

$$U_i^\ell := (u_i^\ell)_{(\ell, \{\})} ;$$

for Level $l = l_{\max}, \dots, 1$ (l_{\max} ist maximales Level) **do**

for Jeden Knoten $N \in \mathcal{I}(T)$ mit $\text{Level}(N) = l$ **do**

Berechne die Singulärwertzerlegung von $Y_{ex}^{(N, \mathcal{N}_C)}$ und definiere anhand der Links-Singulärvektoren $u_1^N, \dots, u_{r_N}^N$ für alle $i \in \{1, \dots, r_N\}$ die Funktionen

$$U_i^N := (u_i^N)_{(N, \{\})} ;$$

Berechne für $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(N)$ die Einträge des Transfertensors \mathcal{B}^N gemäß der Formel

$$\mathcal{B}_{i,j,k}^N := \langle U_i^N, U_j^{\mathcal{N}_1} \cdot U_k^{\mathcal{N}_2} \rangle ;$$

Berechne für $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(N_0)$ die Einträge des Transfertensors \mathcal{B}^{N_0} gemäß der Formel

$$\mathcal{B}_{1,j,k}^{N_0} := \langle Y_{ex}, U_j^{N_1} \cdot U_k^{N_2} \rangle ;$$

if $Y_{ex} \notin \mathcal{H}$ -Tucker $((r_N)_{N \in T})$ **then** (Orthogonalisierung der Funktionen)

for Jeden Knoten $N \in \mathcal{I}(T) \setminus \{N_0\}$ mit

\hat{N} als Mutterknoten und $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(N)$ als Kindknoten **do**

Assembliere die Matrix $\mathbb{B} \in \mathbb{R}^{r_{N_1} r_{N_2} \times r_N}$ mit den Einträgen

$$\mathbb{B}_{(j-1)r_{N_2}+k,i} := \mathcal{B}_{i,j,k}^N ;$$

Bestimme die QR-Zerlegung, $\mathbb{B} := QR$;

Überschreibe alle Einträge der Transfertensoren

$$\begin{aligned} \mathcal{B}_{i,j,k}^N &:= Q_{(j-1)r_{N_2}+k,i} ; \\ \mathcal{B}_{i,j,k}^{\hat{N}} &:= \begin{cases} \sum_{j'=1}^{r_{N_1}} \mathcal{B}_{i,j',k}^{\hat{N}} R_{j,j'} & \text{für } (N, \cdot) = \text{succ}(\hat{N}) \\ \sum_{k'=1}^{r_{N_2}} \mathcal{B}_{i,j,k'}^{\hat{N}} R_{k,k'} & \text{für } (\cdot, N) = \text{succ}(\hat{N}) \end{cases} ; \end{aligned}$$

Übergibt man dem Algorithmus anstelle eines Tensors $Y \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ einen Tensor $\widehat{Y} \in \mathcal{H}\text{-Tucker}((r'_{\mathcal{N}})_{\mathcal{N} \in T})$ mit höherem hierarchischen Rang, also mit

$$\begin{aligned} r'_{\mathcal{N}} &\geq r_{\mathcal{N}} \quad \text{für alle } \mathcal{N} \in T \\ \text{und } r'_{\mathcal{N}^*} &> r_{\mathcal{N}^*} \quad \text{für ein } \mathcal{N}^* \in T, \end{aligned}$$

so liefert die Ausgabe ebenfalls ein Tupel $\left((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^{\ell})_{\ell \in \mathcal{L}(T)} \right)$. Dann wird \widehat{Y} durch die Rekonstruktion $\Psi \left(\left((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^{\ell})_{\ell \in \mathcal{L}(T)} \right) \right)$ jedoch nur approximiert. Grasedyck gibt in diesem Fall die Abschätzung

$$\begin{aligned} \left\| \widehat{Y} - \Psi \left((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^{\ell})_{\ell \in \mathcal{L}(T)} \right) \right\|_F &\leq \sqrt{\sum_{\mathcal{N} \in \mathcal{I}(T)} \sum_{i > r_{\mathcal{N}}} \sigma_i^2(\widehat{Y}^{(\mathcal{N}, \mathcal{N}^c)})} \\ &\leq \sqrt{2d-2} \operatorname{dist} \left(\widehat{Y}, \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T}) \right) \end{aligned}$$

für die Güte der Approximation an, siehe [Gra10, Theorem 3.11]. Wird dem Algorithmus andererseits ein Tensor mit einem geringeren hierarchischen Rang übergeben, also mit

$$\begin{aligned} r'_{\mathcal{N}} &\leq r_{\mathcal{N}} \quad \text{für alle } \mathcal{N} \in T \\ \text{und } r'_{\mathcal{N}^*} &< r_{\mathcal{N}^*} \quad \text{für ein } \mathcal{N}^* \in T, \end{aligned}$$

so liefert die Ausgabe ein Tupel, welches den Tensor gemäß Definition 4 rekonstruiert, jedoch ist in diesem Fall Annahme 2 nicht weiter gültig.

Nachdem wir nun einen Algorithmus vorgestellt haben, welcher uns aus einem Tensor im hierarchischen Tucker-Format ein Tupel $\left((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^{\ell})_{\ell \in \mathcal{L}(T)} \right)$ bestimmt, durch das der Tensor rekonstruiert wird, wenden wir uns erneut der Projektion in den Tangentialraum $\mathcal{T}_Y \mathcal{M}$ der Mannigfaltigkeit $\mathcal{M} = \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ zu.

Korollar 15. *Sei $Y \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ ein Tensor im hierarchischen Tucker-Format, welcher durch das Tupel $\left((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^{\ell})_{\ell \in \mathcal{L}(T)} \right)$ dargestellt wird und $B \in \mathbb{R}^{N_1 \times \dots \times N_d}$ ein beliebiger Tensor, dann existiert unter den Bedingungen von Korollar 13 ein Tupel $\left((\delta \mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (\delta U^{\ell})_{\ell \in \mathcal{L}(T)} \right)$, so dass die Projektion von B in den Tangentialraum $\mathcal{T}_Y \mathcal{M}$ für $(\mathcal{N}_1, \mathcal{N}_2) = \operatorname{succ}(\mathcal{N}_0)$ die Gleichung*

$$P(Y)B = \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \delta \mathcal{B}_{1,j,l}^{\mathcal{N}_0} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} + \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{1,j,l}^{\mathcal{N}_0} \delta U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} + \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{1,j,l}^{\mathcal{N}_0} U_j^{\mathcal{N}_1} \cdot \delta U_l^{\mathcal{N}_2}$$

erfüllt. Für das Tupel $\left((\delta \mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (\delta U^{\ell})_{\ell \in \mathcal{L}(T)} \right)$ gelten die Gleichungen

$$\delta \mathcal{B}_{1,j,l}^{\mathcal{N}_0} = \left((U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2})^{(\mathcal{N}_0, \{\})} \right)^T B^{(\mathcal{N}_0, \{\})}$$

und für $\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\}$ mit $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$

$$\delta U_i^{\mathcal{N}} = \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} + \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i,j,l}^{\mathcal{N}} \delta U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} + \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i,j,l}^{\mathcal{N}} U_j^{\mathcal{N}_1} \cdot \delta U_l^{\mathcal{N}_2}.$$

Für $j \in \{1, \dots, r_{\mathcal{N}_1}\}$ und $l \in \{1, \dots, r_{\mathcal{N}_2}\}$ sind die Einträge des Tensors $\delta \mathcal{B}^{\mathcal{N}}$ durch

$$(\delta \mathcal{B}_{1,j,l}^{\mathcal{N}} \dots \delta \mathcal{B}_{r_{\mathcal{N}},j,l}^{\mathcal{N}}) = \left((U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2})^{(\mathcal{N}, \{\})} \right)^T (P^{\mathcal{N}})^{\perp} B^{(\mathcal{N}, \mathcal{N}_C)} (\mathbb{S}^{\mathcal{N}+})^T$$

beschrieben. Dargestellt durch die Matrix

$$\delta \mathbb{U}^{\ell} := \left((\delta U_1^{\ell})^{(\ell, \{\})} \mid \dots \mid (\delta U_{r_{\ell}}^{\ell})^{(\ell, \{\})} \right)$$

erfüllen die Funktionen $\delta U_1^{\ell}, \dots, \delta U_{r_{\ell}}^{\ell}$ für $\ell \in \mathcal{L}(T)$ die Gleichung

$$\delta \mathbb{U}^{\ell} = (P^{\ell})^{\perp} B^{(\ell, \ell_C)} (\mathbb{S}^{\ell+})^T.$$

Beweis. Die Gleichung für $\delta \mathcal{B}^{\mathcal{N}_0}$ folgt unmittelbar aus Korollar 13. Für die verbleibenden Tensoren $\delta \mathcal{B}^{\mathcal{N}}$ mit $\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\}$ formulieren wir zunächst die entsprechende Gleichung aus Korollar 13 um und erhalten auf diese Weise

$$\begin{aligned} \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} &= \sum_{i'=1}^{r_{\mathcal{N}}} \left\langle B, \mathcal{S}(Y, \mathcal{N}, i') \cdot \mathcal{P}_{\mathcal{N}}^{\perp}(U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}) \right\rangle w_{i,i'}^{\mathcal{N}} \\ &= \sum_{i'=1}^{r_{\mathcal{N}}} \left\langle \left\langle B, \mathcal{S}(Y, \mathcal{N}, i') \right\rangle, \mathcal{P}_{\mathcal{N}}^{\perp}(U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}) \right\rangle w_{i,i'}^{\mathcal{N}} \\ &= \sum_{i'=1}^{r_{\mathcal{N}}} \left\langle \mathcal{P}_{\mathcal{N}}^{\perp} \left\langle B, \mathcal{S}(Y, \mathcal{N}, i') \right\rangle, U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \right\rangle w_{i,i'}^{\mathcal{N}} \\ &= \sum_{i'=1}^{r_{\mathcal{N}}} \left\langle U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}, \mathcal{P}_{\mathcal{N}}^{\perp} \left\langle B, \mathcal{S}(Y, \mathcal{N}, i') \right\rangle \right\rangle w_{i,i'}^{\mathcal{N}} \end{aligned}$$

für $i \in \{1, \dots, r_{\mathcal{N}}\}, j \in \{1, \dots, r_{\mathcal{N}_1}\}, l \in \{1, \dots, r_{\mathcal{N}_2}\}$. Mit den Matrix-Formalismen gilt für die Lastmatrix $M^{\mathcal{N}}$ aus (4.20)

$$M^{\mathcal{N}} = (\mathbb{S}^{\mathcal{N}})^T \mathbb{S}^{\mathcal{N}},$$

so dass wir für die Inverse

$$W^{\mathcal{N}} = \left((\mathbb{S}^{\mathcal{N}})^T \mathbb{S}^{\mathcal{N}} \right)^{-1}$$

erhalten. Sei nun

$$s_i := \left\langle U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}, \mathcal{P}_{\mathcal{N}}^{\perp} \left\langle B, \mathcal{S}(Y, \mathcal{N}, i) \right\rangle \right\rangle \in \mathbb{R},$$

dann folgen aus

$$(s_1 \cdots s_{r_{\mathcal{N}}}) = \left((U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2})^{(\mathcal{N}, \{\})} \right)^T (P^{\mathcal{N}})^{\perp} B^{(\mathcal{N}, \mathcal{N}_C)} \mathbb{S}^{\mathcal{N}}$$

und

$$\mathbb{S}^{\mathcal{N}} W^{\mathcal{N}} = \mathbb{S}^{\mathcal{N}} \left((\mathbb{S}^{\mathcal{N}})^T \mathbb{S}^{\mathcal{N}} \right)^{-1} = \left(\mathbb{S}^{\mathcal{N}^+} \right)^T$$

die Gleichungen

$$\begin{aligned} (\delta \mathcal{B}_{1,j,l}^{\mathcal{N}} \dots \delta \mathcal{B}_{r_{\mathcal{N}},j,l}^{\mathcal{N}}) &= (s_1 \cdots s_{r_{\mathcal{N}}}) W^{\mathcal{N}} \\ &= \left((U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2})^{(\mathcal{N}, \{\})} \right)^T (P^{\mathcal{N}})^{\perp} B^{(\mathcal{N}, \mathcal{N}_C)} \left(\mathbb{S}^{\mathcal{N}^+} \right)^T. \end{aligned}$$

Für $\ell \in \mathcal{L}(T)$ und $i \in \{1, \dots, r_{\ell}\}$ definieren wir

$$v_i := \left(\mathcal{P}_{\ell}^{\perp} \langle B, \mathcal{S}(Y, \mathcal{N}, i) \rangle \right)^{(\mathcal{N}, \{\})} \in \mathbb{R}^{\tau(\mathcal{N})}.$$

Dann gilt

$$v_i = (P^{\ell})^{\perp} B^{(\mathcal{N}, \mathcal{N}_C)} (\mathcal{S}(Y, \mathcal{N}, i))^{(\mathcal{N}_C, \{\})}.$$

Zusammen mit $\mathbb{S}^{\ell} W^{\ell} = \left(\mathbb{S}^{\ell^+} \right)^T$ erhalten wir

$$\delta U^{\ell} = (v_1 | \cdots | v_{r_{\ell}}) = (P^{\ell})^{\perp} B^{(\ell, \ell_C)} \left(\mathbb{S}^{\ell^+} \right)^T.$$

■

Mit den Vorbereitungen aus Korollar 15 sind wir nun in der Lage, eine Formel zu beweisen, welche die Projektion in den Tangentialraum ohne eine Rekursion angibt.

Korollar 16. *Unter den Bedingungen von Korollar 15 erfüllt die Projektion eines Tensors $B \in \mathbb{R}^{N_1 \times \dots \times N_d}$ in den Tangentialraum $\mathcal{T}_Y \mathcal{M}$ mit $(\mathcal{N}_1^0, \mathcal{N}_2^0) = \text{succ}(\mathcal{N}_0)$ die Gleichung*

$$\begin{aligned} P(Y)B &= \left(\left((U^{\mathcal{N}_1^0} \odot U^{\mathcal{N}_2^0}) \left((U^{\mathcal{N}_1^0} \odot U^{\mathcal{N}_2^0})^T B^{(\mathcal{N}_0, \{\})} \right) \right)_{(\mathcal{N}_0, \{\})} \right. \\ &\quad + \sum_{\substack{\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\} \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} \left((U^{\mathcal{N}_1} \odot U^{\mathcal{N}_2}) \left((U^{\mathcal{N}_1} \odot U^{\mathcal{N}_2})^T (P^{\mathcal{N}})^{\perp} B^{(\mathcal{N}, \mathcal{N}_C)} \left(\mathbb{S}^{\mathcal{N}} \mathbb{S}^{\mathcal{N}^+} \right)^T \right) \right)_{(\mathcal{N}, \mathcal{N}_C)} \\ &\quad \left. + \sum_{\ell \in \mathcal{L}(T)} \left((P^{\ell})^{\perp} B^{(\ell, \ell_C)} \left(\mathbb{S}^{\ell} \mathbb{S}^{\ell^+} \right)^T \right)_{(\ell, \ell_C)} \right). \end{aligned}$$

Beweis. Nach Lemma 10 aus Kapitel 4 erhalten wir durch das Auflösen der Rekursion in Korollar 15 die Gleichung

$$P(Y)B = \sum_{j=1}^{r_{\mathcal{N}_1^0}} \sum_{l=1}^{r_{\mathcal{N}_2^0}} U_j^{\mathcal{N}_1^0} \cdot U_l^{\mathcal{N}_2^0} \delta \mathcal{B}_{1,j,l}^{\mathcal{N}_0} \quad (5.6)$$

$$+ \sum_{\substack{\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\} \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \cdot \sum_{i=1}^{r_{\mathcal{N}}} \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} \mathcal{S}(Y, \mathcal{N}, i) \quad (5.7)$$

$$+ \sum_{\ell \in \mathcal{L}(T)} \sum_{i=1}^{r_{\ell}} \delta U_i^{\ell} \cdot \mathcal{S}(Y, \ell, i). \quad (5.8)$$

Wir betrachten nun die Terme aus Zeile (5.6) genauer, setzen für $\delta \mathcal{B}^{\mathcal{N}_0}$ die Formel aus Korollar 15 ein und verwenden den binären Matrix-Operator \odot , um die einzelnen Terme der Summe zu einem Term zusammenzufassen.

$$\begin{aligned} & \sum_{j=1}^{r_{\mathcal{N}_1^0}} \sum_{l=1}^{r_{\mathcal{N}_2^0}} U_j^{\mathcal{N}_1^0} \cdot U_l^{\mathcal{N}_2^0} \delta \mathcal{B}_{1,j,l}^{\mathcal{N}_0} \\ &= \sum_{j=1}^{r_{\mathcal{N}_1^0}} \sum_{l=1}^{r_{\mathcal{N}_2^0}} U_j^{\mathcal{N}_1^0} \cdot U_l^{\mathcal{N}_2^0} \left(\left(U_j^{\mathcal{N}_1^0} \cdot U_l^{\mathcal{N}_2^0} \right)^{(\mathcal{N}_0, \{\})} \right)^T B^{(\mathcal{N}_0, \{\})} \\ &= \left(\sum_{j=1}^{r_{\mathcal{N}_1^0}} \sum_{l=1}^{r_{\mathcal{N}_2^0}} \left(U_j^{\mathcal{N}_1^0} \cdot U_l^{\mathcal{N}_2^0} \right)^{(\mathcal{N}_0, \{\})} \left(\left(U_j^{\mathcal{N}_1^0} \cdot U_l^{\mathcal{N}_2^0} \right)^{(\mathcal{N}_0, \{\})} \right)^T B^{(\mathcal{N}_0, \{\})} \right)_{(\mathcal{N}_0, \{\})} \\ &= \left(\left(\mathbb{U}^{\mathcal{N}_1^0} \odot \mathbb{U}^{\mathcal{N}_2^0} \right) \left(\mathbb{U}^{\mathcal{N}_1^0} \odot \mathbb{U}^{\mathcal{N}_2^0} \right)^T B^{(\mathcal{N}_0, \{\})} \right)_{(\mathcal{N}_0, \{\})} \end{aligned}$$

Die Terme aus Zeile (5.7) behandeln wir durch ein ähnliches Vorgehen und erhalten auf diese Weise eine Darstellung, in der mit Ausnahme der Wurzel lediglich über die inneren Knoten des Dimensionsbaumes aufsummiert wird.

$$\begin{aligned} & \sum_{\substack{\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\} \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \cdot \sum_{i=1}^{r_{\mathcal{N}}} \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} \mathcal{S}(Y, \mathcal{N}, i) \\ &= \sum_{\substack{\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\} \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} \left(\sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \left(U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \right)^{(\mathcal{N}, \{\})} \sum_{i=1}^{r_{\mathcal{N}}} \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} \left(\mathcal{S}(Y, \mathcal{N}, i)^{(\mathcal{N}_C, \{\})} \right)^T \right)_{(\mathcal{N}, \mathcal{N}_C)} \\ &= \sum_{\substack{\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\} \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} \left(\sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \left(\left(U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \right)^{(\mathcal{N}, \{\})} \right) \left(\delta \mathcal{B}_{1,j,l}^{\mathcal{N}} \dots \delta \mathcal{B}_{r_{\mathcal{N}},j,l}^{\mathcal{N}} \right) \left(\mathbb{S}^{\mathcal{N}} \right)^T \right)_{(\mathcal{N}, \mathcal{N}_C)} \end{aligned}$$

Setzen wir nun die Formel

$$(\delta \mathcal{B}_{1,j,l}^{\mathcal{N}} \dots \delta \mathcal{B}_{r_{\mathcal{N}},j,l}^{\mathcal{N}}) = \left((U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2})^{(\mathcal{N}, \{\})} \right)^T (P^{\mathcal{N}})^\perp B^{(\mathcal{N}, \mathcal{N}_C)} (\mathbb{S}^{\mathcal{N}^+})^T$$

aus Korollar 15 ein, so erhalten wir

$$\begin{aligned} & \sum_{\substack{\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\} \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \cdot \sum_{i=1}^{r_{\mathcal{N}}} \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} \mathcal{S}(Y, \mathcal{N}, i) \\ &= \sum_{\substack{\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\} \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} \left((U^{\mathcal{N}_1} \odot U^{\mathcal{N}_2}) (U^{\mathcal{N}_1} \odot U^{\mathcal{N}_2})^T (P^{\mathcal{N}})^\perp B^{(\mathcal{N}, \mathcal{N}_C)} (\mathbb{S}^{\mathcal{N}} \mathbb{S}^{\mathcal{N}^+})^T \right)_{(\mathcal{N}, \mathcal{N}_C)}. \end{aligned}$$

Zuletzt überführen wir Zeile (5.8) in Matrix-Schreibweise, setzen anschließend für δU^ℓ die Formel aus Korollar 15 ein und erhalten die gewünschte Aussage

$$\begin{aligned} \sum_{\ell \in \mathcal{L}(T)} \sum_{i=1}^{r_\ell} \delta U_i^\ell \cdot \mathcal{S}(Y, \ell, i) &= \sum_{\ell \in \mathcal{L}(T)} \left(\sum_{i=1}^{r_\ell} (\delta U_i^\ell)^{(\ell, \{\})} (\mathcal{S}(Y, \ell, i)^{(\ell_C, \{\})})^T \right)_{(\ell, \ell_C)} \\ &= \sum_{\ell \in \mathcal{L}(T)} \left(\delta U^\ell (\mathbb{S}^\ell)^T \right)_{(\ell, \ell_C)} \\ &= \sum_{\ell \in \mathcal{L}(T)} \left((P^\ell)^\perp B^{(\ell, \ell_C)} (\mathbb{S}^{\ell^+})^T (\mathbb{S}^\ell)^T \right)_{(\ell, \ell_C)} \\ &= \sum_{\ell \in \mathcal{L}(T)} \left((P^\ell)^\perp B^{(\ell, \ell_C)} (\mathbb{S}^\ell \mathbb{S}^{\ell^+})^T \right)_{(\ell, \ell_C)}. \end{aligned}$$

■

Als Abschluss dieses Kapitels beweisen wir den Beweisschritt (b) \Rightarrow (a) von Lemma 9 aus Abschnitt 3.3. Durch die Ergebnisse dieses Kapitels verkürzt sich dieser Beweisschritt erheblich, deshalb hatten wir darauf hingewiesen, den Rest des Beweises an dieser Stelle nachzutragen.

Beweis der Rückrichtung von Lemma 9. Angenommen (b) ist erfüllt und es gäbe einen Knoten $\mathcal{N}^* \in T \setminus \{\mathcal{N}_0\}$, so dass die Funktionen

$$\langle Y, U_1^{\mathcal{N}^*} \rangle, \dots, \langle Y, U_{r_{\mathcal{N}^*}}^{\mathcal{N}^*} \rangle$$

linear abhängig sind. Dann folgt aus der linearen Abhängigkeit jedoch $\text{rang}(\mathbb{S}^{\mathcal{N}^*}) < r_{\mathcal{N}^*}$ und daraus $\text{rang}(Y^{(\mathcal{N}^*, \mathcal{N}_C^*)}) < r_{\mathcal{N}^*}$. Nach Lemma 14 wäre $Y \notin \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$. Dies ist jedoch ein Widerspruch zu der Voraussetzung $Y \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$. Somit folgt (b) \Rightarrow (a). ■

6. Differentialgeometrie von Tensoren im hierarchischen Tucker-Format

6.1 Interpretation als Mannigfaltigkeit

Dieses Kapitel beruht auf der Arbeit [UV13] von Uschmajew und Vandereycken und hat das Ziel, zu zeigen, dass es sich bei der Menge \mathcal{H} -Tucker $((r_{\mathcal{N}})_{\mathcal{N} \in T})$ um eine Mannigfaltigkeit handelt. Darüber hinaus bestimmen wir den Tangentialraum der Mannigfaltigkeit $\mathcal{M} = \mathcal{H}$ -Tucker $((r_{\mathcal{N}})_{\mathcal{N} \in T})$.

In Abschnitt 3.3 haben wir einen Tensor im hierarchischen Tucker-Format dadurch definiert, dass ein Tupel $\left((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^{\ell})_{\ell \in \mathcal{L}(T)} \right)$ existiert, durch das der Tensor gemäß Definition 4 dargestellt werden kann und welches den Annahmen 1 und 2 genügt. Diese Annahmen sind an die Basisfunktionen der Darstellung gerichtet und somit nur indirekt an die Transfertensoren des Tupels. Das erste Theorem dieses Abschnitts gibt an, welche Bedingungen an die Transfertensoren und an die Basisfunktionen an den Blättern des Dimensionsbaumes zu stellen sind, so dass die Rekonstruktion einen Tensor im hierarchischen Tucker-Format erzeugt. Als Vorbereitung für dieses Theorem definieren wir für einen Transfertensor $\mathcal{B}^{\mathcal{N}} \in \mathbb{R}^{r_{\mathcal{N}} \times r_{\mathcal{N}_1} \times r_{\mathcal{N}_2}}$ die Matrizen $\mathbb{B}_0^{\mathcal{N}} \in \mathbb{R}^{r_{\mathcal{N}} \times (r_{\mathcal{N}_1} r_{\mathcal{N}_2})}$, $\mathbb{B}_1^{\mathcal{N}} \in \mathbb{R}^{r_{\mathcal{N}_1} \times (r_{\mathcal{N}} r_{\mathcal{N}_2})}$ und $\mathbb{B}_2^{\mathcal{N}} \in \mathbb{R}^{r_{\mathcal{N}_2} \times (r_{\mathcal{N}} r_{\mathcal{N}_1})}$ anhand

$$\left(\mathbb{B}_0^{\mathcal{N}} \right)_{i, (l-1)r_{\mathcal{N}_1} + j} = \left(\mathbb{B}_1^{\mathcal{N}} \right)_{j, (l-1)r_{\mathcal{N}} + i} = \left(\mathbb{B}_2^{\mathcal{N}} \right)_{l, (j-1)r_{\mathcal{N}} + i} = \mathcal{B}_{i,j,l}^{\mathcal{N}}.$$

Diese Matrizen bestehen allesamt aus den Einträgen des Transfertensors; sie werden nur in unterschiedlicher Weise zu Matrizen angeordnet.

Theorem 17. *Ein Tensor $Y \in \mathbb{R}^{N_1 \times \dots \times N_d}$ liegt genau dann in der Menge \mathcal{H} -Tucker $((r_{\mathcal{N}})_{\mathcal{N} \in T})$, wenn ein Tupel $\left((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^{\ell})_{\ell \in \mathcal{L}(T)} \right)$ existiert, durch das Y*

gemäß Definition 4 dargestellt werden kann und welches die folgenden zwei Bedingungen erfüllt:

(B1) Die Transfertensoren $\mathcal{B}^{\mathcal{N}}$ haben für alle $\mathcal{N} \in \mathcal{I}(T)$ vollen multilinearen Rang, das heißt

$$\text{rang}(\mathbb{B}_0^{\mathcal{N}}) = r_{\mathcal{N}} \quad , \quad \text{rang}(\mathbb{B}_1^{\mathcal{N}}) = r_{\mathcal{N}_1} \quad , \quad \text{rang}(\mathbb{B}_2^{\mathcal{N}}) = r_{\mathcal{N}_2}.$$

(B2) Für alle $\ell \in \mathcal{L}(T)$ gilt, dass die Sequenz $U_1^\ell, \dots, U_{r_\ell}^\ell$ linear unabhängig ist.

Beweis. Siehe [UV13, Theorem 3.7]. ■

Die Aussage des Theorems wirkt auf den ersten Anschein widersprüchlich, da sich die Bedingungen des Theorems von den Annahmen aus Abschnitt 3.3 unterscheiden. Tatsächlich kann man sogar zeigen, dass ein Tupel, welches die Annahmen aus Abschnitt 3.3 erfüllt, ebenfalls die Bedingungen aus Theorem 17 erfüllt, dass aber ein Tupel, welches den Bedingungen aus Theorem 17 genügt, nicht notwendigerweise die Annahmen aus Abschnitt 3.3 erfüllen muss. Dieser Sachverhalt begründet jedoch keinen Widerspruch, da für einen Tensor im hierarchischen Tucker-Format nicht nur ein Tupel existiert, durch welches dieser Tensor entsprechend der Definition 4 dargestellt werden kann und den Annahmen 1 und 2 aus Abschnitt 3.3 oder den Bedingungen (B1) und (B2) aus Theorem 17 genügt, sondern gar eine ganze Familie. Liegt ein Tupel vor, welches den Bedingungen des Theorems genügt, so kann man daraus den vollständigen Tensor rekonstruieren, Algorithmus 1 aus Kapitel 5 anwenden und bekommt auf diese Weise ein Tupel, welches sowohl den Annahmen 1 und 2 aus Abschnitt 3.3 als auch den Bedingungen (B1) und (B2) aus Theorem 17 genügt. Der Grund, weshalb wir in den zurückliegenden Kapiteln nur Tupel betrachtet hatten, welche die Annahmen 1 und 2 erfüllen, ist, dass die Orthogonalität der Basisfunktionen bei der Herleitung der Bewegungsgleichungen in Kapitel 4 eine essentielle Rolle spielt.

Wir möchten nun eine formale Beschreibung für die Menge aller Tupel $\left((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^\ell)_{\ell \in \mathcal{L}(T)} \right)$ finden, welche gemäß Definition 4 beziehungsweise anhand der Rekonstruktionsformel aus Lemma 5 einen Tensor im hierarchischen Tucker-Format erzeugen. Zu diesem Zweck definieren wir für einen Transfertensor $\mathcal{B}^{\mathcal{N}}$ die Menge

$$\mathbb{R}_*^{r_{\mathcal{N}} \times r_{\mathcal{N}_1} \times r_{\mathcal{N}_2}} := \left\{ \mathcal{B} \in \mathbb{R}^{r_{\mathcal{N}} \times r_{\mathcal{N}_1} \times r_{\mathcal{N}_2}} : \mathcal{B} \text{ hat vollen multilinearen Rang} \right\}$$

mit $\mathcal{N} \in \mathcal{I}(T)$ und $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$. Diese Menge ist eine dichte und offene Teilmenge des $\mathbb{R}^{r_{\mathcal{N}} \times r_{\mathcal{N}_1} \times r_{\mathcal{N}_2}}$ und wird durch die Nebenbedingung zu einer $r_{\mathcal{N}} r_{\mathcal{N}_1} r_{\mathcal{N}_2}$ -dimensionalen Untermannigfaltigkeit, siehe [Usc13, Satz 5.2]. Für eine Basismatrix $\mathbb{U}^{\{\omega\}}$ mit $\{\omega\} \in \mathcal{L}(T)$ definieren wir die Menge

$$\mathbb{R}_*^{N_\omega \times r_{\{\omega\}}} := \left\{ \mathbb{U} \in \mathbb{R}^{N_\omega \times r_{\{\omega\}}} : \text{rang}(\mathbb{U}) = r_{\{\omega\}} \right\},$$

welche eine glatte $N_\omega r_{\{\omega\}}$ -dimensionale Untermannigfaltigkeit des $\mathbb{R}^{N_\omega \times r_{\{\omega\}}}$ ist, siehe [Lee12, Example 1.28]. Als Kreuzprodukt über diese Mannigfaltigkeiten definieren wir die Menge

$$\mathcal{M}_* := \left(\prod_{\mathcal{N} \in \mathcal{I}(T), (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})} \mathbb{R}_*^{r_{\mathcal{N}} \times r_{\mathcal{N}_1} \times r_{\mathcal{N}_2}} \right) \times \left(\prod_{\{\omega\} \in \mathcal{L}(T)} \mathbb{R}_*^{N_\omega \times r_{\{\omega\}}} \right).$$

Aufgrund der unterliegenden Produkttopologie ist auch die Menge \mathcal{M}_* eine glatte Mannigfaltigkeit der Dimension

$$\sum_{\substack{\mathcal{N} \in \mathcal{I}(T) \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} r_{\mathcal{N}} r_{\mathcal{N}_1} r_{\mathcal{N}_2} + \sum_{\omega=1}^d N_\omega r_{\{\omega\}},$$

siehe [BG68, Abschnitt 1.2.(a)]. Definieren wir mit Hilfe von Abbildung Ψ aus Lemma 5 die Abbildung

$$\begin{aligned} \Psi^* : \mathcal{M}_* &\rightarrow \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T}) \\ \Psi^* \left(\left((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^\ell)_{\ell \in \mathcal{L}(T)} \right) \right) &= \Psi \left(\left((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^\ell)_{\ell \in \mathcal{L}(T)} \right) \right), \end{aligned}$$

so folgt mit Theorem 17 einerseits die Wohldefiniertheit und andererseits die Surjektivität der Abbildung. Die Injektivität der Abbildung liegt hingegen nicht vor; für einen fest gewählten Tensor im hierarchischen Tucker-Format existiert kein **eindeutiges** Element aus der Menge \mathcal{M}_* , welches den Tensor gemäß Definition 4 und unter Berücksichtigung der Bedingungen (B1) und (B2) aus Theorem 17 rekonstruiert, sondern eine ganze Familie. Für ein $x \in \mathcal{M}_*$ lassen sich alle Elemente, welche denselben Tensor rekonstruieren, durch die Äquivalenzklasse

$$[x] = \left\{ y \in \mathcal{M}_* : \Psi^*(y) = \Psi^*(x) \right\}$$

beschreiben. Worin sich die Elemente einer Äquivalenzklasse unterscheiden dürfen, so dass sie einen fest gewählten Tensor im hierarchischen Tucker-Format erzeugen, gibt die nächste Proposition an.

Proposition 18. *Seien*

$$x = \left((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^\ell)_{\ell \in \mathcal{L}(T)} \right) \in \mathcal{M}_* \quad \text{und} \quad y = \left((\tilde{\mathcal{B}}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (\tilde{U}^\ell)_{\ell \in \mathcal{L}(T)} \right) \in \mathcal{M}_*.$$

Dann gilt

$$\Psi^*(x) = \Psi^*(y)$$

genau dann, wenn für alle $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ invertierbare Matrizen $\mathbb{A}^{\mathcal{N}} \in \mathbb{R}^{r_{\mathcal{N}} \times r_{\mathcal{N}}}$ mit Einträgen

$$\hat{a}_{i,j}^{\mathcal{N}} = (\mathbb{A}^{\mathcal{N}})_{i,j} \quad \text{und} \quad \check{a}_{i,j}^{\mathcal{N}} = \left((\mathbb{A}^{\mathcal{N}})^{-1} \right)_{i,j}$$

existieren, so dass

$$\begin{aligned} \tilde{U}_i^\ell &= \sum_{i'=1}^{r_\ell} \hat{a}_{i',i}^\ell U_{i'}^\ell && \text{für alle } \ell \in \mathcal{L}(T) \\ \text{und } \tilde{\mathcal{B}}_{i,j,l}^\mathcal{N} &= \sum_{i'=1}^{r_\mathcal{N}} \sum_{j'=1}^{r_{\mathcal{N}_1}} \sum_{l'=1}^{r_{\mathcal{N}_2}} \hat{a}_{i',i}^\mathcal{N} \check{a}_{j,j'}^{\mathcal{N}_1} \check{a}_{l,l'}^{\mathcal{N}_2} \mathcal{B}_{i',j',l'}^\mathcal{N} && \text{für alle } \mathcal{N} \in \mathcal{I}(T). \end{aligned}$$

Beweis. Siehe [UV13, Proposition 3.9]. ■

Wie die folgenden Argumente zeigen, resultiert die Äquivalenzklasse $[x]$ als Orbit von x unter der Aktion einer Lie-Gruppe. Dazu definieren wir die Gruppe $(\mathcal{G}, *)$ anhand der Menge

$$\begin{aligned} \mathcal{G} &= \left\{ (\mathbb{A}^\mathcal{N})_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}} : \mathbb{A}^\mathcal{N} \in \mathbb{R}^{r_\mathcal{N} \times r_\mathcal{N}} \text{ ist invertierbar} \right\} \\ &= \prod_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}} \text{GL}(r_\mathcal{N}, \mathbb{R}), \end{aligned}$$

sowie der zweistelligen Verknüpfung

$$\begin{aligned} * : \mathcal{G} \times \mathcal{G} &\rightarrow \mathcal{G} \\ (\mathbb{A}^\mathcal{N})_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}} * (\mathbb{B}^\mathcal{N})_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}} &= (\mathbb{C}^\mathcal{N})_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}} \quad \text{mit } \mathbb{C}^\mathcal{N} = \mathbb{A}^\mathcal{N} \mathbb{B}^\mathcal{N}. \end{aligned}$$

Aufgrund der Konstruktion übertragen sich die Eigenschaften der Gruppe $\text{GL}(r_\mathcal{N}, \mathbb{R})$ für $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ komponentenweise auf $(\mathcal{G}, *)$. So ist es nicht verwunderlich, dass auch $(\mathcal{G}, *)$ eine Lie-Gruppe ist; die Verknüpfung $*$ erfüllt die Gruppenaxiome und als Kreuzprodukt über differenzierbare Mannigfaltigkeiten ist auch die Menge \mathcal{G} eine differenzierbare Mannigfaltigkeit der Dimension

$$\sum_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}} r_\mathcal{N}^2.$$

Des Weiteren definieren wir auf \mathcal{M}_* die Rechtsaktion

$$\begin{aligned} \theta : \mathcal{M}_* \times \mathcal{G} &\rightarrow \mathcal{M}_* \\ \theta \left(\left((\tilde{\mathcal{B}}^\mathcal{N})_{\mathcal{N} \in \mathcal{I}(T)}, (\tilde{U}^\ell)_{\ell \in \mathcal{L}(T)} \right), (\mathbb{A}^\mathcal{N})_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}} \right) &= \left((\mathcal{B}^\mathcal{N})_{\mathcal{N} \in \mathcal{I}(T)}, (U^\ell)_{\ell \in \mathcal{L}(T)} \right) \end{aligned}$$

mit $\left((\mathcal{B}^\mathcal{N})_{\mathcal{N} \in \mathcal{I}(T)}, (U^\ell)_{\ell \in \mathcal{L}(T)} \right)$ wie aus Proposition 18. Dann gilt für den Orbit $\mathcal{G}(x)$ von $x \in \mathcal{M}_*$ unter \mathcal{G} einerseits

$$\mathcal{G}(x) = \left\{ \theta \left(x, (\mathbb{A}^\mathcal{N})_{\mathcal{N} \in T} \right) : (\mathbb{A}^\mathcal{N})_{\mathcal{N} \in T} \in \mathcal{G} \right\}$$

und als Konsequenz von Proposition 18 andererseits

$$\mathcal{G}(x) = [x].$$

Die Elemente aus dem Orbit $\mathcal{G}(x)$ sind also genau die Elemente aus \mathcal{M}_* , welche gemäß Definition 4 und unter Berücksichtigung der Bedingungen (B1) und (B2) aus Theorem 17 den Tensor $\Psi^*(x)$ rekonstruieren. In dem Quotientenraum

$$\mathcal{M}_*/\mathcal{G} = \{[x] : x \in \mathcal{M}_*\}$$

ist jedes Element eine Menge; alle Elemente aus \mathcal{M}_* , welche einen fest gewählten Tensor rekonstruieren, definieren ein Element des Quotientenraumes. Wie das nachfolgende Theorem zeigt, ist der Quotientenraum mit der Struktur einer glatten Mannigfaltigkeit versehen.

Theorem 19. *Der Quotientenraum $\mathcal{M}_*/\mathcal{G}$ ist eine topologische Mannigfaltigkeit, welche eine eindeutige glatte Struktur aufweist und die Dimension*

$$\dim(\mathcal{M}_*/\mathcal{G}) = \sum_{\{\omega\} \in \mathcal{L}(T)} r_{\{\omega\}} N_\omega + \sum_{\substack{\mathcal{N} \in \mathcal{I}(T) \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} r_{\mathcal{N}} (r_{\mathcal{N}_1} r_{\mathcal{N}_2} - r_{\mathcal{N}})$$

besitzt. Darüber hinaus ist die Quotientenabbildung

$$\begin{aligned} \pi : \mathcal{M}_* &\rightarrow \mathcal{M}_*/\mathcal{G} \\ x &\mapsto [x]. \end{aligned}$$

eine glatte Submersion.

Unter einer Submersion verstehen wir eine differenzierbare Abbildung zwischen zwei Mannigfaltigkeiten, deren totales Differential an jeder Stelle surjektiv ist, siehe [Lee12, Kapitel 4].

Beweis. Der Beweis ist in [UV13, Theorem 4.3] zu finden und nutzt ein substanzielles Resultat der Differentialgeometrie. Wir beschränken uns in dieser Arbeit darauf, die Beweisschritte zu skizzieren.

In [UV13, Lemma 4.2] zeigen Uschmajew und Vandereycken, dass die Rechtsaktion θ eigentlich ist. Das bedeutet, dass das Urbild $\theta^{-1}(K)$ jeder kompakten Menge $K \subseteq \mathcal{M}_*$ erneut kompakt ist. Darüber hinaus ist die Rechtsaktion θ frei; das einzige Element der Gruppe, welches ein beliebiges Element aus \mathcal{M}_* unverändert lässt, ist das neutrale Element $(I_{r_{\mathcal{N}}})_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}}$. Da \mathcal{G} eine Lie-Gruppe ist, welche frei, glatt und eigentlich auf \mathcal{M}_* wirkt, sind wir in der Lage, ein bekanntes Theorem aus der Differentialgeometrie anzuwenden, siehe Theorem 21.10 (Quotient Manifold Theorem) aus [Lee12]. Es folgt einerseits, dass die Quotientenabbildung π eine Submersion ist und andererseits, dass der Quotientenraum

$\mathcal{M}_*/\mathcal{G}$ eine topologische Mannigfaltigkeit ist, welche mit einer eindeutigen glatten Struktur versehen ist und die Dimension

$$\dim(\mathcal{M}_*/\mathcal{G}) = \dim(\mathcal{M}_*) - \dim(\mathcal{G})$$

besitzt. Setzen wir für $\dim(\mathcal{M}_*)$ und $\dim(\mathcal{G})$ die Dimensionen ein, so folgt die Behauptung. ■

Bilden wir die Komposition von Ψ^* nach der Inversen der Quotientenabbildung, so erhalten wir mit

$$\begin{aligned} \Psi^{**} : \mathcal{M}_*/\mathcal{G} &\rightarrow \mathcal{M}_* \rightarrow \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T}) \subset \mathbb{R}^{N_1 \times \dots \times N_d} \\ \Psi^{**}([x]) &= \Psi^* \circ \pi^{-1} \end{aligned}$$

eine wohldefinierte und zugleich bijektive Abbildung. In [UV13, Theorem 4.11] zeigen Uschmajew und Vandereycken, dass die Menge $\mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ eine zu $\mathcal{M}_*/\mathcal{G}$ diffeomorphe Untermannigfaltigkeit des $\mathbb{R}^{N_1 \times \dots \times N_d}$ ist.

Theorem 20. *Die Menge der Tensoren im hierarchischen Tucker-Format $\mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ ist eine differenzierbare Untermannigfaltigkeit des $\mathbb{R}^{N_1 \times \dots \times N_d}$. Sie besitzt die Dimension*

$$\sum_{\{\omega\} \in \mathcal{L}(T)} r_{\{\omega\}} N_{\omega} + \sum_{\substack{\mathcal{N} \in \mathcal{I}(T) \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} r_{\mathcal{N}} (r_{\mathcal{N}_1} r_{\mathcal{N}_2} - r_{\mathcal{N}}).$$

Beweis. Der Beweis ist in [UV13, Theorem 4.11] geführt, deshalb skizzieren wir lediglich die Beweisschritte.

In dem Beweis von Theorem 4.11 aus [UV13] zeigen Uschmajew und Vandereycken die Homöomorphie von Ψ^{**} . Darüber hinaus wird in [UV13, Lemma 4.10] nachgewiesen, dass Ψ^{**} eine Immersion ist. Ohne genauer zu erklären, was es mit einer Immersion¹ auf sich hat, folgt unter diesen Bedingungen, dass Ψ^{**} eine glatte Einbettung ist. De facto ist das Bild von Ψ^{**} eine Untermannigfaltigkeit des $\mathbb{R}^{N_1 \times \dots \times N_d}$ [Lee12, Proposition 5.2], welche aufgrund der Bijektivität von Ψ^{**} dieselbe Dimension besitzt wie der Quotientenraum. ■

Mit Theorem 20 haben wir eines der zwei Ziele dieses Kapitels erfüllt: Beruhend auf der Arbeit von Uschmajew und Vandereycken haben wir nachgewiesen, dass es sich bei der Menge $\mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ um eine Untermannigfaltigkeit des $\mathbb{R}^{N_1 \times \dots \times N_d}$ handelt. Wir tragen nun den ausstehenden Beweis von Lemma 10 nach und bestimmen die Elemente des Tangentialraumes.

¹Die Definition einer Immersion ist beispielsweise in [Lee12, Kapitel 4] zu finden

Beweis von Lemma 10. Als Voraussetzung für das Lemma ist mit $Y \in \mathcal{H}$ -Tucker $((r_{\mathcal{N}})_{\mathcal{N} \in T})$ ein Tensor im hierarchischen Tucker-Format gegeben, welcher durch das Tupel $\left((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^{\ell})_{\ell \in \mathcal{L}(T)} \right)$ dargestellt werde. Mit den Resultaten dieses Kapitels gilt einerseits

$$x = \left((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^{\ell})_{\ell \in \mathcal{L}(T)} \right) \in \mathcal{M}_*$$

und andererseits

$$\Psi^*(x) = Y.$$

Sowohl die Menge \mathcal{M}_* als auch die Menge \mathcal{H} -Tucker $((r_{\mathcal{N}})_{\mathcal{N} \in T})$ sind mit der Struktur einer glatten Mannigfaltigkeit versehen. Darüber hinaus sind beide Mengen offene Teilmengen euklidischer Räume. Deshalb ist die Abbildung

$$\begin{aligned} D\Psi_x^* : \mathcal{T}_x \mathcal{M}_* &\rightarrow \mathcal{T}_{\Psi^*(x)} \mathcal{H}\text{-Tucker } ((r_{\mathcal{N}})_{\mathcal{N} \in T}) \\ D\Psi_x^*(\delta x) &= \sum_{j=1}^{r_{\mathcal{N}_1^0}} \sum_{l=1}^{r_{\mathcal{N}_2^0}} \delta \mathcal{B}_{1,j,l}^{\mathcal{N}_0} \frac{\partial \Psi^*}{\partial \mathcal{B}_{1,j,l}^{\mathcal{N}_0}} \Big|_{((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^{\ell})_{\ell \in \mathcal{L}(T)})=x} \\ &+ \sum_{\substack{\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\} \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} \sum_{i=1}^{r_{\mathcal{N}}} \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} \frac{\partial \Psi^*}{\partial \mathcal{B}_{i,j,l}^{\mathcal{N}}} \Big|_{((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^{\ell})_{\ell \in \mathcal{L}(T)})=x} \\ &+ \sum_{\ell \in \mathcal{L}(T)} \sum_{i=1}^{r_{\ell}} \delta U_i^{\ell} \frac{\partial \Psi^*}{\partial U_i^{\ell}} \Big|_{((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^{\ell})_{\ell \in \mathcal{L}(T)})=x} \end{aligned}$$

mit $\delta x = \left((\delta \mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (\delta U^{\ell})_{\ell \in \mathcal{L}(T)} \right)$ und $(\mathcal{N}_1^0, \mathcal{N}_2^0) = \text{succ}(\mathcal{N}_0)$ das totale Differential von Abbildung Ψ^* .

Zusammen mit den Gleichungen

$$\begin{aligned} \frac{\partial \Psi^*}{\partial \mathcal{B}_{1,j,l}^{\mathcal{N}_0}} \Big|_{((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^{\ell})_{\ell \in \mathcal{L}(T)})=x} &= U_j^{\mathcal{N}_1^0} \cdot U_l^{\mathcal{N}_2^0}, \\ \frac{\partial \Psi^*}{\partial \mathcal{B}_{i,j,l}^{\mathcal{N}}} \Big|_{((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^{\ell})_{\ell \in \mathcal{L}(T)})=x} &= \langle Y, U_i^{\mathcal{N}} \rangle \cdot U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \\ &= \mathcal{S}(Y, \mathcal{N}, i) \cdot U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \quad \text{für } (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N}) \\ \text{und } \frac{\partial \Psi^*}{\partial U_i^{\ell}} \Big|_{((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^{\ell})_{\ell \in \mathcal{L}(T)})=x} &= \langle Y, U_i^{\ell} \rangle = \mathcal{S}(Y, \ell, i), \end{aligned}$$

erhalten wir für das totale Differential die Darstellung

$$\begin{aligned}
D\Psi_x^*(\delta x) &= \sum_{j=1}^{r_{\mathcal{N}_1^0}} \sum_{l=1}^{r_{\mathcal{N}_2^0}} \delta \mathcal{B}_{1,j,l}^{\mathcal{N}_0} U_j^{\mathcal{N}_1^0} \cdot U_l^{\mathcal{N}_2^0} \\
&+ \sum_{\substack{\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\} \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} \sum_{i=1}^{r_{\mathcal{N}}} \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} \mathcal{S}(Y, \mathcal{N}, i) \cdot U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \\
&+ \sum_{\ell \in \mathcal{L}(T)} \sum_{i=1}^{r_\ell} \delta U_i^\ell \cdot \mathcal{S}(Y, \ell, i).
\end{aligned}$$

Somit folgt für den Tangentialraum

$$\begin{aligned}
\mathcal{T}_Y \mathcal{M} &= \mathcal{T}_{\Psi^*(x)} \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T}) \\
&= \left\{ D\Psi_x^*(\delta x) : \delta x \in \mathcal{T}_x \mathcal{M}_* \right\}.
\end{aligned}$$

Da \mathcal{M}_* eine dichte und offene Teilmenge von

$$M := \prod_{\substack{\mathcal{N} \in \mathcal{I}(T) \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} \mathbb{R}^{r_{\mathcal{N}} \times r_{\mathcal{N}_1} \times r_{\mathcal{N}_2}} \times \prod_{\{\omega\} \in \mathcal{L}(T)} \mathbb{R}^{N_\omega \times r_{\{\omega\}}}$$

ist, folgt $M \cong \mathcal{T}_x \mathcal{M}_*$ und damit die Behauptung

$$\mathcal{T}_Y \mathcal{M} = \left\{ D\Psi_x^*(\delta x) : \delta x \in M \right\},$$

siehe auch [UV13, Abschnitt 4.3]. ■

6.2 Krümmung der Mannigfaltigkeit

In diesem Abschnitt leiten wir eine Abschätzung der Krümmung der Mannigfaltigkeit $\mathcal{M} = \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ her. Unser Vorgehen orientiert sich dabei an den Techniken von Koch und Lubich: In [KL07, Lemma 4.2] bestimmen Koch und Lubich eine Abschätzung der Krümmung der Mannigfaltigkeit

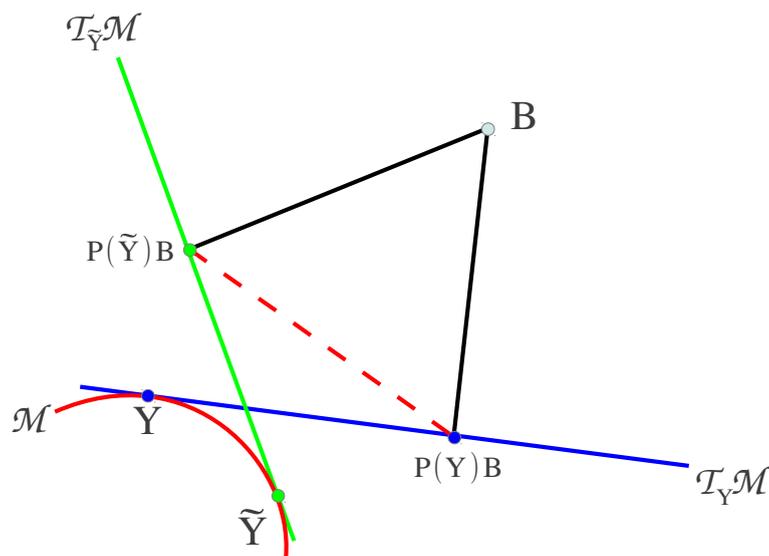
$$\mathcal{M}_r = \left\{ Y \in \mathbb{R}^{m \times n} : \text{rang}(Y) = r \right\}.$$

Darüber hinaus leiten sie in [KL10, Lemma 3.2] eine Abschätzung der Krümmung der Mannigfaltigkeit Tucker (r_1, \dots, r_d) her.

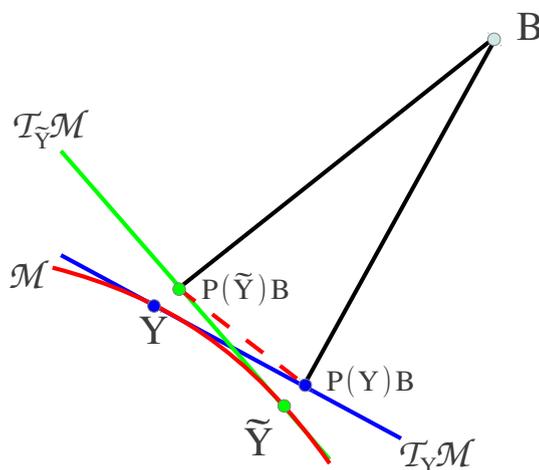
Zu Beginn dieses Kapitels wollen wir zunächst klären, was wir unter der „Krümmung einer Mannigfaltigkeit“ verstehen. Anschließend stellen wir in Theorem 22 das Hauptresultat dieses Kapitels vor, eine Abschätzung der Krümmung der Mannigfaltigkeit $\mathcal{M} = \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$. Da das Resultat in seiner Aussage und mit den Voraussetzungen sehr technisch ist, stellen wir in Korollar 23 eine schwächere aber eingängigere Abschätzung vor.

Die Abschätzung der Krümmung der Mannigfaltigkeit ist ein wesentlicher Bestandteil für die *a posteriori* Fehleranalyse bei der Approximation von zeitabhängigen Problemen durch das hierarchische Tucker-Format.

Um ein Maß für die Krümmung einer Mannigfaltigkeit festzulegen, bedienen wir uns der Orthogonalprojektion in den Tangentialraum der Mannigfaltigkeit. Seien Y und \tilde{Y} Punkte auf der Mannigfaltigkeit und B ein beliebiger Punkt. So können wir aus dem Abstand zwischen der Orthogonalprojektion von B in den Tangentialraum von Y und der Orthogonalprojektion von B in den Tangentialraum von \tilde{Y} ableiten, inwieweit die Menge lokal von einem linearen Raum abweicht. Diese Abweichung interpretieren wir als die Krümmung der Mannigfaltigkeit.



(a) Mannigfaltigkeit mit „großer“ Krümmung



(b) Mannigfaltigkeit mit „geringer“ Krümmung

Abbildung 6.1: Mannigfaltigkeiten unterschiedlicher Krümmung

In Abbildung 6.1 sind zwei Beispiele zu erkennen, worin jeweils die Mannigfaltigkeit in rot und die Verbindungsstrecke zwischen den Projektionen als gestrichelte Linie aufgetragen ist. Teilbild 6.1(a) zeigt eine Mannigfaltigkeit mit großer Krümmung, wohingegen die Mannigfaltigkeit in Teilbild 6.1(b) geringer gekrümmt ist. Die Punkte Y, \tilde{Y} und B haben in beiden Bildern dieselbe Position. Wie aus Teilbild 6.1(a) zu entnehmen ist, folgt aus einer großen Krümmung ein großer Abstand zwischen den Orthogonalprojektionen $P(Y)B$ und $P(\tilde{Y})B$; ist die Mannigfaltigkeit hingegen gering gekrümmt, wie in Teilbild 6.1(b), so liegen die Projektionen von B nahe beieinander.

In unserem Fall ist die Mannigfaltigkeit die Menge der Tensoren im hierarchischen Tucker-Format, $Y, \tilde{Y} \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ und $B \in \mathbb{R}^{N_1 \times \dots \times N_d}$. Da der Abstand zwischen den Orthogonalprojektionen nicht nur von der Mannigfaltigkeit abhängt, sondern auch davon, wie die Tensoren Y, \tilde{Y} und B gewählt sind, ist es das Ziel dieses Kapitels, für gegebene Y und \tilde{Y} eine Konstante $C_{Y, \tilde{Y}}$ zu finden, welche die Abschätzung

$$\|(P(Y) - P(\tilde{Y}))B\|_F \leq C_{Y, \tilde{Y}} \|Y - \tilde{Y}\|_F \|B\|_F$$

erfüllt.

Das nächste Lemma dient als Vorbereitung und zielt darauf ab, aus Informationen über die Singulärwerte der Matrifizierung eines Tensors $Y \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$, sowie dem Abstand zu einem anderen Tensor $\tilde{Y} \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$, Informationen über die Singulärwerte der Matrifizierung von \tilde{Y} zu erhalten. Ein ähnliches Resultat wird von Koch und Lubich im Beweis von [KL10, Lemma 3.2] für Tensoren im Tucker-Format gezeigt. In Lemma 21 übertragen wir die Vorgehensweise von Koch und Lubich auf Tensoren im hierarchischen Tucker-Format.

Lemma 21. *Seien $Y, \tilde{Y} \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ Tensoren im hierarchischen Tucker-Format und $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ mit*

$$\sigma_{r_{\mathcal{N}}}(Y^{(\mathcal{N}, \mathcal{N}_C)}) \geq \rho \quad \text{und} \quad \|Y - \tilde{Y}\|_F \leq \delta,$$

wobei $\sigma_{r_{\mathcal{N}}}(Y^{(\mathcal{N}, \mathcal{N}_C)})$ den $r_{\mathcal{N}}$ -ten Singulärwert der Matrix $Y^{(\mathcal{N}, \mathcal{N}_C)}$ bezeichnet (siehe Seite 47). Dann gilt

$$\sigma_{r_{\mathcal{N}}}(\tilde{Y}^{(\mathcal{N}, \mathcal{N}_C)}) \geq \rho - \delta.$$

Gilt ferner $\rho > \delta$, so erfüllt die Pseudoinverse $\tilde{\mathbb{S}}^{\mathcal{N}^+}$ der Single-Hole-Matrix $\tilde{\mathbb{S}}^{\mathcal{N}}$ von \tilde{Y} die Abschätzung

$$\|\tilde{\mathbb{S}}^{\mathcal{N}^+}\|_2 \leq \frac{1}{\rho - \delta}.$$

Wir weisen darauf hin, dass Lubich, Rohwedder, Schneider und Vandereycken in dem Beweis von [LRSV13, Lemma 4.5] ähnliche Abschätzungen beweisen.

Beweis von Lemma 21. Aus [HJ90, Seite 448] wissen wir, dass zwei Matrizen $A, B \in \mathbb{R}^{m \times n}$ stets die Abschätzung

$$\|A - B\|_F \geq \sqrt{\sum_{i=1}^{\min(m,n)} (\sigma_i(A) - \sigma_i(B))^2}$$

erfüllen. Wenden wir diese Abschätzung nun auf die Matrifizierungen $Y^{(\mathcal{N}, \mathcal{N}_C)}$ und $\tilde{Y}^{(\mathcal{N}, \mathcal{N}_C)}$ an einem Knoten $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ an, so erhalten wir

$$\begin{aligned} \delta = \|Y - \tilde{Y}\|_F &= \|Y^{(\mathcal{N}, \mathcal{N}_C)} - \tilde{Y}^{(\mathcal{N}, \mathcal{N}_C)}\|_F \geq \sqrt{\sum_{i=1}^{r_{\mathcal{N}}} (\sigma_i(Y^{(\mathcal{N}, \mathcal{N}_C)}) - \sigma_i(\tilde{Y}^{(\mathcal{N}, \mathcal{N}_C)}))^2} \\ &\geq |\sigma_{r_{\mathcal{N}}}(Y^{(\mathcal{N}, \mathcal{N}_C)}) - \sigma_{r_{\mathcal{N}}}(\tilde{Y}^{(\mathcal{N}, \mathcal{N}_C)})|. \end{aligned}$$

Daraus folgt

$$\begin{aligned} \sigma_{r_{\mathcal{N}}}(\tilde{Y}^{(\mathcal{N}, \mathcal{N}_C)}) &\geq \sigma_{r_{\mathcal{N}}}(Y^{(\mathcal{N}, \mathcal{N}_C)}) - |\sigma_{r_{\mathcal{N}}}(Y^{(\mathcal{N}, \mathcal{N}_C)}) - \sigma_{r_{\mathcal{N}}}(\tilde{Y}^{(\mathcal{N}, \mathcal{N}_C)})| \\ &\geq \rho - \delta. \end{aligned}$$

Setzen wir nun $\rho > \delta$ voraus, so kann die Pseudoinverse $\tilde{\mathbb{S}}^{\mathcal{N}^+}$ der Single-Hole-Matrix $\tilde{\mathbb{S}}^{\mathcal{N}}$ durch

$$\|\tilde{\mathbb{S}}^{\mathcal{N}^+}\|_2 = \left(\sigma_{r_{\mathcal{N}}}(\tilde{\mathbb{S}}^{\mathcal{N}})\right)^{-1} = \left(\sigma_{r_{\mathcal{N}}}(\tilde{Y}^{(\mathcal{N}, \mathcal{N}_C)})\right)^{-1} \leq (\rho - \delta)^{-1}$$

abgeschätzt werden. ■

Um die gewünschte Abschätzung der Krümmung der Mannigfaltigkeit zu erhalten, benötigen wir eine Voraussetzung, für die wir einen eigenen Begriff definieren.

Definition 8 (Wegzusammenhang). *Das geordnete Paar $(Y, \tilde{Y}) \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T}) \times \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ heißt **wegzusammenhängend** bezüglich der ODE (6.1), wenn das Randwertproblem*

$$\begin{aligned} \dot{X}(\tau) &= \sum_{k=0}^{\infty} (P(X(\tau)) - P(Y))^k P(Y) (\tilde{Y} - Y) \\ X(0) &= Y \quad , \quad X(1) = \tilde{Y} \end{aligned} \tag{6.1}$$

auf dem Intervall $[0, 1]$ eine Lösung besitzt.

Mit Lemma 21 und Definition 8 sind wir in der Lage, das Hauptresultat dieses Kapitels vorzustellen.

Theorem 22. Seien $(Y, \tilde{Y}) \in (\mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T}))^2$ Tensoren im hierarchischen Tucker-Format, welche wegzusammenhängend bezüglich der ODE (6.1) sind. Des Weiteren seien $(\rho_{\mathcal{N}})_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}}$ und $(c_{\mathcal{N}})_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}}$ Familien positiver, reellwertiger Zahlen, so dass für alle $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ die Relationen

$$\sigma_{r_{\mathcal{N}}}(Y^{(\mathcal{N}, \mathcal{N}_C)}) \geq \rho_{\mathcal{N}} > 0 \quad \text{und} \quad \delta := \|Y - \tilde{Y}\|_F \leq c_{\mathcal{N}} \rho_{\mathcal{N}} \quad (6.2)$$

erfüllt sind. Ferner sei \tilde{c} eine Konstante, welche den Ungleichungen

$$\sum_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}} 8 c_{\mathcal{N}} (\tilde{c}(1 - \tilde{c} - c_{\mathcal{N}}))^{-1} \leq 1 \quad \text{und} \quad \tilde{c} + c_{\mathcal{N}} < 1$$

genügt und

$$C_{Y, \tilde{Y}} := \sum_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}} 8 (\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}}))^{-1}.$$

Dann gelten für jeden Tensor $B \in \mathbb{R}^{N_1 \times \dots \times N_d}$ die Abschätzungen

$$\|(P(Y) - P(\tilde{Y}))B\|_F \leq C_{Y, \tilde{Y}} \|Y - \tilde{Y}\|_F \|B\|_F \quad (6.3)$$

$$\|P^\perp(Y)(\tilde{Y} - Y)\|_F \leq \frac{C_{Y, \tilde{Y}}}{1 - \tilde{c}} \|Y - \tilde{Y}\|_F^2. \quad (6.4)$$

Wir weisen darauf hin, dass Lubich, Rohwedder, Schneider und Vandereycken in [LRSV13, Lemma 4.5] mit ähnlichen Argumenten eine Abschätzung der Krümmung der Mannigfaltigkeit $\mathcal{M} = \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ herleiten.

Beweis von Theorem 22. Der Beweis zu Theorem 22 ist lang und sehr technisch. Aus diesem Grund haben wir den Beweis in sechs Teile untergliedert.

Teil 1: In diesem Teil zeigen wir, dass unter den Voraussetzungen des Theorems für alle $\tau \in [0, 1]$ stets

$$X(\tau) \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T}) \quad \text{und} \quad Z(\tau) \perp \mathcal{T}_Y \mathcal{M}$$

existieren, so dass sich jeder Tensor auf der Verbindungsgerade zwischen Y und \tilde{Y} als Summe von $X(\tau)$ und $Z(\tau)$ darstellen lässt, das heißt

$$Y + \tau(\tilde{Y} - Y) = X(\tau) + Z(\tau). \quad (6.5)$$

Darüber hinaus finden wir eine Abschätzung für $\dot{X}(\tau)$.

Wir nehmen zunächst an, dass es $X(\tau) \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ und $Z(\tau) \perp \mathcal{T}_Y \mathcal{M}$ gäbe, welche (6.5) erfüllen und leiten anhand dessen eine differential-algebraische Gleichung für $X(\tau)$ her. Dann weisen wir die Existenz einer Lösung dieser Gleichung nach. Zuletzt zeigen wir, dass aus der Existenz einer Lösung der differential-algebraischen Gleichung unmittelbar die Existenz von $X(\tau) \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ und $Z(\tau) \perp \mathcal{T}_Y \mathcal{M}$ folgt, welche (6.5) erfüllen.

Im Folgenden bezeichnen wir mit Δ die Projektion der Differenz von Y und \tilde{Y} in den Tangentialraum von Y :

$$\Delta := P(Y)(\tilde{Y} - Y) \in \mathcal{T}_Y\mathcal{M}.$$

Es gilt $\|\Delta\|_F \leq \|\tilde{Y} - Y\|_F = \delta$. Projizieren wir nun beide Seiten von (6.5) in den Tangentialraum $\mathcal{T}_Y\mathcal{M}$, so erhalten wir

$$P(Y)(X(\tau) - Y) = \tau\Delta. \quad (6.6)$$

Differenzieren nach τ ergibt

$$P(Y)\dot{X}(\tau) = \Delta. \quad (6.7)$$

Dadurch, dass die Ableitung $\dot{X}(\tau)$ ein Element des Tangentialraumes $\mathcal{T}_{X(\tau)}\mathcal{M}$ ist, gilt

$$\dot{X}(\tau) = P(X(\tau))\dot{X}(\tau) \quad (6.8)$$

und zusammen mit (6.7) erhalten wir die Gleichungen

$$\begin{aligned} \dot{X}(\tau) &= P(X(\tau))\dot{X}(\tau) + \underbrace{\Delta - P(Y)(\dot{X}(\tau))}_{=0} \\ &= \Delta + (P(X(\tau)) - P(Y))\dot{X}(\tau). \end{aligned} \quad (6.9)$$

Wie sich in Teil 6 des Beweises herausstellen wird, lässt sich mit den Voraussetzungen des Theorems die Operatornorm von $P(X(\tau)) - P(Y)$ für alle $\tau \in [0, 1]$ stets durch

$$\|P(X(\tau)) - P(Y)\|_F \leq \tilde{c} \quad (6.10)$$

abschätzen, so dass wir mit (6.9) für $\dot{X}(\tau)$ die Abschätzung

$$\begin{aligned} \|\dot{X}(\tau)\|_F &\leq \|\Delta\|_F + \|(P(X(\tau)) - P(Y))\dot{X}(\tau)\|_F \\ &\leq \|\Delta\|_F + \|P(X(\tau)) - P(Y)\|_F \|\dot{X}(\tau)\|_F \\ &\leq \|\Delta\|_F + \tilde{c} \|\dot{X}(\tau)\|_F \end{aligned}$$

erhalten und damit

$$\|\dot{X}(\tau)\|_F \leq \delta(1 - \tilde{c})^{-1}. \quad (6.11)$$

Um die Existenz von $X(\tau) \in \mathcal{H}$ -Tucker $((r_{\mathcal{N}})_{\mathcal{N} \in \mathcal{T}})$ und $Z(\tau) \perp \mathcal{T}_Y\mathcal{M}$ zu zeigen, welche die Gleichung

$$Y + \tau(\tilde{Y} - Y) = X(\tau) + Z(\tau)$$

erfüllen, weisen wir zunächst nach, dass die differential-algebraische Gleichung (6.9) stets eine Lösung besitzt.

Mit (6.10) und der Voraussetzung $\tilde{c} < 1$ folgt für alle $\tau \in [0, 1]$ die Konvergenz der

Neumann-Reihe zu dem Operator $P(X(\tau)) - P(Y)$ und damit die Invertierbarkeit des Operators

$$I - [P(X(\tau)) - P(Y)],$$

siehe [Wer11, Satz II.1.11]. Definieren wir nun

$$G(X(\tau)) := \left(I - [P(X(\tau)) - P(Y)] \right)^{-1} \Delta,$$

so folgt einerseits

$$G(X(\tau)) = \Delta + \sum_{k=1}^{\infty} [P(X(\tau)) - P(Y)]^k \Delta$$

und andererseits aus (6.9)

$$\dot{X}(\tau) = G(X(\tau)). \quad (6.12)$$

Mit Abschätzung (6.11), sowie Argumenten aus der Theorie der gewöhnliche Differentialgleichungen folgt die Existenz von mindestens einer Lösung $X(\tau)$, welche (6.12) beziehungsweise die differential-algebraische Gleichung (6.9) zum Anfangswert $X(0) = Y$ löst. Mit der Voraussetzung, dass das geordnete Paar (Y, \tilde{Y}) wegzusammenhängend bezüglich der ODE (6.1) ist, folgt darüber hinaus

$$X(1) = \tilde{Y}.$$

Wir zeigen nun, dass $X(\tau)$ sowohl (6.7), als auch (6.8) erfüllt. Die Projektion von (6.9) in den Tangentialraum von $X(\tau)$ liefert

$$P(X(\tau))(P(Y)\dot{X}(\tau) - \Delta) = 0. \quad (6.13)$$

Im Folgenden sei

$$T := \|P^\perp(X(\tau))(P(Y)\dot{X}(\tau) - \Delta)\|_F.$$

Wir zeigen nun die Ungleichung

$$T \leq \tilde{c} T,$$

woraufhin mit $\tilde{c} < 1$ unmittelbar die Gleichungen

$$0 = T = \|P^\perp(X(\tau))(P(Y)\dot{X}(\tau) - \Delta)\|_F \quad (6.14)$$

folgen. Mit $\Delta = P(Y)\Delta$ gelten die Abschätzungen

$$\begin{aligned} T &= \|P^\perp(X(\tau)) (P(Y)\dot{X}(\tau) - P(Y)\Delta)\|_F \\ &= \|[I - P(X(\tau))] P(Y) (\dot{X}(\tau) - \Delta)\|_F \\ &= \|[P(Y) - P(X(\tau))] (P(Y)\dot{X}(\tau) - \Delta)\|_F \\ &= \|[P(Y) - P(X(\tau))] \underbrace{[P(X(\tau)) + P^\perp(X(\tau))]}_{=I} (P(Y)\dot{X}(\tau) - \Delta)\|_F \end{aligned}$$

$$\begin{aligned}
&= \left\| [P(Y) - P(X(\tau))] P^\perp(X(\tau)) (P(Y)\dot{X}(\tau) - \Delta) \right\|_F \\
&\leq \|P(X(\tau)) - P(Y)\|_F \|P^\perp(X(\tau)) (P(Y)\dot{X}(\tau) - \Delta)\|_F \\
&\leq \tilde{c} \|P^\perp(X(\tau)) (P(Y)\dot{X}(\tau) - \Delta)\|_F \\
&= \tilde{c} T
\end{aligned}$$

und somit (6.14). Zusammen mit (6.13) erhalten wir

$$P(Y)\dot{X}(\tau) = \Delta. \quad (6.7)$$

Setzen wir (6.7) in die differential-algebraische Gleichung (6.9) ein, so folgt unmittelbar $P(X)\dot{X}(\tau) = \dot{X}(\tau)$. Integrieren wir nun beide Seiten von (6.7), so erhalten wir Gleichung (6.6) und mit

$$Z(\tau) := \tau P^\perp(Y)(\tilde{Y} - Y) - P^\perp(Y)X(\tau) \perp \mathcal{T}_Y \mathcal{M}$$

folgt die Existenz der Zerlegung

$$\begin{aligned}
X(\tau) + Z(\tau) &= X(\tau) + \tau P^\perp(Y)(\tilde{Y} - Y) - P^\perp(Y)X(\tau) \\
&= Y + P(Y)(X(\tau) - Y) + \tau P^\perp(Y)(\tilde{Y} - Y) \\
&= Y + \tau P(Y)(\tilde{Y} - Y) + \tau P^\perp(Y)(\tilde{Y} - Y) \\
&= Y + \tau(\tilde{Y} - Y).
\end{aligned}$$

Teil 2: Hier betrachten wir von $X(\tau) \in \mathcal{H}$ -Tucker $((r_{\mathcal{N}})_{\mathcal{N} \in T})$ die Darstellung durch das Tupel $\left((\mathcal{B}^{\mathcal{N}}(\tau))_{\mathcal{N} \in \mathcal{I}(T)}, (U^\ell(\tau))_{\ell \in \mathcal{L}(T)} \right)$ und leiten einige Hilfsabschätzungen her. Mit dem Hauptsatz der Differential- und Integralrechnung und $X(0) = Y$ erhalten wir

$$\begin{aligned}
\|X(\tau) - Y\|_F &= \left\| \int_0^\tau \dot{X}(s) ds \right\|_F \\
&\leq \int_0^\tau \|\dot{X}(s)\|_F ds \\
&\leq \tau \max_{s \in [0, \tau]} \|\dot{X}(s)\|_F.
\end{aligned}$$

Zusammen mit (6.11) folgt für $\tau \in [0, 1]$ die Abschätzung

$$\|X(\tau) - Y\|_F \leq \delta(1 - \tilde{c})^{-1}.$$

Im Folgenden sei $\mathbb{U}^{\mathcal{N}}(\tau)$ die orthonormale Basismatrix zu $X(\tau)$ und $\mathbb{S}^{\mathcal{N}}(\tau)$ die Single-Hole-Matrix, also

$$(X(\tau))^{(\mathcal{N}, \mathcal{N}_C)} = \mathbb{U}^{\mathcal{N}}(\tau) (\mathbb{S}^{\mathcal{N}}(\tau))^T$$

für $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$. Dann erfüllt die Pseudoinverse von $\mathbb{S}^{\mathcal{N}}(\tau)$ nach Lemma 21 für alle $\tau \in [0, 1]$ die Abschätzung

$$\|\mathbb{S}^{\mathcal{N}}(\tau)^+\|_2 \leq \frac{1}{\rho_{\mathcal{N}} - (1 - \tilde{c})^{-1}\delta} \leq \frac{1}{\rho_{\mathcal{N}} - (1 - \tilde{c})^{-1}c_{\mathcal{N}}\rho_{\mathcal{N}}} = \frac{1 - \tilde{c}}{\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}})}. \quad (6.15)$$

Aus Gründen der Übersichtlichkeit verzichten wir in Teil 2 des Beweises fortan auf das Argument τ . Wir weisen jedoch darauf hin, dass es sich sowohl bei den Matrizen $\mathbb{U}^{\mathcal{N}}$ und $\mathbb{S}^{\mathcal{N}}$ um parameterabhängige Objekte handelt, als auch bei dem Tensor X . Mit der Identität $(\mathbb{U}^{\mathcal{N}})^T \mathbb{U}^{\mathcal{N}} = I$, der Projektion auf das orthogonale Komplement $P^{\mathcal{N}\perp} = I - \mathbb{U}^{\mathcal{N}}(\mathbb{U}^{\mathcal{N}})^T$, sowie der Gauge-Bedingung $(\mathbb{U}^{\mathcal{N}})^T \dot{\mathbb{U}}^{\mathcal{N}} = 0$ erhalten wir für den Term $\dot{\mathbb{U}}^{\mathcal{N}}\mathbb{S}^{\mathcal{N}}$ die Darstellung

$$\begin{aligned} \dot{\mathbb{U}}^{\mathcal{N}}\mathbb{S}^{\mathcal{N}} &= \dot{\mathbb{U}}^{\mathcal{N}}(\mathbb{S}^{\mathcal{N}})^T \\ &\quad + \mathbb{U}^{\mathcal{N}}(\dot{\mathbb{S}}^{\mathcal{N}})^T - \mathbb{U}^{\mathcal{N}}(\dot{\mathbb{S}}^{\mathcal{N}})^T - \mathbb{U}^{\mathcal{N}}(\mathbb{U}^{\mathcal{N}})^T \dot{\mathbb{U}}^{\mathcal{N}}(\mathbb{S}^{\mathcal{N}})^T \\ &= \dot{\mathbb{U}}^{\mathcal{N}}(\mathbb{S}^{\mathcal{N}})^T + \mathbb{U}^{\mathcal{N}}(\dot{\mathbb{S}}^{\mathcal{N}})^T \\ &\quad - \mathbb{U}^{\mathcal{N}}(\mathbb{U}^{\mathcal{N}})^T \mathbb{U}^{\mathcal{N}}(\dot{\mathbb{S}}^{\mathcal{N}})^T - \mathbb{U}^{\mathcal{N}}(\mathbb{U}^{\mathcal{N}})^T \dot{\mathbb{U}}^{\mathcal{N}}(\mathbb{S}^{\mathcal{N}})^T \\ &= (I - \mathbb{U}^{\mathcal{N}}(\mathbb{U}^{\mathcal{N}})^T) (\dot{\mathbb{U}}^{\mathcal{N}}(\mathbb{S}^{\mathcal{N}})^T + \mathbb{U}^{\mathcal{N}}(\dot{\mathbb{S}}^{\mathcal{N}})^T) \\ &= P^{\mathcal{N}\perp} \dot{X}^{(\mathcal{N}, \mathcal{N}_C)} \end{aligned}$$

und damit

$$\|\dot{\mathbb{U}}^{\mathcal{N}}(\mathbb{S}^{\mathcal{N}})^T\|_F \leq \|P^{\mathcal{N}\perp}\|_2 \cdot \|\dot{X}^{(\mathcal{N}, \mathcal{N}_C)}\|_F \leq \delta(1 - \tilde{c})^{-1}.$$

Zusammen mit (6.15) folgt für die Ableitung der orthonormalen Basismatrix

$$\begin{aligned} \|\dot{\mathbb{U}}^{\mathcal{N}}\|_F &= \|\dot{\mathbb{U}}^{\mathcal{N}}(\mathbb{S}^{\mathcal{N}^+} \mathbb{S}^{\mathcal{N}})^T\|_F \\ &= \|\dot{\mathbb{U}}^{\mathcal{N}}(\mathbb{S}^{\mathcal{N}})^T (\mathbb{S}^{\mathcal{N}^+})^T\|_F \\ &\leq \|\dot{\mathbb{U}}^{\mathcal{N}}(\mathbb{S}^{\mathcal{N}})^T\|_F \|\mathbb{S}^{\mathcal{N}^+}\|_2 \\ &\leq \delta(1 - \tilde{c})^{-1} \frac{1 - \tilde{c}}{\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}})} \\ &= \frac{\delta}{\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}})}. \end{aligned}$$

Mit dieser Ungleichung sind wir in der Lage, die Ableitung der Projektion auf das orthogonale Komplement zu beschränken:

$$\begin{aligned} \|\dot{P}^{\mathcal{N}\perp}\|_F &= \left\| \frac{d}{d\tau} [I - \mathbb{U}^{\mathcal{N}}(\mathbb{U}^{\mathcal{N}})^T] \right\|_F \\ &= \|\dot{\mathbb{U}}^{\mathcal{N}}(\mathbb{U}^{\mathcal{N}})^T + \mathbb{U}^{\mathcal{N}}(\dot{\mathbb{U}}^{\mathcal{N}})^T\|_F \\ &\leq 2 \|\mathbb{U}^{\mathcal{N}}\|_2 \|\dot{\mathbb{U}}^{\mathcal{N}}\|_F \\ &\leq \frac{2\delta}{\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}})}. \end{aligned}$$

Um eine Abschätzung für die Ableitung der Single-Hole-Matrix zu finden, nutzen wir erneut die Identität $(\mathbb{U}^{\mathcal{N}})^T \mathbb{U}^{\mathcal{N}} = I$ und die Gauge-Bedingung $(\mathbb{U}^{\mathcal{N}})^T \dot{\mathbb{U}}^{\mathcal{N}} = 0$ und erhalten

$$\begin{aligned}
\|(\dot{\mathbb{S}}^{\mathcal{N}})^T\|_F &= \|(\mathbb{U}^{\mathcal{N}})^T \dot{\mathbb{U}}^{\mathcal{N}} (\mathbb{S}^{\mathcal{N}})^T + (\mathbb{U}^{\mathcal{N}})^T \mathbb{U}^{\mathcal{N}} (\dot{\mathbb{S}}^{\mathcal{N}})^T\|_F \\
&= \|(\mathbb{U}^{\mathcal{N}})^T [\dot{\mathbb{U}}^{\mathcal{N}} (\mathbb{S}^{\mathcal{N}})^T + \mathbb{U}^{\mathcal{N}} (\dot{\mathbb{S}}^{\mathcal{N}})^T]\|_F \\
&= \|(\mathbb{U}^{\mathcal{N}})^T \dot{X}^{(\mathcal{N}, \mathcal{N}_C)}\|_F \\
&\leq \|(\mathbb{U}^{\mathcal{N}})^T\|_2 \|\dot{X}^{(\mathcal{N}, \mathcal{N}_C)}\|_F \\
&\leq \delta(1 - \tilde{c})^{-1}.
\end{aligned} \tag{6.16}$$

Als letzte Abschätzung in diesem Teil betrachten wir die Zeitableitung des Terms $(\mathbb{S}^{\mathcal{N}})^T \mathbb{S}^{\mathcal{N}+}$. Mit den Hilfsmatrizen

$$\begin{aligned}
T_1 &:= \dot{\mathbb{S}}^{\mathcal{N}} \mathbb{S}^{\mathcal{N}+} \\
\text{und } T_2 &:= \mathbb{S}^{\mathcal{N}} \left[\frac{d}{d\tau} [((\mathbb{S}^{\mathcal{N}})^T \mathbb{S}^{\mathcal{N}})^{-1}] \right] (\mathbb{S}^{\mathcal{N}})^T
\end{aligned}$$

erhalten wir die Darstellung

$$\begin{aligned}
\frac{d}{d\tau} [(\mathbb{S}^{\mathcal{N}})^T \mathbb{S}^{\mathcal{N}+}] &= (\dot{\mathbb{S}}^{\mathcal{N}})^T [(\mathbb{S}^{\mathcal{N}})^T \mathbb{S}^{\mathcal{N}}]^{-1} \mathbb{S}^{\mathcal{N}} \\
&\quad + (\mathbb{S}^{\mathcal{N}})^T \left[\frac{d}{d\tau} [((\mathbb{S}^{\mathcal{N}})^T \mathbb{S}^{\mathcal{N}})^{-1}] \right] \mathbb{S}^{\mathcal{N}} \\
&\quad + (\mathbb{S}^{\mathcal{N}})^T [(\mathbb{S}^{\mathcal{N}})^T \mathbb{S}^{\mathcal{N}}]^{-1} \dot{\mathbb{S}}^{\mathcal{N}} \\
&= T_1 + T_2 + T_1^T.
\end{aligned}$$

Um den Term $(\mathbb{S}^{\mathcal{N}})^T \mathbb{S}^{\mathcal{N}+}$ zu beschränken, genügt es also, Abschätzungen für die Hilfsmatrizen T_1 und T_2 zu finden. Für T_1 gilt mit (6.15) und (6.16)

$$\begin{aligned}
\|T_1\|_F &= \|\dot{\mathbb{S}}^{\mathcal{N}} \mathbb{S}^{\mathcal{N}+}\|_F \\
&\leq \|\dot{\mathbb{S}}^{\mathcal{N}}\|_F \|\mathbb{S}^{\mathcal{N}+}\|_2 \\
&\leq \delta(1 - \tilde{c})^{-1} \frac{1 - \tilde{c}}{\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}})} \\
&= \frac{\delta}{\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}})}.
\end{aligned}$$

Um die Hilfsmatrix T_2 abzuschätzen, benutzen wir die Substitution

$$\frac{d}{d\tau} [((\mathbb{S}^{\mathcal{N}})^T \mathbb{S}^{\mathcal{N}})^{-1}] = -[(\mathbb{S}^{\mathcal{N}})^T \mathbb{S}^{\mathcal{N}}]^{-1} \left[\frac{d}{d\tau} [(\mathbb{S}^{\mathcal{N}})^T \mathbb{S}^{\mathcal{N}}] \right] [(\mathbb{S}^{\mathcal{N}})^T \mathbb{S}^{\mathcal{N}}]^{-1}$$

und erhalten für T_2 die Darstellung

$$\begin{aligned}
T_2 &= -\mathbb{S}^{\mathcal{N}} \left[(\mathbb{S}^{\mathcal{N}})^T \mathbb{S}^{\mathcal{N}} \right]^{-1} \left[\frac{d}{d\tau} \left[(\mathbb{S}^{\mathcal{N}})^T \mathbb{S}^{\mathcal{N}} \right] \right] \left[(\mathbb{S}^{\mathcal{N}})^T \mathbb{S}^{\mathcal{N}} \right]^{-1} (\mathbb{S}^{\mathcal{N}})^T \\
&= -\mathbb{S}^{\mathcal{N}} \left[(\mathbb{S}^{\mathcal{N}})^T \mathbb{S}^{\mathcal{N}} \right]^{-1} (\dot{\mathbb{S}}^{\mathcal{N}})^T \mathbb{S}^{\mathcal{N}} \left[(\mathbb{S}^{\mathcal{N}})^T \mathbb{S}^{\mathcal{N}} \right]^{-1} (\mathbb{S}^{\mathcal{N}})^T \\
&\quad - \mathbb{S}^{\mathcal{N}} \left[(\mathbb{S}^{\mathcal{N}})^T \mathbb{S}^{\mathcal{N}} \right]^{-1} (\mathbb{S}^{\mathcal{N}})^T \dot{\mathbb{S}}^{\mathcal{N}} \left[(\mathbb{S}^{\mathcal{N}})^T \mathbb{S}^{\mathcal{N}} \right]^{-1} (\mathbb{S}^{\mathcal{N}})^T \\
&= - \left[\dot{\mathbb{S}}^{\mathcal{N}} \mathbb{S}^{\mathcal{N}^+} \right]^T \mathbb{S}^{\mathcal{N}} \mathbb{S}^{\mathcal{N}^+} - \left[\mathbb{S}^{\mathcal{N}} \mathbb{S}^{\mathcal{N}^+} \right]^T \left[\dot{\mathbb{S}}^{\mathcal{N}} \mathbb{S}^{\mathcal{N}^+} \right] \\
&= -T_1^T \mathbb{S}^{\mathcal{N}} \mathbb{S}^{\mathcal{N}^+} - \left[\mathbb{S}^{\mathcal{N}} \mathbb{S}^{\mathcal{N}^+} \right]^T T_1
\end{aligned}$$

und mit $\|\mathbb{S}^{\mathcal{N}} \mathbb{S}^{\mathcal{N}^+}\|_2 = 1$ die Abschätzung

$$\|T_2\|_F \leq 2 \|T_1\|_F \|\mathbb{S}^{\mathcal{N}} \mathbb{S}^{\mathcal{N}^+}\|_2 \leq \frac{2\delta}{\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}})}.$$

Insgesamt erhalten wir also

$$\left\| \frac{d}{d\tau} \left[(\mathbb{S}^{\mathcal{N}})^T \mathbb{S}^{\mathcal{N}^+} \right] \right\|_F \leq \|T_1\|_F + \|T_2\|_F + \|T_1^T\|_F \leq \frac{4\delta}{\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}})}.$$

Teil 3: Um eine Abschätzung für die Krümmung der Mannigfaltigkeit herzuleiten, nutzen wir den Hauptsatz der Differential- und Integralrechnung und erhalten

$$\begin{aligned}
[P(\tilde{Y}) - P(Y)]B &= [P(X(1)) - P(X(0))]B \\
&= \int_0^1 \frac{d}{d\tau} [P(X(\tau))B] d\tau.
\end{aligned} \tag{6.17}$$

In (6.17) setzen wir nun für die Projektion $P(X(\tau))B$ die Darstellung aus Korollar 16 auf Seite 54 ein und gelangen auf diese Weise zu der Abschätzung

$$\begin{aligned}
&\|(P(\tilde{Y}) - P(Y))B\|_F \\
&\leq \max_{\tau \in [0,1]} \left\| \frac{d}{d\tau} \left[\left(\mathbb{U}^{\mathcal{N}_1^0}(\tau) \odot \mathbb{U}^{\mathcal{N}_2^0}(\tau) \right) \left(\mathbb{U}^{\mathcal{N}_1^0}(\tau) \odot \mathbb{U}^{\mathcal{N}_2^0}(\tau) \right)^T B^{(\mathcal{N}_0, \{\})} \right] \right\|_F
\end{aligned} \tag{6.18}$$

$$\begin{aligned}
&+ \sum_{\substack{\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\} \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} \max_{\tau \in [0,1]} \left\| \frac{d}{d\tau} \left[\mathbb{W}^{(\mathcal{N}_1, \mathcal{N}_2)}(\tau) P^{\mathcal{N}}(\tau)^\perp B^{(\mathcal{N}, \mathcal{N}_C)} \left(\mathbb{S}^{\mathcal{N}}(\tau) \mathbb{S}^{\mathcal{N}^+}(\tau) \right)^T \right] \right\|_F
\end{aligned} \tag{6.19}$$

$$\begin{aligned}
&+ \sum_{\ell \in \mathcal{L}(T)} \max_{\tau \in [0,1]} \left\| \frac{d}{d\tau} \left[P^\ell(\tau)^\perp B^{(\ell, \ell_C)} \left(\mathbb{S}^\ell(\tau) \mathbb{S}^{\ell^+}(\tau) \right)^T \right] \right\|_F,
\end{aligned} \tag{6.20}$$

wobei die Matrix $\mathbb{W}^{(\mathcal{N}_1, \mathcal{N}_2)}(\tau)$ definiert ist als

$$\mathbb{W}^{(\mathcal{N}_1, \mathcal{N}_2)}(\tau) := (\mathbb{U}^{\mathcal{N}_1}(\tau) \odot \mathbb{U}^{\mathcal{N}_2}(\tau)) (\mathbb{U}^{\mathcal{N}_1}(\tau) \odot \mathbb{U}^{\mathcal{N}_2}(\tau))^T$$

und $(\mathcal{N}_1^0, \mathcal{N}_2^0) = \text{succ}(\mathcal{N}_0)$.

Teil 4: In diesem Teil des Beweises benutzen wir die Hilfsabschätzungen aus Teil 2, um die Terme (6.18), (6.19) und (6.20) aus Teil 3 abzuschätzen.

Für einen inneren Knoten $\mathcal{N} \in \mathcal{I}(T)$ mit $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$ gilt

$$\begin{aligned}
\left\| \frac{d}{d\tau} \mathbb{W}^{(\mathcal{N}_1, \mathcal{N}_2)}(\tau) \right\|_2 &= \left\| \frac{d}{d\tau} \left[(\mathbb{U}^{\mathcal{N}_1}(\tau) \odot \mathbb{U}^{\mathcal{N}_2}(\tau)) (\mathbb{U}^{\mathcal{N}_1}(\tau) \odot \mathbb{U}^{\mathcal{N}_2}(\tau))^T \right] \right\|_2 \\
&\leq 2 \left\| \frac{d}{d\tau} [\mathbb{U}^{\mathcal{N}_1}(\tau) \odot \mathbb{U}^{\mathcal{N}_2}(\tau)] \right\|_2 \\
&\leq 2 \left\| \dot{\mathbb{U}}^{\mathcal{N}_1}(\tau) \odot \mathbb{U}^{\mathcal{N}_2}(\tau) \right\|_2 + 2 \left\| \mathbb{U}^{\mathcal{N}_1}(\tau) \odot \dot{\mathbb{U}}^{\mathcal{N}_2}(\tau) \right\|_2 \\
&\leq 2 \left\| \dot{\mathbb{U}}^{\mathcal{N}_1}(\tau) \right\|_2 + 2 \left\| \dot{\mathbb{U}}^{\mathcal{N}_2}(\tau) \right\|_2 \\
&\leq \frac{2\delta}{\rho_{\mathcal{N}_1}(1 - \tilde{c} - c_{\mathcal{N}_1})} + \frac{2\delta}{\rho_{\mathcal{N}_2}(1 - \tilde{c} - c_{\mathcal{N}_2})}. \tag{6.21}
\end{aligned}$$

Damit folgt für (6.18) mit einem beliebigen Tensor $B \in \mathbb{R}^{N_1 \times \dots \times N_d}$ die Abschätzung

$$\begin{aligned}
&\left\| \frac{d}{d\tau} \left[\left(\mathbb{U}^{\mathcal{N}_1^0}(\tau) \odot \mathbb{U}^{\mathcal{N}_2^0}(\tau) \right) \left(\mathbb{U}^{\mathcal{N}_1^0}(\tau) \odot \mathbb{U}^{\mathcal{N}_2^0}(\tau) \right)^T B^{(\mathcal{N}_0, \{\})} \right] \right\|_F \\
&\leq \frac{2\delta}{\rho_{\mathcal{N}_1^0}(1 - \tilde{c} - c_{\mathcal{N}_1^0})} \|B\|_F + \frac{2\delta}{\rho_{\mathcal{N}_2^0}(1 - \tilde{c} - c_{\mathcal{N}_2^0})} \|B\|_F. \tag{6.22}
\end{aligned}$$

Um den Term (6.19) abzuschätzen, nehmen wir nun $\mathcal{N} \neq \mathcal{N}_0$ an und erhalten mit (6.21), sowie den Abschätzungen aus Teil 2

$$\begin{aligned}
&\left\| \frac{d}{d\tau} \left[\mathbb{W}^{(\mathcal{N}_1, \mathcal{N}_2)}(\tau) P^{\mathcal{N}}(\tau)^\perp B^{(\mathcal{N}, \mathcal{N}_C)} (\mathbb{S}^{\mathcal{N}}(\tau) \mathbb{S}^{\mathcal{N}}(\tau)^+)^T \right] \right\|_F \\
&\leq \left\| \frac{d}{d\tau} [\mathbb{W}^{(\mathcal{N}_1, \mathcal{N}_2)}(\tau)] \right\|_2 \|B\|_F \left\| (\mathbb{S}^{\mathcal{N}}(\tau) \mathbb{S}^{\mathcal{N}}(\tau)^+)^T \right\|_2 \\
&\quad + \left\| \mathbb{W}^{(\mathcal{N}_1, \mathcal{N}_2)}(\tau) \right\|_2 \left\| \frac{d}{d\tau} [P^{\mathcal{N}}(\tau)^\perp] \right\|_F \|B\|_F \left\| (\mathbb{S}^{\mathcal{N}}(\tau) \mathbb{S}^{\mathcal{N}}(\tau)^+)^T \right\|_2 \\
&\quad + \left\| \mathbb{W}^{(\mathcal{N}_1, \mathcal{N}_2)}(\tau) \right\|_2 \|B\|_F \left\| \frac{d}{d\tau} [(\mathbb{S}^{\mathcal{N}}(\tau) \mathbb{S}^{\mathcal{N}}(\tau)^+)^T] \right\|_2 \\
&\leq \left(\frac{2\delta}{\rho_{\mathcal{N}_1}(1 - \tilde{c} - c_{\mathcal{N}_1})} + \frac{2\delta}{\rho_{\mathcal{N}_2}(1 - \tilde{c} - c_{\mathcal{N}_2})} + \frac{6\delta}{\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}})} \right) \|B\|_F. \tag{6.23}
\end{aligned}$$

Zuletzt in diesem Teil des Beweises betrachten für ein Blatt $\ell \in \mathcal{L}(T)$ des Dimensionsbaumes den Term (6.20) und nutzen erneut die Resultate aus Teil 2. Man erhält

$$\begin{aligned}
&\left\| \frac{d}{d\tau} \left[P^\ell(\tau)^\perp B^{(\ell, \ell_C)} (\mathbb{S}^\ell(\tau) \mathbb{S}^\ell(\tau)^+) \right] \right\|_F \\
&\leq \left\| \frac{d}{d\tau} [P^\ell(\tau)^\perp] \right\|_F \|B\|_F \left\| (\mathbb{S}^\ell(\tau) \mathbb{S}^\ell(\tau)^+)^T \right\|_2 \\
&\quad + \|B\|_F \left\| \frac{d}{d\tau} [(\mathbb{S}^\ell(\tau) \mathbb{S}^\ell(\tau)^+)^T] \right\|_2 \\
&\leq \frac{6\delta}{\rho_\ell(1 - \tilde{c} - c_\ell)} \|B\|_F. \tag{6.24}
\end{aligned}$$

Teil 5: Mit den vorhergehenden Beweisteilen sind wir nun in der Lage, (6.3) zu zeigen. Das Einsetzen der Abschätzungen (6.22), (6.23) und (6.24) aus Teil 4 in (6.18), (6.19) beziehungsweise (6.20) aus Teil 3 ergibt

$$\begin{aligned}
& \| (P(\tilde{Y}) - P(Y)) B \|_F \\
& \leq \frac{2\delta}{\rho_{\mathcal{N}_1^0}(1 - \tilde{c} - c_{\mathcal{N}_1^0})} \|B\|_F + \frac{2\delta}{\rho_{\mathcal{N}_2^0}(1 - \tilde{c} - c_{\mathcal{N}_2^0})} \|B\|_F \\
& \quad + \sum_{\substack{\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\} \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} \left(\frac{2\delta}{\rho_{\mathcal{N}_1}(1 - \tilde{c} - c_{\mathcal{N}_1})} + \frac{2\delta}{\rho_{\mathcal{N}_2}(1 - \tilde{c} - c_{\mathcal{N}_2})} + \frac{6\delta}{\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}})} \right) \|B\|_F \\
& \quad + \sum_{\ell \in \mathcal{L}(T)} \frac{6\delta}{\rho_\ell(1 - \tilde{c} - c_\ell)} \|B\|_F \\
& = \sum_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}} \frac{8\delta}{\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}})} \|B\|_F \\
& = C_{Y, \tilde{Y}} \|Y - \tilde{Y}\|_F \|B\|_F,
\end{aligned}$$

mit $(\mathcal{N}_1^0, \mathcal{N}_2^0) = \text{succ}(\mathcal{N}_0)$ als den Sukzessoren der Wurzel des Dimensionsbaumes.

Um die zweite Aussage des Theorems zu beweisen, benutzen wir erneut den Hauptsatz der Differential- und Integralrechnung und erhalten zusammen mit $P(X(\tau))\dot{X}(\tau) = \dot{X}(\tau)$ die Gleichungen

$$\begin{aligned}
P^\perp(Y)(\tilde{Y} - Y) &= \int_0^1 (I - P(Y))\dot{X} d\tau \\
&= \int_0^1 (P(X(\tau)) - P(Y))\dot{X} d\tau. \tag{6.25}
\end{aligned}$$

Mit (6.3), sowie (6.11) sind wir nun in der Lage, die rechte Seite von Gleichung (6.25) abzuschätzen und erhalten auf diese Weise die zweite Aussage des Theorems:

$$\begin{aligned}
\|P^\perp(Y)(\tilde{Y} - Y)\|_F &\leq \max_{\tau \in [0,1]} \|P(X(\tau)) - P(Y)\|_F \cdot \|\dot{X}(\tau)\|_F \\
&\leq \sum_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}} \frac{8\delta}{\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}})} \delta(1 - \tilde{c})^{-1} \\
&= \sum_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}} \frac{8\delta^2}{\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}})(1 - \tilde{c})} \\
&= \frac{C_{Y, \tilde{Y}}}{1 - \tilde{c}} \|Y - \tilde{Y}\|_F^2.
\end{aligned}$$

Teil 6: In Teil 1 des Beweises hatten wir angenommen, dass für alle $0 \leq \tau \leq 1$ die Ungleichung

$$\|P(X(\tau)) - P(Y)\|_F \leq \tilde{c} \tag{6.10}$$

stets erfüllt ist. In dem letzten Teil des Beweises zeigen wir, dass diese Annahme mit den Voraussetzungen des Theorems immer erfüllt ist. Um das zu zeigen, nehmen wir an, dass (6.10) nicht erfüllt ist und leiten damit einen Widerspruch her.

Dadurch, dass $\mathcal{M} = \mathcal{H}$ -Tucker $((r_{\mathcal{N}})_{\mathcal{N} \in T})$ eine glatte Mannigfaltigkeit ist, existieren $0 < \tau^* < 1$ und ein Tensor $B^* \in \mathbb{R}^{N_1 \times \dots \times N_d}$, so dass

$$\begin{aligned} \|P(X(s)) - P(Y)\|_F &\leq \tilde{c} < 1 \\ \text{und } \|(P(X(\tau^*)) - P(Y))B^*\|_F &= \tilde{c} \|B^*\|_F \end{aligned} \quad (6.26)$$

für alle $0 \leq s \leq \tau^*$ erfüllt ist. Unter diesen Gegebenheiten behalten die Abschätzungen aus Teil 2 und Teil 4 für $\tau \in [0, \tau^*]$ ihre Gültigkeit. In Teil 3 betrachten wir dann anstelle des Integrals über das Intervall $[0, 1]$ das Integral über das Intervall $[0, \tau^*]$ und erhalten analog zum Vorgehen aus Teil 5 des Beweises die Abschätzung

$$\|(P(X(\tau^*)) - P(Y))B\|_F \leq \tau^* \sum_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}} \frac{8\delta}{\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}})} \|B\|_F$$

für einen beliebigen Tensor $B \in \mathbb{R}^{N_1 \times \dots \times N_d}$. Dies ist für $B = B^*$ und den Voraussetzungen

$$\sum_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}} \frac{8c_{\mathcal{N}}}{1 - \tilde{c} - c_{\mathcal{N}}} \leq \tilde{c} \quad \text{und} \quad \delta \leq c_{\mathcal{N}} \rho_{\mathcal{N}}$$

ein Widerspruch zu (6.26), da

$$\begin{aligned} \|(P(X(\tau^*)) - P(Y))B^*\|_F &\leq \tau^* \sum_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}} \frac{8\delta}{\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}})} \|B^*\|_F \\ &\leq \tau^* \tilde{c} \|B^*\|_F \\ &\stackrel{!}{<} \tilde{c} \|B^*\|_F. \end{aligned}$$

■

Die Voraussetzung des Wegzusammenhangs bezüglich der ODE (6.1) ist in Theorem 22 ein Hilfsmittel, um zu zeigen, dass eine Kurve $X(t) \in \mathcal{H}$ -Tucker $((r_{\mathcal{N}})_{\mathcal{N} \in T})$ existiert, welche die Punkte Y und \tilde{Y} verbindet und deren Ableitung beschränkt ist, siehe Abschätzung (6.11). Wir haben die Vermutung, dass das Theorem auch unter einer schwächeren Annahme Gültigkeit besitzt. Es bleibt deshalb auch nach dieser Arbeit Gegenstand der Forschung, die richtigen Argumente aus der Differentialgeometrie beziehungsweise die richtigen Annahmen zu finden, so dass das Theorem auch ohne die Voraussetzung des Wegzusammenhangs gültig ist.

Um Theorem 22 zu vereinfachen, nehmen wir

$$c_{\mathcal{N}} = c \text{ für alle } \mathcal{N} \in T \setminus \{\mathcal{N}_0\} \quad (6.27)$$

an und versuchen diese Konstante zu maximieren, so dass das Theorem auf Tensoren Y und \tilde{Y} anwendbar ist, welche einen großen Abstand voneinander haben, jedoch noch immer die Relation

$$\|Y - \tilde{Y}\|_F \leq c \rho_{\mathcal{N}}$$

für alle $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ erfüllen. Weil ein Dimensionsbaum ein voller Binärbaum ist, besitzt dieser stets $2d - 1$ Knoten. Dadurch vereinfachen sich die Nebenbedingungen von Theorem 22 durch Annahme (6.27) zu

$$\tilde{c} + c < 1 \quad (6.28a)$$

$$\text{und } \frac{8c(2d-2)}{\tilde{c}(1-\tilde{c}-c)} \leq 1. \quad (6.28b)$$

Unter Berücksichtigung von Nebenbedingung (6.28a) vergrößert sich der Wert des Terms

$$\frac{8c(2d-2)}{\tilde{c}(1-\tilde{c}-c)}$$

für einen größer werdenden Wert von c . Das Maximum, das wir suchen, wird also für

$$\frac{8c(2d-2)}{\tilde{c}(1-\tilde{c}-c)} = 1 \quad (6.29)$$

angenommen. Wir betrachten nun den Wert von c in Abhängigkeit von \tilde{c} und erhalten mit (6.29) die Funktion

$$c(\tilde{c}) = \frac{-\tilde{c}^2 + \tilde{c}}{8(2d-2) + \tilde{c}}. \quad (6.30)$$

Diese Funktion besitzt ein positives Maximum an der Stelle

$$\tilde{c} = -8(2d-2) + \sqrt{8^2(2d-2)^2 + 8(2d-2)}.$$

Gehen wir nun ferner davon aus, dass die Tensoren eine hohe Dimension besitzen und betrachten den Grenzwert von \tilde{c} für $d \rightarrow \infty$, so erhalten wir den Wert $\tilde{c} = \frac{1}{2}$. Setzen wir diesen Wert wiederum in (6.30) ein, so erhalten wir $c = (64d - 62)^{-1}$.

Mit der Parameterwahl

$$\tilde{c} = \frac{1}{2} \quad \text{und} \quad c = (64d - 62)^{-1}$$

sind beide Nebenbedingungen, (6.28a) und (6.28b), erfüllt und Theorem 22 lässt sich wie folgt vereinfachen.

Korollar 23. Seien $(Y, \tilde{Y}) \in (\mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T}))^2$ Tensoren im hierarchischen Tucker-Format, welche wegzusammenhängend bezüglich der ODE (6.1) sind. Ferner sei $\rho > 0$, so dass für alle $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ die Ungleichungen

$$\sigma_{r_{\mathcal{N}}}(Y^{(\mathcal{N}, \mathcal{N}_C)}) \geq \rho > 0 \quad \text{und} \quad \|Y - \tilde{Y}\|_F \leq (64d - 62)^{-1} \rho$$

erfüllt sind und $\beta := (32d - 31) \rho^{-1}$.

Dann gelten für jeden Tensor $B \in \mathbb{R}^{N_1 \times \dots \times N_d}$ die Abschätzungen

$$\begin{aligned} \|(P(Y) - P(\tilde{Y}))B\|_F &\leq \beta \|Y - \tilde{Y}\|_F \|B\|_F \\ \|P^\perp(Y)(\tilde{Y} - Y)\|_F &\leq 2\beta \|Y - \tilde{Y}\|_F^2. \end{aligned}$$

Beweis. Die Abschätzungen resultieren aus Theorem 22 durch das Einsetzen der Parameterwerte $\tilde{c} = \frac{1}{2}$ und $c = (64d - 62)^{-1}$. \blacksquare

In Teil 1 des Beweises von Theorem 22 haben wir gezeigt, dass unter gewissen Voraussetzungen zu zwei Tensoren $Y, \tilde{Y} \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ für $\tau \in [0, 1]$ stets

$$X(\tau) \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T}) \quad \text{und} \quad Z(\tau) \perp \mathcal{T}_Y \mathcal{M}$$

existieren, so dass sich jeder Tensor auf der Verbindungsgerade zwischen Y und \tilde{Y} durch die Gleichung

$$Y + \tau(\tilde{Y} - Y) = X(\tau) + Z(\tau) \tag{6.31}$$

darstellen lässt. Mit der Annahme, dass das geordnete Paar (Y, \tilde{Y}) wegzusammenhängend bezüglich der ODE (6.1) ist, konnten wir dann auf die Gleichungen

$$X(1) = \tilde{Y} \tag{6.32a}$$

$$\text{und} \quad Z(1) = 0 \tag{6.32b}$$

schließen. Die Existenz von $X(\tau) \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ und $Z(\tau) \perp \mathcal{T}_Y \mathcal{M}$, welche (6.31) erfüllen, lässt sich mit den Argumenten aus Teil 1 des Beweises von Theorem 22 auch ohne die Voraussetzung des Wegzusammenhangs bezüglich der ODE (6.1) zeigen. Dann besitzen die Gleichungen (6.32a) und (6.32b) jedoch nicht notwendigerweise Gültigkeit.

Korollar 24. Seien $Y, \tilde{Y} \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ Tensoren im hierarchischen Tucker-Format und $\rho > 0$, so dass für alle $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ die Ungleichungen

$$\sigma_{r_{\mathcal{N}}}(Y^{(\mathcal{N}, \mathcal{N}_C)}) \geq \rho > 0 \quad \text{und} \quad \delta := \|Y - \tilde{Y}\|_F \leq (64d - 62)^{-1}\rho$$

erfüllt sind. Dann existieren für $\tau \in [0, 1]$ parameterabhängige Tensoren $Z(\tau) \perp \mathcal{T}_Y \mathcal{M}$ und $X(\tau) \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ mit

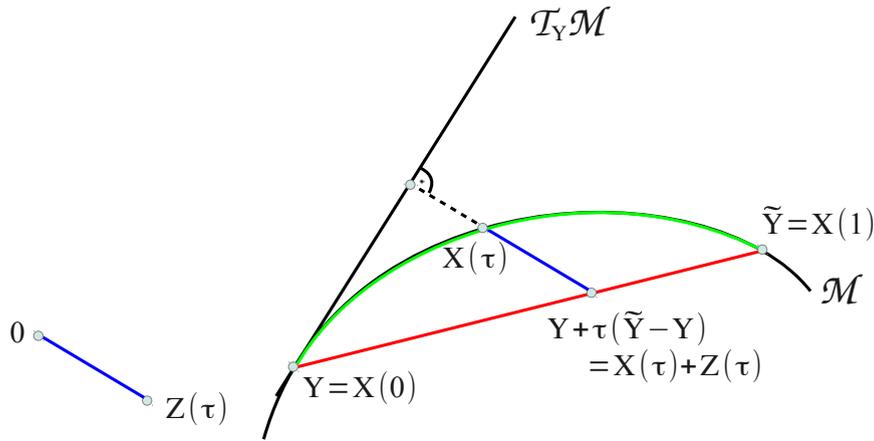
$$X(0) = Y \quad \text{und} \quad \|\dot{X}(\tau)\|_F \leq 2\delta,$$

so dass

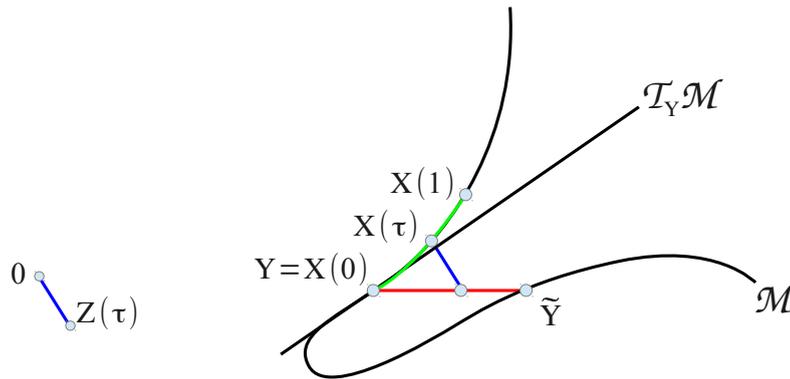
$$Y + \tau(\tilde{Y} - Y) = X(\tau) + Z(\tau).$$

Beweis. Siehe Teil 1 des Beweises von Theorem 22 mit $\tilde{c} = \frac{1}{2}$ und $c = (64d - 62)^{-1}$. ■

Abbildung 6.2 illustriert diesen Sachverhalt und zeigt zwei Beispiele, welche die Bedeutung des Wegzusammenhangs bezüglich der ODE (6.1) verdeutlichen.



(a) Mit Wegzusammenhang



(b) Ohne Wegzusammenhang

Abbildung 6.2: Illustration des Wegzusammenhangs bezüglich der ODE (6.1)

Teilbild 6.2(a) zeigt ein Beispiel, in dem das geordnete Paar (Y, \tilde{Y}) wegzusammenhängend bezüglich der ODE (6.1) ist. In Teilbild 6.2(b) hingegen besteht kein Wegzusammenhang bezüglich der ODE (6.1), denn $\tilde{Y} \neq X(1)$. Die Trajektorie, welche der parameterwertige Tensor $X(\tau)$ gemäß der zugrundeliegenden Differentialgleichung entlangläuft, ist in beiden Teilbildern in grün dargestellt.

7. A posteriori Fehleranalyse der Niedrigrangapproximation

In Kapitel 4 haben wir für eine Funktion $F : \mathbb{R}^{N_1 \times \dots \times N_d} \rightarrow \mathbb{R}^{N_1 \times \dots \times N_d}$ die Lösung des hochdimensionalen Anfangswertproblems

$$\dot{Y}_{ex}(t) = F(Y_{ex}(t)) \quad , \quad Y(0) = Y_0 \in \mathbb{R}^{N_1 \times \dots \times N_d} \quad , \quad t \geq 0 \quad (7.1)$$

durch eine Niedrigrangapproximation $Y(t)$ approximiert, welche auf der Mannigfaltigkeit $\mathcal{M} = \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ verläuft. Zu diesem Zweck haben wir in Theorem 11 mittels des Variationsprinzips von Dirac-Frenkel Bewegungsgleichungen für das Tupel $\left((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^\ell)_{\ell \in \mathcal{L}(T)} \right)$ hergeleitet, durch welches $Y(t)$ gemäß Definition 4 dargestellt wird. Ziel dieses Vorgehens war, die Anzahl an Unbekannten des hochdimensionalen Systems zu reduzieren in der Hoffnung, dass der Fehler zwischen der exakten Lösung $Y_{ex}(t)$ und der Modellreduktion $Y(t)$ klein ist. In diesem Kapitel entwickeln wir für den Fehler der Niedrigrangapproximation eine *a posteriori* Fehleranalyse.

Wir weisen darauf hin, dass Lubich, Rohwedder, Schneider und Vandereycken in [LRSV13] mit ähnlichen Argumenten eine *a posteriori* Fehleranalyse der Niedrigrangapproximation bestimmen. Die Forschungsarbeit dieser Ausarbeitung geschah jedoch unabhängig zu der Arbeit von Lubich, Rohwedder, Schneider und Vandereycken. Das Vorgehen unserer Arbeit ist an den Techniken von Koch und Lubich in [KL10] angelehnt. In [KL10, Theorem 4.4] leiten Koch und Lubich eine Abschätzung für den Fehler der Niedrigrangapproximation der Lösung eines hochdimensionalen Anfangswertproblems in der Mannigfaltigkeit Tucker (r_1, \dots, r_d) her. Wir übertragen diese Vorgehensweise, um in Theorem 25 und Korollar 26 eine Abschätzung für den Fehler der Niedrigrangapproximation in der Mannigfaltigkeit $\mathcal{M} = \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ zu bestimmen.

Im Folgenden bezeichnen wir mit $X(t)$ die Bestapproximation der exakten Lösung $Y_{ex}(t)$

in der Mannigfaltigkeit $\mathcal{M} = \mathcal{H}$ -Tucker $((r_{\mathcal{N}})_{\mathcal{N} \in T})$, das heißt zu jedem Zeitpunkt $t \in [0, T]$ gilt

$$\|X(t) - Y_{ex}(t)\|_F \leq \|\widehat{X} - Y_{ex}(t)\|_F \quad \text{für alle } \widehat{X} \in \mathcal{M}.$$

Theorem 25. *Wir setzen auf kompakten Zeitintervallen, also für $t \in [0, T]$, die folgenden Annahmen voraus:*

(A1) *Die Abweichung zwischen Niedrigrangapproximation und Bestapproximation zum Zeitpunkt $t = 0$ ist durch*

$$\|Y(0) - X(0)\|_F \leq \frac{1}{2\beta}$$

beschränkt, wobei $\beta = (32d - 31)\rho^{-1}$ wie in Korollar 23 auf Seite 78.

(A2) *$X(t)$ ist stetig differenzierbar.*

(A3) *Es existiert eine Konstante $\mu > 0$, so dass die Abschätzungen*

$$\|F(Y_{ex}(t))\|_F \leq \mu \quad \text{und} \quad \|F(X(t))\|_F \leq \mu$$

erfüllt sind.

(A4) *Es existieren Konstanten $\ell, \lambda > 0$, so dass die Lipschitz-Bedingungen*

$$\begin{aligned} \|F(Y_{ex}(t)) - F(X(t))\|_F &\leq \ell \|Y_{ex}(t) - X(t)\|_F \\ \text{und} \quad \|F(Y(t)) - F(X(t))\|_F &\leq \lambda \|Y(t) - X(t)\|_F \end{aligned}$$

erfüllt sind.

(A5) *Die Singulärwerte der Matrifizierung $X^{(\mathcal{N}, \mathcal{N}_C)}$ sind für alle $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ durch*

$$\sigma_{r_{\mathcal{N}}}\left(X^{(\mathcal{N}, \mathcal{N}_C)}(t)\right) \geq \rho > 0$$

nach unten beschränkt.

(A6) *Der Abstand der exakten Lösung $Y_{ex}(t)$ zur Bestapproximation $X(t)$ ist durch*

$$\delta(t) := \|X(t) - Y_{ex}(t)\|_F \leq \frac{1}{2\beta}$$

beschränkt.

(A7) *Des Weiteren sei die Abschätzung*

$$R(t) := \|Y(0) - X(0)\|_F e^{(\beta\mu+\lambda)t} + (\ell + 2\mu\beta) \int_0^t \delta(s) e^{(\beta\mu+\lambda)(t-s)} ds \leq \frac{1}{2\beta}$$

gegeben.

(A8) *Darüber hinaus sei das geordnete Paar $(Y(t), X(t))$ für alle $t \in [0, T]$ wegzusammenhängend bezüglich der ODE (6.1).*

Dann gelten die Abschätzungen

$$\|Y(t) - X(t)\|_F \leq R(t) \quad (7.2)$$

$$\text{und } \|Y(t) - Y_{ex}(t)\|_F \leq \delta(t) + R(t). \quad (7.3)$$

Ehe wir den Beweis des Theorems präsentieren, wollen wir auf die Bedeutung der Annahmen eingehen. Die Annahmen (A1), (A5), (A7) und (A8) gewährleisten, dass wir in dem Beweis des Theorems die Abschätzung über die Krümmung der Mannigfaltigkeit aus Korollar 23 auf die Tensoren $Y(t)$ und $X(t)$ anwenden können. Zu diesem Zweck müssen beide Tensoren nah genug beieinander liegen. Annahme (A1) stellt sicher, dass die Tensoren $Y(t)$ und $X(t)$ zum Zeitpunkt $t = 0$ nicht bereits zu weit voneinander entfernt sind. Mit Hilfe von Annahme (A7) schließen wir dann im Laufe des Beweises darauf, dass die Tensoren $Y(t)$ und $X(t)$ auf dem gesamten Zeitintervall nah genug beieinander liegen. Um Korollar 23 anwenden zu können, ist es ebenfalls erforderlich, dass die Singulärwerte der Matrifizierung der Bestapproximation nach unten beschränkt sind und dass das geordnete Paar $(Y(t), X(t))$ wegzusammenhängend bezüglich der ODE (6.1) ist. Diese Eigenschaften setzen wir in den Annahmen (A5) und (A8) voraus. Dass die Bestapproximation stetig differenzierbar ist, wie wir es in Annahme (A2) voraussetzen, benötigen wir im Beweis des Theorems, wenn wir die Ableitung der Bestapproximation betrachten. Geht das Anfangswertproblem (7.1) aus der Diskretisierung einer partiellen Differentialgleichung hervor, so hängt die Konstante μ aus Annahme (A3) von den Operatoren der partiellen Differentialgleichung, sowie von der gewählten Diskretisierung ab. Aus der Theorie der gewöhnlichen Differentialgleichungen ist bekannt, dass Lipschitz-Bedingungen die Existenz und Eindeutigkeit der Lösung eines Anfangswertproblems sicherstellen [Wal00, Kapitel 2 §6], daher ist es legitim, solche Bedingungen in Annahme (A4) vorauszusetzen. Annahme (A6) verwenden wir im Beweis dazu, dass sich manche Terme auf eine sinnvolle Weise kürzen.

Beweis von Theorem 25. Wir untergliedern den Beweis in drei Teile. Diese Unterteilung fördert einerseits die Lesbarkeit des Beweises und andererseits sind wir dadurch in der Lage, die Teile 1 und 3 im Beweis des nächsten Korollars wiederverwerten zu können. In den Teilen 1 und 2 beweisen wir die *a posteriori* Fehleranalyse zunächst unter der Annahme, dass die linke Seite von (7.2) die Schranke

$$L(t) := \|Y(t) - X(t)\|_F \leq \frac{1}{2\beta} \quad (7.4)$$

besitzt. In Teil 3 zeigen wir die Gültigkeit von Ungleichung (7.4), wenn die rechte Seite von (7.2)

$$R(t) = d_0 e^{(\beta\mu+\lambda)t} + (\ell + 2\mu\beta) \int_0^t \delta(s) e^{(\beta\mu+\lambda)(t-s)} ds \quad (7.5)$$

durch $\frac{1}{2\beta}$ nach oben beschränkt ist, so wie es in Annahme (A7) gefordert ist.

Teil 1: Im ersten Teil des Beweises leiten wir eine Abschätzung für den Term

$$D(t) := [P'(X(t)) \cdot (Y_{ex}(t) - X(t))] \dot{X}(t)$$

her und finden eine alternative Darstellung für den Term $\dot{X}(t)$.

Weil $X(t)$ die Bestapproximation von $Y_{ex}(t)$ ist, liegt die Differenz $Y_{ex}(t) - X(t)$ orthogonal zum Tangentialraum $\mathcal{T}_{X\mathcal{M}}$,

$$P(X(t))(Y_{ex}(t) - X(t)) = 0.$$

Differenzieren ergibt

$$\begin{aligned} 0 &= \frac{d}{dt} [P(X(t))(Y_{ex}(t) - X(t))] \\ &= \underbrace{[P'(X(t)) \cdot (Y_{ex}(t) - X(t))] \dot{X}(t)}_{=: H(t)} + P(X(t))(\dot{Y}_{ex}(t) - \dot{X}(t)). \end{aligned}$$

Dabei ist der Term $H(t)$ als Grenzwert

$$H(t) = \lim_{h \rightarrow 0} \frac{1}{h} \left([(P(X(t+h)) - P(X(t)))(Y_{ex}(t) - X(t))] \right)$$

zu verstehen. Mit $\dot{X}(t) = P(X(t))\dot{X}(t)$ erhalten wir die Gleichungen

$$\begin{aligned} \dot{X}(t) &= P(X(t))\dot{Y}_{ex}(t) + [P'(X(t)) \cdot (Y_{ex}(t) - X(t))] \dot{X}(t) \\ &= P(X(t))\dot{Y}_{ex}(t) + D(t) \end{aligned} \tag{7.6}$$

und damit eine alternative Darstellung für $\dot{X}(t)$. Da es sich bei der Menge \mathcal{H} -Tucker $((r_{\mathcal{N}})_{\mathcal{N} \in T})$ um eine glatte Mannigfaltigkeit handelt, besteht für hinreichend kleines $h > 0$ ein Wegzusammenhang bezüglich der ODE (6.1) für das geordnete Paar $(X(t), X(t+h))$ und wir erhalten unter Verwendung von Korollar 23 für einen beliebigen Tensor $B \in \mathbb{R}^{N_1 \times \dots \times N_d}$ die Abschätzung

$$\begin{aligned} \left\| (P'(X(t)) \cdot B) \dot{X}(t) \right\|_F &= \left\| \frac{d}{dt} [P(X(t))B] \right\|_F \\ &= \lim_{h \rightarrow 0} \frac{1}{h} \left\| (P(X(t+h)) - P(X(t)))B \right\|_F \\ &\leq \lim_{h \rightarrow 0} \frac{1}{h} \beta \|X(t+h) - X(t)\|_F \|B\|_F \\ &= \beta \|\dot{X}(t)\|_F \|B\|_F. \end{aligned}$$

Setzen wir nun $B = Y_{ex}(t) - X(t)$, so führt das auf die Ungleichung

$$\|D(t)\|_F \leq \beta \|\dot{X}(t)\|_F \|Y_{ex}(t) - X(t)\|_F. \tag{7.7}$$

Durch das Einsetzen von Gleichung (7.6) in (7.7) erhalten wir

$$\begin{aligned}\|D(t)\|_F &\leq \beta \|P(X(t))\dot{Y}_{ex}(t) + D(t)\|_F \|Y_{ex}(t) - X(t)\|_F \\ &\leq \beta \delta(t) \left(\|P(X(t))\dot{Y}_{ex}(t)\|_F + \|D(t)\|_F \right).\end{aligned}$$

An dieser Stelle nutzen wir Annahme (A6), $\delta(t) \leq \frac{1}{2\beta}$, und bekommen

$$\|D(t)\|_F \leq \beta \delta(t) \|P(X(t))\dot{Y}_{ex}(t)\|_F + \frac{1}{2}\|D(t)\|_F.$$

Dies führt wiederum zu der Ungleichung

$$\|D(t)\|_F \leq 2\beta \delta(t) \|P(X(t))\dot{Y}_{ex}(t)\|_F.$$

Verwenden wir nun Annahme (A3), $\|\dot{Y}_{ex}(t)\|_F = \|F(Y_{ex}(t))\|_F \leq \mu$, so führt das zu der gewünschten Abschätzung von Teil 1 des Beweises

$$\|D(t)\|_F \leq 2\mu\beta\delta(t).$$

Teil 2: In diesem Beweisabschnitt zeigen wir, dass die Abschätzungen des Theorems unter der Annahme (7.4) stets erfüllt sind.

Zu diesem Zweck betrachten wir die Differenz $\dot{Y}(t) - \dot{X}(t)$ und substituieren die Terme durch

$$\dot{Y}(t) = P(Y(t))F(Y(t))$$

und (7.6). Auf diese Weise erhalten wir

$$\begin{aligned}\dot{Y}(t) - \dot{X}(t) &= P(Y(t))F(Y(t)) - P(X(t))\dot{Y}_{ex}(t) - D(t) \\ &= [P(Y(t)) - P(X(t))]F(X(t)) + P(X(t))[F(X(t)) - F(Y_{ex}(t))] \\ &\quad + P(Y(t))[F(Y(t)) - F(X(t))] - D(t).\end{aligned}$$

Die folgenden Hilfsabschätzungen dienen dazu, eine obere Schranke für das Skalarprodukt $\langle \dot{Y}(t) - \dot{X}(t), Y(t) - X(t) \rangle$ zu finden.

Als Erstes betrachten wir das Skalarprodukt $\langle [P(Y(t)) - P(X(t))]F(X(t)), Y(t) - X(t) \rangle$. An dieser und nur an dieser Stelle machen wir von der Annahme (7.4), $L(t) \leq \frac{1}{2\beta}$, Gebrauch und sind somit in der Lage, Korollar 23 anzuwenden

$$\begin{aligned}&\langle [P(Y(t)) - P(X(t))]F(X(t)), Y(t) - X(t) \rangle \\ &\leq \| [P(Y(t)) - P(X(t))]F(X(t)) \|_F \|Y(t) - X(t)\|_F \\ &\leq \beta \|Y(t) - X(t)\|_F^2 \|F(X(t))\|_F \\ &\stackrel{(A3)}{\leq} \beta \mu L^2(t).\end{aligned}\tag{7.8}$$

Bei der zweiten und dritten Hilfsabschätzung verwenden wir Annahme (A4) und erhalten so

$$\begin{aligned}
& \langle P(X(t))[F(X(t)) - F(Y_{ex}(t))], Y(t) - X(t) \rangle \\
& \leq \|F(X(t)) - F(Y_{ex}(t))\|_F \|Y(t) - X(t)\|_F \\
& \leq \ell \|X(t) - Y_{ex}(t)\|_F \|Y(t) - X(t)\|_F \\
& = \ell \delta(t) L(t)
\end{aligned}$$

und

$$\begin{aligned}
& \langle P(Y(t))[F(Y(t)) - F(X(t))], Y(t) - X(t) \rangle \\
& \leq \|F(Y(t)) - F(X(t))\|_F \|Y(t) - X(t)\|_F \\
& \leq \lambda \|Y(t) - X(t)\|_F^2 \\
& = \lambda L^2(t).
\end{aligned}$$

Als letzte Hilfsabschätzung betrachten wir

$$\begin{aligned}
\langle D(t), Y(t) - X(t) \rangle & \leq \|D(t)\|_F \|Y(t) - X(t)\|_F \\
& \leq 2 \mu \beta \delta(t) L(t).
\end{aligned}$$

Mit diesen Hilfsabschätzungen sind wir in der Lage, den Term $L(t)\dot{L}(t)$ auf die folgende Weise abzuschätzen

$$\begin{aligned}
L(t)\dot{L}(t) & = \|Y(t) - X(t)\|_F \frac{d}{dt} \|Y(t) - X(t)\|_F \\
& = \frac{1}{2} \frac{d}{dt} \|Y(t) - X(t)\|_F^2 \\
& = \langle \dot{Y}(t) - \dot{X}(t), Y(t) - X(t) \rangle \\
& \leq \beta \mu L^2(t) + \ell \delta(t) L(t) + \lambda L^2(t) + 2 \mu \beta \delta(t) L(t)
\end{aligned}$$

und schließen damit auf die Ungleichung

$$\dot{L}(t) \leq (\beta\mu + \lambda) L(t) + (\ell + 2\mu\beta) \delta(t).$$

Wenden wir nun das Lemma von Gronwall an, so erhalten wir (7.2),

$$L(t) \leq L(0)e^{(\beta\mu+\lambda)t} + (\ell + 2\mu\beta) \int_0^t \delta(s) e^{(\beta\mu+\lambda)(t-s)} ds.$$

Abschätzung (7.3) folgt unmittelbar aus (7.2) durch das Anwenden der Dreiecksungleichung

$$\begin{aligned}
\|Y(t) - Y_{ex}(t)\|_F &= \|Y(t) - X(t) + X(t) - Y_{ex}(t)\|_F \\
&\leq \|X(t) - Y_{ex}(t)\|_F + \|Y(t) - X(t)\|_F \\
&\leq \delta(t) + L(0)e^{(\beta\mu+\lambda)t} + (\ell + 2\mu\beta) \int_0^t \delta(s) e^{(\beta\mu+\lambda)(t-s)} ds.
\end{aligned}$$

Teil 3: In den zurückliegenden Beweisteilen haben wir gezeigt, dass die Abschätzung $L(t) \leq R(t)$, also (7.2) und damit auch (7.3) so lang Gültigkeit besitzen, wie $L(t) \leq \frac{1}{2\beta}$ erfüllt ist. Dann lässt sich Korollar 23 anwenden, welches in Teil 2 nötig ist, um Ungleichung (7.8) zu beweisen. Zum Zeitpunkt $t = 0$ haben wir in Annahme (A1) vorausgesetzt, dass die Abweichung zwischen Niedrigrangapproximation und Bestapproximation durch $\|Y(0) - X(0)\|_F \leq \frac{1}{2\beta}$ beschränkt ist. Demnach gilt $L(0) = R(0) \leq \frac{1}{2\beta}$. Des Weiteren sind wir in Annahme (A7) davon ausgegangen, dass $R(t)$ für alle $t \in [0, T]$ durch $\frac{1}{2\beta}$ nach oben beschränkt ist. Nun könnte aber zu einem Zeitpunkt $0 < t^* < T$ die folgende Situation eintreten

$$L(t) \leq \frac{1}{2\beta} \text{ für alle } t \in [0, t^*], \quad L(t^*) = R(t^*) = \frac{1}{2\beta} \quad \text{und} \quad \left. \frac{d}{dt} [L(t)] \right|_{t=t^*} > 0.$$

In dieser Situation würden die Abschätzungen (7.2) und (7.3) keine Gültigkeit mehr besitzen, obgleich $R(t)$ für alle $t \in [0, T]$ durch $\frac{1}{2\beta}$ nach oben beschränkt ist. Es könnte also zu einem hinreichend kleinen $\varepsilon > 0$ ein $0 < \theta \leq T - t^*$ existieren, so dass

$$L(t) \leq \frac{1}{2\beta} + \varepsilon \text{ für alle } t \in [0, t^* + \theta] \quad \text{und} \quad L(t^* + \theta) = \frac{1}{2\beta} + \varepsilon. \quad (7.9)$$

Wir zeigen nun, dass (7.9) nicht eintreten kann, indem wir anstelle von Korollar 23 das allgemeinere Theorem 22 verwenden. Zu diesem Zweck wählen wir als Parameter für Theorem 22

$$\begin{aligned}
\tilde{c} &= -8(2d-2) + \sqrt{8^2(2d-2)^2 + 8(2d-2)} < \frac{1}{2} \quad (\text{da } d < \infty), \\
c_{\mathcal{N}} &= \frac{1}{2} - \tilde{c} + (64d - 62)^{-1} \\
\text{und } \rho_{\mathcal{N}} &= \rho.
\end{aligned}$$

Diese Parameter erfüllen die Nebenbedingungen des Theorems. Darüber hinaus lässt sich das Theorem im Gegensatz zu Korollar 23 für $\varepsilon \leq (\frac{1}{2} - \tilde{c})\rho$ auf den Term $\| [P(Y(t)) - P(X(t))] F(X(t)) \|_F$ anwenden, so dass wir für $t \leq t^* + \theta$ die Abschätzung

$$\begin{aligned}
&\| [P(Y(t)) - P(X(t))] F(X(t)) \|_F \\
&\leq \sum_{\mathcal{N} \in \mathcal{T} \setminus \{\mathcal{N}_0\}} \frac{8}{\rho (1 - \tilde{c} - (\frac{1}{2} + \tilde{c} + (64d - 62)^{-1}))} \|Y(t) - X(t)\|_F \|F(X(t))\|_F \\
&= (32d - 31) \rho^{-1} \|Y(t) - X(t)\|_F \|F(X(t))\|_F \\
&= \beta \|Y(t) - X(t)\|_F \|F(X(t))\|_F
\end{aligned}$$

und damit Ungleichung (7.8) erhalten. Nun gilt aber gemäß Teil 2 des Beweises die Ungleichung $L(t) \leq R(t)$ und damit $L(t) \leq \frac{1}{2\beta}$. Dies ist ein Widerspruch zu (7.9). ■

Die *a posteriori* Fehleranalyse in Theorem 25 lässt sich insbesondere für Anfangswertprobleme modifizieren, welche von einem dissipativen Operator erzeugt werden. So gelangen wir in Korollar 26 an eine Abschätzung, welche unter bestimmten Umständen stärker ist. Die Fehlerabschätzung in Korollar 26 unterscheidet sich in den Voraussetzungen geringfügig von denen in Theorem 25. Aus diesem Grund haben wir die Ergänzung in den Voraussetzungen in grün dargestellt und die Änderungen gegenüber Theorem 25 in rot.

Korollar 26. *Wir setzen auf kompakten Zeitintervallen, also für $t \in [0, T]$, die folgenden Annahmen voraus:*

(A1) *Die Abweichung zwischen Niedrigrangapproximation und Bestapproximation zum Zeitpunkt $t = 0$ ist durch*

$$\|Y(0) - X(0)\|_F \leq \frac{1}{2\beta}$$

beschränkt, wobei $\beta = (32d - 31)\rho^{-1}$ wie in Korollar 23 auf Seite 78.

(A2) *$X(t)$ ist stetig differenzierbar.*

(A3) *Es existiert eine Konstante $\mu > 0$, so dass die Abschätzungen*

$$\|F(Y_{ex}(t))\|_F \leq \mu \text{ und } \|F(X(t))\|_F \leq \mu \text{ und } \|F(Y(t))\|_F \leq \mu$$

erfüllt sind.

(A4) *Es existieren Konstanten $\ell > 0$ und $\lambda \in \mathbb{R}$, so dass die Lipschitz-Bedingungen*

$$\|F(Y_{ex}(t)) - F(X(t))\|_F \leq \ell \|Y_{ex}(t) - X(t)\|_F$$

$$\text{und } \langle F(Y(t)) - F(X(t)), Y(t) - X(t) \rangle \leq \lambda \|Y(t) - X(t)\|_F^2$$

erfüllt sind.

(A5) *Die Singulärwerte der Matrifizierung $X^{(\mathcal{N}, \mathcal{N}_C)}$ sind für alle $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ durch*

$$\sigma_{r_{\mathcal{N}}} \left(X^{(\mathcal{N}, \mathcal{N}_C)}(t) \right) \geq \rho > 0$$

nach unten beschränkt.

(A6) *Der Abstand der exakten Lösung $Y_{ex}(t)$ zur Bestapproximation $X(t)$ ist durch*

$$\delta(t) := \|X(t) - Y_{ex}(t)\|_F \leq \frac{1}{2\beta}$$

beschränkt.

(A7) Des Weiteren sei die Abschätzung

$$R(t) := \|Y(0) - X(0)\|_F e^{(5\beta\mu+\lambda)t} + (\ell + 2\mu\beta) \int_0^t \delta(s) e^{(5\beta\mu+\lambda)(t-s)} ds \leq \frac{1}{2\beta}$$

gegeben.

(A8) Darüber hinaus sei das geordnete Paar $(Y(t), X(t))$ für alle $t \in [0, T]$ wegzusammenhängend bezüglich der ODE (6.1).

Dann gelten die Abschätzungen

$$\|Y(t) - X(t)\|_F \leq R(t) \tag{7.10}$$

$$\text{und } \|Y(t) - Y_{ex}(t)\|_F \leq \delta(t) + R(t). \tag{7.11}$$

Auf den ersten Anschein wirkt die Fehleranalyse von Korollar 26 pessimistischer als die von Theorem 25, jedoch kann in Korollar 26 der Parameter λ auch negative Werte annehmen und somit in bestimmten Fällen erheblich bessere Abschätzungen liefern. Das ist insbesondere dann der Fall, wenn das Anfangswertproblem (7.1) von einem dissipativen Operator erzeugt wird.

Wir weisen darauf hin, dass Lubich, Rohwedder, Schneider und Vandereycken in [LRSV13, Theorem 5.3] identische Abschätzungen beweisen.

Beweis von Korollar 26. Auch dieser Beweis lässt sich in drei Teile untergliedern, wobei Teil 1 und Teil 3 analog zu den entsprechenden Teilen aus dem Beweis von Theorem 25 sind und deswegen nicht extra ausformuliert werden. Stattdessen beginnen wir den Beweis damit, lediglich die Resultate dieser Beweisteile anzuführen.

Aus Teil 1 aus dem Beweis von Theorem 25 wissen wir

$$\dot{X}(t) = P(X(t))\dot{Y}_{ex}(t) + D(t) \tag{7.12}$$

mit $D(t) = P'(X(t)) \cdot (Y_{ex}(t) - X(t))\dot{X}(t)$ und $\|D(t)\|_F \leq 2\mu\beta\delta(t)$.

Mit den Argumenten aus Teil 3 von Theorem 25 lässt sich auch hier zeigen, dass es legitim ist mit den Voraussetzungen des Korollars anzunehmen, dass die linke Seite von (7.10) durch

$$L(t) := \|Y(t) - X(t)\|_F \leq \frac{1}{2\beta}$$

beschränkt ist.

Subtrahieren wir (7.12) von $\dot{Y}(t) = P(Y(t))F(Y(t))$, so erhalten wir

$$\begin{aligned}
& \dot{Y}(t) - \dot{X}(t) \\
&= P(Y(t))F(Y(t)) - P(X(t))\dot{Y}_{ex}(t) - D(t) \\
&= [P(Y(t)) - P(X(t))]F(X(t)) + P(X(t))[F(X(t)) - F(Y_{ex}(t))] \\
&\quad + P(Y(t))(F(Y(t)) - F(X(t))) - D(t) \\
&= [P(Y(t)) - P(X(t))]F(X(t)) + P(X(t))[F(X(t)) - F(Y_{ex}(t))] \quad (7.13) \\
&\quad + [F(Y(t)) - F(X(t))] - P^\perp(Y(t))[F(Y(t)) - F(X(t))] - D(t).
\end{aligned}$$

Entgegen zu Teil 2 im Beweis von Theorem 25 nutzen wir in (7.13) die Identität

$$P(Y(t)) = I - P^\perp(Y(t)).$$

Aus Teil 2 im Beweis von Theorem 25 wissen wir

$$\begin{aligned}
\langle [P(Y(t)) - P(X(t))]F(X(t)), Y(t) - X(t) \rangle &\leq \beta \mu L^2(t), \\
\langle P(X(t))[F(X(t)) - F(Y_{ex}(t))], Y(t) - X(t) \rangle &\leq \ell \delta(t) L(t), \\
\text{und } \langle D(t), Y(t) - X(t) \rangle &\leq 2 \mu \beta \delta(t) L(t).
\end{aligned}$$

Darüber hinaus gilt mit Annahme (A4)

$$\langle F(Y(t)) - F(X(t)), Y(t) - X(t) \rangle \leq \lambda L^2(t)$$

und mit Annahme (A3) und Korollar 23

$$\begin{aligned}
& \langle P^\perp(Y(t))[F(Y(t)) - F(X(t))], Y(t) - X(t) \rangle \\
&= \langle F(Y(t)) - F(X(t)), P^\perp(Y(t))Y(t) - X(t) \rangle \\
&\leq \|F(Y(t)) - F(X(t))\|_F \|P^\perp(Y(t))Y(t) - X(t)\|_F \\
&\leq (\|F(Y(t))\|_F + \|F(X(t))\|_F) 2 \beta \|Y(t) - X(t)\|_F^2 \\
&\leq 4 \beta \mu L^2(t).
\end{aligned}$$

Damit erhalten wir die Ungleichung

$$\begin{aligned}
L(t)\dot{L}(t) &= \langle \dot{Y}(t) - \dot{X}(t), Y(t) - X(t) \rangle \\
&\leq \beta \mu L^2(t) + \ell \delta(t) L(t) + \lambda L^2(t) + 4 \beta \mu L^2(t) + 2 \mu \beta \delta(t) L(t)
\end{aligned}$$

und schließen auf

$$\dot{L}(t) \leq (5\beta\mu + \lambda) L(t) + (\ell + 2\mu\beta) \delta(t).$$

Wenden wir nun das Lemma von Gronwall an, so erhalten wir (7.10). Analog zu Teil 2 aus dem Beweis von Theorem 25 folgt (7.11) aus (7.10) durch das Anwenden der Dreiecksungleichung auf $\|Y(t) - X(t) + X(t) - Y_{ex}(t)\|_F$. \blacksquare

In Kapitel 4 haben wir nicht nur Formeln entwickelt, um die Lösung von hochdimensionalen Anfangswertproblemen im hierarchischen Tucker-Format anzunähern, sondern auch eine Möglichkeit angeführt, einen zeitabhängigen Tensor $A(t) \in \mathbb{R}^{N_1 \times \dots \times N_d}$ in der Mannigfaltigkeit $\mathcal{M} = \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ zu approximieren. Zu diesem Zweck haben wir die Approximation $Y(t) \approx A(t)$ als Lösung des Anfangswertproblems

$$\dot{Y}(t) = P(Y(t))\dot{A}(t), \quad Y(0) = Y_0 \in \mathcal{M}, \quad t \geq 0$$

definiert. In Korollar 27 leiten wir eine Abschätzung für den Fehler der zeitabhängigen Approximation her. Unser Vorgehen orientiert sich an den Techniken von Koch und Lubich in [KL10, Theorem 4.1].

Im Folgenden bezeichne $X(t)$ die Bestapproximation von $A(t)$ in der Menge $\mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$, das heißt zu jedem Zeitpunkt $t \in [0, T]$ gilt

$$\|X(t) - A(t)\|_F \leq \|\hat{X} - A(t)\|_F \quad \text{für alle } \hat{X} \in \mathcal{M}.$$

Korollar 27. *Wir setzen auf kompakten Zeitintervallen, also für $t \in [0, T]$, die folgenden Annahmen voraus:*

(A1) *Die Abweichung zwischen Niedrigrangapproximation und Bestapproximation zum Zeitpunkt $t = 0$ ist durch*

$$d_0 := \|Y(0) - X(0)\|_F \leq \frac{1}{2\beta}$$

beschränkt, wobei $\beta := (32d - 31)\rho^{-1}$ wie in Korollar 23 auf Seite 78.

(A2) *$X(t)$ ist stetig differenzierbar.*

(A3) *Es existiert eine Konstante $\mu > 0$, so dass die Abschätzung $\|\dot{A}(t)\|_F \leq \mu$ erfüllt ist.*

(A4) *Die Singulärwerte der Matrifizierung $X^{(\mathcal{N}, \mathcal{N}_C)}$ sind für alle $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ durch*

$$\sigma_{r_{\mathcal{N}}}\left(X^{(\mathcal{N}, \mathcal{N}_C)}(t)\right) \geq \rho > 0$$

nach unten beschränkt.

(A5) *Der Abstand des zeitabhängigen Tensors $A(t)$ zur Bestapproximation $X(t)$ ist durch*

$$\delta(t) := \|X(t) - A(t)\|_F \leq \frac{1}{2\beta}$$

beschränkt.

Dann gelten die Abschätzungen

$$\|Y(t) - X(t)\|_F \leq d_0 e^{\beta\mu t} + 2\mu\beta \int_0^t \delta(s) e^{\beta\mu(t-s)} ds \quad (7.14)$$

$$\text{und } \|Y(t) - A(t)\|_F \leq \delta(t) + d_0 e^{\beta\mu t} + 2\mu\beta \int_0^t \delta(s) e^{\beta\mu(t-s)} ds. \quad (7.15)$$

Ein identisches Resultat wird von Lubich, Rohwedder, Schneider und Vandereycken in [LRSV13, Theorem 5.1] hergeleitet.

Beweis von Korollar 27. Der Beweis lässt sich in zwei Teile untergliedern. Setzen wir $Y_{ex}(t) = A(t)$, so ist Teil 1 analog zu Teil 1 des Beweises von Theorem 25 und es gilt

$$\dot{X}(t) = P(X(t))\dot{A}(t) + D(t) \quad (7.16)$$

mit $D(t) = P'(X(t)) \cdot (A(t) - X(t))\dot{X}(t)$ und $\|D(t)\|_F \leq 2\mu\beta\delta(t)$.

Betrachten wir nun die Differenz zwischen (7.16) und $\dot{Y}(t) = P(Y(t))\dot{A}(t)$, so führt das auf die Gleichung

$$\begin{aligned} \dot{Y}(t) - \dot{X}(t) &= P(Y(t))\dot{A}(t) - P(X(t))\dot{A}(t) - D(t) \\ &= (P(Y(t)) - P(X(t)))\dot{A}(t) - D(t). \end{aligned} \quad (7.17)$$

Darüber hinaus gelten unter Verwendung von Korollar 23 und Annahme (A3) die Abschätzungen

$$\begin{aligned} &\langle (P(Y(t)) - P(X(t)))\dot{A}(t), Y(t) - X(t) \rangle \\ &\leq \| (P(Y(t)) - P(X(t)))\dot{A}(t) \|_F \|Y(t) - X(t)\|_F \\ &\leq \beta\mu \|Y(t) - X(t)\|_F^2 \end{aligned} \quad (7.18)$$

und

$$\begin{aligned} \langle D(t), Y(t) - X(t) \rangle &\leq \|D(t)\|_F \|Y(t) - X(t)\|_F \\ &\leq 2\mu\beta\delta(t) \|Y(t) - X(t)\|_F. \end{aligned} \quad (7.19)$$

Mittels (7.17), sowie den Abschätzungen (7.18) und (7.19) schließen wir auf

$$\begin{aligned} &\left[\frac{d}{dt} \|Y(t) - X(t)\|_F \right] \|Y(t) - X(t)\|_F \\ &= \langle \dot{Y}(t) - \dot{X}(t), Y(t) - X(t) \rangle \\ &= \langle [P(Y(t)) - P(X(t))]\dot{A}(t), Y(t) - X(t) \rangle + \langle D(t), Y(t) - X(t) \rangle \\ &\leq \beta\mu \|Y(t) - X(t)\|_F^2 + 2\mu\beta\delta(t) \|Y(t) - X(t)\|_F \end{aligned}$$

und es folgt

$$\frac{d}{dt} \|Y(t) - X(t)\|_F \leq \beta\mu \|Y(t) - X(t)\|_F + 2\mu\beta\delta(t).$$

Unter Gebrauch des Lemmas von Gronwall, erhalten wir Abschätzung (7.14). Ungleichung (7.15) folgt unmittelbar aus (7.14) durch das Anwenden der Dreiecksungleichung auf den Term $\|Y(t) - A(t) + A(t) - X(t)\|_F$.

■

8. Zeitabhängige Approximation der Black-Scholes-Gleichung im \mathcal{H} -Tucker-Format

Die Black-Scholes-Gleichung zählt zu den wohl berühmtesten Gleichungen in der Mathematik und geht zurück auf die Wirtschaftswissenschaftler Fischer Black und Myron Samuel Scholes in [BS73]. Unabhängig von Black und Scholes wurden zeitgleich ähnliche Resultate von Robert Carhart Merton in [Mer73] veröffentlicht. Im Jahre 1997 wurden Scholes und Merton für ihre Errungenschaften im Bereich der Wirtschaftswissenschaften durch den Nobelpreis geehrt, Black starb zwei Jahre zuvor.

8.1 Die Black-Scholes-Gleichung

Die Black-Scholes-Gleichung ist eine partielle Differentialgleichung deren Lösung unter idealisierten Annahmen den Wert einer Option angibt. Im Zuge dieser Arbeit beschränken wir uns auf sogenannte europäische Put-Optionen. Eine Put-Option gibt dem Käufer der Option das Recht, aber nicht die Verpflichtung, einen Basiswert zu einem festgelegten Zeitpunkt T und zu einem festgelegten Preis K , dem sogenannten Ausübungspreis, zu verkaufen. Dabei kann ein Basiswert beispielsweise ein Rohstoff, eine Aktie oder eine Währung sein. Durch den Kauf einer solchen Option kann sich der Käufer gegen Kurschwankungen des Basiswerts absichern. Zur Herleitung der Gleichung haben Black und Scholes die folgenden Modellannahmen vorausgesetzt:

- Der Kurs S des Basiswerts folgt einer geometrischen Brownschen Bewegung, das heißt

$$S_t = S_0 \exp \left(\left(\mu - \frac{\sigma^2}{2} \right) t + \sigma W_t \right)$$

mit $\mu \in \mathbb{R}$ als Driftparameter, $\sigma \in \mathbb{R}$ als Volatilität und W_t als ein Wiener-Prozess.

- Für Kredite und Geldanlagen existiert ein einziger risikoneutraler Zinssatz $r \geq 0$.
- Auf den Basiswert erfolgen keine Dividendenzahlungen.
- Der Markt ist frei von Arbitrage, liquide und friktionslos.
- Der Basiswert kann zu beliebigen Zeitpunkten gehandelt werden und ist beliebig teilbar.
- Leerverkäufe sind möglich.

Die Modellannahmen werden in [GJ10, Abschnitt 4.2.1], [Sey09, Abschnitt 1.2] und [KK01, Abschnitt II.2] genauer erklärt. Um eine Gleichung für den Wert der Option zu erhalten, wandten Black und Scholes eine Duplikationsstrategie an. Diese Strategie sieht vor, dass der Wert der Option dem Wert eines Portfolios gleichgesetzt wird; ein Portfolio, welches aus Anteilen des Basiswerts, sowie aus Anteilen einer festverzinslichen Geldanlage besteht und einer selbst-finanzierenden Handelsstrategie unterworfen ist. Unter diesen Voraussetzungen ist es unter Verwendung der Itô-Formel möglich, die **Black-Scholes-Gleichung**

$$\partial_t V(t, S) = -\frac{1}{2} \sigma^2 S^2 \partial_S^2 V(t, S) - r S \partial_S V(t, S) + r V(t, S), \quad t \in [0, T], \quad S \geq 0$$

herzuleiten¹. Die Lösung $V(t, S)$ gibt den Wert der Option zur Zeit t in Abhängigkeit vom Basiswert an. Am Verfallstag der Option, also zum Zeitpunkt $t = T$, ist der Wert der Option durch die sogenannte **Auszahlungsfunktion**

$$V(T, S) = (K - S)^+ = \begin{cases} K - S & \text{für } K \geq S \\ 0 & \text{für } K < S \end{cases}$$

gegeben. Zusammen mit der Auszahlungsfunktion definiert die Black-Scholes-Gleichung ein Endwertproblem. Das bedeutet, dass die partielle Differentialgleichung „rückwärts“ in der Zeit gelöst werden muss.

Neben Optionen mit einem einzigen Basiswert gibt es auch sogenannte **Basket-Optionen**, deren Wert von den Kursen mehrerer Basiswerte abhängt. Die Option gibt dann dem Käufer beispielsweise das Recht, aber nicht die Verpflichtung, ein Aktienpaket zu einem festgelegten Preis zu verkaufen. Um den Preis einer europäischen Basket-Put-Option anhand der Lösung einer partiellen Differentialgleichung zu bestimmen, ist es nötig, die Modellannahmen anzupassen. Liegen d Basiswerte vor, so setzen wir voraus, dass sich deren Kurse gemäß der multivariaten geometrischen Brownschen Bewegung entwickeln, siehe [Sey09, Abschnitt 6.2]. Unter diesen Modellannahmen lässt sich das Vorgehen zur Herleitung einer partiellen Differentialgleichung auf Basket-Optionen erweitern und es folgt, dass der Preis

¹Nähere Informationen zur Herleitung der Gleichung sind in [KK01, Abschnitt III.3] und [GJ10, Abschnitt 4.2] zu finden. Die Itô-Formel, sowie dessen Herleitung werden in [Øks07, Abschnitt 4.1 und 4.2] erklärt.

$V(t, S_1, \dots, S_d)$ einer Basket-Option für alle $(t, S_1, \dots, S_d) \in [0, T) \times [0, \infty)^d$ der **multivariaten Black-Scholes-Gleichung**

$$\begin{aligned} \partial_t V(t, S_1, \dots, S_d) &= -\frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \sigma_i \sigma_j \rho_{ij} S_i S_j \partial_{S_i} \partial_{S_j} V(t, S_1, \dots, S_d) \\ &\quad - r \sum_{i=1}^d S_i \partial_{S_i} V(t, S_1, \dots, S_d) \\ &\quad + r V(t, S_1, \dots, S_d) \end{aligned}$$

genügt, wobei ρ_{ij} die Einträge der sogenannten Korrelationsmatrix

$$\Sigma = (\rho_{ij})_{i,j=1}^d \in \mathbb{R}^{d \times d}$$

sind. Die Korrelationsmatrix ist stets symmetrisch. Bei Basket-Optionen lautet die Endbedingung nun

$$V(T, S_1, \dots, S_d) = \left(K - \sum_{i=1}^d \alpha_i S_i \right)^+$$

mit Koeffizienten $\alpha_i > 0$. Durch die Transformation $\tau = T - t$ erhalten wir für $(t, S_1, \dots, S_d) \in (0, T] \times [0, \infty)^d$ die partielle Differentialgleichung

$$\left. \begin{aligned} \partial_\tau V(\tau, S_1, \dots, S_d) &= \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \sigma_i \sigma_j \rho_{ij} S_i S_j \partial_{S_i} \partial_{S_j} V(\tau, S_1, \dots, S_d) \\ &\quad + r \sum_{i=1}^d S_i \partial_{S_i} V(\tau, S_1, \dots, S_d) \\ &\quad - r V(\tau, S_1, \dots, S_d) \end{aligned} \right\} \quad (\text{CP})$$

mit Anfangsbedingung

$$V(0, S_1, \dots, S_d) = \left(K - \sum_{i=1}^d \alpha_i S_i \right)^+ \quad \text{für } (S_1, \dots, S_d) \in [0, \infty)^d.$$

Auf dem Rand

$$\partial\Omega = \left\{ (S_1, \dots, S_d) \in [0, \infty)^d : S_i = 0 \text{ für mindestens ein } i \in \{1, \dots, d\} \right\}$$

genügt die partielle Differentialgleichung in (CP) den sogenannten natürlichen Randbedingungen. Das bedeutet vereinfacht, dass es nicht nötig ist, explizite Randbedingungen zu setzen; die Raumableitungen, welche Randbedingungen erfordern würden, werden in der Differentialgleichung für ein $i \in \{1, \dots, d\}$ mit $S_i = 0$ multipliziert und verschwinden somit auf dem Rand. Genauere Informationen zum Lösungsverhalten von (CP) auf dem Rand sind in [PA05] zu finden.

Aus der Theorie über parabolische Differentialgleichungen folgt die Existenz und Eindeu-

tigkeit einer Lösung, welche für alle $(t, S_1, \dots, S_d) \in (0, T] \times [0, \infty)^d$ einmal stetig differenzierbar in der Zeit und zweimal stetig differenzierbar in den Variablen S_1, \dots, S_d ist, siehe [Fri08] und [PA05, Abschnitt 2.3]. Darüber hinaus zeigen Pironneau und Achdou in [PA05, Abschnitt 2.3.3] die Existenz und Eindeutigkeit einer schwachen Lösung in einem dafür geeigneten Funktionenraum.

Eine ausführliche Herleitung der multivariaten Black-Scholes-Gleichung findet man in [Sey09, Abschnitt 6.2] und [GJ10, Abschnitt 4.5.4].

8.2 Formulierung des diskreten Anfangswertproblems

Zur numerischen Lösung von (CP) existieren verschiedene Verfahren. Unser Vorgehen orientiert sich an der vertikalen Linienmethode, siehe [HB09, Abschnitt 98]. Diese sieht vor, dass erst der Ortsbereich diskretisiert wird und anschließend die Zeit. Ausgehend von der parabolischen Differentialgleichung (CP) ändern wir in verschiedenen Approximationsschritten die Problemstellung derart, dass wir zuletzt ein diskretes Anfangswertproblem erhalten.

Ehe wir eine diskrete Darstellung für den Ortsbereich finden, ist es nötig, das Gebiet zu beschränken. Für $0 < S_{max} < \infty$ betrachten wir die parabolische Differentialgleichung (CP) fortan in der halboffenen Menge

$$\Omega = [0, S_{max})^d.$$

Durch die Restriktion entstehen künstliche Ränder. Um die Eindeutigkeit einer Lösung auf Ω sicherzustellen, ist es erforderlich, die künstlichen Ränder mit Randbedingungen zu versehen. Dadurch, dass die Lösung von (CP) für alle $\tau \geq 0$ und $i \in \{1, \dots, d\}$ die Gleichung

$$\lim_{S_i \rightarrow \infty} V(\tau, S_1, \dots, S_d) = 0$$

erfüllt, siehe [Sey09, Abschnitt 4.4], fordern wir auf dem künstlichen Rand

$$\partial\Omega = \left\{ (S_1, \dots, S_d) \in [0, S_{max}]^d : S_i = S_{max} \text{ für mindestens ein } i \in \{1, \dots, d\} \right\}$$

homogene Dirichlet-Randbedingungen und erhalten auf diese Weise das Anfangs- Randwertproblem

$$\left. \begin{aligned}
\partial_\tau \tilde{V}(\tau, S_1, \dots, S_d) &= \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \sigma_i \sigma_j \rho_{ij} S_i S_j \partial_{S_i} \partial_{S_j} \tilde{V}(\tau, S_1, \dots, S_d) \\
&\quad + r \sum_{i=1}^d S_i \partial_{S_i} \tilde{V}(\tau, S_1, \dots, S_d) \\
&\quad - r \tilde{V}(\tau, S_1, \dots, S_d) \quad \text{für } \tau > 0, (S_1, \dots, S_d) \in \Omega, \\
\tilde{V}(0, S_1, \dots, S_d) &= \left(K - \sum_{i=1}^d \alpha_i S_i \right)^+ \quad \text{für } (S_1, \dots, S_d) \in \Omega, \\
\tilde{V}(\tau, S_1, \dots, S_d) &= 0 \quad \text{für } \tau \geq 0, (S_1, \dots, S_d) \in \partial\Omega.
\end{aligned} \right\} \quad (\text{ARP})$$

Die Lösung $\tilde{V}(\tau, S_1, \dots, S_d)$ von (ARP) ist eine Approximation der Lösung $V(\tau, S_1, \dots, S_d)$ von (CP). Kangro und Nicolaidis geben in [KN01] in Abhängigkeit von der Variable S_{max} und anderen Parametern eine Abschätzung für die Güte dieser Approximation an.

Als nächsten Approximationsschritt wählen wir die Finite-Differenzen-Methode. Mit der Finite-Differenzen-Methode sind wir in der Lage, die Lösung von (ARP) durch die Lösung eines diskreten Anfangswertproblems zu approximieren, siehe [LT05, Kapitel 9]. Zu diesem Zweck definieren wir für ein $N \in \mathbb{N}$ mit $N > 1$ und $h = \frac{S_{max}}{N}$ das Gitter

$$\Omega_N = \left\{ (x_1, \dots, x_d) : x_i \in \{0, h, \dots, (N-1)h\} \text{ für } i \in \{1, \dots, d\} \right\}.$$

Während die Variablen S_i für $i \in \{1, \dots, d\}$ beliebige Werte im Intervall $[0, S_{max})$ annehmen kann, nimmt die Variable x_i nur diskrete Werte an. Für einen Tensor beziehungsweise eine multivariate Funktion $Y : \Omega_N \rightarrow \mathbb{R}$ definieren wir die Differenzenoperatoren

$$\tilde{\partial}_i Y(x_1, \dots, x_d) = \begin{cases} 0 & \text{für } x_i = 0, \\ \frac{Y(x_1, \dots, x_i + h, \dots, x_d) - Y(x_1, \dots, x_i - h, \dots, x_d)}{2h} & \text{für } x_i \in \{h, \dots, (N-2)h\}, \\ \frac{-Y(x_1, \dots, x_i - h, \dots, x_d)}{2h} & \text{für } x_i = (N-1)h \end{cases}$$

und

$$\tilde{\partial}_i^2 Y(x_1, \dots, x_d) = \begin{cases} 0 & \text{für } x_i = 0, \\ \frac{Y(x_1, \dots, x_i + h, \dots, x_d) - 2Y(x_1, \dots, x_i, \dots, x_d) + Y(x_1, \dots, x_i - h, \dots, x_d)}{h^2} & \text{für } x_i \in \{h, \dots, (N-2)h\}, \\ \frac{-2Y(x_1, \dots, x_i, \dots, x_d) + Y(x_1, \dots, x_i - h, \dots, x_d)}{h^2} & \text{für } x_i = (N-1)h. \end{cases}$$

Die Differenzenoperatoren $\tilde{\partial}_i$ und $\tilde{\partial}_i^2$ approximieren die partielle Ableitungen ∂_i und ∂_i^2 auf dem diskreten Gebiet Ω_N , siehe [KA00, Abschnitt 1.2].

Sei $\mathcal{A} : \mathbb{R}^{N \times \dots \times N} \rightarrow \mathbb{R}^{N \times \dots \times N}$ ein linearer Operator mit

$$\begin{aligned} \mathcal{A} Y(x_1, \dots, x_d) &= \frac{1}{2} \sum_{j=1}^d \sigma_j^2 x_j^2 \tilde{\partial}_j^2 Y(x_1, \dots, x_d) \\ &\quad + \sum_{j_1=1}^d \sum_{j_2=1}^{j_1-1} \sigma_{j_1} \sigma_{j_2} \rho_{j_1 j_2} x_{j_1} x_{j_2} \tilde{\partial}_{j_1} \tilde{\partial}_{j_2} Y(x_1, \dots, x_d) \\ &\quad + r \sum_{j=1}^d x_j \tilde{\partial}_j Y(x_1, \dots, x_d) \\ &\quad - r Y(x_1, \dots, x_d), \end{aligned}$$

dann approximiert unter geeigneten Regularitätsbedingungen die Lösung des Anfangswertproblems

$$\left. \begin{aligned} \partial_\tau Y_{\text{AWP}}(\tau) &= \mathcal{A} Y_{\text{AWP}}(\tau) \\ Y_{\text{AWP}}(0) &= \left(K - \sum_{j=1}^d \alpha_j x_j \right)^+ \end{aligned} \right\} \quad (\text{AWP})$$

für $\tau \geq 0$ die Lösung von (ARP). Die Güte der Approximation $Y_{\text{AWP}}(\tau)$ ist aus der Theorie über die Numerik partieller Differentialgleichungen bekannt, siehe [LT05, Abschnitt 9.2]. Alternativ könnte man auch mit der Finite-Elemente-Methode aus dem Anfangs-Randwertproblem (ARP) ein Anfangswertproblem erhalten, siehe [HRSW13, Abschnitt 8.4.2].

Aus naiver Sicht würde man nun das Anfangswertproblem (AWP) einem numerischen Löser für gewöhnliche Differentialgleichungen übergeben, wie beispielsweise `ode45` in Matlab. Dadurch, dass die gewöhnliche Differentialgleichung aus der Diskretisierung einer partiellen Differentialgleichung hervorgeht, wächst die Anzahl an Unbekannten bei der Berechnung

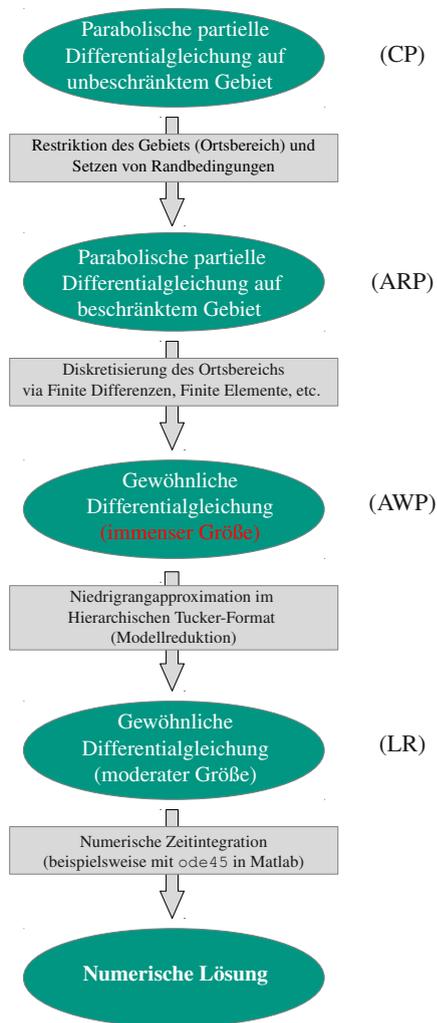
der Zeitableitung exponentiell mit der Dimension, so dass Standardverfahren bereits für kleine Werte von d scheitern. Aus diesem Grund betrachten wir die Niedrigrangapproximation in der Menge $\mathcal{M} = \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$. Anstelle von (AWP) betrachten wir auf einem kompakten Zeitintervall das Anfangswertproblem

$$\dot{Y}_{\text{HT}}(\tau) = P(Y_{\text{HT}}(\tau)) \mathcal{A} Y_{\text{HT}}(\tau) \quad , \quad Y(0) = Y_0 \in \mathcal{M}. \quad (\text{LR})$$

Dabei ist $P(Y_{\text{HT}})$ der orthogonale Projektor auf den Tangentialraum $\mathcal{T}_{Y_{\text{HT}}}\mathcal{M}$. Um eine Approximation der Auszahlungsfunktion in der Menge $\mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ zu erhalten, welche dem Anfangswertproblem (LR) als Startwert Y_0 dient, eignet sich Algorithmus 1 auf Seite 51. Verwenden wir zur Berechnung der Zeitableitung in (LR) die Bewegungsgleichungen aus Theorem 11 auf Seite 36, so ist ein Gleichungssystem mit

$$\sum_{\substack{\mathcal{N} \in \mathcal{I}(T) \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} r_{\mathcal{N}} r_{\mathcal{N}_1} r_{\mathcal{N}_2} + \sum_{\{j\} \in \mathcal{L}(T)} r_{\{j\}} N_j$$

Unbekannten zu lösen. Die Berechnung der Zeitableitung in (AWP) erfordert das Auswerten einer Funktion, deren Argument ein Tensor mit N^d Einträgen ist. Ungeachtet der Nichtlinearität wird durch die Niedrigrangapproximation eine Situation erreicht, in der eine numerische Lösung mit gängiger Computer-Hardware möglich ist, wohingegen das numerische Lösen von (AWP) durch das exponentielle Ansteigen an Unbekannten an mangelndem Computer-Speicher in hohen Dimensionen scheitert. Selbst wenn ausreichend Computer-Speicher vorhanden wäre, so würde mit der Dimension auch der numerische Aufwand zur Zeitintegration exponentiell steigen.



Ausgehend von der multivariaten Black-Scholes-Gleichung haben wir in diesem Abschnitt in verschiedenen Approximationsschritten die Problemstellung so modifiziert, dass wir letztendlich mit (LR) ein diskretes Anfangswertproblem moderater Größe erhalten. Abbildung 8.1 illustriert die einzelnen Schritte unserer Vorgehensweise. Der nächste Abschnitt dieser Arbeit ist der numerischen Lösung einer vierdimensionalen Black-Scholes-Gleichung gewidmet.

Abbildung 8.1:
Approximationsschritte

8.3 Numerische Lösung

Dieser Abschnitt befasst sich mit der numerischen Lösung der Anfangswertprobleme (AWP) und (LR) aus Abschnitt 8.2. Wir betrachten die Anfangswertprobleme in vier Dimensionen auf dem Gitter Ω_{50} mit $S_{max} = 50$ und Gitterweite $h = 1$. Als Parameter verwenden wir die Werte $K = 20$, $r = 0.01$, $\sigma_1 = 0.2$, $\sigma_2 = 0.5$, $\sigma_3 = 0.3$, $\sigma_4 = 0.4$, $\alpha_1 = \dots = \alpha_4 = 1$, sowie die Korrelationsmatrix

$$\Sigma = \begin{pmatrix} 1 & 0.8 & 0.7 & 0.5 \\ 0.8 & 1 & 0.6 & 0.7 \\ 0.7 & 0.6 & 1 & 0.8 \\ 0.5 & 0.7 & 0.8 & 1 \end{pmatrix}.$$

Für die Zeitintegration wurde in sämtlichen Beispielen der Matlab-Integrator `ode45` mit einer absoluten Fehlertoleranz von 10^{-9} und einer relativen Fehlertoleranz von 10^{-3} verwendet. Wir haben uns für den Matlab-Integrator entschieden, weil er aufgrund seiner Schrittweitensteuerung die Genauigkeit der numerischen Approximation zu der vorgegebenen Toleranz gewährleistet.

Referenzlösung

Um die Qualität der Niedrigrangapproximation bewerten zu können, bestimmen wir zunächst eine Referenzlösung für das Anfangswertproblem (AWP). Zu diesem Zweck lösen wir das Anfangswertproblem (AWP) numerisch und erhalten eine Approximation $\tilde{Y}_{\text{AWP}}(\tau)$. Diese Approximation betrachten wir als Referenzlösung. Das Lösen von (AWP) auf dem Zeitintervall $[0, 3]$ nahm mit dem Matlab-Integrator insgesamt 2 Tage und 23 Minuten in Anspruch². Zur Berechnung der Zeitableitung wird eine lineare Funktion ausgewertet, deren Argument ein Tensor mit 6'250'000 Einträgen ist.

²Die numerischen Tests wurden auf einem Computer mit 32 Gigabyte Arbeitsspeicher und acht Prozessorkernen durchgeführt. Die Prozessorkerne sind mit 2.3 Ghz getaktet.

In Abbildung 8.2 sind zu verschiedenen Zeitpunkten zwei Darstellungen der vierdimensionalen numerischen Lösung $\tilde{Y}_{AWP}(\tau)$ von (AWP) zu sehen.

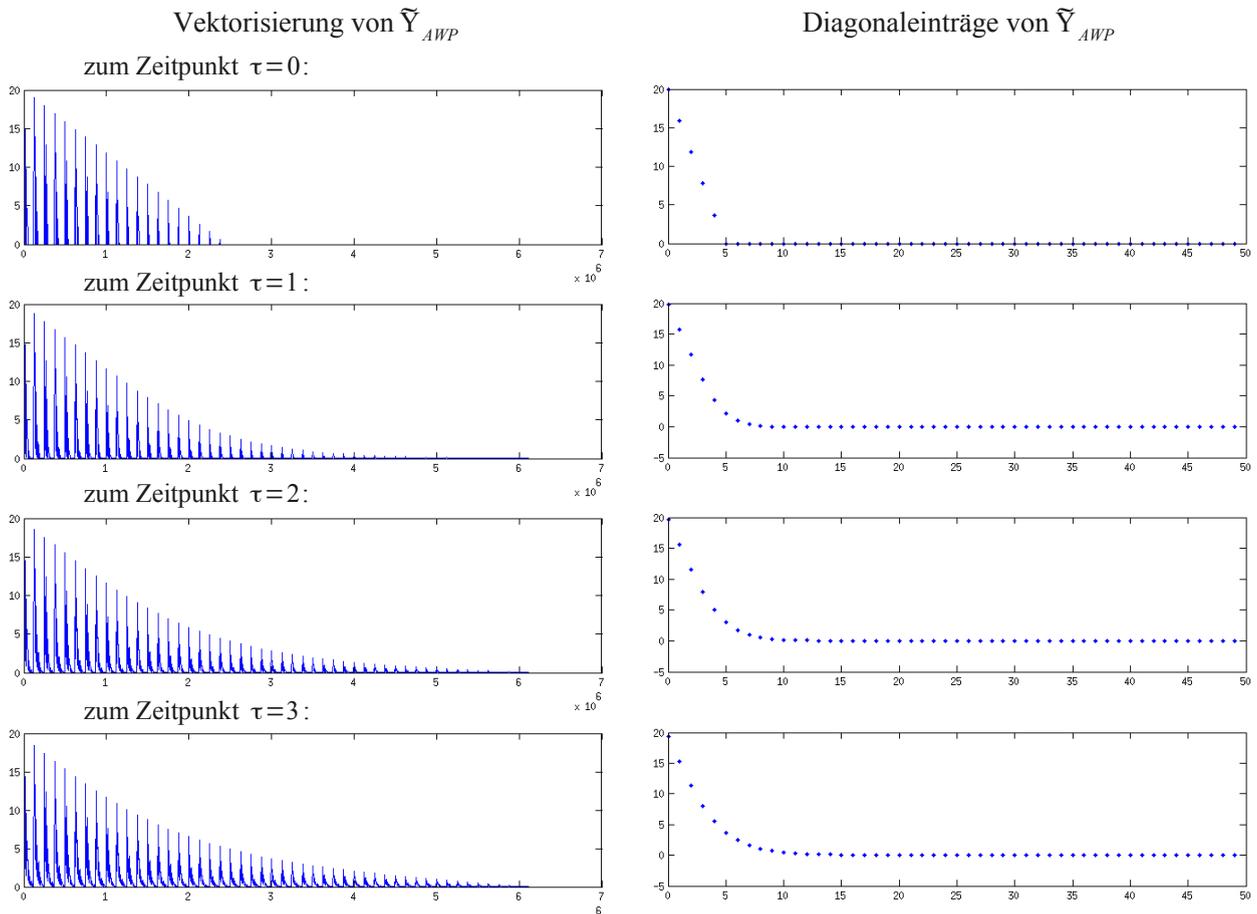


Abbildung 8.2: Numerische Lösung von (AWP)

Die Bilder auf der linken Seite von Abbildung 8.2 zeigen die Vektorisierung der Lösung zu verschiedenen Zeitpunkten. In den Plots auf der rechten Seite sind die „Diagonaleinträge“ des Tensors, also die Einträge

$$\tilde{Y}_{AWP}(\tau, x_1, x_2, x_3, x_4) \quad \text{mit } x_1 = x_2 = x_3 = x_4 = x \text{ und } x \in \{0, \dots, 49\},$$

eingetragen. Zum Zeitpunkt $\tau = 0$ kann man im Plot, welcher die Diagonaleinträge zeigt, deutlich den „Knick“ der Auszahlungsfunktion erkennen. Die vektorisierte Darstellung des Tensors zeigt, dass zu diesem Zeitpunkt die meisten Einträge des Tensors gleich 0 sind. Mit voranschreitender Zeit ist in den Plots sowohl ein Abflachen als auch ein Ausschmieren der Auszahlungsfunktion zu beobachten. Zum Zeitpunkt $\tau = 3$ zeigt der Plot auf der rechten Seite, dass die meisten Diagonaleinträge des Tensors noch immer nahe der 0 sind. Wie in vektorisierter Form zu erkennen ist, gilt dies jedoch nicht für alle Einträge des Tensors.

Niedrigrangapproximation in der Mannigfaltigkeit \mathcal{H} -Tucker $((r_{\mathcal{N}})_{\mathcal{N} \in T})$

Für die Niedrigrangapproximation der Lösung von (AWP) im hierarchischen Tucker-Format, ist es nötig, die Parameter der Darstellung festzulegen. Dazu zählt zum Einen der Dimensionsbaum der Darstellung und zum Anderen der hierarchische Rang. Für die numerischen Beispiele verwenden wir einen perfekten Binärbaum als Dimensionsbaum, siehe [PD08, Abschnitt 3.5.2].

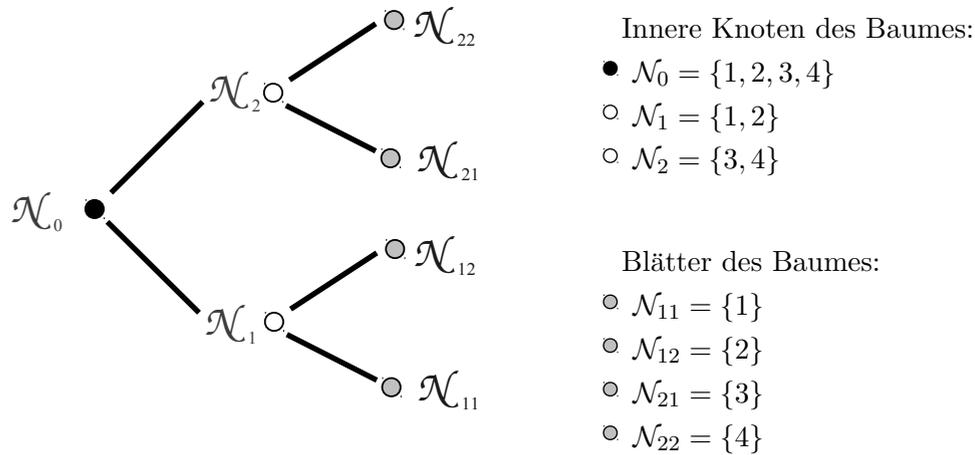


Abbildung 8.3: Dimensionsbaum
des numerischen Beispiels

Wie in Abbildung 8.3 zu erkennen ist, liegen sämtliche Blätter des Dimensionsbaumes auf demselben Level. Vereinfachend nehmen wir für das numerische Beispiel an, dass jeder Knoten des Dimensionsbaumes den gleichen Rang r_{HT} besitzt, also die Gleichung

$$r_{\mathcal{N}} = r_{HT}$$

für alle $\mathcal{N} \in T$ erfüllt. Diesem Rang weisen wir zunächst den Wert $r_{HT} = 13$ zu. Das Tupel $\left((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^{\ell})_{\ell \in \mathcal{L}(T)} \right)$ besteht somit aus insgesamt 7'163 Einträgen.

Abbildung 8.4³ zeigt schematisch die Einträge der Basisfunktionen und Transfertensoren eines Tensors aus der Menge \mathcal{H} -Tucker $((13)_{\mathcal{N} \in T})$. Die Basisfunktionen an den Blättern des Dimensionsbaumes sind darin als die korrespondierenden Basismatrizen dargestellt.

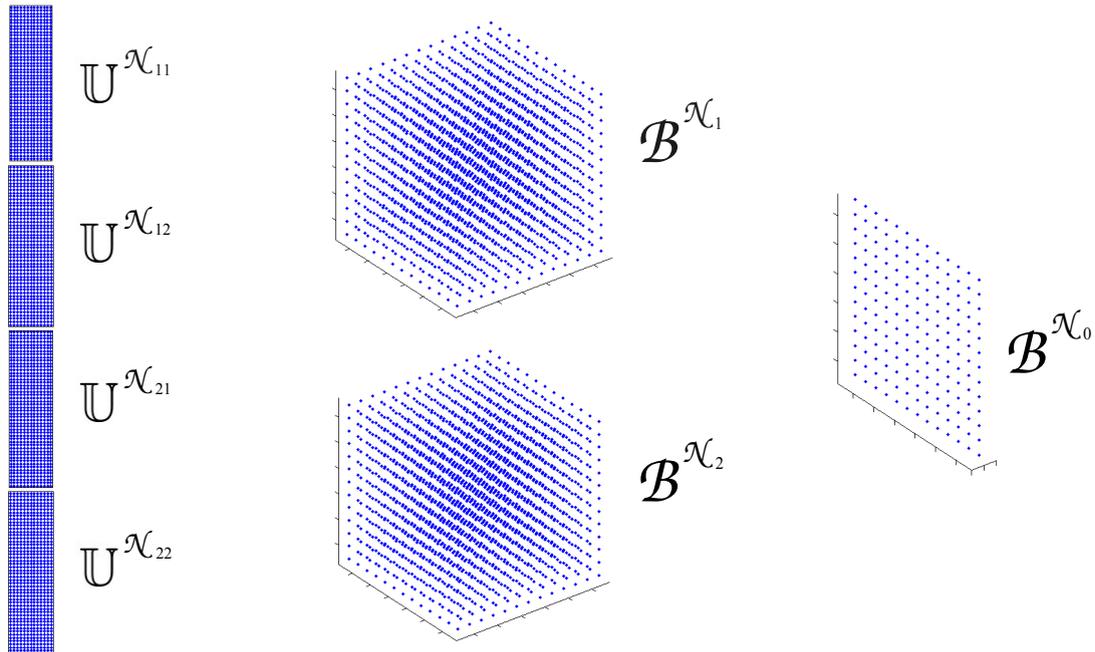


Abbildung 8.4: Tupel $((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^{\ell})_{\ell \in \mathcal{L}(T)})$

Wie in Abbildung 8.4 zu erkennen ist, besteht die Darstellung aus insgesamt sieben Objekten: Vier Basismatrizen, zwei dreidimensionale Transfertensoren und ein zweidimensionaler Transfertensor an der Wurzel des Dimensionsbaumes. Dadurch, dass die meisten Objekte des Tupels $((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^{\ell})_{\ell \in \mathcal{L}(T)})$ lediglich zweidimensional sind, genügen 7'163 Freiheitsgrade zur Darstellung eines vierdimensionalen Tensors mit insgesamt 6'250'000 Einträgen. Das entspricht einer Kompressionsrate von 99.89 %.

Das numerische Lösen von (LR) auf dem Zeitintervall $[0, 3]$ nahm mit dem Matlab-Integrator 2 Stunden und 12 Minuten in Anspruch. Im Vergleich zur numerischen Lösung von (AWP) bringt diese Berechnung eine Zeitersparnis von 95.45 %. In den numerischen Tests zeigt sich, dass die Orthonormalität der Basisfunktionen zu der gewählten Fehlertoleranz in der numerischen Approximation nahezu vollständig erhalten bleibt, vergleiche mit Lemma 12 auf Seite 37. Verwendet man ein Verfahren ohne Schrittweitensteuerung, so empfehlen wir entweder nach jedem Zeitintegrationsschritt die Orthonormalisierung der Basisfunktionen mit Algorithmus 3 aus [Gra10] oder ein Gauß-Runge-Kutta-Verfahren. Die numerische Approximation eines Gauß-Runge-Kutta-Verfahrens erhält per se die Orthonormalität in der numerischen Approximation, siehe [HLW06, Abschnitt IV.9.1] und [Coo87].

³Abbildung 8.4 wurde unter Verwendung der `htucker` toolbox [TK12] angefertigt, siehe auch [KT12].

Nachdem das numerische Lösen von (LR) sowohl einen deutlich geringeren Speicherbedarf, als auch eine deutlich geringere Rechenzeit mit sich bringt, stellt sich die Frage, wie gut die numerische Lösung $\tilde{Y}_{HT}(\tau)$ von (LR) die Referenzlösung $\tilde{Y}_{AWP}(\tau)$ approximiert. Abbildung 8.5 zeigt zu verschiedenen Zeitpunkten die Abweichung zwischen $\tilde{Y}_{HT}(\tau)$ und $\tilde{Y}_{AWP}(\tau)$ im Absolutbetrag.

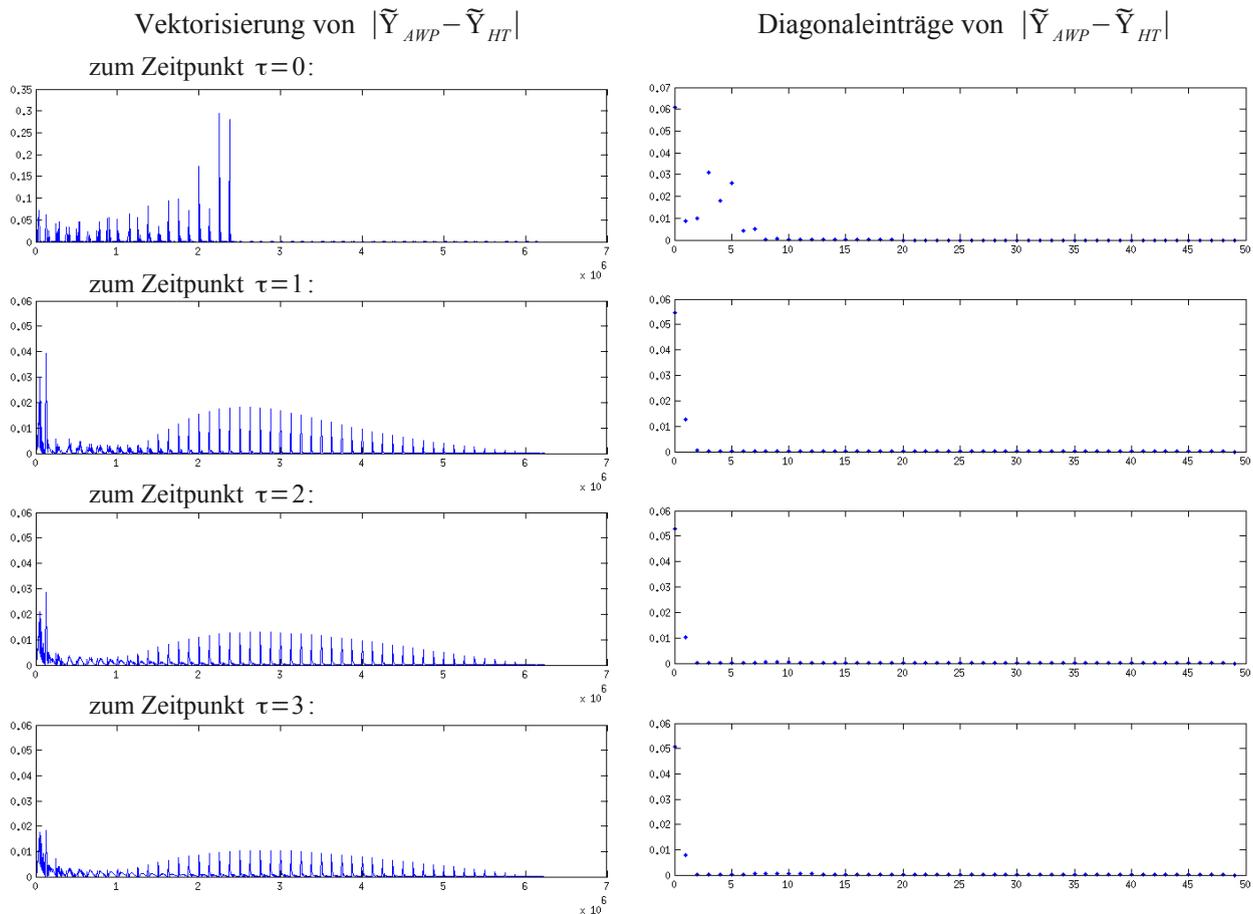


Abbildung 8.5: Abweichung zwischen $\tilde{Y}_{HT}(\tau)$ und $\tilde{Y}_{AWP}(\tau)$

Analog zu Abbildung 8.2 zeigen die Plots auf der linken Seite von Abbildung 8.5 die Abweichungen in vektorisierter Form, wohingegen in den Plots auf der rechten Seite die Abweichungen in den Diagonaleinträgen der Tensoren zu erkennen ist. Die Abweichung zum Zeitpunkt $\tau = 0$ resultiert aus der Approximation der Auszahlungsfunktion durch Algorithmus 1 auf Seite 51. Unmittelbar nach Beginn der Zeitintegration ist ein rapides Abfallen des Fehlers zu beobachten. Wie in der vektorisierten Darstellung zu erkennen ist, liegt die maximale Abweichung zum Zeitpunkt $\tau = 0$ noch in der Größenordnung von 0.35, während die maximale Abweichung für die anderen Zeitpunkte in der Größenordnung von 0.06 liegt. Betrachten wir die Differenz der Tensoren in den Diagonaleinträgen, so stellen wir fest, dass der Fehler für sämtliche Zeitpunkte im Bereich nahe des Ursprungs lokali-

siert ist. Die Ursache hierfür sehen wir darin, dass die Referenzlösung im Ursprung für alle Zeitpunkte den größten Eintrag besitzt.

Weil sich durch das Erhöhen des hierarchischen Rangs ebenfalls die Anzahl an Freiheitsgraden in der Darstellung erhöht, würde man annehmen, dass die Approximation zu einem Rang $r_{HT} > 13$ zu einer Approximation mit einer höheren Güte führen würde. In den Simulationen stellen wir jedoch fest, dass für $r_{HT} = 14$ im Intervall $[0, 3]$ ein Zeitpunkt τ^* auftritt, an dem die numerische Lösung von (LR) instabil wird. Zu jenem Zeitpunkt erniedrigt sich an einen Knoten \mathcal{N}^* der Rang der Single-Hole-Matrix $\mathbb{S}^{\mathcal{N}^*}(\tau^*)$ und damit der hierarchische Rang der Darstellung. Dies führt dazu, dass in den Bewegungsgleichungen die Lastmatrix $M^{\mathcal{N}^*}(\tau^*)$ singular wird. Unter diesen Bedingungen ist es nicht mehr möglich, die Zeitableitung anhand der Formeln aus Theorem 11 zu bestimmen. Um diese Problematik zu vermeiden, empfiehlt es sich, den hierarchischen Rang zu einem geeigneten Zeitpunkt $\tau < \tau^*$ zu erniedrigen und damit die Instabilität zu umgehen. Ein Verfahren, welches den hierarchischen Rang adaptiv steuert, setzt eine Strategie mit den folgenden Komponenten voraus:

- (K1) Die Strategie legt fest, an welchen Zeitpunkten eine Änderung des hierarchischen Rangs stattfindet.
- (K2) An jenen Zeitpunkten wird ebenfalls festgelegt, in welcher Weise sich der hierarchische Rang ändert.
- (K3) Zu einem Zeitpunkt, an dem sich der hierarchische Rang ändert, kann eine Approximation in der „veränderten“ Mannigfaltigkeit bestimmt werden. Unter dieser Mannigfaltigkeit verstehen wir die Menge an Tensoren im hierarchischen Tucker-Format, jedoch zum veränderten hierarchischen Rang.

Zum gegenwärtigen Stand sind uns keine Verfahren bekannt, welche den hierarchischen Rang adaptiv steuern. Während es für das Erniedrigen des hierarchischen Rangs einfach zu sein scheint, eine Heuristik für (K1), (K2) und (K3) zu finden, wirft das Erhöhen des hierarchischen Rangs vor allem die Frage auf, wie die Komponente (K3) umzusetzen ist. Im Rahmen dieser Arbeit betrachten wir die adaptive Steuerung des hierarchischen Rangs nicht näher.

Niedrigrangapproximation in der Mannigfaltigkeit Tucker (r_1, \dots, r_d)

Das Tucker-Format aus Abschnitt 3.2 ist ein Tensor-Format, welches in seiner Darstellung dem hierarchischen Tucker-Format sehr ähnlich ist. Für $d = 2$ stimmen die beiden Formate sogar überein. Der wesentliche Unterschied zwischen den Formaten besteht darin, dass die Basisfunktionen im hierarchischen Tucker-Format für Knoten im Inneren des Dimensionsbaumes erneut einer hierarchischen Zerlegung genügen. Aufgrund der großen Ähnlichkeit der beiden Tensor-Formate wollen wir an dieser Stelle die numerischen Tests durch die Niedrigrangapproximation in der Mannigfaltigkeit Tucker (r_1, \dots, r_d) ergänzen.

Mit $\tilde{P}(Y_{\text{Tu}})$ als den orthogonalen Projektor auf den Tangentialraum der Mannigfaltigkeit lautet das Anfangswertproblem der Niedrigrangapproximation

$$\dot{Y}_{\text{Tu}}(\tau) = \tilde{P}(Y_{\text{Tu}}(\tau)) \mathcal{A} Y_{\text{Tu}}(\tau), Y_{\text{Tu}}(0) = Y_0 \in \text{Tucker}(r_1, \dots, r_d), \tau \geq 0. \quad (8.1)$$

In [KL10, Lemma 3.1] leiten Koch und Lubich die Formel für die orthogonale Projektion auf den Tangentialraum der Mannigfaltigkeit Tucker (r_1, \dots, r_d) her. Um die Niedrigrangapproximation im Tucker-Format in adäquater Weise mit der Niedrigrangapproximation im hierarchischen Tucker-Format zu vergleichen, verwenden wir den Tucker-Rang $r_1 = r_2 = r_3 = r_4 = 13$. Ehe wir eine numerische Lösung des Anfangswertproblems (8.1) bestimmen, ist es nötig, eine Approximation der Auszahlungsfunktion in der Menge Tucker $(13, 13, 13, 13)$ zu bestimmen, welche dem Anfangswertproblem (8.1) als Startwert dient. Zu diesem Zweck verwenden wir die Funktion `tucker_als` der Matlab Tensor Toolbox [BK⁺12] von Bader und Kolda, siehe auch [BK06]. Diese Funktion bestimmt mittels einer kleinsten Quadrate Methode die Approximation eines Tensors in der Menge der Tensoren im Tucker-Format, siehe [KS08, Algorithmus 1]. Mit dem Matlab-Integrator `ode45` nahm das Lösen des Anfangswertproblems (8.1) auf dem Intervall $[0, 3]$ 15 Minuten in Anspruch. Mit den Formeln aus [KL10, Lemma 3.1] besteht die Zeitableitung des Core-Tensors und der univariaten Basisfunktionen aus insgesamt 31'161 Unbekannten. Im Vergleich zum hierarchischen Tucker-Format resultiert die deutlich höhere Anzahl an Freiheitsgraden im Tucker-Format daraus, dass die Anzahl an Einträgen des Core-Tensors exponentiell mit der Dimension wächst. Dies führt dazu, dass das Tucker-Format in sehr hohen Dimensionen keine Alternative zum hierarchischen Tucker-Format darstellt.

Auf den ersten Anschein wirkt die Zeit für die Berechnung der Niedrigrangapproximation in der Menge \mathcal{H} -Tucker $((13)_{\mathcal{N} \in T})$ nachteilig, obgleich die Zeitableitung in (LR) nur 7'163 Freiheitsgrade besitzt. Wir vermuten die Gründe hierfür in der Programmiersprache Matlab. Um Schleifen in der Implementierung der Zeitableitung von (LR) zu umgehen, erzeugen wir bei der Berechnung der Zeitableitung eine Reihe von Hilfstensoren. Es ist zu vermuten, dass das Assemblieren beziehungsweise das Reservieren und Freigeben von Speicherbereich für diese Objekte der Grund für die schlechtere Laufzeit ist. Um diese Effekte zu vermeiden, empfehlen wir, auf eine andere Programmiersprache wie C/C++ auszuweichen.

Abbildung 8.6 zeigt die Abweichung zwischen der numerischen Lösung $\tilde{Y}_{Tu}(\tau)$ von (8.1) und der Referenzlösung $\tilde{Y}_{AWP}(\tau)$ von (4.1) im Absolutbetrag.

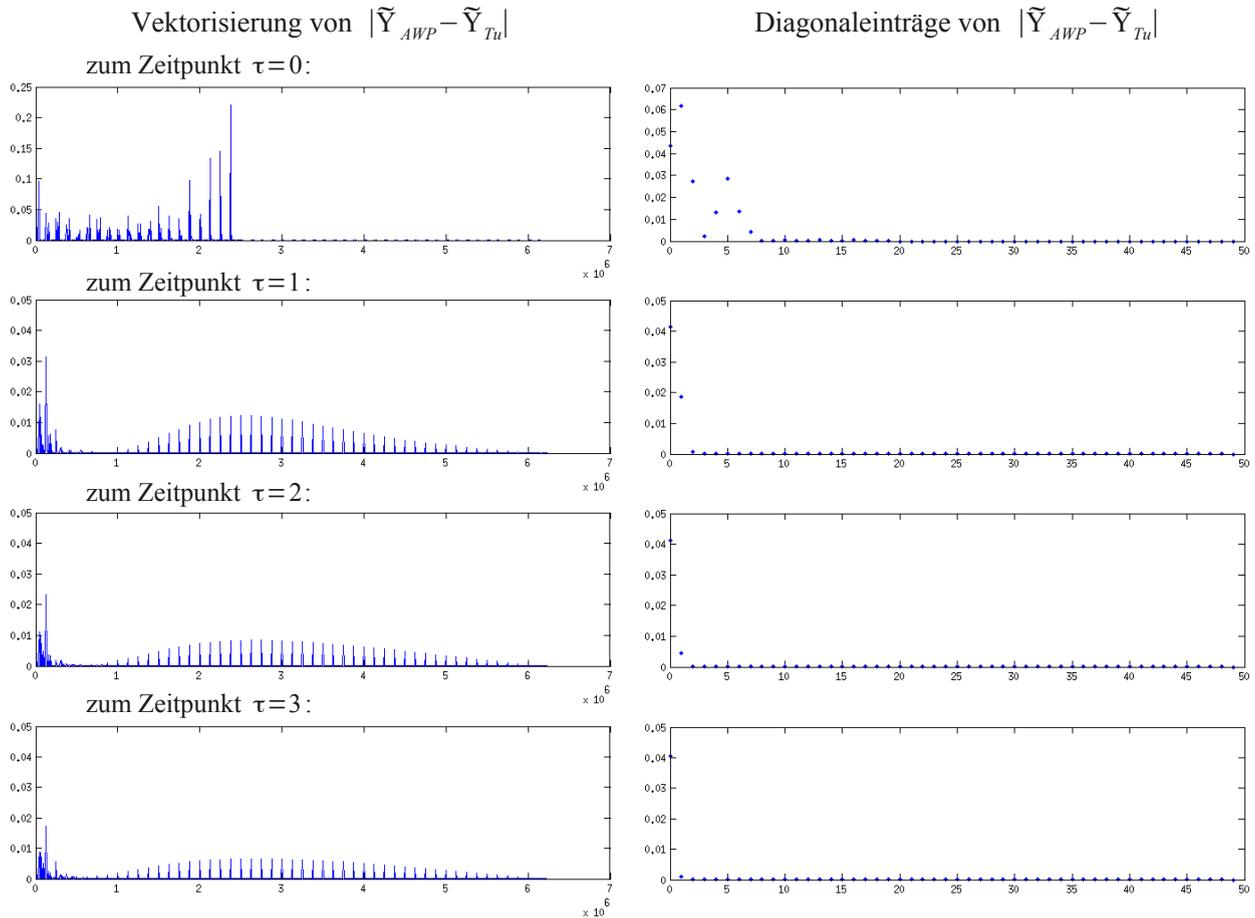


Abbildung 8.6: Abweichung zwischen $\tilde{Y}_{AWP}(\tau)$ und $\tilde{Y}_{Tu}(\tau)$

Analog zu Abbildung 8.5 zeigen die Plots auf der linken Seite von Abbildung 8.6 die Abweichung in vektorisierter Form, während auf den Plots der rechten Seite die Abweichung in den Diagonaleinträgen zu sehen ist. Vergleichen wir nun Abbildung 8.5 mit Abbildung 8.6, so fällt zum Zeitpunkt $\tau = 0$ auf, dass die Approximation der Auszahlungsfunktion in der Menge Tucker (13, 13, 13, 13) eine höhere Genauigkeit aufweist. Darüber hinaus zeigt die Abweichung zwischen $\tilde{Y}_{AWP}(\tau)$ und $\tilde{Y}_{Tu}(\tau)$ qualitativ ein ähnliches Verhalten wie die Abweichung zwischen $\tilde{Y}_{AWP}(\tau)$ und $Y_{HT}(\tau)$: Zu Beginn der Zeitintegration fällt der Fehler sehr schnell ab und bezüglich der Diagonaleinträge ist der Fehler für sämtliche Zeitpunkte im Bereich nahe des Ursprungs lokalisiert.

In Abbildung 8.7 ist sowohl die Differenz zwischen $\tilde{Y}_{AWP}(\tau)$ und $\tilde{Y}_{Tu}(\tau)$ als auch die Differenz zwischen $\tilde{Y}_{AWP}(\tau)$ und $\tilde{Y}_{HT}(\tau)$ in der Frobeniusnorm eingezeichnet.

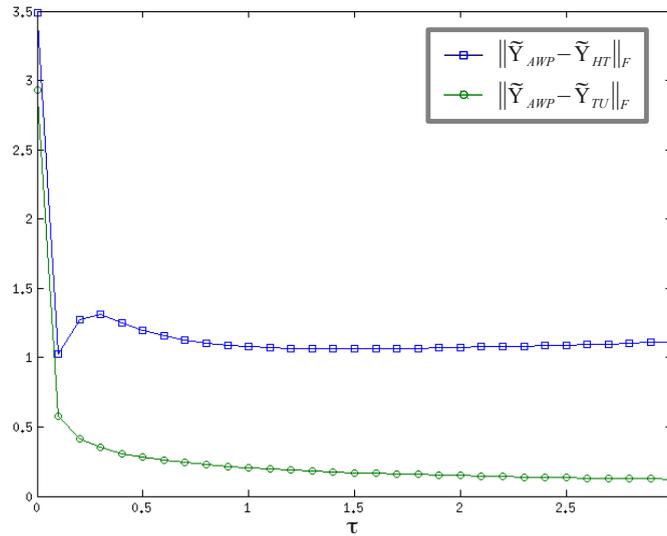


Abbildung 8.7: Abweichung in der Frobeniusnorm

Es sei darauf hingewiesen, dass die Frobeniusnorm in unserem Beispiel über insgesamt 6'250'000 Einträge aufsummiert. Vor diesem Hintergrund ist die Güte beider Niedrigrangapproximationen als positiv zu bewerten. Aufgrund der Mengenrelation

$$\mathcal{H}\text{-Tucker}((13)_{\mathcal{N} \in T}) \subseteq \text{Tucker}(13, 13, 13, 13)$$

überrascht es nicht, dass die Approximation $\tilde{Y}_{\text{TU}}(\tau)$ eine höhere Genauigkeit aufweist als $\tilde{Y}_{\text{HT}}(\tau)$. Es sollte jedoch nicht vergessen werden, dass zur Darstellung eines Tensors in der Menge $\mathcal{H}\text{-Tucker}((13)_{\mathcal{N} \in T})$ deutlich weniger Freiheitsgrade erforderlich sind.

	$\mathbb{R}^{50 \times 50 \times 50 \times 50}$ (volle Darstellung)	Tucker (13, 13, 13, 13)	$\mathcal{H}\text{-Tucker}((13)_{\mathcal{N} \in T})$
# Freiheitsgrade	6'250'000	31'161	7'163

Auswirkungen von schwacher Korrelation

Das hierarchische Tucker-Format ist insbesondere dann zur Darstellung von hochdimensionalen Tensoren geeignet, wenn gewisse Richtungen eine schwache Korrelation aufweisen. In diesem Fall kann der Tensor beziehungsweise die Basisfunktion eines Knotens im Inneren des Dimensionsbaumes durch wenige Basisfunktionen der Kindknoten dargestellt werden. Bei der Black-Scholes-Gleichung liefert die Korrelationsmatrix Σ ein Maß für die

Koppelung der einzelnen Richtungen in der Lösung des Anfangswertproblems (AWP). Die in den bisherigen Beispielen gewählte Korrelationsmatrix

$$\Sigma = \begin{pmatrix} 1 & 0.8 & 0.7 & 0.5 \\ 0.8 & 1 & 0.6 & 0.7 \\ 0.7 & 0.6 & 1 & 0.8 \\ 0.5 & 0.7 & 0.8 & 1 \end{pmatrix}$$

begünstigt nicht wirklich die hierarchische Struktur. In unserem Beispiel verzweigt sich die Wurzel des Dimensionsbaumes zu den Knoten $\mathcal{N}_1 = \{1, 2\}$ und $\mathcal{N}_2 = \{3, 4\}$. Demnach wäre das hierarchische Tucker-Format dann zur Darstellung eines Tensors geeignet, wenn der jeweilige Tensor zwischen den Dimensionen 1, 2 und 3, 4 eine schwache Korrelation besitzt. Um das zu illustrieren wählen wir nun die Korrelationsmatrix

$$\Sigma = \begin{pmatrix} 1 & 0.8 & 0.01 \cdot 0.7 & 0.01 \cdot 0.5 \\ 0.8 & 1 & 0.01 \cdot 0.6 & 0.01 \cdot 0.7 \\ 0.01 \cdot 0.7 & 0.01 \cdot 0.6 & 1 & 0.8 \\ 0.01 \cdot 0.5 & 0.01 \cdot 0.7 & 0.8 & 1 \end{pmatrix}$$

und untersuchen vor diesem Hintergrund erneut die Güte der numerischen Approximationen $\tilde{Y}_{\text{HT}}(\tau)$ und $\tilde{Y}_{\text{Tu}}(\tau)$. In der Modellierung könnte eine solche Situation beispielsweise dann auftreten, wenn die zugrundeliegenden Basiswerte in der multivariaten Black-Scholes-Gleichung Aktien aus unterschiedlichen Branchen repräsentieren. Wäre beispielsweise Basiswert 1 und 2 jeweils eine Aktie aus dem Automobilbereich und Aktie 3 und 4 jeweils eine Aktie aus dem Bereich der Telekommunikation, so würde man erwarten, dass die Basiswerte 1 und 2 mit den Basiswerten 3 und 4 nur schwach korreliert sind.

In Abbildung 8.8 ist die Differenz zwischen $\tilde{Y}_{\text{AWP}}(\tau)$ und $\tilde{Y}_{\text{Tu}}(\tau)$ als auch zwischen $\tilde{Y}_{\text{AWP}}(\tau)$ und $\tilde{Y}_{\text{HT}}(\tau)$ in der Frobeniusnorm eingezeichnet.

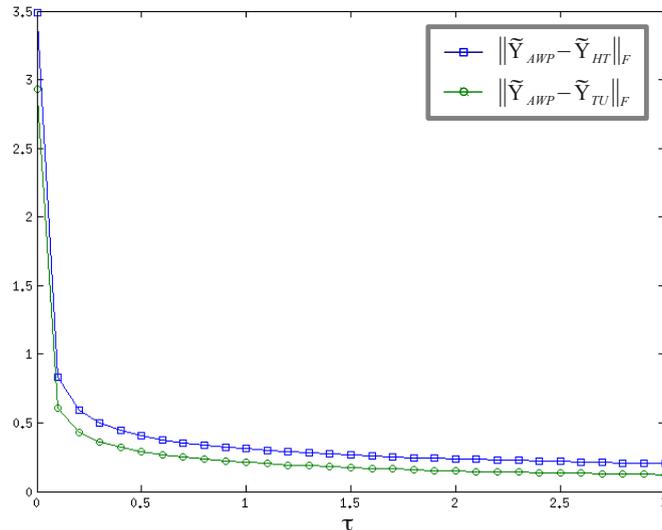


Abbildung 8.8: Abweichung in der Frobeniusnorm

Während die Abweichung von $\tilde{Y}_{HT}(\tau)$ zur Referenzlösung $\tilde{Y}_{AWP}(\tau)$ bezüglich der Frobeniusnorm durch die neue Korrelationsmatrix deutlich kleiner wird, ändert sich die Differenz zwischen $\tilde{Y}_{AWP}(\tau)$ und $\tilde{Y}_{Tu}(\tau)$ nur unwesentlich. Für wachsendes τ beobachten wir in der neuen Situation nun sogar eine leicht abnehmende Abweichung zwischen $\tilde{Y}_{HT}(\tau)$ und der Referenzlösung $\tilde{Y}_{AWP}(\tau)$ bezüglich der Frobeniusnorm.

Die Änderung in der vektorisierten Darstellung beziehungsweise in den Diagonaleinträgen der Abweichungen sind so gering, dass kein Unterschied zu den Abbildungen (8.5) und (8.6) zu erkennen ist. Wir sehen deshalb davon ab, diese Abbildungen zu der neuen Korrelationsmatrix erneut zu zeigen.

Ergebnis und Ausblick

Die Beispiele in diesem Kapitel unterstreichen die Vorzüge der zeitabhängigen Approximation im hierarchischen Tucker-Format. Im Fall der diskretisierten Black-Scholes-Gleichung konnten wir auf diese Weise eine Situation erreichen, in welcher der Fehler trotz einer Kompressionsrate von 99.98 % klein ist. Als Abschluss dieses Kapitels möchten wir auf ein offenes Problem hinweisen. Betrachten wir ein hochdimensionales Anfangswertproblem mit einem Anfangswert, der, wie im Fall der Black-Scholes-Gleichung, als eine Funktion vorliegt. Mit den bekannten Verfahren bestünde der erste Schritt zur zeitabhängigen Approximation im hierarchischen Tucker-Format darin, die Funktion an jeder Position auszuwerten und die Einträge in einem Tensor abzuspeichern. Dieses Vorgehen könnte jedoch daran scheitern, dass der Tensor so viele Einträge besitzt, dass er nicht mehr gespeichert werden kann. Und selbst wenn dies gelingen würde, so müsste in einem zweiten Schritt ein Verfahren wie Algorithmus 1 auf Seite 51 eine Approximation in der Menge $\mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ bestimmen. Algorithmus 1 verwendet zur Berechnung der Approxi-

mation insgesamt $2d-2$ Singulärwertzerlegungen, welche allesamt auf Matrizen angewandt werden, welche dieselbe Anzahl an Einträgen besitzen, wie der hochdimensionale Tensor. Die Durchführung solcher Zerlegungen würde folglich noch mehr Speicherbedarf erfordern und immense numerische Kosten verursachen. Obgleich das Lösen der Bewegungsgleichungen aus Kapitel 4 technisch möglich wäre, würde das Vorhaben aufgrund der fehlenden Approximation des Startwerts in der Mannigfaltigkeit \mathcal{H} -Tucker $((r_{\mathcal{N}})_{\mathcal{N} \in T})$ scheitern. Es sollte deshalb ein Gegenstand künftiger Forschung sein, einen Algorithmus zu entwerfen, welcher zu einem Tensor, der ausschließlich über eine Funktion definiert ist, eine Approximation in der Menge \mathcal{H} -Tucker $((r_{\mathcal{N}})_{\mathcal{N} \in T})$ bestimmt und zwar auf eine Weise, welche es nicht erfordert, den hochdimensionalen Tensor zu assemblieren.

Literaturverzeichnis

- [AJ12] Andreas Arnold und Tobias Jahnke: *On the approximation of high-dimensional differential equations in the hierarchical Tucker format*. Preprint, Karlsruher Institut für Technologie, 2012.
- [Bel61] Richard E. Bellman: *Adaptive control processes - A guided tour*. Princeton University Press, Princeton, New Jersey, 1961.
- [BG68] Richard L. Bishop und Samuel I. Golberg: *Tensor Analysis on Manifolds*. Dover Publications, New York, 1968.
- [BK06] Brett W. Bader und Tamara G. Kolda: *Algorithm 862: MATLAB tensor classes for fast algorithm prototyping*. ACM Transactions on Mathematical Software, 32(4):635–653, 2006.
- [BK⁺12] Brett W. Bader, Tamara G. Kolda *et al.*: *MATLAB Tensor Toolbox Version 2.5*. Verfügbar online, Januar 2012. <http://www.sandia.gov/~tgkolda/TensorToolbox/>.
- [BS73] Fischer Black und Myron Scholes: *The Pricing of Options and Corporate Liabilities*. The Journal of Political Economy, 81(3):637–654, 1973.
- [CC70] Douglas J. Carroll und Jih Jie Chang: *Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition*. Psychometrika, 35(3):283–319, 1970.
- [CL10] Dajana Conte und Christian Lubich: *An Error Analysis of The Multi-configuration Time-dependent Hartree Method of Quantum Dynamics*. Mathematical Modelling and Numerical Analysis, 44(4):759–780, 2010.
- [Coo87] G. J. Cooper: *Stability of Runge-Kutta Methods for Trajectory Problems*. IMA Journal of Numerical Analysis, 7:1–13, 1987.
- [DH08] Peter Deuffhard und Andreas Hohmann: *Numerische Mathematik I: Eine algorithmisch orientierte Einführung*. de Gruyter, Berlin, Boston, 4. Auflage, 2008.
- [Dir30] Paul A. M. Dirac: *Note on Exchange Phenomena in the Thomas Atom*. Mathematical Proceedings of the Cambridge Philosophical Society, 26:376–385, 1930.

- [DLDMV00] Lieven De Lathauwer, Bart De Moor und Joos Vandewalle: *A Multilinear Singular Value Decomposition*. SIAM Journal on Matrix Analysis and Applications, 21(4):1253–1278, 2000.
- [dSL08] Vin de Silva und Lek-Heng Lim: *Tensor Rank and the Ill-Posedness of the Best Low-Rank Approximation Problem*. SIAM Journal on Matrix Analysis and Applications, 30(3):1084–1127, 2008.
- [EY36] Carl Eckart und Gale Young: *The approximation of one matrix by another of lower rank*. Psychometrika, 1(3):211–218, 1936.
- [Fre34] Jakov I. Frenkel: *Wave mechanics: Advanced general theory*. The Clarendon Press, Oxford, 1934.
- [Fri08] Avner Friedman: *Partial Differential Equations of Parabolic Type*. Dover Publications, Mineola, New York, 2008.
- [GHS87] Gene H. Golub, Alan Hoffman und Gilbert W. Stewart: *A generalization of the Eckart-Young-Mirsky matrix approximation theorem*. Linear Algebra and its Applications, 88/89(0):317–327, 1987.
- [GJ10] Michael Günther und Ansgar Jüngel: *Finanzderivate mit MATLAB: Mathematische Modellierung und numerische Simulation*. Vieweg+Teubner, Wiesbaden, 2010.
- [Gra10] Lars Grasedyck: *Hierarchical Singular Value Decomposition of Tensors*. SIAM journal on matrix analysis and applications, 31:2029–2054, 2010.
- [Gra11] Lars Grasedyck: *Hierarchical Low Rank Approximation of Tensors and Multivariate Functions*. Skript zur Sommerschule über "Hierarchical Tensor Approximation" am Hausdorff Research Institute for Mathematics, 2011.
- [Hac12] Wolfgang Hackbusch: *Tensor Spaces and Numerical Tensor Calculus*. Springer Series in Computational Mathematics. Springer, Berlin, Heidelberg, 2012.
- [Har70] Richard A. Harshman: *Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis*. UCLA Working Papers in Phonetics, 16(1):1–84, 1970.
- [HB09] Martin Hanke-Bourgeois: *Grundlagen der Numerischen Mathematik und des Wissenschaftlichen Rechnens*. Vieweg+Teubner, Wiesbaden, 3. Auflage, 2009.
- [Hit27a] Frank L. Hitchcock: *The expression of a tensor or a polyadic as a sum of products*. Journal of Mathematics and Physics, 6:164–189, 1927.
- [Hit27b] Frank L. Hitchcock: *Multiple invariants and generalized rank of a p -way matrix or tensor*. Journal of Mathematics and Physics, 7:39–79, 1927.

- [HJ90] Roger A. Horn und Charles R. Johnson: *Matrix analysis*. Cambridge University Press, Cambridge, 1990.
- [HK09] Wolfgang Hackbusch und Stefan Kühn: *A New Scheme for the Tensor Representation*. Journal of Fourier Analysis and Applications, 15:706–722, 2009.
- [HLW06] Ernst Hairer, Christian Lubich und Gerhard Wanner: *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*, Band 31 der Reihe *Springer Series in Computational Mathematics*. Springer, Berlin, Heidelberg, 2. Auflage, 2006.
- [HRS12] Sebastian Holtz, Thorsten Rohwedder und Reinhold Schneider: *On manifolds of tensors of fixed TT-rank*. Numerische Mathematik, 120(4):701–731, 2012.
- [HRSW13] Norbert Hilber, Oleg Reichmann, Christoph Schwab und Christoph Winter: *Computational Methods for Quantitative Finance: Finite Element Methods for Derivative Pricing*. Springer Finance. Springer, Berlin, Heidelberg, 2013.
- [JH08] Tobias Jahnke und Wilhelm Huisinga: *A Dynamical Low-Rank Approach to the Chemical Master Equation*. Bulletin of Mathematical Biology, 70:2283–2302, 2008.
- [KA00] Peter Knabner und Lutz Angermann: *Numerik partieller Differentialgleichungen: Eine anwendungsorientierte Einführung*. Springer, Berlin, 2000.
- [KB09] Tamara G. Kolda und Brett W. Bader: *Tensor decompositions and applications*. SIAM Review, 51:455–500, 2009.
- [KK01] Ralf Korn und Elke Korn: *Optionsbewertung und Portfolio-Optimierung: Moderne Methoden der Finanzmathematik*. Vieweg+Teubner, Braunschweig, 2. Auflage, 2001.
- [KL07] Othmar Koch und Christian Lubich: *Dynamical Low-Rank Approximation*. SIAM Journal on Matrix Analysis and Applications, 29:434–454, 2007.
- [KL10] Othmar Koch und Christian Lubich: *Dynamical tensor approximation*. SIAM Journal on Matrix Analysis and Applications, 31:2360–2375, 2010.
- [KN01] Raul Kangro und Roy Nicolaides: *Far Field Boundary Conditions for Black-Scholes Equations*. SIAM Journal on Numerical Analysis, 38(4):1357–1368, 2001.
- [KS08] Tamara G. Kolda und Jimeng Sun: *Scalable Tensor Decompositions for Multi-aspect Data Mining*. In: *ICDM 2008: Proceedings of the 8th IEEE International Conference on Data Mining*, Seiten 363–372, Dezember 2008.
- [KT12] Daniel Kressner und Christine Tobler: *htucker - A Matlab toolbox for tensors in hierarchical Tucker format*. Technischer Bericht, Eidgenössische Technische Hochschule Zürich, 2012.

- [Lee03] John M. Lee: *Introduction to Smooth Manifolds*, Band 218 der Reihe *Graduate Texts in Mathematics*. Springer, New York, 2003.
- [Lee12] John M. Lee: *Introduction to Smooth Manifolds*, Band 218 der Reihe *Graduate Texts in Mathematics*. Springer, New York, 2. Auflage, 2012.
- [LRSV13] Christian Lubich, Thorsten Rohwedder, Reinhold Schneider und Bart Vandereycken: *Dynamical approximation by hierarchical Tucker and tensor-train tensors*. *SIAM Journal on Matrix Analysis and Applications*, 34:470–494, 2013.
- [LT05] Stig Larsson und Vidar Thomée: *Partielle Differentialgleichungen und numerische Methoden*. Springer, Berlin, 2005.
- [Lub05] Christian Lubich: *On variational approximations in quantum molecular dynamics*. *Mathematics of Computation*, 74(250):765–779, 2005.
- [Lub08] Christian Lubich: *From Quantum to Classical Molecular Dynamics: Reduced Models and Numerical Analysis*. Zürich lectures in advanced mathematics. European Mathematical Society, Zürich, 2008.
- [Man08] Uwe Manthe: *A multilayer multiconfigurational time-dependent Hartree approach for quantum dynamics on general potential energy surfaces*. *Journal of Chemical Physics*, 128:164116, 2008.
- [Mer73] Robert Merton: *Theory of Rational Option Pricing*. *The Bell Journal of Economics and Management Science*, 4(1):141–183, 1973.
- [NL08] Achim Nonnenmacher und Christian Lubich: *Dynamical low-rank approximation: applications and numerical experiments*. *Mathematics and Computers in Simulation*, 79:1346–1357, 2008.
- [Øks07] Bernt Øksendal: *Stochastic Differential Equations: An Introduction with Applications*. Springer, New York, 6. Auflage, 2007.
- [Ose11] Ivan V. Oseledets: *Tensor-Train Decomposition*. *SIAM Journal on Scientific Computing*, 33(5):2295–2317, 2011.
- [OT10] Ivan V. Oseledets und Eugene Tyrtyshnikov: *TT-cross approximation for multidimensional arrays*. *Linear Algebra and its Applications*, 432(1):70–88, 2010.
- [PA05] Olivier Pironneau und Yves Achdou: *Computational methods for option pricing*. Society for Industrial and Applied Mathematics, Philadelphia, 2005.
- [PD08] Gustav Pomberger und Heinz Dobler: *Algorithmen und Datenstrukturen: Eine systematische Einführung in die Programmierung*. it informatik. Pearson Studium, München, 2008.

- [Sey09] Rüdiger U. Seydel: *Tools for Computational Finance*. Springer, London, 2009.
- [TK12] Christine Tobler und Daniel Kressner: *Hierarchical Tucker Toolbox*. Verfügbar online, Februar 2012. http://www.sam.math.ethz.ch/NLAGroup/htucker_toolbox.html.
- [Usc13] André Uschmajew: *Zur Theorie der Niedrigrangapproximation in Tensorprodukten von Hilberträumen*. Dissertation, Technische Universität Berlin, 2013.
- [UV13] André Uschmajew und Bart Vandereycken: *The geometry of algorithms using hierarchical tensors*. Linear Algebra and its Applications, 439:133–166, 2013.
- [VM11] Oriol Vendrell und Hans-Dieter Meyer: *Multilayer multi-configuration time-dependent Hartree method: implementation and applications to a Henon-Heiles Hamiltonian and to pyrazine*. Journal of Chemical Physics, 134(4):044135, 2011.
- [Wal00] Wolfgang Walter: *Gewöhnliche Differentialgleichungen: Eine Einführung*. Springer, Berlin, 7. Auflage, 2000.
- [Wer11] Dirk Werner: *Funktionalanalysis*. Springer, Berlin, 2011.
- [WT03] Haobin Wang und Michael Thoss: *Multilayer formulation of the multiconfiguration time-dependent Hartree theory*. Journal of Chemical Physics, 119:1289–1299, 2003.