



Several Approaches to Break the Curse of Dimensionality

Dissertation

*zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)*

vorgelegt dem Rat der
Fakultät für Mathematik und Informatik
der Friedrich-Schiller-Universität Jena

von Dipl.-Math. Markus Weimar
geboren am 28. Februar 1986 in Weimar

Gutachter:

- 1.) Prof. Dr. Erich Novak (Jena) – *summa cum laude*
- 2.) Prof. Dr. Aicke Hinrichs (Rostock) – *summa cum laude*
- 3.) Prof. Dr. Henryk Woźniakowski (New York, Warschau) – *summa cum laude*

Tag der öffentlichen Verteidigung: 13.05.2013

Acknowledgements

I would like to express my deepest gratitude to my supervisor Professor Dr. Erich Novak for numerous hints, suggestions and remarks during the preparation of this work. Furthermore, I would like to thank all the members of the research groups “Theoretical numerics” and “Function spaces” in Jena for supporting me during the times of my Diploma thesis and my Ph.D. studies. Finally, I like to thank our friends from the IBC community for many fruitful discussions at several conferences during the last years.

Table of contents

Zusammenfassung	iii
Preface	xi
1 Preliminaries	1
1.1 Basic notation	1
1.2 General problem	3
1.3 Algorithms and cost model	6
1.4 Notions of tractability	8
2 Properties and tools for special problem classes	11
2.1 Lower bounds on linear subspaces	11
2.2 Linearity and compactness	13
2.3 General Hilbert space problems	16
2.3.1 Singular value decomposition	16
2.3.2 Optimal algorithm	17
2.3.3 Polynomial tractability	19
2.4 Tensor product problems	22
2.4.1 Definition and simple properties	22
2.4.2 Eigenpairs and the optimal algorithm	24
2.4.3 Complexity	26
2.5 Reproducing kernel Hilbert spaces	28
2.5.1 Definition and properties	29
2.5.2 Examples: Integration and approximation problems	31
3 Problems on Hilbert spaces with scaled norms	35
3.1 Definitions, eigenpairs and the optimal algorithm	35
3.2 Complexity	37
3.2.1 Polynomial tractability	37
3.2.2 Weak tractability and the curse	40
3.2.3 Normalized errors	48
3.3 Examples	49

4	Problems on function spaces with weighted norms	53
4.1	The concept of weighted spaces	54
4.1.1	Weighted Banach spaces of smooth functions	54
4.1.2	Weighted Hilbert spaces and weighted RKHS	55
4.2	Uniform approximation in reproducing kernel Hilbert spaces	59
4.3	Uniform approximation in Banach spaces of smooth functions	63
4.3.1	Lower bounds for spaces of low-degree polynomials	63
4.3.2	Complexity results via embeddings	67
4.3.3	Conclusions and applications	72
4.4	Possible extensions and further results	75
4.4.1	L_p -approximation	75
4.4.2	Uniform approximation in the weighted Sobolev space	80
5	Problems on Hilbert spaces with (anti)symmetry conditions	83
5.1	Basic definitions related to (anti)symmetry	83
5.1.1	Hilbert function spaces	84
5.1.2	Tensor products of Hilbert function spaces	86
5.1.3	Arbitrary tensor product Hilbert spaces	90
5.2	Optimal algorithms for (anti)symmetric problems	91
5.3	Complexity of (anti)symmetric problems	95
5.3.1	Symmetric problems (absolute errors)	97
5.3.2	Symmetric problems (normalized errors)	102
5.3.3	Antisymmetric problems (absolute errors)	103
5.3.4	Antisymmetric problems (normalized errors)	109
5.4	Applications	110
5.4.1	Toy examples	110
5.4.2	Wavefunctions	112
	Bibliography	115

Zusammenfassung

In den modernen Wissenschaften gewinnt die effiziente numerische Behandlung hochdimensionaler Aufgabenstellungen zunehmend an Bedeutung. Eine fundamentale Erkenntnis der informationsbasierten Komplexitätstheorie¹ ist es, dass die Kompliziertheit eines Problems durch die Konvergenzrate allein nicht hinreichend beschrieben werden kann. Beispielsweise haben Novak und Woźniakowski [NW09] eindrucksvoll gezeigt, dass Probleme existieren, deren algorithmische Lösung trotz beliebig guter Konvergenzrate exponentiell viele Informationsoperationen benötigt. In so einem Fall spricht man vom *Fluch der Dimension* bzw. dem *Curse of Dimensionality*. Bei der analytischen Untersuchung numerischer Probleme stellt sich heraus, dass dieser Fluch oftmals durch die Ausnutzung zusätzlicher Struktureigenschaften überwunden werden kann. Die vorliegende Arbeit strebt an, verschiedene solche Ansätze vorzustellen.

Ein numerisches Problem S wird beschrieben durch eine Folge linearer, kompakter Operatoren S_d zwischen normierten Räumen \mathcal{F}_d und \mathcal{G}_d , wobei $d \in \mathbb{N}$. Im Allgemeinen besteht das Ziel darin, zu jedem S_d Näherungsverfahren $A_{n,d}$ zu finden, welche höchstens $n \in \mathbb{N}_0$ Informationsoperationen verwenden und dabei den sogenannten *worst-case Fehler*

$$\Delta^{\text{wor}}(A_{n,d}; S_d) = \sup_{\|f\|_{\mathcal{F}_d} \leq 1} \|S_d(f) - A_{n,d}(f)\|_{\mathcal{G}_d}$$

minimieren. Besonderes Interesse gilt Problemstellungen, die auf Tensorproduktstrukturen basieren, sowie einfach zu implementierenden, linearen Approximationen $A_{n,d}$. Die minimale Anzahl an Informationsfunktionalen, die benötigt werden, um ein Problem S bis auf einen Fehler $\varepsilon > 0$ zu lösen, wird als Informationskomplexität²

$$n(\varepsilon, d; S_d) = \min \{n \in \mathbb{N}_0 \mid \exists A_{n,d}: \Delta^{\text{wor}}(A_{n,d}; S_d) \leq \varepsilon\}, \quad \varepsilon > 0, d \in \mathbb{N},$$

bezeichnet. Wächst diese Zahl exponentiell mit der Dimension d , so leidet S am Fluch der Dimension. Ist $n(\varepsilon, d; S_d)$ dagegen weder exponentiell in d noch in ε^{-1} , so nennen wir das Problem S *weakly tractable*. Insbesondere sprechen wir von

¹engl.: *information based complexity (IBC)*

²engl.: *information complexity*

polynomial tractability, falls sich $n(\varepsilon, d; S_d)$ durch ein Polynom in ε^{-1} und d nach oben beschränken lässt:

$$n(\varepsilon, d; S_d) \leq C \varepsilon^{-p} d^q \quad \text{für gewisse } C, p > 0, q \geq 0 \quad \text{und alle } \varepsilon \in (0, 1], d \in \mathbb{N}.$$

Gilt diese Ungleichung sogar mit $q = 0$, so heißt S *strongly polynomially tractable*.

Im Folgenden stellen wir die drei in dieser Arbeit untersuchten Ansätze Struktureigenschaften auszunutzen vor und fassen die wichtigsten daraus abgeleiteten Komplexitätsaussagen kurz zusammen.

Eine der simpelsten Klassen von Problemen S ist die Menge aller linearen, kompakten Operatoren zwischen Tensorprodukt-Hilberträumen. Speziell die Komplexität von Tensorproduktproblemen $S_d = \bigotimes_{k=1}^d S_1: H_d \rightarrow \mathcal{G}_d$, erzeugt durch $S_1: H_1 \rightarrow \mathcal{G}_1$, ist dabei gut verstanden. Sie hängt von der nicht-wachsend angeordneten Folge $\lambda = (\lambda_m)_{m \in \mathbb{N}}$ der Quadrate der singulären Zahlen des zugrundeliegenden Operators S_1 ab. Insbesondere ist $S = (S_d)_{d \in \mathbb{N}}$ nicht polynomially tractable, falls $\lambda_1 \geq 1$ und $\lambda_2 > 0$. Ist λ_1 echt größer 1 und $\lambda_2 > 0$ oder $\lambda_1 \geq \lambda_2 = 1$, so gilt sogar der Fluch der Dimension; vergleiche dazu [Theorem 2.11](#).

Ein erster Ansatz, ein solches Problem zu modifizieren, besteht darin, das innere Produkt des Grundraums H_d zu skalieren und S_d auf dem so entstehenden Hilbertraum \mathcal{F}_d zu betrachten:

$$\langle \cdot, \cdot \rangle_{\mathcal{F}_d} = \frac{1}{s_d} \langle \cdot, \cdot \rangle_{H_d} \quad \text{für ein } s_d > 0 \quad \text{und alle } d \in \mathbb{N}.$$

Das mit Faktoren der Folge $s = (s_d)_{d \in \mathbb{N}}$ skalierte Problem bezeichnen wir mit $S_{(s)} = (S_{d,s_d}: \mathcal{F}_d \rightarrow \mathcal{G}_d)_{d \in \mathbb{N}}$. Wir betrachten die worst-case Komplexität von $S_{(s)}$ bezüglich des absoluten Fehlerkriteriums und beweisen folgenden

Satz 1. Mit den obigen Bezeichnungen und unter der Annahme $\lambda_2 > 0$ sind die folgenden Aussagen äquivalent:

- (I) $S_{(s)}$ ist strongly polynomially tractable.
- (II) $S_{(s)}$ ist polynomially tractable.
- (III) Es existiert ein $\tau \in (0, \infty)$, sodass $\lambda \in \ell_\tau$ und $\sup_{d \in \mathbb{N}} s_d \|\lambda\|_{\ell_\tau}^d < \infty$.
- (IV) Es existiert ein $\varrho \in (0, \infty)$, sodass $\lambda \in \ell_\varrho$ und $\limsup_{d \rightarrow \infty} s_d^{1/d} < \frac{1}{\lambda_1}$.

Falls eine der Bedingungen zutrifft, ist der *exponent of strong polynomial tractability* gegeben durch $p^* = \inf\{2\tau \mid \tau \text{ erfüllt Bedingung (III)}\}$.

Vergleiche dazu [Theorem 3.2](#) in [Abschnitt 3.2.1](#). Bemerkenswert an diesem Resultat ist, dass analog zu unskalierten Problemen polynomial tractability bereits strong polynomial tractability impliziert, ungeachtet der vielfältigen Möglichkeiten, die Folge $(s_d)_{d \in \mathbb{N}}$ zu wählen.

Die schwächere Eigenschaft weak tractability und der Fluch der Dimension lassen sich charakterisieren, sofern wir ein gewisses asymptotisches Verhalten des initialen Fehlers $\varepsilon_d^{\text{init}} = \sqrt{s_d \cdot \lambda_1^d}$ unterstellen; siehe [Theorem 3.7](#) in [Abschnitt 3.2.2](#).

Satz 2. Wir betrachten das skalierte Tensorproduktproblem $S_{(s)} = (S_{d,s_d})_{d \in \mathbb{N}}$ bezüglich des absoluten Fehlerkriteriums und setzen $\lambda_2 > 0$ voraus. Ferner

- sei $\ln(\varepsilon_d^{\text{init}}) \notin o(d)$ für $d \rightarrow \infty$. Dann gilt der Fluch der Dimension.
- gelte $\varepsilon_d^{\text{init}} \in \Theta(d^\alpha)$ für $d \rightarrow \infty$ und ein $\alpha \geq 0$.
 - Falls $\lambda_1 = \lambda_2$, so leidet $S_{(s)}$ unter dem Fluch der Dimension.
 - Ist $\lambda_1 > \lambda_2$, so ist das Problem $S_{(s)}$ genau dann weakly tractable, wenn $\lambda_n \in o(\ln^{-2(1+\alpha)} n)$ für $n \rightarrow \infty$ gilt.
- gelte $\varepsilon_d^{\text{init}} \rightarrow 0$ für d gegen unendlich. Dann kann der Fluch der Dimension nicht auftreten. Außerdem ist $S_{(s)}$ genau dann weakly tractable, wenn
 - (i) $\lambda_1 = \lambda_2$ und $\lambda_n \in o(\ln^{-2} n)$ für $n \rightarrow \infty$ und $\varepsilon_d^{\text{init}} \in o(1/d)$ für $d \rightarrow \infty$, oder wenn
 - (ii) $\lambda_1 > \lambda_2$ und $\lambda_n \in o(\ln^{-2} n)$ für $n \rightarrow \infty$ gilt.

Besonderes Interesse kommt hier dem Parameter α zu, welcher das polynomielle Wachstum des initialen Fehlers kontrolliert. Er schlägt sich im Falle $\lambda_1 > \lambda_2$ unmittelbar in der Charakterisierung für weak tractability nieder. Außerdem ist die Bedingung $\varepsilon_d^{\text{init}} \in o(1/d)$ für $d \rightarrow \infty$ im dritten Teil des Satzes überraschend. Da der initiale Fehler unskalierter Probleme entweder durchweg exponentiell fallen oder wachsen kann, oder in jeder Dimension den Wert eins annimmt, treten diese Phänomene in der klassischen Theorie, d.h. im Fall $s_d = 1$ für alle $d \in \mathbb{N}$, nicht auf.

Ein weiterer Ansatz, den Fluch der Dimension zu brechen, bezieht sich auf Probleme über Funktionenräumen. Hier nutzen wir a priori gegebenes Wissen über den Einfluss einzelner Variablen auf die Funktionen im Grundraum, um diese effizient approximieren zu können. Dazu stattdessen wir die Grundräume mit gewichteten Normen aus. Insbesondere Probleme auf Funktionenräumen mit Hilbertraumstruktur und Produktgewichten wurden in den letzten Jahren ausgiebig untersucht. Aufgabenstellungen, bei denen die Grund- und/oder Zielräume allgemeinere Banachräume sind, wurden dagegen aus Sicht der informationsbasierten Komplexitätstheorie weitaus seltener studiert.

In dieser Abhandlung betrachten wir unter anderem das Problem der gleichmäßigen Approximation glatter Funktionen:

$$\text{App} = (\text{App}_d: F_d^\gamma \rightarrow L_\infty([0, 1]^d))_{d \in \mathbb{N}} \quad \text{mit} \quad \text{App}_d(f) = f \quad \text{für} \quad d \in \mathbb{N},$$

wobei wir $F_d^\gamma = \{f: [0, 1]^d \rightarrow \mathbb{R} \mid f \in C^\infty([0, 1]^d) \text{ mit } \|f\|_{F_d^\gamma} < \infty\}$ als Grundraum ansetzen und mit der gewichteten Norm

$$\|f\|_{F_d^\gamma} = \sup_{\alpha \in \mathbb{N}_0^d} \frac{1}{\gamma_\alpha} \|D^\alpha f\|_{L_\infty([0, 1]^d)}$$

ausstatten. Hierbei werden die Zahlen $\gamma_{\alpha} = \prod_{j=1}^d (\gamma_{d,j})^{\alpha_j}$ für jedes $\alpha \in \mathbb{N}_0^d$ mithilfe der Folge $C_{\gamma} \geq \gamma_{d,1} \geq \dots \geq \gamma_{d,d} > 0$ definiert und als *Produktgewichte* bezeichnet. Es stellt sich heraus, dass die Komplexität des Approximationsproblems von gewissen Summierbarkeitseigenschaften dieser Folge abhängt, die ebenfalls für gewichtete Hilbertraumprobleme eine entscheidende Rolle spielen. Wir definieren

$$p(\gamma) = \inf \left\{ \kappa > 0 \left| \limsup_{d \rightarrow \infty} \sum_{j=1}^d (\gamma_{d,j})^{\kappa} < \infty \right. \right\} \quad \text{sowie}$$

$$q(\gamma) = \inf \left\{ \kappa > 0 \left| \limsup_{d \rightarrow \infty} \sum_{j=1}^d (\gamma_{d,j})^{\kappa} / \ln(d+1) < \infty \right. \right\}$$

und beweisen den folgenden

Satz 3. Mit den obigen Bezeichnungen gilt bzgl. des absoluten Fehlerkriteriums:

- Ist das Problem App polynomially tractable, so folgt $q(\gamma) \leq 1$. Weiterhin impliziert strong polynomial tractability die Bedingung $p(\gamma) \leq 1$.
- Gilt $q(\gamma) < 1$ oder sogar $p(\gamma) < 1$, so ist App polynomially beziehungsweise sogar strongly polynomially tractable.

Genauer gesagt beweisen wir diese notwendigen und hinreichenden Kriterien für eine ganze Klasse gewichteter Banachräume, welche gewisse Einbettungsbedingungen erfüllen; siehe dazu [Proposition 4.6](#) und [Proposition 4.7](#). Der Raum F_d^{γ} bildet dabei einen Spezialfall, welcher Ergebnisse von Novak und Woźniakowski [NW09] verallgemeinert. Weiterhin zeigen wir, dass die hinreichenden Bedingungen $q(\gamma) < 1$ bzw. $p(\gamma) < 1$ für polynomial bzw. strong polynomial tractability des L_{∞} -Approximationsproblems auf einem unverankerten Sobolevraum \mathcal{H}_d^{γ} auch notwendig sind; siehe dazu [Theorem 4.18](#).

Weak tractability sowie der Fluch der Dimension lassen sich wie folgt charakterisieren.

Satz 4. Für $\text{App} = (\text{App}_d)_{d \in \mathbb{N}}$ sind die folgenden Aussagen äquivalent:

- (i) Das Problem ist weakly tractable.
- (ii) Der Fluch der Dimension gilt nicht.
- (iii) Für alle $\kappa > 0$ gilt $\lim_{d \rightarrow \infty} \frac{1}{d} \sum_{j=1}^d (\gamma_{d,j})^{\kappa} = 0$.
- (iv) Es existiert ein $\kappa \in (0, 1)$, sodass $\lim_{d \rightarrow \infty} \frac{1}{d} \sum_{j=1}^d (\gamma_{d,j})^{\kappa} = 0$.

Dies folgt unmittelbar aus unserem etwas allgemeineren [Theorem 4.9](#). Bemerkenswert sind hierbei die nicht-triviale Implikation (ii) \Rightarrow (i) sowie die Charakterisierung (iv), welche typischerweise bei Problemen auf Hilberträumen mit Produktgewichten anzutreffen ist.

Unser dritter Ansatz basiert schließlich auf dem Ausnutzen gewisser Symmetrieeigenschaften der Problemelemente im Grundraum. Dazu betrachten wir abermals Tensorproduktprobleme $S = (S_d: H_d \rightarrow \mathcal{G}_d)_{d \in \mathbb{N}}$ zwischen Hilberträumen, schränken diese aber auf geeignete Teilräume, welche ausschließlich aus (anti-)symmetrischen Elementen bestehen, ein. Dieses Konzept lässt sich besonders anschaulich am Spezialfall von Problemen über Funktionenräumen erläutern. Für $d \in \mathbb{N}$ und $I \subseteq \{1, \dots, d\}$ bezeichne \mathcal{S}_I die Menge aller Permutationen π der Koordinatenmenge $\{1, \dots, d\}$, welche die Elemente aus $I^c = \{1, \dots, d\} \setminus I$ unverändert lassen. Dann heißt eine reelle Funktion $f \in H_d = H_1 \otimes \dots \otimes H_1$ auf $[0, 1]^d$ genau dann *I-symmetrisch*, wenn

$$f(\mathbf{x}) = f(\boldsymbol{\pi}(\mathbf{x})) \quad \text{für jedes } \mathbf{x} \in [0, 1]^d \quad \text{und alle } \pi \in \mathcal{S}_I$$

gilt. Erfüllt f hingegen die Gleichung $f(\mathbf{x}) = (-1)^{|\pi|} f(\boldsymbol{\pi}(\mathbf{x}))$, so sprechen wir von *I-Antisymmetrie*. Die entsprechenden linearen Teilräume des Grundraums H_d , welche genau jene Funktionen enthalten, bezeichnen wir mit $\mathfrak{S}_I(H_d)$ bzw. $\mathfrak{A}_I(H_d)$. Besonders antisymmetrische Funktionen, d.h. solche, deren Funktionswert das Vorzeichen wechselt, sobald zwei Koordinaten ihres Arguments vertauscht werden, haben sich in der Vergangenheit als praktisch relevant erwiesen; siehe bspw. [Abschnitt 5.4.2](#). Die Einschränkung eines Tensorproduktproblems $S = (S_d: H_d \rightarrow \mathcal{G}_d)_{d \in \mathbb{N}}$ auf die Räume $P_{I_d}(H_d)$, $d \in \mathbb{N}$, wird im Folgenden mit $S_I = (S_{d,I_d})_{d \in \mathbb{N}}$ bezeichnet. Hierbei setzen wir die Art der Symmetrie $P \in \{\mathfrak{S}, \mathfrak{A}\}$ sowie eine Folge $(I_d)_{d \in \mathbb{N}}$ von Teilmengen der Koordinaten als fixiert voraus.

Da sich die Operatoren S_{d,I_d} für jedes $d \in \mathbb{N}$ als Verknüpfung von S_d mit einer geeigneten Orthogonalprojektion auffassen lassen, besteht ein enger Zusammenhang zwischen den singulären Zahlen von S_d und denen der Einschränkungen S_{d,I_d} . Diese Zahlen bestimmen maßgeblich den minimalen worst-case Fehler des Problems S_I und ermöglichen weiterhin die Konstruktion eines optimalen (linearen) Algorithmus, welcher diesen realisiert; siehe dazu [Theorem 5.4](#).

Ausgehend davon lassen sich Aussagen ableiten, welche die informationsbasierte Komplexität von S_I mithilfe der Quadrate der singulären Zahlen von S_1 sowie der Menge der unterstellten (Anti-)Symmetriebedingungen beschreiben. Der Einfachheit halber beschränken wir uns hier wieder auf das absolute Fehlerkriterium und beginnen mit dem Fall symmetrischer Probleme; vergleiche [Theorem 5.10](#).

Satz 5 (Polynomial tractability, $P = \mathfrak{S}$). Es sei $S_1: H_1 \rightarrow \mathcal{G}_1$ ein kompakter linearer Operator zwischen Hilberträumen und $\lambda = (\lambda_m)_{m \in \mathbb{N}}$ die nicht-wachsend angeordnete Folge der Eigenwerte von $W_1 = S_1^\dagger S_1$. Es gelte $\lambda_2 > 0$ und für $d > 1$ sei ferner $\emptyset \neq I_d \subseteq \{1, \dots, d\}$ gegeben. Wir betrachten die Einschränkung $S_I = (S_{d,I_d})_{d \in \mathbb{N}}$ des Tensorproduktproblems $S = (S_d: H_d \rightarrow \mathcal{G}_d)_{d \in \mathbb{N}}$ auf die I_d -symmetrischen Teilräume $\mathfrak{S}_{I_d}(H_d) \subset H_d$, $d \in \mathbb{N}$. Dann ist S_I genau dann strongly polynomially tractable, wenn $\lambda \in \ell_\tau$ für ein $\tau \in (0, \infty)$ und entweder

- $\lambda_1 < 1$, oder

- $1 = \lambda_1 > \lambda_2$ und $(d - \#I_d) \in \mathcal{O}(1)$ für $d \rightarrow \infty$

gilt. Im Falle $\lambda_1 \leq 1$ ist das Problem genau dann polynomially tractable, wenn $\lambda \in \ell_\tau$ für ein $\tau \in (0, \infty)$ und entweder

- $\lambda_1 < 1$, oder
- $\lambda_1 = 1$ und $(d - \#I_d) \in \mathcal{O}(\ln d)$ für $d \rightarrow \infty$

gilt.

Für den Fall $\lambda_1 > 1$ ist keine hinreichende Bedingung für polynomial tractability bekannt. Unsere Ergebnisse zeigen allerdings, dass die Bedingungen $\lambda \in \ell_\tau$ und $(d - \#I_d) \in \mathcal{O}(\ln d)$ auch hier notwendig sind. Zusammenfassend ergibt sich, dass beispielsweise für $\lambda_1 = \lambda_2 = 1$ ausreichend viele zusätzliche Symmetrieannahmen den Fluch der Dimension brechen können, unter dem das ursprüngliche Problem S leidet; siehe auch [Theorem 2.11](#).

Die etwas kompliziertere Komplexitätsanalyse antisymmetrischer Probleme zeigt, dass hier sogar noch wesentlich schwächere Voraussetzungen genügen, um polynomial tractability schlussfolgern zu können und damit den Fluch zu vermeiden. Dies lässt sich unter anderem mit der Struktur des initialen Fehlers dieser Probleme begründen. Unsere Komplexitätsaussagen lassen sich, wie in [Theorem 5.16](#) in [Abschnitt 5.3](#) dargestellt, folgendermaßen vereinfacht zusammenfassen:

Satz 6 (Polynomial tractability, $P = \mathfrak{A}$). Es sei $S_1: H_1 \rightarrow \mathcal{G}_1$ ein kompakter linearer Operator zwischen Hilberträumen und $\lambda = (\lambda_m)_{m \in \mathbb{N}}$ die nicht-wachsend angeordnete Folge der Eigenwerte von $W_1 = S_1^\dagger S_1$. Es gelte $\lambda_2 > 0$ und für $d > 1$ sei ferner $\emptyset \neq I_d \subseteq \{1, \dots, d\}$ gegeben. Wir betrachten die Einschränkung $S_I = (S_{d,I_d})_{d \in \mathbb{N}}$ des Tensorproduktproblems $S = (S_d: H_d \rightarrow \mathcal{G}_d)_{d \in \mathbb{N}}$ auf die I_d -antisymmetrischen Teilräume $\mathfrak{A}_{I_d}(H_d) \subset H_d$, $d \in \mathbb{N}$. Im Falle $\lambda_1 < 1$ sind die nachstehenden Aussagen äquivalent:

- S_I ist strongly polynomially tractable.
- S_I ist polynomially tractable.
- Es existiert eine Konstante $\tau \in (0, \infty)$, sodass $\lambda \in \ell_\tau$.

Darüber hinaus gelten die selben Äquivalenzen, falls $\lambda_1 \geq 1$ ist und die Anzahl der Antisymmetriebedingungen $\#I_d$ linear mit der Dimension d wächst.

An diesem Resultat ist zu erkennen, dass antisymmetrische Tensorproduktprobleme eine deutlich geringere Komplexität aufweisen als ihre symmetrischen Gegenstücke, welche wiederum wesentlich einfacher sind als volle Tensorproduktprobleme, sofern wir genügend (Anti-)Symmetrie fordern. Es existieren allerdings natürliche Beispiele, die zeigen, dass selbst vollständig antisymmetrische Probleme nicht notwendigerweise trivial oder polynomially tractable sein müssen; siehe dazu [Abschnitt 5.4.1](#).

Die vorliegende Arbeit ist folgendermaßen gegliedert: Im ersten Kapitel erläutern wir neben der verwendeten Notation die allgemeine Problemstellung sowie das zugrundeliegende Kostenmodell. Zusätzlich werden hier die betrachteten Komplexitätsklassen formal definiert.

[Kapitel 2](#) widmet sich dann speziellen Klassen numerischer Probleme sowie elementaren Hilfsmitteln, welche benötigt werden, um diese zu behandeln. Insbesondere stellen wir hier die Singulärwertzerlegung³ kompakter Operatoren zwischen Hilberträumen detailliert vor. Sie bildet in vielen Fällen die Grundlage der Konstruktion optimaler Algorithmen und ist daher für den Rest der Arbeit von fundamentaler Bedeutung. Darüber hinaus werden Tensorproduktstrukturen in Hilberträumen diskutiert und wohlbekannte Komplexitätsaussagen für darauf basierende Problemklassen wiederholt. Schließlich gehen wir kurz auf Hilberträume mit reproduzierendem Kern⁴ ein.

In den ersten beiden Abschnitten von [Kapitel 3](#) leiten wir die in [Satz 1](#) und [Satz 2](#) beschriebenen Charakterisierungen der verschiedenen Komplexitätsklassen für skalierte Tensorproduktprobleme zwischen Hilberträumen her. Darüber hinaus folgern wir aus diesen Aussagen die entsprechenden Charakterisierungen für das normalisierte Fehlerkriterium in [Abschnitt 3.2.3](#). Es stellt sich heraus, dass die Skalierungsfaktoren in diesem Fall irrelevant werden. Neben der Angabe des optimalen Algorithmus sowie dessen worst-case Fehlers wird außerdem gezeigt, dass diese neuen, bisher unveröffentlichten Ergebnisse die bekannte Theorie in natürlicher Weise verallgemeinern. Abschließend illustrieren wir die gewonnenen Ergebnisse an zwei einfachen Beispielen.

Das vierte Kapitel befasst sich mit Problemen über Funktionenräumen mit gewichteten Normen. Hier stellen wir zunächst das Konzept gewichteter Normen ausführlich vor und illustrieren dies dann am Beispiel des unverankerten Sobolev Raums \mathcal{H}_d^γ mit Produktgewichten. Für eben diesen Raum präsentieren wir einen Algorithmus $A_{n,d}^*$, der geeignete obere Fehlerschranken für das L_∞ -Approximationsproblem einhält. Zusammen mit entsprechenden unteren Schranken, welche für Teilräume einfacher Polynome bewiesen werden, folgern wir schließlich die in [Satz 3](#) und [Satz 4](#) angesprochenen Komplexitätsaussagen für eine ganze Skala produktgewichteter Banachräume mithilfe einfacher Einbettungen. Der letzte Teilabschnitt dieses Kapitels, [Abschnitt 4.4](#), zeigt dann verschiedene Möglichkeiten auf, die entwickelten Methoden zu verallgemeinern, sodass entsprechende Resultate auch für das L_p -Approximationsproblem, $1 \leq p < \infty$, auf geeigneten Räumen gewonnen werden können. Weiterhin zeigen wir hier, dass der Algorithmus $A_{n,d}^*$ im wesentlichen optimal auf \mathcal{H}_d^γ operiert. Grundlage dafür bildet ein Resultat von Kuo, Wasilkowski und Woźniakowski [[KWW08](#)], welches das L_∞ -Approximationsproblem auf kernreproduzierenden Hilberträumen mit dem L_2 -Approximationsproblem in Verbindung setzt.

Einige der in diesem Kapitel präsentierten Aussagen wurden bereits (in teilweise

³engl.: *singular value decomposition (SVD)*

⁴engl.: *reproducing kernel Hilbert spaces (RKHSs)*

schwächerer Form) in der Arbeit [Wei12b] veröffentlicht. Auf Verallgemeinerungen und neue Resultate weisen wir an den entsprechenden Stellen explizit hin.

Kapitel 5 behandelt schließlich Probleme mit (Anti-)Symmetriebedingungen. Dazu definieren wir (Anti-)Symmetrie zunächst in Hilbert-Funktionenräumen. Wir gehen speziell auf Tensorproduktstrukturen ein und folgern grundlegende Eigenschaften der entsprechenden Projektoren und Teilräume. Diese Eigenschaften dienen am Ende von Abschnitt 5.1 als Hilfsmittel für die Verallgemeinerung des (Anti-)Symmetriebegriffs auf abstrakte Tensorprodukthilberträume. Anschließend definieren wir (anti-)symmetrische numerische Probleme $S_I = (S_{d,I_d})_{d \in \mathbb{N}}$ durch die Restriktion eines gegebenen Tensorproduktproblems $S = (S_d)_{d \in \mathbb{N}}$ auf die entsprechenden Unterräume (anti-)symmetrischer Elemente des Grundraums. Wir zeigen die Kommutativität der Operatoren S_d mit gewissen Projektoren und schließen so auf optimale Algorithmen und deren worst-case Fehler. In Abschnitt 5.3 analysieren wir ausgehend davon die Komplexität (anti-)symmetrischer numerischer Problemstellungen. Wir differenzieren zwischen Symmetrie und Antisymmetrie sowie zwischen dem absoluten und dem normalisierten Fehlerkriterium. Insbesondere Satz 5 und Satz 6 werden hier bewiesen. Im letzten Teilabschnitt dieses Kapitels diskutieren wir verschiedene Anwendungen. Zum einen illustrieren wir an einfachen Beispielen, dass zusätzliches Wissen um (Anti-)Symmetriebedingungen die Komplexität drastisch verringern kann. Zum anderen wenden wir uns praktisch relevanteren Problemstellungen zu und demonstrieren mögliche Anwendungen dieser neuen Theorie auf das Approximationsproblem für Wellenfunktionen, die in Modellen der Quantenmechanik und der theoretischen Chemie von Bedeutung sind.

Ein Großteil der Ergebnisse dieses Kapitels wurde in [Wei12a] veröffentlicht. Wir verwenden in der vorliegenden Arbeit jedoch stellenweise andere Beweismethoden, welche geringfügige Verallgemeinerungen ermöglichen.

Im Anhang der Abhandlung befinden sich das Literaturverzeichnis sowie ein akademischer Lebenslauf und die ehrenwörtliche Erklärung des Autors. Innerhalb eines jeden Kapitels werden Formelnummern durchgehend nummeriert. Außerdem verwenden wir hier eine fortlaufende Nummerierung für Lemmata (Lemma), Bemerkungen (Remark), Beispiele (Example), Hilfsresultate (Proposition) und Hauptsätze (Theorem); also beispielweise Proposition 3.1, gefolgt von Theorem 3.2 und Lemma 3.3. Die Symbole \square und \blacksquare zeigen darüber hinaus das Ende von Bemerkungen und Beispielen bzw. von Beweisen an.

Preface

In modern science the efficient numerical treatment of high-dimensional problems becomes more and more important. A fundamental insight of the theory of *information-based complexity* (IBC for short) is that the computational hardness of a problem can not be described properly only by the rate of convergence. An impressive example that illustrates this fact was given recently by Novak and Woźniakowski [NW09]. They studied a problem for which an exponential number of information operations is needed in order to reduce the initial error, although there exist algorithms which provide an arbitrary large rate of convergence. Problems that yield this exponential dependence are said to suffer from the *curse of dimensionality*. While analyzing numerical problems it turns out that we can often vanquish this curse by exploiting additional structural properties. The aim of this thesis is to present several approaches of this type.

A numerical problem S is given by a sequence of compact linear operators S_d acting between normed spaces \mathcal{F}_d and \mathcal{G}_d , where $d \in \mathbb{N}$. In general we seek for algorithms $A_{n,d}$ that approximate S_d while using at most $n \in \mathbb{N}_0$ pieces of information on the input elements $f \in \mathcal{F}_d$. The quality of this approximation is measured by the so-called *worst case error*

$$\Delta^{\text{wor}}(A_{n,d}; S_d) = \sup_{\|f\|_{\mathcal{F}_d} \leq 1} \|S_d(f) - A_{n,d}(f)\|_{\mathcal{G}_d}$$

which we try to minimize. Problems based on tensor product structures, as well as linear algorithms that are easy to implement, are of particular interest. The minimal number of information operations needed to solve a given problem S to within a threshold $\varepsilon > 0$ is called *information complexity*:

$$n(\varepsilon, d; S_d) = \min \{n \in \mathbb{N}_0 \mid \exists A_{n,d}: \Delta^{\text{wor}}(A_{n,d}; S_d) \leq \varepsilon\}, \quad \varepsilon > 0, d \in \mathbb{N}.$$

If this quantity grows exponentially fast with the dimension d then S suffers from the curse of dimensionality. In the case where $n(\varepsilon, d; S_d)$ is neither exponential in d , nor in ε^{-1} , the problem S is said to be *weakly tractable*. A special case is described by the notion of *polynomial tractability* for which the information complexity needs to be bounded from above by a polynomial in d and ε^{-1} , i.e.

$$n(\varepsilon, d; S_d) \leq C \varepsilon^{-p} d^q \quad \text{for some } C, p > 0, q \geq 0 \quad \text{and all } \varepsilon \in (0, 1], d \in \mathbb{N}.$$

If the latter inequality is valid even for $q = 0$ then S is called *strongly polynomially tractable*.

Next we present the three approaches to exploit structural properties we study in this thesis and we briefly summarize our main complexity results.

A rather simple class of problems S is given by the set of all compact linear operators between tensor products of Hilbert spaces. Especially the complexity of tensor product problems $S_d = \bigotimes_{k=1}^d S_1: H_d \rightarrow \mathcal{G}_d$, induced by some operator $S_1: H_1 \rightarrow \mathcal{G}_1$, is well-understood. It depends on the non-increasingly ordered sequence $\lambda = (\lambda_m)_{m \in \mathbb{N}}$ of the squares of the singular values of the underlying operator S_1 . In particular, it is well-known that $S = (S_d)_{d \in \mathbb{N}}$ is not polynomially tractable if we have $\lambda_1 \geq 1$ and $\lambda_2 > 0$. Actually, we are faced with the curse of dimensionality if λ_1 is strictly larger than 1 and $\lambda_2 > 0$, or if $\lambda_1 \geq \lambda_2 = 1$; cf. [Theorem 2.11](#).

A first approach to modify such a problem is to scale the inner products of the source spaces H_d , $d \in \mathbb{N}$. We set

$$\langle \cdot, \cdot \rangle_{\mathcal{F}_d} = \frac{1}{s_d} \langle \cdot, \cdot \rangle_{H_d} \quad \text{for some } s_d > 0 \quad \text{and all } d \in \mathbb{N}$$

and investigate the complexity of the problem operators S_d interpreted as mappings between the Hilbert spaces \mathcal{F}_d and \mathcal{G}_d , $d \in \mathbb{N}$. The resulting problem, scaled by factors from the sequence $s = (s_d)_{d \in \mathbb{N}}$, then is denoted by $S_{(s)} = (S_{d,s_d}: \mathcal{F}_d \rightarrow \mathcal{G}_d)_{d \in \mathbb{N}}$. We study the worst case setting with respect to the absolute error criterion and prove

Theorem 1. Using the introduced notation and assuming that $\lambda_2 > 0$ the following assertions are equivalent:

- (I) $S_{(s)}$ is strongly polynomially tractable.
- (II) $S_{(s)}$ is polynomially tractable.
- (III) There exists $\tau \in (0, \infty)$ such that $\lambda \in \ell_\tau$ and $\sup_{d \in \mathbb{N}} s_d \|\lambda\|_{\ell_\tau}^d < \infty$.
- (IV) There exists $\varrho \in (0, \infty)$ such that $\lambda \in \ell_\varrho$ and $\limsup_{d \rightarrow \infty} s_d^{1/d} < \frac{1}{\lambda_1}$.

If one of these (and hence all) conditions applies then the *exponent of strong polynomial tractability* is given by $p^* = \inf\{2\tau \mid \tau \text{ fulfills condition (III)}\}$.

We refer to [Theorem 3.2](#) in [Section 3.2.1](#). It is remarkable that similar to unscaled problems polynomial tractability of the problem $S_{(s)}$ already implies strong polynomial tractability, despite the fact that we can choose the sequence of scaling factors $(s_d)_{d \in \mathbb{N}}$ completely arbitrary.

The less restrictive property weak tractability and the curse of dimensionality can be characterized, provided that we additionally assume a certain asymptotic behavior of the initial error $\varepsilon_d^{\text{init}} = \sqrt{s_d \cdot \lambda_1^d}$; see [Theorem 3.7](#) in [Section 3.2.2](#).

Theorem 2. We study the scaled tensor product problem $S_{(s)} = (S_{d,s_d})_{d \in \mathbb{N}}$ in the worst case setting w.r.t. the absolute error criterion and assume $\lambda_2 > 0$. Moreover,

- let $\ln(\varepsilon_d^{\text{init}}) \notin o(d)$, as $d \rightarrow \infty$. Then we have the curse of dimensionality.
- let $\varepsilon_d^{\text{init}} \in \Theta(d^\alpha)$, as $d \rightarrow \infty$, for some $\alpha \geq 0$.
 - If $\lambda_1 = \lambda_2$ then $S_{(s)}$ suffers from the curse of dimensionality.
 - In the case $\lambda_1 > \lambda_2$ the problem $S_{(s)}$ is weakly tractable if and only if $\lambda_n \in o(\ln^{-2(1+\alpha)} n)$, as $n \rightarrow \infty$.
- let $\varepsilon_d^{\text{init}} \rightarrow 0$, as d approaches infinity. Then we are never faced with the curse of dimensionality. Furthermore, $S_{(s)}$ is weakly tractable if and only if
 - (i) $\lambda_1 = \lambda_2$ and $\lambda_n \in o(\ln^{-2} n)$, as $n \rightarrow \infty$, and $\varepsilon_d^{\text{init}} \in o(1/d)$, as $d \rightarrow \infty$,
or
 - (ii) $\lambda_1 > \lambda_2$ and $\lambda_n \in o(\ln^{-2} n)$, as $n \rightarrow \infty$.

Here the parameter α that controls the polynomial growth of the initial error is of particular interest. In the case where $\lambda_1 > \lambda_2$ it directly enters the condition for the characterization of weak tractability. Moreover, the condition $\varepsilon_d^{\text{init}} \in o(1/d)$, as $d \rightarrow \infty$, in the third part of the theorem is quite surprising. Since for unscaled problems the initial error only can grow or decline exponentially, or it equals one in any dimension, these phenomena can not occur in the classical theory, i.e. in the case where $s_d = 1$ for all $d \in \mathbb{N}$.

Another approach to overcome the curse of dimensionality is related to problems defined between function spaces. Here we can make use of some a priori given knowledge about the influence of certain (groups of) variables on the functions in the source space, in order to approximate them efficiently. To this end, we endow these spaces with weighted norms. During the last years especially problems on function spaces that yield a Hilbert space structure, equipped with so-called product weights, attracted a lot of attention. Problems where the source and/or target spaces are allowed to be more general Banach spaces were studied less frequently within the IBC community.

Among other things, in this thesis we consider the uniform approximation problem

$$\text{App} = (\text{App}_d: F_d^\gamma \rightarrow L_\infty([0, 1]^d))_{d \in \mathbb{N}} \quad \text{with} \quad \text{App}_d(f) = f \quad \text{for} \quad d \in \mathbb{N}$$

defined on certain classes of smooth functions

$$F_d^\gamma = \{f: [0, 1]^d \rightarrow \mathbb{R} \mid f \in C^\infty([0, 1]^d) \text{ with } \|f\|_{F_d^\gamma} < \infty\}$$

which are endowed with the weighted norms

$$\|f\|_{F_d^\gamma} = \sup_{\alpha \in \mathbb{N}_0^d} \frac{1}{\gamma_\alpha} \|D^\alpha f\|_{L_\infty([0, 1]^d)}.$$

Here for every $\alpha \in \mathbb{N}_0^d$, $d \in \mathbb{N}$, the *product weights* $\gamma_\alpha = \prod_{j=1}^d (\gamma_{d,j})^{\alpha_j}$ are constructed out of a uniformly bounded sequence $C_\gamma \geq \gamma_{d,1} \geq \dots \geq \gamma_{d,d} > 0$ of so-called generator weights. It turns out that the complexity of the approximation problem depends on certain summability properties of these generators which also play an important role when dealing with problems on product-weighted Hilbert spaces. We define the quantities

$$p(\gamma) = \inf \left\{ \kappa > 0 \mid \limsup_{d \rightarrow \infty} \sum_{j=1}^d (\gamma_{d,j})^\kappa < \infty \right\}, \quad \text{as well as}$$

$$q(\gamma) = \inf \left\{ \kappa > 0 \mid \limsup_{d \rightarrow \infty} \sum_{j=1}^d (\gamma_{d,j})^\kappa / \ln(d+1) < \infty \right\},$$

and prove the following

Theorem 3. For the worst case setting w.r.t. the absolute error criterion we have:

- If the problem App is polynomially tractable then $q(\gamma) \leq 1$. Moreover strong polynomial tractability implies the condition $p(\gamma) \leq 1$.
- If $q(\gamma) < 1$ or even $p(\gamma) < 1$ then App is polynomially tractable or even strongly polynomially tractable, respectively.

In fact, we show these necessary and sufficient criteria for a whole scale of weighted Banach spaces that fulfill certain embedding conditions; see [Proposition 4.6](#) and [Proposition 4.7](#) for details. The source space F_d^γ as defined above appears as a special case within this scale. On the other hand, it generalizes a space considered by Novak und Woźniakowski [[NW09](#)]. In addition, we prove that the sufficient conditions $q(\gamma) < 1$ and $p(\gamma) < 1$ are also necessary for (strong) polynomial tractability of the L_∞ -approximation problem defined on a certain unanchored Sobolev space \mathcal{H}_d^γ ; cf. [Theorem 4.18](#).

Weak tractability and the curse of dimensionality can be characterized as follows.

Theorem 4. For $\text{App} = (\text{App}_d)_{d \in \mathbb{N}}$ the following assertions are equivalent:

- (i) The problem is weakly tractable.
- (ii) The curse of dimensionality is not present.
- (iii) For all $\kappa > 0$ we have $\lim_{d \rightarrow \infty} \frac{1}{d} \sum_{j=1}^d (\gamma_{d,j})^\kappa = 0$.
- (iv) There exists $\kappa \in (0, 1)$ such that $\lim_{d \rightarrow \infty} \frac{1}{d} \sum_{j=1}^d (\gamma_{d,j})^\kappa = 0$.

This immediately follows from our [Theorem 4.9](#) in which we discuss a more general situation. Note that the implication (ii) \Rightarrow (i) is not trivial. Moreover,

the condition (iv) is typical for problems defined on Hilbert spaces equipped with product weights.

Finally, our third approach to vanquish the curse is based on exploiting certain symmetry properties of the elements in the source space. For this purpose we again consider tensor product problems $S = (S_d: H_d \rightarrow \mathcal{G}_d)_{d \in \mathbb{N}}$ between Hilbert spaces. But now we restrict them to suitable subspaces which solely consist of (anti)symmetric elements. We illustrate this concept by considering the special case of problems defined between function spaces.

For $d \in \mathbb{N}$ and $I \subseteq \{1, \dots, d\}$ let \mathcal{S}_I denote the collection of all permutations π of the coordinate set $\{1, \dots, d\}$ that leave the complement $I^c = \{1, \dots, d\} \setminus I$ of I fixed. Then a real-valued function $f \in H_d = H_1 \otimes \dots \otimes H_1$ on $[0, 1]^d$ is called *I-symmetric* if

$$f(\mathbf{x}) = f(\boldsymbol{\pi}(\mathbf{x})) \quad \text{for every } \mathbf{x} \in [0, 1]^d \quad \text{and all } \pi \in \mathcal{S}_I.$$

In contrast, f is called *I-antisymmetric* if the equality $f(\mathbf{x}) = (-1)^{|\pi|} f(\boldsymbol{\pi}(\mathbf{x}))$ holds true for every \mathbf{x} and π . In what follows we denote the corresponding linear subspaces of H_d that exclusively contain symmetric or antisymmetric functions by $\mathfrak{S}_I(H_d)$ and $\mathfrak{A}_I(H_d)$, respectively. Particularly antisymmetric functions, i.e. functions that change their sign when we exchange the variables x_i and x_j , $i, j \in I$, turned out to be of some practical interest; see, e.g., [Section 5.4.2](#). For the restriction of a given tensor product problem $S = (S_d: H_d \rightarrow \mathcal{G}_d)_{d \in \mathbb{N}}$ to the subspaces $P_{I_d}(H_d)$, $d \in \mathbb{N}$, we write $S_I = (S_{d, I_d})_{d \in \mathbb{N}}$. Here the kind of symmetry $P \in \{\mathfrak{S}, \mathfrak{A}\}$, as well as a sequence $(I_d)_{d \in \mathbb{N}}$ of subsets of the coordinates, is assumed to be fixed.

Since for $d \in \mathbb{N}$ the operators S_{d, I_d} can be interpreted as a composition of S_d with suitable orthogonal projections, there exists a close relation of the singular values of S_d with the corresponding singular values of the restricted operators S_{d, I_d} . These numbers essentially determine the minimal worst case error of the problem S_I . This knowledge furthermore allows the construction of an optimal (linear) algorithm that realizes this error; cf. [Theorem 5.4](#).

Consequently, we can conclude assertions that relate the information complexity of S_I to the squares of the singular values of S_1 and to the number of (anti)symmetry conditions we impose. For the sake of simplicity we restrict ourselves again to the absolute error criterion and start by discussing the case of symmetric problems; see [Theorem 5.10](#).

Theorem 5 (Polynomial tractability, $P = \mathfrak{S}$). Let $S_1: H_1 \rightarrow \mathcal{G}_1$ denote a compact linear operator between Hilbert spaces and let $\lambda = (\lambda_m)_{m \in \mathbb{N}}$ be the sequence of eigenvalues of $W_1 = S_1^\dagger S_1$ w.r.t. a non-increasing ordering. Assume $\lambda_2 > 0$ and for $d > 1$ let $\emptyset \neq I_d \subseteq \{1, \dots, d\}$ be fixed. We consider the restriction $S_I = (S_{d, I_d})_{d \in \mathbb{N}}$ of the tensor product problem $S = (S_d: H_d \rightarrow \mathcal{G}_d)_{d \in \mathbb{N}}$ to the I_d -symmetric subspaces $\mathfrak{S}_{I_d}(H_d) \subset H_d$, $d \in \mathbb{N}$. Then S_I is strongly polynomially tractable if and only if $\lambda \in \ell_\tau$ for some $\tau \in (0, \infty)$ and

- $\lambda_1 < 1$, or

- $1 = \lambda_1 > \lambda_2$ and $(d - \#I_d) \in \mathcal{O}(1)$, as $d \rightarrow \infty$.

Moreover, provided that $\lambda_1 \leq 1$ the problem is polynomially tractable if and only if $\lambda \in \ell_\tau$ for some $\tau \in (0, \infty)$ and

- $\lambda_1 < 1$, or
- $\lambda_1 = 1$ and $(d - \#I_d) \in \mathcal{O}(\ln d)$, as $d \rightarrow \infty$.

It remains the open problem to find sufficient conditions for polynomial tractability in the case $\lambda_1 > 1$. However, our results show that the conditions $\lambda \in \ell_\tau$ and $(d - \#I_d) \in \mathcal{O}(\ln d)$ are necessary in this situation, too. In conclusion, we see that imposing sufficiently many additional symmetry assumptions, we can avoid the curse of dimensionality which we are faced with e.g. in the case $\lambda_1 = \lambda_2 = 1$; see also [Theorem 2.11](#).

The complexity analysis of antisymmetric problems is more demanding. On the other hand, it turns out that here even weaker conditions are sufficient to conclude polynomial tractability and thus to vanquish the curse. One of the reasons is the structure of the initial error which is more complicated in this case. Similar to [Theorem 5.16](#) in [Section 5.3](#) we can summarize the main results on the complexity as follows:

Theorem 6 (Polynomial tractability, $P = \mathfrak{A}$). Let $S_1: H_1 \rightarrow \mathcal{G}_1$ denote a compact linear operator between Hilbert spaces and let $\lambda = (\lambda_m)_{m \in \mathbb{N}}$ be the sequence of eigenvalues of $W_1 = S_1^\dagger S_1$ w.r.t. a non-increasing ordering. Assume $\lambda_2 > 0$ and for $d > 1$ let $\emptyset \neq I_d \subseteq \{1, \dots, d\}$ be fixed. We consider the restriction $S_I = (S_{d, I_d})_{d \in \mathbb{N}}$ of the tensor product problem $S = (S_d: H_d \rightarrow \mathcal{G}_d)_{d \in \mathbb{N}}$ to the I_d -antisymmetric subspaces $\mathfrak{A}_{I_d}(H_d) \subset H_d$, $d \in \mathbb{N}$. Then for the case $\lambda_1 < 1$ the following statements are equivalent:

- S_I is strongly polynomially tractable.
- S_I is polynomially tractable.
- There exists a constant $\tau \in (0, \infty)$ such that $\lambda \in \ell_\tau$.

Moreover, the same equivalences hold true if $\lambda_1 \geq 1$ and the number of antisymmetric coordinates $\#I_d$ grows linearly with the dimension d .

Clearly, these assertions show that antisymmetric tensor product problems are significantly easier than their symmetric counterparts which on their part possess a lower information complexity than entire tensor product problems, as long as we impose enough (anti)symmetry conditions. On the other hand, there exist quite natural examples which show that even fully antisymmetric problems are not necessarily trivial or polynomially tractable, in general. For details we refer to [Section 5.4.1](#).

Let us briefly explain the structure of the present thesis. In the first chapter we settle some notational conventions and we define the abstract problem we are faced with in IBC. Furthermore, here we introduce the used cost model and recall the formal definitions of several complexity categories.

In [Chapter 2](#) we discuss special classes of numerical problems, as well as elementary tools that we need to handle them. In particular, here we give a detailed introduction to the singular value decomposition (SVD) of compact operators between Hilbert spaces. In many cases it builds the basis for the construction of optimal algorithms. Hence it is of fundamental importance for the rest of our work. In addition, we discuss tensor product structures in Hilbert spaces and recall some well-known complexity assertions for problems related to this concept. Finally, we briefly introduce so-called reproducing kernel Hilbert spaces (RKHSs) and collect some of their properties.

In the first two sections of the third chapter we derive the characterizations of the different types of tractability of scaled tensor product problems between Hilbert spaces we presented in [Theorem 1](#) and [Theorem 2](#) above. Moreover, from them we conclude a complete characterization for the normalized error criterion in [Section 3.2.3](#). It turns out that here the scaling factors become irrelevant. Apart from formulas of the optimal algorithm and its worst case error, we additionally show that these new assertions generalize the known theory in a quite natural way. We conclude this chapter by the application of the obtained results to two simple examples.

[Chapter 4](#) then deals with problems on function spaces endowed with weighted norms. Here we explain the concept of weighted spaces in full detail and illustrate it using the example of some unanchored Sobolev \mathcal{H}_d^γ space equipped with product weights. For the uniform approximation problem on this space we present an algorithm $A_{n,d}^*$ that satisfies suitable upper error bounds. Together with corresponding lower bounds, which we prove for spaces of low-degree polynomials, the application of simple embedding arguments then leads us to complexity assertions for a whole scale of product-weighted Banach spaces. In particular, these assertions cover the results for the space F_d^γ stated in [Theorem 3](#) and [Theorem 4](#). Finally, the last section within this chapter, [Section 4.4](#), presents some generalizations of the techniques developed before. Among other things, here we show how to handle L_p -approximation problems, where $1 \leq p < \infty$, defined on suitable spaces. Moreover, we show that the algorithm $A_{n,d}^*$ is essentially optimal for L_∞ -approximation on \mathcal{H}_d^γ . For the proof we make use of arguments due to Kuo, Wasilkowski and Woźniakowski [[KWW08](#)] that relate the uniform approximation problem in the worst case setting, defined on quite general reproducing kernel Hilbert spaces, to a certain average case L_2 -approximation problem.

Some of the results presented in this chapter were already published in [[Wei12b](#)]. However, we were able to partially improve these assertions. We will explicitly emphasize generalizations and new results at the appropriate points.

Finally, [Chapter 5](#) is devoted to problems with (anti)symmetry conditions. We

start with the definition of (anti)symmetry in Hilbert function spaces. In particular, we focus our attention to tensor product structures and conclude fundamental properties of the respective projections and subspaces. At the end of [Section 5.1](#) we use these properties in order to generalize the notion of (anti)symmetry to tensor products of abstract Hilbert spaces. Afterwards we define (anti)symmetric numerical problems $S_I = (S_{d,I_d})_{d \in \mathbb{N}}$ by the restriction of a given tensor product problem $S = (S_d)_{d \in \mathbb{N}}$ to the subspaces of (anti)symmetric elements in the source spaces. We prove the commutativity of the operators S_d with certain projections and conclude formulas for optimal algorithms and their worst case errors. This in hand, in [Section 5.3](#) we then discuss the complexity of (anti)symmetric numerical problems. We distinguish between symmetric and antisymmetric problems, as well as between the absolute and the normalized error criterion. Here we particularly derive the proofs of [Theorem 5](#) and [Theorem 6](#). The chapter is concluded by a section which is devoted to several applications. On the one hand, we use simple examples to show that the additional knowledge about (anti)symmetry conditions can dramatically reduce the information complexity. On the other hand, we also discuss more advanced problems that play a role in computational practice. To this end, we illustrate the application of this new theory to the approximation problem of so-called wavefunctions that arise in certain models of quantum mechanics and theoretical chemistry.

A major part of the results proven in this chapter was published in [[Wei12a](#)]. However, at some points we use different proof techniques that allow slight generalizations.

Within every chapter formulas are numbered consecutively. Moreover, we use a sequential numbering for lemmata, remarks, examples, propositions, and theorems; e.g. [Proposition 3.1](#) is followed by [Theorem 3.2](#) and [Lemma 3.3](#). The symbols \square and \blacksquare are used to indicate the end of remarks and examples, as well as of proofs, respectively.

CHAPTER 1

Preliminaries

Apart from introducing some notational conventions, the aim of this first chapter is to define the general objects of interest in *information based complexity* (IBC). We give an abstract formulation of the general problem in [Section 1.2](#). Afterwards we introduce some classes of algorithms and discuss the used cost model in [Section 1.3](#). Finally, in [Section 1.4](#), we recall the notions of tractability, as well as the definition of the curse of dimensionality.

1.1 Basic notation

As usual we denote by \mathbb{N} the natural numbers and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ are all non-negative integers. Moreover, \mathbb{R} denotes the real line and \mathbb{R}^d ($d \in \mathbb{N}$) is the collection of all points $\mathbf{x} = (x_1, \dots, x_d)$ in the d -dimensional Euclidean space. Given a real number $y > 0$ the symbol $\lfloor y \rfloor$ means the largest $n \in \mathbb{N}_0$ such that $n \leq y$ and we define $\lceil y \rceil$ to be the smallest number $m \in \mathbb{N}$ with $y \leq m$. The value of the Riemann zeta function at some $z > 1$ is denoted by $\zeta(z) = \sum_{n=1}^{\infty} n^{-z}$.

If $\mathbf{k} = (k_1, \dots, k_d) \in \mathbb{N}_0^d$ is a multi-index then $|\mathbf{k}| = \sum_{i=1}^d k_i$ stands for its length. Furthermore, we use the common notation $\mathbf{x}^{\mathbf{k}} = x_1^{k_1} \cdot \dots \cdot x_d^{k_d}$. For $\boldsymbol{\alpha} \in \mathbb{N}_0^d$ partial derivatives of d -variate functions are denoted by $D^{\boldsymbol{\alpha}}$, i.e.

$$D^{\boldsymbol{\alpha}} f = \frac{\partial^{|\boldsymbol{\alpha}|} f}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}}.$$

Derivatives of univariate functions g are indicated as g' , g'' , \dots , $g^{(n)}$. For real numbers $a < b$ half-open intervals are symbolized by $[a, b)$, and $[a, b]^d$ stands for the Cartesian product $\times_{i=1}^d [a, b] = [a, b] \times \dots \times [a, b]$. If $\mathbf{x} \in \mathbb{R}^d$ belongs to $[a, b]^d$ then the value of the characteristic (or indicator) function $\chi_{[a, b]^d}(\mathbf{x})$ of this set equals 1. Otherwise we define $\chi_{[a, b]^d}(\mathbf{x}) = 0$. Similarly the Kronecker delta function $\delta_{i, j}$ is one if the two objects i and j coincide and $\delta_{i, j} = 0$ when they differ from each other.

In what follows we assume that the reader has a fundamental knowledge in measure theory and probability theory as it can be found, e.g., in the textbooks of

Bauer [Bau01, Bau96]. We write λ^d for the Lebesgue measure in \mathbb{R}^d and use the symbols \mathbb{P} and \mathbb{E} for probabilities and expectations, respectively. We use $\#I$ to denote the cardinality of a finite set I . As usual the sum over an empty index set I is to be interpreted as zero whereas empty products equal 1 by definition.

Throughout the whole thesis we assume that the reader is familiar with the basic concepts in functional analysis such as, e.g., complete normed spaces (Banach spaces), weak derivatives or tensor products. For a comprehensive introduction we refer to the textbooks of Triebel [Tri92] and Yosida [Yos80]. The norm in some space F is denoted by $\|\cdot\|_F$. We write $B_r(F) = \{f \in F \mid \|f\|_F \leq r\}$ for centered, closed balls of radius $r \geq 0$ in normed spaces F . Moreover, we use ∂M for the boundary and $\text{int}(M)$ for the interior of a set M . Consequently $\mathcal{B}(F) = B_1(F) = \text{int}(B_1(F)) \cup \partial B_1(F)$ denotes the unit ball in F . For the class of all bounded linear operators between normed spaces F and G we write $\mathcal{L}(F, G)$. The subset of all compact operators is denoted by $\mathcal{K}(F, G)$. We say a space F is (continuously) embedded into another space G with norm C if the operator norm of $\text{id}: F \rightarrow G$, $f \mapsto \text{id}(f) = f$, equals $C \in [0, \infty)$. In this case we write $F \hookrightarrow G$ and $\|\text{id}\|_{\mathcal{L}(F, G)} = C$. We use the symbol $\langle \cdot, \cdot \rangle_H$ for the inner product in the case of a Hilbert spaces H . Moreover, we write M^\perp for the orthogonal complement of some linear subspace $M \subset H$ and we use \oplus to denote the orthogonal sum with respect to $\langle \cdot, \cdot \rangle_H$.

If $(\mathcal{X}, \mathfrak{a}, \mu)$ is an arbitrary measure space and $0 < p \leq \infty$ then we use the symbol $L_p(\mathcal{X}, \mathfrak{a}, \mu)$ for the classical Lebesgue spaces. Hence, if $p < \infty$ then we deal with the set of (equivalence classes of) μ -measurable functions $f: \mathcal{X} \rightarrow \mathbb{R}$ for which the norm⁵

$$\|f\|_{L_p(\mathcal{X}, \mathfrak{a}, \mu)} = \left(\int_{\mathcal{X}} |f(x)|^p \, d\mu(x) \right)^{1/p}$$

is finite. Moreover, $L_\infty(\mathcal{X}, \mathfrak{a}, \mu)$ is the space (of classes) of μ -essentially bounded functions on \mathcal{X} , equipped with the norm

$$\|f\|_{L_\infty(\mathcal{X}, \mathfrak{a}, \mu)} = \text{ess-sup}_{x \in \mathcal{X}} |f(x)|.$$

As usual two functions are identified if they coincide μ -almost everywhere on \mathcal{X} and we do not distinguish between functions and their equivalence classes. The following special cases are of particular interest for us.

For a Borel measurable subset $\mathcal{X} = \Omega \subset \mathbb{R}^d$, the Borel sigma algebra $\mathfrak{a} = \Sigma = \Sigma(\Omega)$ and $\mu = \lambda^d$ we use the shorthand $L_p(\Omega) = L_p(\Omega, \Sigma, \lambda^d)$. If in this definition μ does not equal the Lebesgue measure, but is absolute continuous w.r.t. λ^d , and if $\varrho = d\mu/d\lambda^d$ describes a probability density function that is strictly positive (λ^d -a.e.) on Ω , then we write $L_p^\varrho(\Omega)$. On the other hand, for a discrete measure space $(\Gamma, \mathfrak{b}, \nu)$

⁵Actually, in the case $0 < p < 1$ the given formula only provides a quasi-norm, i.e. then we need an additional constant $k > 1$ for the triangle inequality. Since this does not play any role in our applications we do not emphasize this difference in what follows.

on some set Γ with $\nu(\{i\}) = 1$ for each $i \in \Gamma$ we write $\ell_p(\Gamma) = L_p(\Gamma, \mathfrak{b}, \nu)$ and we abbreviate the notation to ℓ_p if $\Gamma = \mathbb{N}$. Keep in mind that in this case the norms simplify to

$$\|\lambda\|_{\ell_p} = \begin{cases} \left(\sum_{m=1}^{\infty} |\lambda_m|^p \right)^{1/p}, & \text{if } 0 < p < \infty, \\ \sup_{m \in \mathbb{N}} |\lambda_m|, & \text{if } p = \infty, \end{cases}$$

where $\lambda = (\lambda_m)_{m \in \mathbb{N}}$ is any real-valued sequence such that the above norm is finite.

Finally, we make use of the Bachmann-Landau notation of asymptotic growth rates. That is, for real-valued functions f and g defined on some subset of the real line we write $f(x) \in \mathcal{O}(g(x))$, as $x \rightarrow a$, if there exists a universal constant $M > 0$ such that the estimate

$$|f(x)| \leq M |g(x)|$$

holds for all x sufficiently close to the point a . If g is non-zero (at least in the neighborhood of a) then this definition equivalently reads

$$\limsup_{x \rightarrow a} \left| \frac{f(x)}{g(x)} \right| < \infty.$$

If we have $f(x) \in \mathcal{O}(g(x))$ and simultaneously $g(x) \in \mathcal{O}(f(x))$, as $x \rightarrow a$, then we write $f(x) \in \Theta(g(x))$, $x \rightarrow a$. Moreover, we say that $f(x) \in o(g(x))$, as $x \rightarrow a$, if for any $\delta > 0$ there exists a neighborhood U of a such that

$$|f(x)| \leq \delta |g(x)|$$

for all $x \in U$. Again this property can be reformulated for non-vanishing g . In this case we have

$$\lim_{x \rightarrow a} \left| \frac{f(x)}{g(x)} \right| = 0.$$

All these three notations will be used especially for sequences $(f_a)_{a \in \mathbb{N}}$ (interpreted as special classes of functions), where we have $a = \infty$.

1.2 General problem

In numerous applications from physics, chemistry, finance, economics, and computer science we are faced with very high dimensional continuous problems which can almost never be solved analytically. Therefore we search for algorithms which approximate the unknown solutions numerically to within a threshold $\varepsilon > 0$.

In general, such a problem is given by a non-trivial *solution operator*

$$S: \tilde{\mathcal{F}} \rightarrow \mathcal{G}, \tag{1.1}$$

mapping a *problem element* f out of a subset $\tilde{\mathcal{F}}$ of some normed space \mathcal{F} onto its *solution* $S(f)$ in some (other) *target space* \mathcal{G} . Often, but not always, $\tilde{\mathcal{F}}$ is assumed to be the unit ball $\mathcal{B}(\mathcal{F})$ in some Banach space \mathcal{F} of multivariate functions $f: \Omega_d \subset \mathbb{R}^d \rightarrow \mathbb{R}$. For the domain of definition Ω_d usually the unit cube $[0, 1]^d$ is taken. Since the dependence on d will play a crucial role in this thesis we concentrate on whole sequences $S = (S_d)_{d \in \mathbb{N}}$ of solution operators, where every

$$S_d: \tilde{\mathcal{F}}_d \rightarrow \mathcal{G}_d, \quad d \in \mathbb{N}, \quad (1.2)$$

is of the form (1.1).

Typically, $\tilde{\mathcal{F}}_d$ is an infinite dimensional subset of the *source space* \mathcal{F}_d and thus we cannot input $f \in \tilde{\mathcal{F}}_d$ directly into the computer. Instead we assume that the input for our algorithms A consists of finitely many cleverly chosen pieces of information which hopefully describe f as well as possible. In [Section 1.3](#) we define different kinds of information operations which lead us to different classes \mathcal{A}_d of algorithms. For now assume $A_d: \tilde{\mathcal{F}}_d \rightarrow \mathcal{G}_d$ to be a fixed element in some class \mathcal{A}_d .

The *local error* $\Delta_{\text{loc}}(f; A_d, S_d)$ of a given algorithm $A_d \in \mathcal{A}_d$ applied to a problem element $f \in \tilde{\mathcal{F}}_d$ is defined as the difference of the exact solution $S_d(f)$ and the approximate solution $A_d(f)$, measured in the norm of the target space \mathcal{G}_d , i.e.

$$\Delta_{\text{loc}}(f; A_d, S_d) = \|S_d(f) - A_d(f)\|_{\mathcal{G}_d}.$$

The latter definition in hand, there are several ways to quantify the quality of A_d .

In the *worst case setting* this is done in terms of the maximal local error of the algorithm among all possible inputs $f \in \tilde{\mathcal{F}}_d$. Hence, by

$$\Delta^{\text{wor}}(A_d; S_d: \tilde{\mathcal{F}}_d \rightarrow \mathcal{G}_d) = \sup_{f \in \tilde{\mathcal{F}}_d} \Delta_{\text{loc}}(f; A_d, S_d)$$

we define the *worst case error* of the algorithm A_d for the problem $S_d: \tilde{\mathcal{F}}_d \rightarrow \mathcal{G}_d$. On the other hand, sometimes it is useful to measure the average performance of a given algorithm on the input set $\tilde{\mathcal{F}}_d$. This corresponds to the so-called *average case setting*. Here we need to assume in addition that $\tilde{\mathcal{F}}_d$ is equipped with a probability measure μ_d . The term

$$\Delta^{\text{avg}}(A_d; S_d: \tilde{\mathcal{F}}_d \rightarrow \mathcal{G}_d) = \left(\int_{\tilde{\mathcal{F}}_d} \Delta_{\text{loc}}(f; A_d, S_d)^2 d\mu_d(f) \right)^{1/2}$$

then denotes the *average case error* of A_d .⁶ Since the worst case setting seems to be much more important we will mainly deal with worst case errors in what follows. However, for some problems there exist close relations to the average case setting.

⁶In fact, μ_d is defined on the Borel sets of $\tilde{\mathcal{F}}_d$ and we need to claim $\Delta_{\text{loc}}(\cdot; A_d, S_d)$ to be a measurable function, but these are only formal issues. See, e.g., [NW08, p. 129] for further details.

One such example will be presented in [Section 4.4.2](#). For the sake of completeness we stress the point that there exist even more settings which are subject to current research. To this end, we mention the *probabilistic* and the *randomized setting* and refer to [\[NW08, Chapter 3.2\]](#) for an extensive discussion.

In numerical analysis one major assumption states that information is expensive. Therefore we are interested in algorithms which solve a given problem within a tolerance ε while using as few as possible pieces of information on the inputs. This property can be captured by the concept of the *n th minimal error*

$$e^{\text{sett}}\left(n, d; S_d: \tilde{\mathcal{F}}_d \rightarrow \mathcal{G}_d\right) = \inf_{A_{n,d} \in \mathcal{A}_d^n} \Delta^{\text{sett}}\left(A_{n,d}; S_d: \tilde{\mathcal{F}}_d \rightarrow \mathcal{G}_d\right)$$

for $\text{sett} \in \{\text{wor}, \text{avg}\}$, $d \in \mathbb{N}$, and $n \in \mathbb{N}_0$, where the infimum is taken over all algorithms in the class

$$\mathcal{A}_d^n = \{A \in \mathcal{A}_d \mid A \text{ uses at most } n \text{ information operations on the input}\}.$$

Consequently, the *initial error*

$$\varepsilon_d^{\text{init,sett}} = e^{\text{sett}}\left(0, d; S_d: \tilde{\mathcal{F}}_d \rightarrow \mathcal{G}_d\right), \quad d \in \mathbb{N},$$

describes the smallest error we can achieve without using any information on the input in a given setting $\text{sett} \in \{\text{wor}, \text{avg}\}$. We will see in [Section 2.2](#) that under mild assumptions this initial error can be attained by the zero algorithm, i.e. by $A_{0,d} \equiv 0 \in \mathcal{G}_d$.

If there is no danger of confusion we abbreviate the above notations and simply write $\Delta^{\text{sett}}(A_d; S_d)$ and $e^{\text{sett}}(n, d; S_d)$, where sett is an element of $\{\text{wor}, \text{avg}\}$, or even only $\Delta(A_d)$ and $e(n, d)$, respectively. Moreover, in [Chapter 4](#) and [Chapter 5](#) it seems to be useful to stress especially the source spaces \mathcal{F}_d the problem elements come from rather than the operator S_d . There we slightly abuse notation and write $e^{\text{wor}}(n, d; \mathcal{F}_d)$ instead of $e^{\text{wor}}(n, d; S_d: \mathcal{B}(\mathcal{F}_d) \rightarrow \mathcal{G}_d)$.

The main goal in the classical theory is to find sharp bounds on the n th minimal error in terms of the amount of information operations. In fact, there is a huge literature where the existence of constants $c_d, C_d > 0$ and $p_d, P_d > 0$ was proven such that estimates of the type

$$\frac{1}{c_d} \cdot n^{-p_d} \leq e(n, d) \leq C_d \cdot n^{-P_d} \quad \text{for all } n \in \mathbb{N}$$

hold for certain problems S in a given setting.⁷ Back then, the respective researchers did not pay much attention to the involved constants c_d and C_d . These numbers can be arbitrary large and in some cases their dependence on d is completely unknown.

⁷Actually, in many cases these estimates hold modulo $\log n$ to some power which usually depends linearly on d . For simplicity we omit these factors because they are not crucial for the following argument.

Instead the attention was focused on the so-called *rate* (or *order*) of convergence, i.e. on proofs which yield $p_d = P_d$. Often this rate tends to zero as d approaches infinity. Therefore these bounds are not meaningful at all for large d . Thus, usually the parameter d was assumed to be a fixed (and reasonably small) constant in this approach. Since we also want to work in huge dimensions a more careful error analysis is needed.

1.3 Algorithms and cost model

For fixed $n \in \mathbb{N}$ and $d \in \mathbb{N}$ an algorithm $A_{n,d} \in \mathcal{A}_d^n$ is modeled as a mapping $\varphi_n: \mathbb{R}^n \rightarrow \mathcal{G}_d$ and a function $N_n: \tilde{\mathcal{F}}_d \rightarrow \mathbb{R}^n$ such that $A_{n,d} = \varphi_n \circ N_n$. For the sake of completeness in the case $n = 0$ we simply assign a constant value $c \in \mathcal{G}_d$ to every element $f \in \tilde{\mathcal{F}}$, i.e. $A_{0,d} \equiv c$, in order to model an algorithm that does not depend on the input at all. If $n > 0$ then the *information map* N_n is given by

$$N_n(f) = (L_1(f), L_2(f), \dots, L_n(f)), \quad f \in \tilde{\mathcal{F}}_d, \quad (1.3)$$

where $L_j \in \Lambda$. Here we distinguish certain classes of information operations Λ . In one case we assume that we are allowed to compute arbitrary continuous linear functionals on the inputs f . Then $\Lambda = \Lambda^{\text{all}}$ coincides with \mathcal{F}_d^* , the dual space of \mathcal{F}_d . If we deal with problem operators S_d defined on function spaces $\tilde{\mathcal{F}}_d$ then often only function evaluations are permitted, i.e. $L_j(f) = f(t^{(j)})$ for a certain fixed $t^{(j)} \in \Omega_d$ in the domain of definition of f . In this case $\Lambda = \Lambda^{\text{std}}$ is called *standard information*. If function evaluation is continuous for all $t \in \Omega_d$ we have $\Lambda^{\text{std}} \subset \Lambda^{\text{all}}$. In particular this is the case when dealing with problems defined on reproducing kernel Hilbert spaces; see [Section 2.5](#). If L_j depends continuously on f but is not necessarily linear then the respective class is denoted by Λ^{cont} . Note that in this case also N_n is continuous and we obviously have $\Lambda^{\text{all}} \subset \Lambda^{\text{cont}}$.

Furthermore, we distinguish between *adaptive* and *non-adaptive* algorithms. The latter case is described above in formula (1.3), where L_j does not depend on the previously computed values $L_1(f), \dots, L_{j-1}(f)$. In contrast, we also discuss algorithms of the form $A_{n,d} = \varphi_n \circ N_n$ with

$$N_n(f) = (L_1(f), L_2(f; y_1), \dots, L_n(f; y_1, \dots, y_{n-1})), \quad f \in \tilde{\mathcal{F}}_d, \quad (1.4)$$

where $y_1 = L_1(f)$ and $y_j = L_j(f; y_1, \dots, y_{j-1})$ for $j = 2, 3, \dots, n$. If N_n is adaptive we restrict ourselves to the case where L_j depends linearly on f , e.g. $L_j(\cdot; y_1, \dots, y_{j-1}) \in \Lambda^{\text{all}}$. Note that in any case N_n is either continuous, or it is constructed out of linear information operations (which may be combined adaptively). Moreover, in all cases of information maps, the mapping φ_n can be chosen arbitrarily.

For upper error bounds small classes of algorithms are most important. The smallest such class under consideration is the family of linear, non-adaptive algo-

rithms of the form

$$A_{n,d}(f) = \sum_{j=1}^n L_j(f) \cdot g_j \quad (1.5)$$

with some $g_j \in \mathcal{G}_d$ and $L_j \in \Lambda^{\text{all}}$ or even $L_j \in \Lambda^{\text{std}}$. We denote this set of algorithms by $\mathcal{A}_d^{n,\text{lin}}(\Lambda)$, where $\Lambda = \Lambda^{\text{all}}$, or $\Lambda = \Lambda^{\text{std}}$, respectively. On the other hand, it is reasonable to prove lower error bounds for preferably large classes of algorithms. The most general families consist of algorithms $A_{n,d} = \varphi_n \circ N_n$, where φ_n is completely arbitrary and N_n either uses non-adaptive continuous or adaptive linear information. We denote the respective classes by $\mathcal{A}_d^{n,\text{cont}}$ and $\mathcal{A}_d^{n,\text{adapt}}$.

One of the most fundamental assumptions in IBC is that we can perform (exact) basic arithmetic operations on elements of the target space \mathcal{G}_d , as well as on real numbers, with unit cost. Formally this means that we work with the *real number model* in contrast to the *bit number model* which is used in some other fields of computational science; see, e.g., [NW08, Section 4.1.3]. Moreover, we assume that information operations on the input are given by certain black box computations which are sometimes called *oracle calls*. Typically the computational costs for information operations are much higher than for simple arithmetic operations since the computation of a function value or a linear functional may require billions of such operations. If we assume that every oracle call has a fixed cost $C \gg 1$ then the total cost of computing the output of an algorithm is proportional to the number of needed information operations.⁸ Therefore it is reasonable to study not only the n th minimal error of a given problem but also the inverse quantity which we call *information complexity*

$$\begin{aligned} n_{\text{abs}}^{\text{set}}(\varepsilon, d; S_d: \tilde{\mathcal{F}}_d \rightarrow \mathcal{G}_d) &= \min \{n \in \mathbb{N}_0 \mid \exists A \in \mathcal{A}_d^n \text{ such that } \Delta^{\text{set}}(A) \leq \varepsilon\} \\ &= \min \{n \in \mathbb{N}_0 \mid e^{\text{set}}(n, d) \leq \varepsilon\}, \end{aligned}$$

where $d \in \mathbb{N}$, $\varepsilon > 0$ and $\text{set} \in \{\text{wor}, \text{avg}\}$. That is, we look at the amount of oracle calls needed to compute an ε -approximation in dimension d . Hence, due to our assumptions this information complexity roughly equals the *total complexity* of a given problem and therefore describes its computational hardness. For a detailed discussion of algorithms and their costs, as well as on the relations of information complexity and total complexity we refer the reader to Section 4.1 in [NW08].

Finally we want to mention that the above definition addresses the *absolute error criterion*. In contrast we will also consider the *normalized error criterion* where we search for the minimal number of information operations needed to improve the initial error by some factor $\varepsilon' > 0$. We denote the corresponding

⁸There also exist approaches in which the cost of an oracle call depends on the parameters of the problem. These attempts stress the point that the computational effort for function evaluations increases with the numbers of (active) variables. See, e.g., [KSWW10a] for details.

information complexity by

$$n_{\text{norm}}^{\text{set}}(\varepsilon', d; S_d: \tilde{\mathcal{F}}_d \rightarrow \mathcal{G}_d) = n_{\text{abs}}^{\text{set}}(\varepsilon' \cdot \varepsilon_d^{\text{init}}, d; S_d: \tilde{\mathcal{F}}_d \rightarrow \mathcal{G}_d), \quad \text{set} \in \{\text{wor}, \text{avg}\}.$$

Obviously both the notions coincide if the problem under consideration is well-scaled. That is, if $\varepsilon_d^{\text{init}} = 1$. Otherwise the problem may be significantly harder with respect to the normalized error criterion, e.g. if $\varepsilon_d^{\text{init}}$ is exponentially small in d . Of course also the converse situation is conceivable. However, note that both the information complexities are always non-increasing in the first argument and we have

$$n_{\text{norm}}^{\text{set}}(1, d; S_d: \tilde{\mathcal{F}}_d \rightarrow \mathcal{G}_d) = n_{\text{abs}}^{\text{set}}(\varepsilon_d^{\text{init}}, d; S_d: \tilde{\mathcal{F}}_d \rightarrow \mathcal{G}_d) = 0 \quad (1.6)$$

for all $d \in \mathbb{N}$ due to the definition of the initial error.

Again we will use shorthands such as $n_{\text{abs}}(\varepsilon, d; S_d)$ or even $n(\varepsilon, d)$ to simplify notation.

1.4 Notions of tractability

As already indicated we strongly believe that it is not sufficient only to study the rate of convergence, i.e. the dependence of $n(\varepsilon, d)$ on ε , to properly describe the computational hardness of a given problem. We also need to incorporate the dependence on the parameter d . Keep in mind that the following definitions equally refer to both, the absolute and the normalized, error criteria. Therefore we simply write $n(\varepsilon, d)$ instead of $n_{\text{abs}}(\varepsilon, d)$ or $n_{\text{norm}}(\varepsilon, d)$ for the information complexity.

When dealing with multivariate problems we often observe the so-called *curse of dimensionality* which goes back to Bellman in the late 1950s; cf. [Bel57]. Given a concrete setting a problem is said to suffer from the curse of dimensionality if the corresponding information complexity $n(\varepsilon, d)$ increases exponentially with the dimension d . That is, for at least one $\varepsilon > 0$ there exist positive constants C and γ which are independent of the dimension such that we have

$$n(\varepsilon, d) \geq C \cdot (1 + \gamma)^d$$

for infinitely many $d \in \mathbb{N}$. More generally, if $n(\varepsilon, d)$ depends exponentially on d or ε^{-1} then we call the problem *intractable*⁹. Otherwise we have *tractability* which goes back to Woźniakowski in the early 1990s; see [Woź94a, Woź94b]. At this time a problem was called tractable if its complexity depends at most polynomially on ε^{-1} and d . Today this is only one case in a whole hierarchy of notions of tractability. We describe these classes starting with the weakest notion.

⁹Formally that means that there exist universal constants $\gamma, C > 0$, as well as sequences $(\varepsilon_k)_{k \in \mathbb{N}}$ and $(d_k)_{k \in \mathbb{N}}$ with $\varepsilon_k \in (0, 1]$ and $d_k \in \mathbb{N}$ for all $k \in \mathbb{N}$, such that $\varepsilon_k^{-1} + d_k \rightarrow \infty$, as $k \rightarrow \infty$, and $n(\varepsilon_k, d_k) \geq C \cdot (1 + \gamma)^{\varepsilon_k^{-1} + d_k}$ for every $k \in \mathbb{N}$. Note that this definition includes the curse as a special case, where $\varepsilon_k \equiv \varepsilon_0$.

If a problem is not intractable then we have *weak tractability* which can be equivalently expressed by

$$\lim_{\varepsilon^{-1}+d \rightarrow \infty} \frac{\ln(n(\varepsilon, d))}{\varepsilon^{-1} + d} = 0,$$

see [GW08, NW08]. Here the limit is taken with respect to all two-dimensional sequences $((\varepsilon_k, d_k))_{k \in \mathbb{N}} \subset (0, 1] \times \mathbb{N}$ such that $\varepsilon_k < \varepsilon_{d_k}^{\text{init}}$ and $\varepsilon_k^{-1} + d_k \rightarrow \infty$, as k approaches infinity. In particular, the latter restriction ensures that $n(\varepsilon, d) \geq 1$. Furthermore, we want to stress the point that weak tractability implies the absence of the curse of dimensionality, but in general the converse is not true. Recently a slightly stronger notion called *uniform weak tractability* has been suggested. We will not follow this line of research and refer to [Sie13].

Since there are many ways to measure the lack of exponential dependence the abstract notion of *generalized* (or *T-*) *tractability* was introduced; see [GW07, GW09] and [NW08, Chapter 8]. Here the essence is to describe the behavior of the information complexity in terms of a multiple of some power of a so-called *tractability function* T depending on ε^{-1} and d . Without going into details we mention that the following classes can be seen as special cases in this general framework.

For the sake of completeness we also introduce the quite recently developed notion of *quasi-polynomial tractability*. A problem is called quasi-polynomially tractable if there are universal constants $C, t > 0$ such that

$$n(\varepsilon, d) \leq C \exp(t(1 + \ln \varepsilon^{-1})(1 + \ln d))$$

for every $\varepsilon \in (0, 1]$ and $d \in \mathbb{N}$. Note that for fixed ε or d this upper bound behaves polynomially in the second argument what somehow justifies the name of this class of problems. For details see [GW11] and [NW12].

Finally, the most important and until now most studied type of tractability is called *polynomial tractability*. We say that a problem is polynomially tractable if there exist absolute constants $C, p > 0$ and $q \geq 0$ such that we can bound the information complexity by

$$n(\varepsilon, d) \leq C \cdot \varepsilon^{-p} \cdot d^q \quad \text{for all } d \in \mathbb{N}, \quad \varepsilon \in (0, 1]. \quad (1.7)$$

If this last inequality holds with $q = 0$, i.e. if we have no dependence on the dimension at all, then the problem is called *strongly polynomially tractable*. In this case the smallest possible constant p in (1.7) is denoted by p^* . It is called the *exponent of strong polynomial tractability*.

If, in contrast, there do not exist constants C, p and q which fulfill (1.7) then the problem is said to be *polynomially intractable*.

Observe that (1.6) shows that, as long as the absolute error criterion is concerned, it is enough to consider $\varepsilon \in (0, \min\{\varepsilon_d^{\text{init}}, 1\}]$ instead of $\varepsilon \in (0, 1]$ in all the above definitions.

CHAPTER 2

Properties and tools for special problem classes

This chapter deals with basic properties of certain classes of problems and algorithms. We state simple consequences obtained from fundamental assumptions on the operators under consideration. Furthermore, we present more or less classical tools used in the framework of information-based complexity to acquire tractability results in a quite general context.

In detail, we begin with a simple lower error bound in a very general setting which will be used on several occasions later on. In [Section 2.2](#) we then show that for our purposes it is reasonable to concentrate mainly on compact problems and linear, non-adaptive algorithms. Moreover, there we derive a formula for the initial error of the problems we are interested in. Afterwards, in [Section 2.3](#), we turn to the important class of problems defined between Hilbert spaces. We recall well-known tools such as the singular value decomposition, conclude optimal algorithms and characterize several types of tractabilities of such problems. In [Section 2.4](#) we restrict ourselves further and assume an additional tensor product structure which will play an important role throughout the rest of this thesis. Finally we conclude this chapter with the discussion of so-called reproducing kernel Hilbert spaces.

The main references for the functional analytic background needed in this part, as well as on the theory of s -numbers (or n -widths, respectively) are the monographs of Pinkus [[Pin85](#)] and Pietsch [[Pie87](#), [Pie07](#)]. For a detailed discussion of applications to tractability questions we refer again to Novak and Woźniakowski [[NW08](#), [NW10](#), [NW12](#)] and to Mathé [[Mat90](#)].

2.1 Lower bounds on linear subspaces

For the purpose of this chapter it is enough to study the worst case setting. In addition, we will only focus on the case where all the problem elements lie in some centered ball of the respective source space. In this first section we present a quite general method to obtain lower bounds on the n th minimal error with respect to a wide class of algorithms. In contrast to the rest of this thesis (where we will restrict ourselves basically to linear and compact problems) we present a result that holds

for any homogeneous operator S between linear normed spaces \mathcal{F} and \mathcal{G} over the field of real numbers. That is, we first only assume that $S(\alpha \cdot f) = \alpha \cdot S(f)$ for every $f \in \mathcal{F}$ and all $\alpha \in \mathbb{R}$.

We start by proving the following (modified) assertion of Borsuk and Ulam for linear normed spaces:

Lemma 2.1 (Borsuk-Ulam). Let V be a linear normed space over \mathbb{R} with $0 < \dim V = s < \infty$ and, moreover, let $N: V \rightarrow \mathbb{R}^n$ be a continuous mapping for some $0 \leq n < s$. Then for all $r \geq 0$ there exists an element $f^* \in V$ with $\|f^* | V\| = r$, such that $N(f^*) = N(-f^*)$.

Proof. Obviously, the cases $n = 0$, i.e. $N \equiv 0$, and $r = 0$ are trivial. Hence, let $n \in \mathbb{N}$ and $r > 0$. Since $\dim V = s$ we find an isomorphism $T: V \rightarrow \mathbb{R}^s$ such that T and T^{-1} are linear and bounded. Hence, for every $r > 0$ the set $\Omega_r = T(\text{int}(B_r(V)))$ is an open, bounded and symmetric subset of \mathbb{R}^s which contains zero. Moreover, the function $g = N \circ T^{-1}: \partial\Omega_r \rightarrow \mathbb{R}^n$ is continuous. From the theorem of Borsuk-Ulam (cf. Deimling [Dei85, Corollary 4.2]) we conclude the existence of some $\mathbf{x}^* \in \partial\Omega_r$ with $g(\mathbf{x}^*) = g(-\mathbf{x}^*)$. The claim now follows by taking $f^* = T^{-1}\mathbf{x}^*$. ■

This result in hand, we can prove a generalization of [Wei12b, Lemma 1].

Proposition 2.2. Suppose S to be a homogeneous operator between linear normed spaces \mathcal{F} and \mathcal{G} . Further assume that $V \subset \mathcal{F}$ is a linear subspace with dimension $s \in \mathbb{N}$ and that there exists a constant $a \geq 0$ such that

$$a \cdot \|f | \mathcal{F}\| \leq \|S(f) | \mathcal{G}\| \quad \text{for all } f \in V. \quad (2.1)$$

Then for every $0 \leq n < s$, any algorithm $A_n \in \mathcal{A}^{n,\text{cont}} \cup \mathcal{A}^{n,\text{adapt}}$, and all $r \geq 0$

$$\Delta^{\text{wor}}(A_n; S: B_r(\mathcal{F}) \rightarrow \mathcal{G}) = \sup_{f \in B_r(\mathcal{F})} \|S(f) - A_n(f) | \mathcal{G}\| \geq a \cdot r. \quad (2.2)$$

In particular, the n th minimal worst case error (among the unit ball $\mathcal{B}(\mathcal{F})$ of \mathcal{F}) satisfies $e^{\text{wor}}(n; S: \mathcal{B}(\mathcal{F}) \rightarrow \mathcal{G}) \geq a$ for all $n < s$.

Proof. It is well-known that for $A_n = \varphi_n \circ N_n \in \mathcal{A}^{n,\text{cont}} \cup \mathcal{A}^{n,\text{adapt}}$ with $n < s$ there exists $f^* \in V$ such that $N_n(f^*) = N_n(-f^*)$ and $\|f^* | \mathcal{F}\| = r$.

Without loss of generality let us again assume $n \in \mathbb{N}$ and $r > 0$ to avoid triviality. Then, for $A_n \in \mathcal{A}^{n,\text{cont}}$, the existence of f^* is a simple conclusion of Lemma 2.1 since in this case N_n is continuous by definition. On the other hand, if $A_n \in \mathcal{A}^{n,\text{adapt}}$ then the proof can be obtained by arguments from linear algebra. We follow the lines of the proof of Werschulz and Woźniakowski [WW09, Theorem 3.1] and search for a nonzero $g \in V$ such that $N_n(g) = 0$, i.e.

$$\begin{aligned} L_1(g) &= 0, \\ L_2(g; 0) &= 0, \\ &\vdots \\ L_n(g; 0, \dots, 0) &= 0. \end{aligned} \quad (2.3)$$

Since $\dim V = s$ every $g \in V$ can be represented uniquely as a linear combination $g = \sum_{m=1}^s c_m b_m$ of at most s linearly independent basis functions b_m of V . Due to the imposed linearity of $L_j(\cdot; 0, \dots, 0)$, $j = 1, \dots, n$, the system (2.3) can be reformulated as a system of n homogeneous linear equations in the $s > n$ unknowns $\mathbf{c} = (c_m)_{m=1}^s \in \mathbb{R}^s$. Consequently, it possesses a non-trivial solution $\mathbf{c}^* = (c_m^*)_{m=1}^s$ which implies the existence of some $g^* \in V \setminus \{0\}$ with $N_n(g^*) = 0$. Since with L_j also N_n is linear, we can easily construct f^* out of g^* .

Anyway, every such f^* satisfies $A_n(f^*) = A_n(-f^*)$. Using the norm properties in the target space \mathcal{G} and the homogeneity of S we obtain (2.2):

$$\begin{aligned} \Delta^{\text{wor}}(A_n; S: B_r(\mathcal{F}) \rightarrow \mathcal{G}) &\geq \max \{ \|S(\pm f^*) - A_n(\pm f^*)\|_{\mathcal{G}} \} \\ &= \max \{ \|S(f^*) \pm A_n(f^*)\|_{\mathcal{G}} \} \\ &\geq \frac{1}{2} (\|S(f^*) + A_n(f^*)\|_{\mathcal{G}} + \|S(f^*) - A_n(f^*)\|_{\mathcal{G}}) \\ &\geq \frac{1}{2} \|2 \cdot S(f^*)\|_{\mathcal{G}} \geq a \|f^*\|_{\mathcal{F}} = a \cdot r. \end{aligned}$$

The remaining implication for the n th minimal error finally follows from the case $r = 1$ by taking the infimum over all $A_n \in \mathcal{A}^{n, \text{cont}} \cup \mathcal{A}^{n, \text{adapt}}$. \blacksquare

At this point we stress that the case $r \neq 1$ in (2.2) might be useful only if we deal with non-homogeneous (and thus non-linear) algorithms A_n . Otherwise we clearly have

$$\Delta^{\text{wor}}(A_n; S: B_r(\mathcal{F}) \rightarrow \mathcal{G}) = r \cdot \Delta^{\text{wor}}(A_n; S: \mathcal{B}(\mathcal{F}) \rightarrow \mathcal{G})$$

for all $r \geq 0$ provided that $S: \mathcal{F} \rightarrow \mathcal{G}$ is homogeneous. The importance of (2.2) for $r \neq 1$ will be made clear in Section 4.3.2 when we deal with embeddings $\mathcal{P} \hookrightarrow \mathcal{F}$. There we conclude a lower bound for the worst case error of S on $\mathcal{B}(\mathcal{F})$ out of a lower bound on $B_r(\mathcal{P})$ using $r = \|\text{id}|_{\mathcal{L}(\mathcal{P}, \mathcal{F})}\|^{-1}$.

2.2 Linearity and compactness

In what follows we will exclusively consider linear continuous problems $S = (S_d)_{d \in \mathbb{N}}$. That is, we assume every solution operator S_d given by (1.2) to be the restriction of a bounded linear mapping between some Banach spaces defined over the field of real numbers.¹⁰ If we assume the set of problem elements $\tilde{\mathcal{F}}_d$ to be some centered ball $B_r(\mathcal{F}_d)$, $r > 0$, in the source space then conversely every bounded mapping S_d that acts linearly on this set¹¹ can be uniquely extended to a continuous linear

¹⁰In fact, for most of the following results completeness is not needed. Many of them even remain valid (at least up to constants) using only quasi-norms or p -norms, but for simplicity we restrict ourselves to the case of Banach spaces. Finally, for the ease of notation, we only consider spaces over \mathbb{R} .

¹¹That means, $S_d(\alpha \cdot f + \beta \cdot g)$ equals $\alpha \cdot S_d(f) + \beta \cdot S_d(g)$ for every convex combination $\alpha \cdot f + \beta \cdot g$ of elements $f, g \in \tilde{\mathcal{F}}_d$.

operator \tilde{S}_d on the whole space \mathcal{F}_d , i.e. $\tilde{S}_d \in \mathcal{L}(\mathcal{F}_d, \mathcal{G}_d)$. From this point of view S_d and \tilde{S}_d can be identified with each other and thus we use the symbol S_d for both of them.

At the first glance the linearity assumption seems to be very restrictive. On the other hand, both the most important problems, namely approximation and integration, are indeed of this type. Moreover, the linear case is much better understood than the non-linear such that an overwhelming percentage of work on IBC was done in this setting. For the sake of completeness we also mention so-called quasilinear problems and refer to [WW07] and [NW12, Chapter 28].

Since we are interested in algorithms which are easy (and cheap) to implement we pay special attention to the family of linear and non-adaptive algorithms $\mathcal{A}_d^{n,\text{lin}}(\Lambda)$; see (1.5). It is well-known that this choice is reasonable for many classes of problems, since it can be shown that under mild assumptions optimal algorithms are indeed linear and non-adaptive. General assertions of this type can be found in Traub, Wasilkowski and Woźniakowski [TWW88], as well as in Novak and Woźniakowski [NW08, Section 4.2]. We do not present these results here explicitly. The reason is that for the problems we are interested in, our assertions already imply the mentioned optimality statements.

Furthermore, we focus on information maps which are linear and continuous, i.e. $\Lambda \subseteq \Lambda^{\text{all}}$. Observe that then $A_{n,d} \in \mathcal{L}(\mathcal{F}_d, \mathcal{G}_d)$ and $\text{rank}(A_{n,d}) \leq n$. Moreover, for $\tilde{\mathcal{F}}_d = \mathcal{B}(\mathcal{F}_d)$ we obtain

$$\Delta^{\text{wor}}(A_{n,d}; S_d) = \Delta^{\text{wor}}\left(A_{n,d}; S_d: \tilde{\mathcal{F}}_d \rightarrow \mathcal{G}_d\right) = \|S_d - A_{n,d} | \mathcal{L}(\mathcal{F}_d, \mathcal{G}_d)\|.$$

It seems natural to ask when problems of this type are solvable at all. We say a problem $S = (S_d)_{d \in \mathbb{N}}$ is *solvable* if for any fixed $d \in \mathbb{N}$ there exists a sequence of algorithms $A_{n,d} \in \mathcal{A}_d^{n,\text{lin}}(\Lambda^{\text{all}})$ such that their worst case errors $\Delta^{\text{wor}}(A_{n,d}; S_d)$ tend to zero as n approaches infinity. Hence, S_d needs to be an element of $\overline{\mathcal{FR}(\mathcal{F}_d, \mathcal{G}_d)}$, the closure of the finite rank operators in $\mathcal{L}(\mathcal{F}_d, \mathcal{G}_d)$, which is a subset of $\mathcal{K}(\mathcal{F}_d, \mathcal{G}_d)$. Therefore solvable problems are necessarily compact such that we can restrict ourselves in the following to $S_d \in \mathcal{K}(\mathcal{F}_d, \mathcal{G}_d)$. Due to the celebrated result of Enflo [Enf73] it is known that the converse is not true in this generality. Indeed, there are compact problems which are not solvable since there exist Banach spaces \mathcal{G}_d which do not satisfy the so-called approximation property. However, the following (incomplete) list shows that in the cases we are interested in every compact problem is solvable:

Proposition 2.3. Let $S = (S_d)_{d \in \mathbb{N}}$ be given such that $S_d \in \mathcal{K}(\mathcal{F}_d, \mathcal{G}_d)$ for all $d \in \mathbb{N}$. Then S is solvable if for every $d \in \mathbb{N}$ one of the following conditions applies:

- The source space \mathcal{F}_d is a Hilbert space, or
- The target space \mathcal{G}_d is a Hilbert space, or
- The target space \mathcal{G}_d is $L_\infty(\mathcal{X}, \mathbf{a}, \mu)$ for an arbitrary measure space $(\mathcal{X}, \mathbf{a}, \mu)$.

Proof. Let $d \in \mathbb{N}$. Given all the above restrictions we note that if we consider the class $\mathcal{A}_d^{n,\text{lin}}(\Lambda^{\text{all}})$ then the numbers $e^{\text{wor}}(n, d; S_d)$, $n \in \mathbb{N}_0$, per definition equal the *linear n -widths* (or *approximation numbers*) $\delta_n(S_d)$ as defined in [Pin85, Definition 7.3]. Up to an index shift these numbers form an s -scale¹² in the sense of Pietsch [Pie07, Section 6.2]. Other important s -scales are the *Gelfand numbers* $c_n(S_d)$ and the *Kolmogorov numbers* $d_n(S_d)$. Without going into details we mention that for any compact operator $S_d \in \mathcal{K}(\mathcal{F}_d, \mathcal{G}_d)$ both these numbers tend to zero as $n \rightarrow \infty$; see Propositions 7.4 and 7.1 in [Pin85]. Hence, to prove solvability it suffices to show that $\delta_n(S_d) \leq \max\{c_n(S_d), d_n(S_d)\}$ for all $n \in \mathbb{N}$. Indeed, if \mathcal{F}_d is a Hilbert space then we have $\delta_n(S_d) = c_n(S_d)$. Furthermore, $\delta_n(S_d) = d_n(S_d)$ if \mathcal{G}_d is a Hilbert space; see, e.g., [Pin85, p. 33]. Finally Proposition 8.13 in [Pin85] shows that the second last equality remains valid if the target space \mathcal{G}_d enjoys the so-called (*metric*) *extension property*. It is known that in particular $L_\infty(\mathcal{X}, \mathbf{a}, \mu)$ has this property; see, e.g., König [Kön86, 1.c.2]. ■

We want to stress that in Proposition 2.3 we do not need to assume the Hilbert spaces to be separable.

Let us conclude this section with a proposition which shows that the zero algorithm $A_0 \equiv 0$ is the optimal choice among all approximations to a given operator $S \in \mathcal{L}(\mathcal{F}, \mathcal{G})$ that do not use any information on the input $f \in \mathcal{F}$. Here \mathcal{F} and \mathcal{G} can be arbitrary normed spaces.

Proposition 2.4. For $S \in \mathcal{L}(\mathcal{F}, \mathcal{G})$ and $A_0 = 0 \in \mathcal{L}(\mathcal{F}, \mathcal{G})$ we have

$$e^{\text{wor}}(0; S: \mathcal{B}(\mathcal{F}) \rightarrow \mathcal{G}) = \Delta^{\text{wor}}(A_0; S: \mathcal{B}(\mathcal{F}) \rightarrow \mathcal{G}) = \|S|_{\mathcal{L}(\mathcal{F}, \mathcal{G})}\|.$$

Consequently, the zero algorithm is optimal for S within the class $\mathcal{A}^{0,\text{cont}} \cup \mathcal{A}^{0,\text{adapt}}$ and the initial worst case error $\varepsilon^{\text{init},\text{wor}}$ of S is given by the its operator norm.

Proof. Obviously the second equality is true by the definition of Δ^{wor} . Moreover, the linear algorithm $A_0 \equiv 0$ is included in every class of algorithms we defined in Section 1.3. This particularly implies $e^{\text{wor}}(0; S) \leq \Delta^{\text{wor}}(A_0) = \|S|_{\mathcal{L}(\mathcal{F}, \mathcal{G})}\|$.

To show the converse inequality, recall that every algorithm A that does not use any information on the input necessarily takes the form $A(f) \equiv g$ for some element $g \in \mathcal{G}$. A calculation similar to that in the proof of Proposition 2.2 yields that

$$\|S(f)|_{\mathcal{G}}\| \leq \max\{\|S(f) - g|_{\mathcal{G}}\|, \|S(-f) - g|_{\mathcal{G}}\|\}$$

holds for every $f \in \mathcal{F}$. Taking the supremum over $f \in \mathcal{B}(\mathcal{F})$ now shows that $\|S|_{\mathcal{L}(\mathcal{F}, \mathcal{G})}\| \leq \Delta^{\text{wor}}(A)$ which implies the desired result since A was chosen arbitrary. ■

We note in passing that the last step in the latter proof crucially depends on the fact that the unit ball $\tilde{\mathcal{F}} = \mathcal{B}(\mathcal{F})$ of the source space \mathcal{F} is symmetric in the sense that $f \in \tilde{\mathcal{F}}$ implies $-f \in \tilde{\mathcal{F}}$.

¹²Note that due to historical reasons there is some notational danger concerning s -numbers versus n -widths. See, e.g., [Pie07, p. 336] for details.

2.3 General Hilbert space problems

In this section we describe the *singular value decomposition* (SVD) which turns out to be the main tool when dealing with problems where both the source and the target spaces are Hilbert spaces. We prove well-known formulas for optimal linear algorithms using continuous linear functionals and calculate their worst case errors. Afterwards, we use the obtained assertions to give characterizations for (strong) polynomial tractability for these problems.

2.3.1 Singular value decomposition

Given any compact operator $T \in \mathcal{K}(\mathcal{F}, \mathcal{G})$ acting between two arbitrary real Hilbert spaces \mathcal{F} and \mathcal{G} we define its *adjoint operator* $T^\dagger: \mathcal{G} \rightarrow \mathcal{F}$ in the usual way by

$$\langle Tf, g \rangle_{\mathcal{G}} = \langle f, T^\dagger g \rangle_{\mathcal{F}}, \quad \text{for all } f \in \mathcal{F}, g \in \mathcal{G}. \quad (2.4)$$

Of course, T^\dagger is always unique and well-defined. For details we refer the reader to Yosida [Yos80, VII.2]. If $\mathcal{F} = \mathcal{G}$ and $T^\dagger = T$, then we say that T is *self-adjoint*. Due to Schauder's Theorem we know that $T^\dagger \in \mathcal{K}(\mathcal{G}, \mathcal{F})$ if and only if $T \in \mathcal{K}(\mathcal{F}, \mathcal{G})$; see, e.g. [Pin85, p. 31]. Hence, it is easily seen that also

$$W = T^\dagger T: \mathcal{F} \rightarrow \mathcal{F}$$

defines a compact operator. Moreover, W is obviously self-adjoint and positive, i.e. $\langle Wf, f \rangle_{\mathcal{F}} \geq 0$ for every $f \in \mathcal{F}$. It is a well-known fact that therefore all the eigenvalues $\lambda_m = \lambda_m(W)$ of W are necessarily real and furthermore non-negative. Following Pinkus [Pin85, p. 64] we denote the sum of the algebraic multiplicities of the non-zero eigenvalues of W by $v = v(W)$. Note that the theory of Riesz-Schauder provides that there are at most countably many non-zero eigenvalues. They are uniformly bounded, each of them has a finite multiplicity and there are no accumulation points but (possibly) zero. See, e.g., Theorem 2 in [Yos80, X.5]. Observe further that in any case $v \leq \dim \mathcal{F} \in \mathbb{N} \cup \{\infty\}$. Let us denote these eigenvalues in a non-increasing ordering subscripted by indices from the set $\mathcal{M} = \{m \in \mathbb{N} \mid m < v + 1\}$,

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq \dots > 0. \quad (2.5)$$

Note that without loss of generality we will always assume the existence of at least one non-trivial eigenvalue, i.e. we explicitly exclude the operator $T \equiv 0$ which ensures that $\mathcal{M} \neq \emptyset$. We denote the corresponding (mutually orthonormal) eigenvectors of W by ϕ_m , $m \in \mathcal{M}$, and refer to $\{(\lambda_m, \phi_m) \mid m \in \mathcal{M}\}$ as the set of *non-trivial eigenpairs* of W . Consequently, for $i, j \in \mathcal{M}$ we have by (2.4)

$$\langle T\phi_i, T\phi_j \rangle_{\mathcal{G}} = \langle \phi_i, T^\dagger T\phi_j \rangle_{\mathcal{F}} = \langle \phi_i, W\phi_j \rangle_{\mathcal{F}} = \langle \phi_i, \lambda_j \phi_j \rangle_{\mathcal{F}} = \delta_{i,j} \cdot \lambda_j. \quad (2.6)$$

If we extend the possibly finite eigenvalue sequence $(\lambda_m)_{m=1}^v$ by taking $\lambda_m = 0$ for all $m > v$ then, clearly, $\lambda = (\lambda_m)_{m \in \mathbb{N}}$ forms a null sequence. Again following Pinkus, we call the square root $\sigma = \sigma(T)$ of $\lambda = \lambda(W)$,

$$\sigma_m = \sqrt{\lambda_m}, \quad m \in \mathbb{N},$$

sequence of the *singular values* of T . The importance of this bunch of definitions comes from the following assertion.

Theorem 2.5 (Singular value decomposition). Let \mathcal{F} and \mathcal{G} be arbitrary Hilbert spaces and $T \in \mathcal{K}(\mathcal{F}, \mathcal{G})$. Then, with the above notations,

$$T = \sum_{m=1}^v \langle \cdot, \phi_m \rangle_{\mathcal{F}} T \phi_m. \quad (2.7)$$

Proof. A detailed proof can be found in the monograph of König [Kön86, 1.b.3]. It is mainly based on the so-called *polar decomposition* of linear continuous operators and the theory of Riesz-Schauder. Actually, the proof deals with complex Hilbert spaces but it literally transfers to the real case. Moreover, only the existence of an orthonormal sequence $(\psi_m)_{m=1}^v$ in \mathcal{G} is shown such that the pointwise equality

$$Tf = \sum_{m=1}^v \sigma_m \langle f, \phi_m \rangle_{\mathcal{F}} \psi_m, \quad f \in \mathcal{F},$$

holds true. However, setting $f = \phi_k$ for $k \in \mathcal{M}$ together with the mutual orthonormality of $(\phi_m)_{m=1}^v$ immediately implies $\sigma_k \psi_k = T \phi_k$ for any k . The claimed identity in $\mathcal{K}(\mathcal{F}, \mathcal{G})$ finally follows from Bessel's inequality. ■

Remark 2.6. Note that again the Hilbert spaces \mathcal{F} and \mathcal{G} do not need to be separable. Nevertheless the image of \mathcal{F} under T is indeed separable, because it is spanned by at most countable many elements $T \phi_m \in \mathcal{G}$. Since the elements of the set $\Phi = \{\phi_m \in \mathcal{F} \mid m \in \mathcal{M}\}$ are mutually orthonormal we can extend Φ to an orthonormal basis (ONB) E of \mathcal{F} . Then (2.7) shows that $\ker T = \Phi^\perp$. Remember that we are only interested in the approximation of the image of T . Hence, we can without loss of generality assume that $E = \Phi$. In other words, even though \mathcal{F} may be non-separable in general we can restrict ourselves to the separable case in what follows. We only need to replace \mathcal{F} by $\overline{\Phi}$, the closure of the orthonormal eigenelements of $W = T^*T$ under $\langle \cdot, \cdot \rangle_{\mathcal{F}}$. □

2.3.2 Optimal algorithm

Observe that by (2.7) we obtained a representation of any operator $T \in \mathcal{K}(\mathcal{F}, \mathcal{G})$ as the limit of related finite rank operators. Therefore we are able to construct n th optimal linear algorithms which only use information from Λ^{all} . This is stated in the following corollary which can be found (slightly modified) as Corollary 4.12 in [NW08].

Corollary 2.7. For $d \in \mathbb{N}$ assume \mathcal{F}_d and \mathcal{G}_d to be arbitrary Hilbert spaces. Further let $S = (S_d)_{d \in \mathbb{N}}$ denote a compact problem acting between these spaces, i.e. $S_d \in \mathcal{K}(\mathcal{F}_d, \mathcal{G}_d)$ for every d . Then for all $d \in \mathbb{N}$ and $n \in \mathbb{N}_0$ the algorithm $A_{n,d}^* \in \mathcal{A}_d^{n,\text{lin}}(\Lambda^{\text{all}})$ given by

$$A_{n,d}^*: \mathcal{F}_d \rightarrow \mathcal{G}_d, \quad f \mapsto A_{n,d}^*(f) = \sum_{m=1}^{\min\{n, v(W_d)\}} \langle f, \phi_{d,m} \rangle_{\mathcal{F}_d} \cdot S_d \phi_{d,m},$$

for S_d is optimal in the class $\mathcal{A}_d^{n,\text{cont}} \cup \mathcal{A}_d^{n,\text{adapt}}$ and we have

$$e^{\text{wor}}(n, d; S_d) = \Delta^{\text{wor}}(A_{n,d}^*; S_d) = \sigma_{d,n+1} = \sqrt{\lambda_{d,n+1}}. \quad (2.8)$$

Here for every $d \in \mathbb{N}$ the singular values $(\sigma_{d,m})_{m \in \mathbb{N}}$, as well as the eigenvectors $(\phi_{d,m})_{m=1}^{v(W_d)}$, are constructed out of $W_d = S_d^\dagger S_d$ as explained above.

Proof. Recall that $\Delta^{\text{wor}}(A_{n,d}^*; S_d)$ equals $\|S_d - A_{n,d}^* | \mathcal{L}(\mathcal{F}_d, \mathcal{G}_d)\|$ for any fixed $d \in \mathbb{N}$ and $n \in \mathbb{N}_0$. Without loss of generality we can assume $n < v = v(W_d)$ since otherwise $A_{n,d}^* = S_d$ due to (2.7). This would imply (2.8) because of $\sigma_{d,m} = 0$ for all $m > v$.

Let $M \in \mathbb{N}$ with $n+1 \leq M \leq v$ and $f \in \mathcal{B}(\mathcal{F}_d)$ be arbitrarily fixed. Then, due to (2.6), the non-increasing ordering of $(\lambda_m)_{m=1}^v$ and Bessel's inequality,

$$\begin{aligned} \left\| \sum_{m=n+1}^M \langle f, \phi_{d,m} \rangle_{\mathcal{F}_d} S_d \phi_{d,m} \Big| \mathcal{G}_d \right\|^2 &= \sum_{m=n+1}^M \langle f, \phi_{d,m} \rangle_{\mathcal{F}_d}^2 \lambda_{d,m} \leq \lambda_{d,n+1} \sum_{m=1}^v \langle f, \phi_{d,m} \rangle_{\mathcal{F}_d}^2 \\ &\leq \lambda_{d,n+1} \|f | \mathcal{F}_d\|^2 \leq \sigma_{d,n+1}^2. \end{aligned}$$

In particular, the choice $f = \phi_{d,n+1}$ shows that the latter estimates are sharp. Anyway, we obtain $\|S_d - A_{n,d}^* | \mathcal{L}(\mathcal{F}_d, \mathcal{G}_d)\| \leq \sigma_{d,n+1}$ which proves

$$e^{\text{wor}}(n, d; S_d) \leq \Delta^{\text{wor}}(A_{n,d}^*; S_d) \leq \sigma_{d,n+1}.$$

To show the converse, i.e. $e^{\text{wor}}(n, d; S_d) \geq \sigma_{d,n+1}$ for $n \in \mathbb{N}_0$ and $d \in \mathbb{N}$, we use Parseval's identity on $V = \text{span}\{\phi_{d,m} | m \leq n+1\} \subset \mathcal{F}_d$ together with the linearity of S_d to obtain

$$\|S_d f | \mathcal{G}_d\|^2 = \sum_{m=1}^{n+1} \langle f, \phi_{d,m} \rangle_{\mathcal{F}_d}^2 \lambda_{d,m} \geq \lambda_{d,n+1} \sum_{m=1}^{n+1} \langle f, \phi_{d,m} \rangle_{\mathcal{F}_d}^2 = \sigma_{d,n+1}^2 \|f | \mathcal{F}_d\|^2$$

for all $f \in V$. The claim now follows from the application of [Proposition 2.2](#) with $a = \sigma_{d,n+1}$. Moreover, [Proposition 2.2](#) also shows that we cannot reduce the error by taking algorithms $A_{n,d} \in (\mathcal{A}_d^{n,\text{cont}} \cup \mathcal{A}_d^{n,\text{adapt}}) \setminus \mathcal{A}_d^{n,\text{lin}}$. \blacksquare

Note that (2.8) together with [Proposition 2.4](#) particularly implies that for $d \in \mathbb{N}$ the initial worst case error of S_d is given by

$$\varepsilon_d^{\text{init,wor}} = \|S_d | \mathcal{L}(\mathcal{F}_d, \mathcal{G}_d)\| = \sigma_{d,1} = \sqrt{\lambda_{d,1}}.$$

2.3.3 Polynomial tractability

As an immediate consequence of (2.8) we can calculate the information complexity of Hilbert space problems in the worst case setting (with respect to the class $\mathcal{A}_d^{n,\text{lin}}(\Lambda^{\text{all}})$) for every $d \in \mathbb{N}$ and $\varepsilon > 0$ by

$$n_{\text{abs}}^{\text{wor}}(\varepsilon, d) = \min \{n \in \mathbb{N}_0 \mid \sigma_{d,n+1} \leq \varepsilon\} = \# \{n \in \mathbb{N} \mid \lambda_{d,n} > \varepsilon^2\} \quad (2.9)$$

for the absolute and by

$$n_{\text{norm}}^{\text{wor}}(\varepsilon, d) = \# \{n \in \mathbb{N} \mid \lambda_{d,n}/\lambda_{d,1} > \varepsilon^2\}, \quad (2.10)$$

for the normalized error criterion, respectively. This observation leads to the following refinement of Theorem 5.1 in Novak and Woźniakowski [NW08] which also can be found in [Wei11]. It gives necessary and sufficient conditions for (strong) polynomial tractability in terms of summability properties of the sequences $(\lambda_{d,m})_{m \in \mathbb{N}}$.

Theorem 2.8. Assume S to be a problem as in Corollary 2.7 and consider the absolute error criterion in the worst case setting.

- If S is polynomially tractable with the constants $C, p > 0$ and $q \geq 0$ then for all $\tau > p/2$ we have

$$C_\tau = \sup_{d \in \mathbb{N}} \frac{1}{d^r} \left(\sum_{i=f(d)}^{\infty} \lambda_{d,i}^\tau \right)^{1/\tau} < \infty, \quad (2.11)$$

where $r = 2q/p$ and $f: \mathbb{N} \rightarrow \mathbb{N}$ with $f(d) = \lceil (1+C)d^q \rceil$. In this case $C_\tau \leq C^{2/p} \cdot \zeta(2\tau/p)^{1/\tau}$.

- If (2.11) is satisfied for some parameters $r \geq 0, \tau > 0$ and a function $f: \mathbb{N} \rightarrow \mathbb{N}$ such that $f(d) = \lceil C (\min \{\varepsilon_d^{\text{init}}, 1\})^{-p} d^q \rceil$, where $C > 0$ and $p, q \geq 0$, then the problem S is polynomially tractable. In detail, we have the bound $n_{\text{abs}}^{\text{wor}}(\varepsilon, d) \leq (C + C_\tau^\tau) \varepsilon^{-\max\{p, 2\tau\}} d^{\max\{q, r\tau\}}$ for any $\varepsilon \in (0, 1]$ and every $d \in \mathbb{N}$.

Proof. If the problem is polynomially tractable then there exist constants $C, p > 0$ and $q \geq 0$ such that for all $d \in \mathbb{N}$ and $\varepsilon \in (0, 1]$

$$n(\varepsilon, d) = n_{\text{abs}}^{\text{wor}}(\varepsilon, d) \leq C \cdot \varepsilon^{-p} \cdot d^q.$$

Formula (2.9) and the non-increasing ordering of $(\lambda_{d,i})_{i \in \mathbb{N}}$ therefore imply

$$\lambda_{d, \lfloor C\varepsilon^{-p}d^q \rfloor + 1} \leq \lambda_{d, n(\varepsilon, d) + 1} \leq \varepsilon^2, \quad \varepsilon \in (0, 1].$$

If we set $i = \lfloor C\varepsilon^{-p}d^q \rfloor + 1$ and vary $\varepsilon \in (0, 1]$ then i takes the values $\lfloor Cd^q \rfloor + 1, \lfloor Cd^q \rfloor + 2$, and so forth. On the other hand, we have $i \leq C\varepsilon^{-p}d^q + 1$ which is

equivalent to $\varepsilon^2 \leq (Cd^q/(i-1))^{2/p}$ if $i \geq 2$. For all $i \geq f(d) = \lceil (1+C)d^q \rceil$ we indeed have $i \geq 2$ and, consequently,

$$\lambda_{d,i} \leq \lambda_{d,n(\varepsilon,d)+1} \leq \varepsilon^2 \leq \left(\frac{Cd^q}{i-1} \right)^{2/p}.$$

Choosing $\tau > p/2 > 0$ we conclude

$$\sum_{i=f(d)}^{\infty} \lambda_{d,i}^{\tau} \leq \sum_{i=f(d)}^{\infty} \left(\frac{Cd^q}{i-1} \right)^{2\tau/p} = (Cd^q)^{2\tau/p} \sum_{i=f(d)-1}^{\infty} \frac{1}{i^{2\tau/p}} \leq (C^{2/p} d^{2q/p})^{\tau} \cdot \zeta \left(\frac{2\tau}{p} \right).$$

for every $d \in \mathbb{N}$. In other words, we have shown (2.11) with $r = 2q/p$, as well as the estimate on C_{τ} .

Conversely, assume now that for some $r \geq 0$ and $\tau > 0$ estimate (2.11) holds true with

$$f(d) = \left\lceil C (\min \{ \varepsilon_d^{\text{init}}, 1 \})^{-p} d^q \right\rceil, \quad \text{where } C > 0 \quad \text{and} \quad p, q \geq 0.$$

That is, we assume $0 < C_{\tau} < \infty$. For $n \geq f(d)$ the ordering of $(\lambda_{d,i})_{i \in \mathbb{N}}$ implies $\sum_{i=f(d)}^n \lambda_{d,i}^{\tau} \geq \lambda_{d,n}^{\tau} \cdot (n - f(d) + 1)$. Hence, for every $d \in \mathbb{N}$ and $n \geq f(d)$

$$\lambda_{d,n} \cdot (n - f(d) + 1)^{1/\tau} \leq \left(\sum_{i=f(d)}^n \lambda_{d,i}^{\tau} \right)^{1/\tau} \leq \left(\sum_{i=f(d)}^{\infty} \lambda_{d,i}^{\tau} \right)^{1/\tau} \leq C_{\tau} d^r,$$

or, respectively, $\lambda_{d,n+1} \leq C_{\tau} d^r \cdot ((n+1) - f(d) + 1)^{-1/\tau}$, for all $n \geq f(d) - 1$. Note that for $\varepsilon \in (0, \min \{ \varepsilon_d^{\text{init}}, 1 \}]$ we have $C_{\tau} d^r \cdot ((n+1) - f(d) + 1)^{-1/\tau} \leq \varepsilon^2$ if and only if

$$n \geq n^* = \left\lceil \left(\frac{C_{\tau} d^r}{\varepsilon^2} \right)^{\tau} \right\rceil + f(d) - 2.$$

In particular, it is $\lambda_{d,n+1} \leq \varepsilon^2$ at least for $n \geq \max \{ n^*, f(d) - 1 \}$. In other words, for every $d \in \mathbb{N}$ and all $\varepsilon \in (0, \min \{ \varepsilon_d^{\text{init}}, 1 \}]$ it is

$$\begin{aligned} n_{\text{abs}}^{\text{wor}}(\varepsilon, d) &\leq \max \{ n^*, f(d) - 1 \} \leq f(d) - 1 + \left(\frac{C_{\tau} d^r}{\varepsilon^2} \right)^{\tau} \\ &\leq C (\min \{ \varepsilon_d^{\text{init}}, 1 \})^{-p} d^q + C_{\tau}^{\tau} \varepsilon^{-2\tau} d^{r\tau} \\ &\leq (C + C_{\tau}^{\tau}) \varepsilon^{-\max \{ p, 2\tau \}} d^{\max \{ q, r\tau \}}. \end{aligned}$$

Thus, the problem is polynomially tractable since $n_{\text{abs}}^{\text{wor}}(\varepsilon, d) = 0$ for $\varepsilon \geq \varepsilon_d^{\text{init}}$. \blacksquare

Let us add some comments on this result. [Theorem 2.8](#) clearly provides a characterization for (strong) polynomial tractability. In comparison to [Theorem 5.1](#) in [\[NW08\]](#) our result yields the essential advantage that the given estimates incorporate the initial error $\varepsilon_d^{\text{init}}$. Hence if $\varepsilon_d^{\text{init}}$ is sufficiently small then we can

conclude polynomial tractability while ignoring a larger set of eigenvalues in the summation (2.11).

Observe that the first statement does not cover any assertion about the initial error itself, since $f(d) \geq 2$. Thus it might happen that we have (strong) polynomial tractability w.r.t. the absolute error criterion, though the largest eigenvalue $\lambda_{d,1} = (\varepsilon_d^{\text{init}})^2$ tends faster to infinity than any polynomial. To give an example, for $d \in \mathbb{N}$ we consider the sequences $(\lambda_{d,m})_{m \in \mathbb{N}}$ defined by

$$\lambda_{d,1} = e^{2d} \quad \text{and} \quad \lambda_{d,m} = \frac{1}{m} \quad \text{for} \quad m \geq 2.$$

Here, obviously, the initial error grows exponentially fast to infinity, but nevertheless the second point of [Theorem 2.8](#) shows that S is strongly polynomially tractable since (2.11) holds with $r = p = q = 0$, and $C = \tau = 2$.

Next we present an analogue of [Theorem 2.8](#) for the normalized error criterion. Again a slightly modified statement can be found in [[NW08](#), Theorem 5.2].

Theorem 2.9. Assume S to be a problem as in [Corollary 2.7](#) and consider the normalized error criterion in the worst case setting.

- If S is polynomially tractable with the constants $C, p > 0$ and $q \geq 0$ then for all $\tau > p/2$ we have

$$C_\tau = \sup_{d \in \mathbb{N}} \frac{1}{d^r} \left(\sum_{i=f(d)}^{\infty} \left(\frac{\lambda_{d,i}}{\lambda_{d,1}} \right)^\tau \right)^{1/\tau} < \infty, \quad (2.12)$$

where $r = 2q/p$ and $f: \mathbb{N} \rightarrow \mathbb{N}$ with $f(d) \equiv 1$. In this case the bound $C_\tau \leq 2^{1/\tau} (1 + C)^{2/p} \zeta(2\tau/p)^{1/\tau}$ holds for any such τ .

- If (2.11) is satisfied for some parameters $r \geq 0, \tau > 0$ and a function $f: \mathbb{N} \rightarrow \mathbb{N}$ such that $f(d) = \lceil C d^q \rceil$, where $C > 0$ and $q \geq 0$, then the problem S is polynomially tractable. If so, then $n_{\text{norm}}^{\text{wor}}(\varepsilon, d) \leq (C + C_\tau^\tau) \varepsilon^{-2\tau} d^{\max\{q, r\tau\}}$ for any $\varepsilon \in (0, 1]$ and every $d \in \mathbb{N}$.

Proof. Due to the strong relation between the absolute and the normalized error criterion, i.e. $n_{\text{norm}}^{\text{wor}}(\varepsilon, d) = n_{\text{abs}}^{\text{wor}}(\varepsilon \cdot \sqrt{\lambda_{d,1}}, d)$ for $\varepsilon \in (0, 1]$ and $d \in \mathbb{N}$, we note that [Theorem 2.9](#) can be shown using essentially the same arguments as in the proof for [Theorem 2.8](#). Indeed, if we replace $\lambda_{d,i}$ by $\lambda_{d,i}/\lambda_{d,1}$ for $i \in \mathbb{N}$ we obtain a scaled problem T with initial error $\varepsilon_d^{\text{init}} = 1$. Now the information complexity of T (w.r.t. the absolute error criterion) equals the information complexity of S w.r.t. normalized errors.¹³ Following the lines of the proof of [Theorem 2.8](#) this shows the

¹³For details we refer to the proof of [Theorem 2.12](#).

second point of [Theorem 2.9](#), where we set $p = 0$. Moreover, we conclude for any $\tau > p/2$ and $d \in \mathbb{N}$

$$\sum_{i=\lceil(1+C)d^q\rceil}^{\infty} \left(\frac{\lambda_{d,i}}{\lambda_{d,1}} \right)^{\tau} \leq (C^{2/p} d^{2q/p})^{\tau} \zeta \left(\frac{2\tau}{p} \right) \leq (1+C)^{2\tau/p} \zeta \left(\frac{2\tau}{p} \right) d^{q \cdot 2\tau/p},$$

provided that S is polynomially tractable with the constants $C, p > 0$ and $q \geq 0$. Furthermore, for any $d \in \mathbb{N}$ we have

$$\sum_{i=1}^{\lceil(1+C)d^q\rceil-1} \left(\frac{\lambda_{d,i}}{\lambda_{d,1}} \right)^{\tau} \leq \lceil(1+C)d^q\rceil - 1 \leq (1+C) d^q \leq (1+C)^{2\tau/p} \zeta \left(\frac{2\tau}{p} \right) d^{q \cdot 2\tau/p}$$

since $\lambda_{d,i} \leq \lambda_{d,1}$, $2\tau/p > 1$ and $\zeta(2\tau/p) > 1$. Consequently, setting $r = 2q/p$ and combining both the previous estimates leads to

$$\frac{1}{d^r} \left(\sum_{i=1}^{\infty} \left(\frac{\lambda_{d,i}}{\lambda_{d,1}} \right)^{\tau} \right)^{1/\tau} \leq 2^{1/\tau} (1+C)^{2/p} \zeta \left(\frac{2\tau}{p} \right)^{1/\tau} \quad \text{for } d \in \mathbb{N}$$

which shows [\(2.12\)](#), as well as the claimed bound on C_{τ} . ■

Obviously [Theorem 2.9](#) again provides a characterization of (strong) polynomially tractability of a given compact Hilbert space problem $S = (S_d)_{d \in \mathbb{N}}$ in terms of summability properties of the eigenvalue sequence $(\lambda_{d,i})_{i \in \mathbb{N}}$ of $W_d = S_d^{\dagger} S_d$.

2.4 Tensor product problems

In the former section we investigated tractability properties of compact Hilbert space problems $S = (S_d)_{d \in \mathbb{N}}$ without assuming any relation between subsequent problem operators S_d . Next we want to consider problems S where every S_d is generated out of one single (univariate) operator S_1 via a d -fold tensor product construction.

2.4.1 Definition and simple properties

We need to recall the concept of tensor product Hilbert spaces first. To this end, we use the approach given in Chapter 2.6 of Kadison and Ringrose [[KR83](#)]. For a comprehensive introduction to more general tensor products in functional analysis we refer to the first chapter of Light and Cheney [[LC85](#)] and to Section 1.3 in Hansen [[Han10](#)].

Without going too much into details, we note that given a finite number of arbitrary Hilbert spaces $H^{(k)}$ with inner products $\langle \cdot, \cdot \rangle_{H^{(k)}}$, $k = 1, \dots, d$, the tensor product space

$$H_d = \bigotimes_{k=1}^d H^{(k)} = H^{(1)} \otimes \dots \otimes H^{(d)}$$

can be identified¹⁴ with the closure of the algebraic tensor product $H_{d,0}$, with respect to a (reasonable cross) norm which is induced by a certain inner product $\langle \cdot, \cdot \rangle_{H_{d,0}}$. Keep in mind that the algebraic tensor product is defined as the quotient of the *free vector space*, i.e. the set of all finite linear combinations of formal objects $f = \bigotimes_{k=1}^d f_k$ with $f_k \in H^{(k)}$, which we call *simple* (or *pure*) *tensors*, by a suitable linear subspace.¹⁵ Moreover, the mentioned inner product on the algebraic tensor product $H_{d,0}$ is defined by

$$\left\langle \bigotimes_{k=1}^d f_k, \bigotimes_{k=1}^d g_k \right\rangle_{H_{d,0}} = \prod_{k=1}^d \langle f_k, g_k \rangle_{H^{(k)}} \quad \text{for } f_k, g_k \in H^{(k)}.$$

By means of continuous (multi-) linear extension this functional uniquely determines the inner product $\langle \cdot, \cdot \rangle_{H_d}$ on H_d . As usual we denote the corresponding norm by $\|\cdot\|_{H_d}$.

Due to the tensor product structure, many useful properties such as completeness and separability of the underlying spaces $H^{(k)}$ are transferred directly to H_d provided that all the $H^{(k)}$ share them. In particular, it is well-known how to construct an orthonormal basis of the tensor product space given an ONB

$$E^{(k)} = \left\{ e_i^{(k)} \in H^{(k)} \mid i \in \mathcal{J}^{(k)} \right\}$$

in each $H^{(k)}$, $k = 1, \dots, d$. Here every $\mathcal{J}^{(k)}$ denotes a (possibly non-countable) abstract index set. Then the set of all d -fold simple tensors given by

$$E_d = \left\{ e_{d,\mathbf{j}} = \bigotimes_{k=1}^d e_{j_k}^{(k)} \mid \mathbf{j} = (j_1, \dots, j_d) \in \mathcal{J}_d = \bigtimes_{k=1}^d \mathcal{J}^{(k)} \right\}$$

builds the desired ONB in H_d ; see [KR83, Theorem 2.6.4].

For the applications we have in mind we will focus our attention on the special case where all the building blocks $H^{(k)}$, $k = 1, \dots, d$, of H_d coincide. In what follows we therefore assume that $H^{(k)} \equiv H_1$ for some Hilbert space H_1 . The respective ONB of H_1 will be denoted by $E_1 = \{e_i \in H_1 \mid i \in \mathcal{J}_1\}$. Then the latter formula for E_d simplifies to

$$E_d = \left\{ e_{d,\mathbf{j}} = \bigotimes_{k=1}^d e_{j_k} \mid \mathbf{j} = (j_1, \dots, j_d) \in \mathcal{J}_d = (\mathcal{J}_1)^d \right\}. \quad (2.13)$$

We are ready to introduce the tensor product problem operators S_d , $d \geq 1$, we are interested in. Thus let $S_1: \mathcal{F}_1 \rightarrow \mathcal{G}_1$ be a compact linear operator between arbitrary Hilbert spaces \mathcal{F}_1 and \mathcal{G}_1 . For $d \geq 2$ we assume $\mathcal{F}_d = H_d$ to be the

¹⁴Note that this association is unique up to some isometric isomorphism.

¹⁵To abbreviate the notation we do not distinguish between simple tensors and their equivalence classes in what follows.

d -fold tensor product space of $H^{(k)} = H_1 = \mathcal{F}_1$, $k = 1, \dots, d$, as explained above. Analogously, we construct the space $\mathcal{G}_d = \bigotimes_{k=1}^d \mathcal{G}_1$ out of d copies of \mathcal{G}_1 . Now Proposition 2.6.12 of [KR83] yields that there exists a uniquely defined linear operator $S_d = \bigotimes_{k=1}^d S_1: \mathcal{F}_d \rightarrow \mathcal{G}_d$ such that

$$S_d \left(\bigotimes_{k=1}^d f_k \right) = \bigotimes_{k=1}^d S_1 f_k, \quad f_k \in \mathcal{F}_1,$$

and we have $\|S_d|_{\mathcal{L}(\mathcal{F}_d, \mathcal{G}_d)}\| = \|S_1|_{\mathcal{L}(\mathcal{F}_1, \mathcal{G}_1)}\|^d < \infty$ for any fixed $d \in \mathbb{N}$. In detail, we define the bounded linear operator $\tilde{S}_d: E_d \rightarrow \mathcal{G}_d$ such that for all $\mathbf{j} \in \mathcal{J}_d$ we have $\tilde{S}_d(e_{d,\mathbf{j}}) = \tilde{S}_d(\bigotimes_{k=1}^d e_{j_k}) = \bigotimes_{k=1}^d S_1(e_{j_k}) \in \mathcal{G}_d$. Then S_d is assumed to be the uniquely defined linear, continuous extension of \tilde{S}_d from E_d to \mathcal{F}_d . Due to the compactness of S_1 it is easy to check that the problem operator S_d is not only bounded but even compact. Moreover, a linear extension argument shows that the adjoint operator S_d^\dagger is given by the d -fold tensor product of S_1^\dagger , i.e. $S_d^\dagger = \bigotimes_{k=1}^d (S_1^\dagger)$, and hence

$$W_d = S_d^\dagger S_d = \left(\bigotimes_{k=1}^d (S_1^\dagger) \right) \left(\bigotimes_{k=1}^d S_1 \right) = \bigotimes_{k=1}^d (S_1^\dagger S_1) = \bigotimes_{k=1}^d W_1; \quad (2.14)$$

cf. [KR83, p. 146].

2.4.2 Eigenpairs and the optimal algorithm

From Section 2.3 we know that for $d \in \mathbb{N}$ the optimal algorithm, as well as the (information) complexity, crucially depends on the singular value decomposition of S_d . Hence, we have to calculate the eigenpairs $(\lambda_{d,i}, \phi_{d,i})$ of the tensor product operator W_d obtained in (2.14). We follow the arguments presented in [NW08, Section 5.2] and claim that these eigenpairs are given by (tensor) products of the non-trivial eigenpairs (λ_m, ϕ_m) , $m \in \mathcal{M}_1 = \{m \in \mathbb{N} \mid m < v(W_1) + 1\}$, of the univariate operator $W_1 = S_1^\dagger S_1$; see (2.5). This is the subject of the following assertion.

Proposition 2.10. For $d \in \mathbb{N}$ the non-trivial eigenpairs of the operator $W_d = S_d^\dagger S_d$ are given by $\left\{ \left(\tilde{\lambda}_{d,\mathbf{m}}, \tilde{\phi}_{d,\mathbf{m}} \right) \mid \mathbf{m} = (m_1, \dots, m_d) \in \mathcal{M}_d = (\mathcal{M}_1)^d \right\}$, where

$$\tilde{\lambda}_{d,\mathbf{m}} = \prod_{k=1}^d \lambda_{m_k} \quad \text{and} \quad \tilde{\phi}_{d,\mathbf{m}} = \bigotimes_{k=1}^d \phi_{m_k}. \quad (2.15)$$

Proof. Obviously, all the $\tilde{\phi}_{d,\mathbf{m}}$'s are mutually orthonormal in \mathcal{F}_d , i.e.

$$\left\langle \tilde{\phi}_{d,\mathbf{i}}, \tilde{\phi}_{d,\mathbf{j}} \right\rangle_{\mathcal{F}_d} = \prod_{k=1}^d \langle \phi_{i_k}, \phi_{j_k} \rangle_{\mathcal{F}_1} = \prod_{k=1}^d \delta_{i_k, j_k} = \delta_{\mathbf{i}, \mathbf{j}}, \quad \mathbf{i}, \mathbf{j} \in \mathcal{M}_d.$$

Furthermore,

$$\begin{aligned} W_d \tilde{\phi}_{d,\mathbf{m}} &= \left(\bigotimes_{k=1}^d W_1 \right) \left(\bigotimes_{k=1}^d \phi_{m_k} \right) = \bigotimes_{k=1}^d (W_1 \phi_{m_k}) \\ &= \bigotimes_{k=1}^d (\lambda_{m_k} \cdot \phi_{m_k}) = \prod_{k=1}^d \lambda_{m_k} \cdot \bigotimes_{k=1}^d \phi_{m_k} = \tilde{\lambda}_{d,\mathbf{m}} \cdot \tilde{\phi}_{d,\mathbf{m}} \end{aligned}$$

shows that $\tilde{\phi}_{d,\mathbf{m}}$, $\mathbf{m} \in \mathcal{M}_d$, is indeed an eigenelement with respect to the strictly positive eigenvalue $\tilde{\lambda}_{d,\mathbf{m}}$ of W_d .

Assume for a moment there exists an eigenpair (μ, η) of W_d with $\mu \neq 0$ which cannot be represented by (2.15). Then, due to the assertions in the former section, η is orthogonal to every other eigenelement $\tilde{\phi}_{d,\mathbf{m}}$, $\mathbf{m} \in \mathcal{M}_d$. Remember that $\Phi_1 = \{\phi_m \in \mathcal{F}_1 \mid m \in \mathcal{M}_1\}$ can be extended to an orthonormal basis $E_1 = \{e_m \mid m \in \mathcal{J}_1\}$ of \mathcal{F}_1 (see Remark 2.6) which can be used to construct an ONB $E_d = \{e_{d,\mathbf{j}} \mid \mathbf{j} \in \mathcal{J}_d = (\mathcal{J}_1)^d\}$ of \mathcal{F}_d given by (2.13). Therefore η can be represented as

$$\begin{aligned} \eta &= \sum_{\mathbf{j} \in \mathcal{J}_d} \langle \eta, e_{d,\mathbf{j}} \rangle_{\mathcal{F}_d} e_{d,\mathbf{j}} = \sum_{\mathbf{j} \in \mathcal{M}_d} \langle \eta, \tilde{\phi}_{d,\mathbf{j}} \rangle_{\mathcal{F}_d} \tilde{\phi}_{d,\mathbf{j}} + \sum_{\mathbf{j} \in \mathcal{J}_d \setminus \mathcal{M}_d} \langle \eta, e_{d,\mathbf{j}} \rangle_{\mathcal{F}_d} e_{d,\mathbf{j}} \\ &= \sum_{\mathbf{j} \in \mathcal{J}_d \setminus \mathcal{M}_d} \langle \eta, e_{d,\mathbf{j}} \rangle_{\mathcal{F}_d} e_{d,\mathbf{j}}, \end{aligned}$$

where each of these sums consists of at most countably many non-vanishing summands and converges unconditionally. Now the boundedness of S_d implies

$$S_d \eta = \sum_{\mathbf{j} \in \mathcal{J}_d \setminus \mathcal{M}_d} \langle \eta, e_{d,\mathbf{j}} \rangle_{\mathcal{F}_d} S_d e_{d,\mathbf{j}} = 0,$$

since each of the tensor products $S_d e_{d,\mathbf{j}} = \bigotimes_{k=1}^d (S_1 e_{j_k})$, $\mathbf{j} \in \mathcal{J}_d \setminus \mathcal{M}_d$, includes at least one factor $S_1 e_{j_k}$ with $j_k \notin \mathcal{M}_1$. These factors need to vanish because the set $\{S_1 e_m = S_1 \phi_m \mid m \in \mathcal{M}_1\}$ builds an ONB of the image of S_1 in \mathcal{G}_1 . Hence, $W_d \eta = S_d^\dagger (S_d \eta) = 0$ which contradicts our assumption. In other words, (2.15) completely describes the eigenpairs of W_d as claimed. \blacksquare

Again the latter proof justifies the restriction to separable spaces \mathcal{F}_1 (and hence also \mathcal{F}_d) in what follows, see Remark 2.6. Thus we can assume that the set of univariate eigenelements Φ_1 already builds an ONB in \mathcal{F}_1 , i.e. that $\Phi_1 = E_1$, and consequently $\Phi_d = \{\tilde{\phi}_{d,\mathbf{m}} \mid \mathbf{m} \in \mathcal{M}_d\}$ builds an ONB in \mathcal{F}_d .

To unify our notation we rearrange the obtained eigenpairs according to a non-increasing ordering of the eigenvalues. To this end, note that $\#\mathcal{M}_d = (\#\mathcal{M}_1)^d$, i.e. we have $v(W_d) = v(W_1)^d$ strictly positive eigenvalues in dimension d . Therefore we define a sequence of bijections $\psi = \psi_d: \{i \in \mathbb{N} \mid i < v(W_1)^d + 1\} \rightarrow \mathcal{M}_d$ such that

$$\lambda_{d,i} = \tilde{\lambda}_{d,\psi(i)} \geq \tilde{\lambda}_{d,\psi(i+1)} \quad \text{for all } 1 \leq i < v(W_1)^d + 1.$$

Consequently the corresponding eigenelements are denoted by $\phi_{d,i} = \tilde{\phi}_{d,\psi(i)}$. Similar to the definitions in [Section 2.3.1](#) we extend the (possibly finite) sequence of eigenvalues by $\lambda_{d,i} = 0$ for $i > v(W_1)^d$. Observe that the largest eigenvalue in dimension d is given by

$$\lambda_{d,1} = \lambda_1^d$$

and thus the initial error is $\varepsilon_d^{\text{init}} = \lambda_1^{d/2}$.

[Proposition 2.10](#) in hand, the optimal algorithm $A_{n,d}^*$ for linear tensor product problems $S = (S_d)_{d \in \mathbb{N}}$ is stated in [Corollary 2.7](#). For $d \in \mathbb{N}$ and $n \in \mathbb{N}_0$ it reads

$$A_{n,d}^*: \mathcal{F}_d \rightarrow \mathcal{G}_d, \quad f \mapsto A_{n,d}^*(f) = \sum_{i=1}^{\min\{n,v(W_d)\}} \langle f, \phi_{d,i} \rangle_{\mathcal{F}_d} \cdot S_d \phi_{d,i}, \quad (2.16)$$

and its worst case error can be expressed in terms of the sequence $\lambda = (\lambda_m)_{m \in \mathbb{N}}$. More precisely, we have $e^{\text{wor}}(n, d; S_d) = \Delta^{\text{wor}}(A_{n,d}^*; S_d) = \sqrt{\lambda_{d,n+1}}$.

We are ready to characterize tractability of such problems in the next subsection.

2.4.3 Complexity

We begin by analyzing the information complexity with respect to the absolute error criterion. Let $S_1: \mathcal{F}_1 \rightarrow \mathcal{G}_1$ denote a compact linear operator between arbitrary Hilbert spaces \mathcal{F}_1 and \mathcal{G}_1 and let $S = (S_d)_{d \in \mathbb{N}}$ be the sequence of d -fold tensor product problems defined in [Section 2.4.1](#). As before the non-increasing sequence of non-negative eigenvalues of the univariate operator $W_1 = S_1^\dagger S_1$ is denoted by $\lambda = (\lambda_m)_{m \in \mathbb{N}}$. At this point we stress that it is reasonable to assume that $\lambda_2 > 0$. Otherwise for every $d \in \mathbb{N}$ there would be only at most one non-vanishing d -dimensional eigenvalue of $W_d = S_d^\dagger S_d$. Hence the problem S_d would be trivial since then $n_{\text{abs}}^{\text{wor}}(\varepsilon, d) \leq 1$ for all $\varepsilon > 0$. Note that $\lambda_2 > 0$ also implies $\lambda_1 > 0$ such that S_1 and S_d are not the zero operator.

We proceed by presenting an assertion which is mainly based on [Theorem 5.5](#) in [Novak and Woźniakowski \[NW08\]](#). The sufficient condition for weak tractability later was given by [Papageorgiou and Petras \[PP09\]](#). Although the results of these authors only refer to linear tensor product problems defined between Hilbert *function* spaces they remain valid even in our more general setting.

Theorem 2.11. Consider the problem $S = (S_d)_{d \in \mathbb{N}}$ as described before. We study the absolute error criterion in the worst case setting.

- Let $\lambda_1 > 1$. Then S suffers from the curse of dimensionality.
- Let $\lambda_1 = 1$. Then
 - S is polynomially intractable. In particular, if $\lambda_2 = 1$ then S suffers from the curse of dimensionality.

- S is weakly tractable if and only if $\lambda_2 < 1$ and $\lambda_n \in o(\ln^{-2} n)$, as $n \rightarrow \infty$.¹⁶
- Let $\lambda_1 < 1$. Then
 - S never suffers from the curse.
 - S is weakly tractable if and only if $\lambda_n \in o(\ln^{-2} n)$, as $n \rightarrow \infty$.
 - S is polynomially tractable if and only if it is strongly polynomially tractable. Moreover, this holds if and only if there exists some $\tau \in (0, \infty)$ such that $\lambda \in \ell_\tau$ and the exponent of strong polynomial tractability is given by

$$p^* = \inf \left\{ 2\tau \mid \sum_{m=1}^{\infty} \lambda_m^\tau \leq 1 \right\}.$$

For the sake of completeness we mention that Theorem 5.5 in [NW08] includes some additional lower bounds on the information complexity in the case $\lambda_1 \geq 1$. For polynomial (in)tractability the main idea of the proof is to apply Theorem 2.8 and to use the product structure of the involved sequences $(\lambda_{d,i})_{i \in \mathbb{N}}$ which are essentially given by Proposition 2.10. We will not provide an explicit proof here. Instead the interested reader is referred to Example 3.9 in Chapter 3 where we conclude all assertions stated in Theorem 2.11 out of a generalized result for *scaled* tensor product problems. To conclude these more general assertions we will exactly follow the mentioned proof sketch.

Many authors in IBC use phrases like “(unweighted) tensor product problems are intractable”. In this regard they refer to the following Theorem for the normalized error criterion which is essentially based on Theorem 5.6 of [NW08], as well as on [PP09]. From our point of view it is not more than a simple consequence of the assertions for absolute errors.

Theorem 2.12. Consider the problem $S = (S_d)_{d \in \mathbb{N}}$ as described above. We study the normalized error criterion in the worst case setting.

- Let $\lambda_1 = \lambda_2$. Then S suffers from the curse of dimensionality.
- Let $\lambda_1 > \lambda_2$. Then
 - S is weakly tractable if and only if $\lambda_n \in o(\ln^{-2} n)$, as $n \rightarrow \infty$.
 - S is polynomially intractable.

Since the subsequent proof technique is typical in this field of research, we include the proof of Theorem 2.12 in full detail.

¹⁶To avoid possible confusions, here and in what follows, $\ln^\alpha n$ means $[\ln(n)]^\alpha$ where $\alpha \in \mathbb{R}$.

Proof. Assume we had already proven [Theorem 2.11](#). Given the problem $S = (S_d)_{d \in \mathbb{N}}$, constructed out of $S_1: \mathcal{F}_1 \rightarrow \mathcal{G}_1$, as well as the associated sequence $(\lambda_m)_{m \in \mathbb{N}}$, we define a new operator $T_1: \mathcal{F}_1 \rightarrow \mathcal{G}_1$ by $f \mapsto T_1 f = 1/\sqrt{\lambda_1} \cdot S_1 f$. Clearly, T_1 is a linear and compact mapping between Hilbert spaces and

$$\langle T_1 f, g \rangle_{\mathcal{G}_1} = \frac{1}{\sqrt{\lambda_1}} \langle S_1 f, g \rangle_{\mathcal{G}_1} = \frac{1}{\sqrt{\lambda_1}} \langle f, S_1^\dagger g \rangle_{\mathcal{F}_1} = \left\langle f, \left(\frac{1}{\sqrt{\lambda_1}} S_1^\dagger \right) g \right\rangle_{\mathcal{F}_1}$$

for $f \in \mathcal{F}_1$ and $g \in \mathcal{G}_1$. Hence, $T_1^\dagger = 1/\sqrt{\lambda_1} \cdot S_1^\dagger$ and the (extended) eigenvalue sequence of $V_1 = W_1(T) = T_1^\dagger T_1 = 1/\lambda_1 \cdot S_1^\dagger S_1 = 1/\lambda_1 \cdot W_1(S)$ is given by $\mu = (\mu_m)_{m \in \mathbb{N}}$, where $\mu_m = \lambda_m/\lambda_1$ for $m \in \mathbb{N}$. For details, see also the arguments used in [Section 3.1](#). Anyway, the mapping T_1 in hand, we can construct the tensor product problem $T = (T_d)_{d \in \mathbb{N}}$ by the usual procedure. Now (2.15) in [Proposition 2.10](#) shows that the corresponding eigenvalues of $V_d = W_d(T) = T_d^\dagger T_d$ are given by

$$\tilde{\mu}_{d,\mathbf{m}} = \prod_{k=1}^d \mu_{m_k} = \frac{1}{\lambda_1^d} \prod_{k=1}^d \mu_{m_k} = \frac{1}{\lambda_{d,1}} \tilde{\lambda}_{d,\mathbf{m}}, \quad \mathbf{m} = (m_1, \dots, m_d) \in \mathcal{M}_d,$$

such that $\mu_{d,i} = 1/\lambda_{d,1} \cdot \lambda_{d,i}$ for $i \in \mathbb{N}$. This yields that the information complexity of S w.r.t. the normalized error criterion coincides with the absolute information complexity of the (scaled) problem T , i.e.

$$\begin{aligned} n_{\text{norm}}(\varepsilon', d; S_d) &= \#\{n \in \mathbb{N}_0 \mid \lambda_{d,n}/\lambda_{d,1} > (\varepsilon')^2\} = \#\{n \in \mathbb{N}_0 \mid \mu_{d,n} > (\varepsilon')^2\} \\ &= n_{\text{abs}}(\varepsilon', d; T_d) \end{aligned}$$

for all $\varepsilon' \in (0, 1]$ and each $d \in \mathbb{N}$. Since $\mu_1 = 1 \geq \mu_2 > 0$ we are allowed to apply [Theorem 2.11](#) for T . Finally the observations that $\mu_2 = 1$ if and only if $\lambda_1 = \lambda_2$, as well as that $\mu_n \in o(\ln^{-2} n)$ (as $n \rightarrow \infty$) if and only if $\lambda = (\lambda_n)_{n \in \mathbb{N}}$ belongs to this class, complete the proof. \blacksquare

2.5 Reproducing kernel Hilbert spaces

When we deal with problems defined on Hilbert *function* spaces H a special kind of Hilbert spaces is of particular interest. The reason is that in practice often only function evaluations rather than information obtained by arbitrary linear functionals are permitted. In order to compare the power of these classes of information operations (Λ^{std} vs. Λ^{all}) from a theoretical point of view, it seems to be useful to investigate conditions which ensure that point evaluation functionals

$$L_y: H \rightarrow \mathbb{R}, \quad f \mapsto L_y(f) = f(y),$$

for all y in the domain of definition Ω of $f \in H$, belong to the class Λ^{all} . Clearly L_y is always linear such that it is enough to ask whether it is also continuous (or bounded, respectively) in f . It turns out that, as long as we restrict ourselves to Hilbert

spaces, this property can be characterized by the existence of a so-called *reproducing kernel* K . If so, then the space H is referred to as a *reproducing kernel Hilbert space* (RKHS for short) and we write $H = \mathcal{H}(K)$. In the present section we collect some basic properties of this concept which we will need later on in [Section 4.1.2](#). The presentation given here is based on the famous paper of Aronszajn [[Aro50](#)], as well as the textbook of Wahba [[Wah90](#)].¹⁷ Standard examples for RKHSs such as Korobov spaces and Sobolev spaces of dominating mixed smoothness can be found in [[NW08](#), Appendix A].

2.5.1 Definition and properties

A (real) Hilbert space H of functions $f: \Omega \rightarrow \mathbb{R}$, equipped with inner product $\langle \cdot, \cdot \rangle_H$, is said to be a reproducing kernel Hilbert space if there exists a function

$$K: \Omega \times \Omega \rightarrow \mathbb{R}$$

such that

- for all fixed $y \in \Omega$ the function $K_y = K(\cdot, y)$ belongs to H , and
- for every $f \in H$ and all $y \in \Omega$ it is

$$L_y(f) = f(y) = \langle f, K_y \rangle_H = \langle f, K(\cdot, y) \rangle_H. \quad (2.17)$$

The second point (2.17) is known as the *reproducing property*. Together with the first point it obviously implies the boundedness of point evaluations on $H = \mathcal{H}(K)$. The converse, i.e. the existence (and uniqueness) of the reproducing kernel K , is a simple consequence of the Riesz representation theorem; see [[Tri92](#), p. 90] or [[Yos80](#), III.6]. Unfortunately the proof of this theorem is non-constructive and therefore it does not provide an explicit method to find the *representer* $K_y = K(\cdot, y)$ of L_y . In fact, given a specific RKHS $\mathcal{H}(K)$ it seems to be a challenging problem to deduce a closed form of its reproducing kernel K . However, as long as we restrict ourselves to separable RKHSs, it is easy to prove that K is given by

$$K(x, y) = \sum_{m \in \mathcal{J}} e_m(x) e_m(y), \quad x, y \in \Omega, \quad (2.18)$$

where $\{e_m: \Omega \rightarrow \mathbb{R} \mid m \in \mathcal{J}\}$ denotes an arbitrary orthonormal basis of $\mathcal{H}(K)$. Furthermore we know that every reproducing kernel K is positive definite. That is, for all $n \in \mathbb{N}$ and any sequence $\mathbf{x} = (x_m)_{m=1}^n \in \Omega^n$ the quadratic form

$$Q_{K;\mathbf{x}}(\xi_1, \dots, \xi_n) = \sum_{i,j=1}^n K(x_i, x_j) \xi_i \xi_j, \quad \boldsymbol{\xi} = (\xi_m)_{m=1}^n \in \mathbb{R}^n, \quad (2.19)$$

¹⁷For the ease of notation (and in contrast to the mentioned references) we restrict ourselves to spaces over \mathbb{R} . Once more the theory can be transferred almost literally to \mathbb{C} .

is a non-negative function of ξ . In particular,

$$K(x, x) \geq 0 \quad \text{and} \quad K(x, y) = K(y, x) \quad \text{for all } x, y \in \Omega.$$

Conversely, Moore showed that every positive definite function K in the above sense uniquely determines a RKHS H admitting K as its reproducing kernel; see [Aro50]. Again it turned out to be a hard problem to conclude a suitable representation of H (and its inner product) for a given function K .

Besides further fascinating properties, we want to focus our attention on products of kernel functions. To this end, for $d \in \mathbb{N}$ let $K^{(k)}$, $k = 1, \dots, d$, denote a finite number of reproducing kernels defined on the sets $\Omega^{(k)} \times \Omega^{(k)}$, respectively. Then we may consider the tensor product

$$K_d = \bigotimes_{k=1}^d K^{(k)} : \left(\bigtimes_{k=1}^d \Omega^{(k)} \right) \times \left(\bigtimes_{k=1}^d \Omega^{(k)} \right) \rightarrow \mathbb{R}, \quad K_d(\mathbf{x}, \mathbf{y}) = \prod_{k=1}^d K^{(k)}(x_k, y_k),$$

where we set $\mathbf{x} = (x_1, \dots, x_d)$ and $\mathbf{y} = (y_1, \dots, y_d)$ with $x_k, y_k \in \Omega^{(k)}$. On the other hand, each kernel $K^{(k)}$ induces a uniquely defined RKHS $\mathcal{H}(K^{(k)})$ which in turn implies the existence of one (and only one) tensor product space $H_d = \bigotimes_{k=1}^d \mathcal{H}(K^{(k)})$ using the arguments presented in Section 2.4. Now it can be checked that H_d itself is a RKHS and its kernel is given by K_d , i.e.

$$H_d = \mathcal{H}(K_d) = \bigotimes_{k=1}^d \mathcal{H}(K^{(k)}). \quad (2.20)$$

The proof of this assertion can be obtained inductively by adding one factor in every step. Then it remains to show that the resulting quadratic forms (2.19) are non-negative again which can be done using a classical result due to Schur.

Note that the whole theory works for arbitrary point sets Ω which turned out to be useful in the context of so-called *support vector machines* which are instances of the more general class of *kernel methods*. However, in IBC special choices such as $\Omega = \Omega_1 = [0, 1]$ (or $\Omega = \mathbb{R}$) are of particular interest. For multivariate problems the standard choice is $\Omega = \Omega_d = \Omega_1^d$ which perfectly fits to the tensor product construction explained before. In this respect the univariate kernels $K^{(k)}$, $k = 1, \dots, d$, are often taken as weighted instances $K_1^{\gamma_{d,k}}$ of some underlying kernel K_1 . A prominent example is given by $K_1^{\gamma_{d,k}}(x, y) = 1 + \gamma_{d,k} \min\{x, y\}$ which leads to an anchored Sobolev space $\tilde{\mathcal{H}}_d^\gamma$ related to the Wiener sheet measure; see, e.g., [KWW08, Section 8] or [Wei12b]. Another example of this type will be discussed in detail within Section 4.1.2.

Finally we mention that the concept of RKHSs was generalized recently to the class of so-called *reproducing kernel Banach spaces (RKBSs)*. For a brief introduction to this topic we refer to Zhang and Zhang [ZZ13].

2.5.2 Examples: Integration and approximation problems

Let us conclude the presentation with some examples which show that the knowledge about the existence of a reproducing kernel can be exploited to obtain complexity assertions for the classical problems of integration and approximation.

Example 2.13 (Worst case error of QMC rules). For $d \in \mathbb{N}$ suppose $\mathcal{H}(K_d)$ to be a RKHS of real-valued functions f defined on some Borel measurable subset Ω_d of \mathbb{R}^d . Consider the solution operator of the integration problem

$$\text{Int}_d^{\varrho_d} : \mathcal{B}(\mathcal{H}(K_d)) \rightarrow \mathbb{R}, \quad f \mapsto \text{Int}_d^{\varrho_d} f = \int_{\Omega_d} f(\mathbf{x}) \varrho_d(\mathbf{x}) \, d\lambda^d(\mathbf{x}),$$

where ϱ_d denotes a probability density function on Ω_d . Let us additionally assume that the function

$$h_d = \int_{\Omega_d} K_d(\cdot, \mathbf{x}) \varrho_d(\mathbf{x}) \, d\lambda^d(\mathbf{x})$$

is well-defined and belongs to $\mathcal{H}(K_d)$. Then it is easy to see that h_d is the representer of the linear functional $\text{Int}_d^{\varrho_d}$, i.e. that $\text{Int}_d^{\varrho_d} f = \langle f, h_d \rangle_{\mathcal{H}(K_d)} < \infty$ for all $f \in \mathcal{H}(K_d)$. Since allowing arbitrary linear functionals to approximate the value of the integral would make the problem trivial we consider cubature rules of the form

$$A_{n,d} f = \sum_{i=1}^n a_i f(\mathbf{x}^{(i)}), \quad n \in \mathbb{N}_0,$$

defined by a priori chosen sample points $\mathbf{x}^{(i)} \in \Omega_d$ and some weights $a_i \in \mathbb{R}$, $i = 1, \dots, n$. Due to (2.17) also the linear operator $A_{n,d}$ possesses a representer in the space $\mathcal{H}(K_d)$. Consequently its worst case error can be computed exactly in terms of the reproducing kernel and the parameters $(a_i)_{i=1}^n$ and $(\mathbf{x}^{(i)})_{i=1}^n$:

$$\begin{aligned} & \Delta^{\text{wor}}(A_{n,d}; \text{Int}_d^{\varrho_d} : \mathcal{B}(\mathcal{H}(K_d)) \rightarrow \mathbb{R})^2 \\ &= \sup_{f \in \mathcal{B}(\mathcal{H}(K_d))} |(\text{Int}_d^{\varrho_d} - A_{n,d})(f)|^2 = \left\| h_d - \sum_{i=1}^n a_i K_d(\cdot, \mathbf{x}^{(i)}) \right\|_{\mathcal{H}(K_d)}^2 \\ &= \int_{\Omega_d^2} K_d(\mathbf{x}, \mathbf{y}) \varrho_d(\mathbf{x}) \varrho_d(\mathbf{y}) \, d\lambda^{2d}(\mathbf{x}, \mathbf{y}) - 2 \sum_{i=1}^n a_i \int_{\Omega_d} K_d(\mathbf{x}, \mathbf{x}^{(i)}) \varrho_d(\mathbf{x}) \, d\lambda^d(\mathbf{x}) \\ & \quad + \sum_{i,j=1}^n a_i a_j K_d(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}). \end{aligned}$$

Choosing special weights a_i (such as $a_i \equiv 1/n$), as well as specific sample points $\mathbf{x}^{(i)}$ (e.g. from a so-called *integration lattice*), we end up with well-studied classes of cubature rules which are known as *quasi-Monte Carlo* (QMC) methods and *lattice rules*, respectively. The common feature of these integration schemes is that their

complexity analysis is mainly based on the presented worst case error formula and thus on the properties of the reproducing kernel K_d . Moreover, the latter expression for $\Delta^{\text{wor}}(A_{n,d})$ plays an important role in *discrepancy theory*.

Various kinds of integration problems are studied in Novak and Woźniakowski [NW10]. For the recent state of the art in discrepancy theory and QMC methods we refer the reader to the monograph of Dick and Pillichshammer [DP10], as well as to the survey article of Dick, Kuo and Sloan [DKS13] and the references therein. An introduction to lattice rules can also be found in the textbook of Sloan and Joe [SJ94]. \square

Our second example shows the relation of reproducing kernels and the singular values for certain approximation operators.

Example 2.14 (Weighted L_2 -approximation). For $d \in \mathbb{N}$ let $\mathcal{H}(K_d)$ be a separable and infinite-dimensional RKHS which is compactly embedded into $L_2^{\varrho_d}(\Omega_d)$. Here ϱ_d again denotes some probability density on $\Omega_d \subseteq \mathbb{R}^d$. Then we may study the approximation problem

$$\text{App}_d^{\varrho_d} : \mathcal{B}(\mathcal{H}(K_d)) \rightarrow L_2^{\varrho_d}(\Omega_d), \quad f \mapsto \text{App}_d^{\varrho_d} f = f, \quad (2.21)$$

in the worst case setting. Since both source and target space are Hilbert spaces we can use the theory developed in Section 2.3 to conclude complexity results with respect to the class Λ^{all} . Therefore we need to analyze the eigenvalues of the compact operator $W_d^{\varrho_d} = (\text{App}_d^{\varrho_d})^\dagger \text{App}_d^{\varrho_d}$. Using the reproducing property (2.17) and the symmetry of K_d we conclude

$$\begin{aligned} (W_d^{\varrho_d} f)(\mathbf{x}) &= \left\langle (\text{App}_d^{\varrho_d})^\dagger \text{App}_d^{\varrho_d} f, K_d(\cdot, \mathbf{x}) \right\rangle_{\mathcal{H}(K_d)} = \langle \text{App}_d^{\varrho_d} f, \text{App}_d^{\varrho_d} K_d(\cdot, \mathbf{x}) \rangle_{L_2^{\varrho_d}(\Omega_d)} \\ &= \int_{\Omega_d} f(\mathbf{y}) K_d(\mathbf{x}, \mathbf{y}) \varrho_d(\mathbf{y}) \, d\lambda^d(\mathbf{y}) \end{aligned}$$

for all $f \in \mathcal{H}(K_d)$ and any $\mathbf{x} \in \Omega_d$. Hence, $W_d^{\varrho_d}$ takes the form of a weighted integral operator against the kernel $K_d(\cdot, \mathbf{y})$ and its non-trivial eigenpairs $\{(\lambda_{d,\varrho_d,i}, \phi_{d,\varrho_d,i}) \mid i \in \mathcal{M}_d\}$ can be found by solving integral equations. Formula (2.18) yields that

$$K_d(\mathbf{x}, \mathbf{x}) = \sum_{i \in \mathcal{M}_d} \phi_{d,\varrho_d,i}(\mathbf{x})^2 < \infty$$

for every $\mathbf{x} \in \Omega_d$ because we know that $\{\phi_{d,\varrho_d,i} \mid i \in \mathcal{M}_d\}$ forms an ONB in $\mathcal{H}(K_d)$. Since $\phi_{d,\varrho_d,i} \in L_2^{\varrho_d}(\Omega_d)$ and $\|\phi_{d,\varrho_d,i} \mid L_2^{\varrho_d}(\Omega_d)\|^2 = \lambda_{d,\varrho_d,i}$ for all $i \in \mathcal{M}_d$, it easily follows that

$$\text{trace } W_d^{\varrho_d} = \sum_{i \in \mathcal{M}_d} \lambda_{d,\varrho_d,i} = \sum_{i \in \mathcal{M}_d} \|\phi_{d,\varrho_d,i} \mid L_2^{\varrho_d}(\Omega_d)\|^2 = \int_{\Omega_d} K_d(\mathbf{x}, \mathbf{x}) \varrho_d(\mathbf{x}) \, d\lambda^d(\mathbf{x}). \quad (2.22)$$

Note that this trace may be finite or infinite depending on the values of K_d on the diagonal $\{(\mathbf{x}, \mathbf{x}) \mid \mathbf{x} \in \Omega_d\}$. It turns out that an infinite trace implies that there is, in general, no (non-trivial) relation of the power of Λ^{all} and Λ^{std} for the given approximation problem. In contrast, it is known that for finite traces there exist close relations of these classes of information operations. In particular, it is possible to conclude bounds on the rate of convergence for Λ^{std} out of corresponding bounds for Λ^{all} . For details we refer to [NW12, Chapter 26].

Finally we note that the finite trace property of $W_d^{\varrho_d}$ immediately implies $\lambda_{d,\varrho_d,i} \in \mathcal{O}(i^{-1})$, as $i \rightarrow \infty$. Hence, if we deal with linear information then we can conclude $e^{\text{wor}}(n, d; \text{App}_d^{\varrho_d}) \in \mathcal{O}(n^{-1/2})$, $n \rightarrow \infty$, directly out of an integrability property of the kernel K_d . \square

In the last example we present a useful relation of reproducing kernels and average case approximation problems.

Example 2.15 (Average case approximation). For $d \in \mathbb{N}$ assume ϱ_d to be some probability density function on $\Omega_d = [0, 1]^d$ and let $K_d: \Omega_d \times \Omega_d \rightarrow \mathbb{R}$ denote a reproducing kernel such that the mapping $\mathbf{x} \mapsto K_d(\mathbf{x}, \mathbf{x})$ belongs to $L_1^{\varrho_d}(\Omega_d)$. That is, suppose (2.22) to be finite. Furthermore, let \mathcal{F}_d denote a separable Banach space of real-valued functions on Ω_d which is continuously embedded into $L_2^{\varrho_d}(\Omega_d)$ and for which function evaluations are continuous. We equip \mathcal{F}_d with a zero-mean Gaussian measure μ_d such that its correlation operator $C_{\mu_d}: \mathcal{F}_d^* \rightarrow \mathcal{F}_d$ applied to point evaluation functionals $L_{\mathbf{x}}$ can be expressed in terms of K_d :

$$K_d(\mathbf{x}, \mathbf{y}) = L_{\mathbf{x}}(C_{\mu_d} L_{\mathbf{y}}) = \int_{\mathcal{F}_d} f(\mathbf{x}) f(\mathbf{y}) d\mu_d(f) \quad \text{for all } \mathbf{x}, \mathbf{y} \in \Omega_d.$$

We stress the point that this is always possible for a suitable choice of \mathcal{F}_d and that our assumptions imply a continuous embedding of the RKHS $\mathcal{H}(K_d)$ (induced by K_d) into \mathcal{F}_d . Consequently also $\text{App}_d^{\varrho_d}: \mathcal{H}(K_d) \rightarrow L_2^{\varrho_d}(\Omega_d)$ is bounded, i.e. continuous; see (2.21). For details and concrete examples the reader is referred to [NW08, Appendix B], [NW10, Section 13.2], and [NW12, Section 24.1].

As in the previous example we want to look for good approximations $A_{n,d}f$ to input functions f in the norm of $L_2^{\varrho_d}(\Omega_d)$. This time we measure the average performance of the algorithm $A_{n,d}$ with respect to μ_d , i.e. we try to minimize

$$\Delta^{\text{avg}}(A_{n,d}; \text{id}_d^{\varrho_d}: \mathcal{F}_d \rightarrow L_2^{\varrho_d}(\Omega_d)) = \left(\int_{\mathcal{F}_d} \|f - A_{n,d}f\|_{L_2^{\varrho_d}(\Omega_d)}^2 d\mu_d(f) \right)^{1/2}.$$

Observe that $\nu_d = \mu_d \circ (\text{id}_d^{\varrho_d})^{-1}$ defines a Gaussian measure on the subset $\text{id}_d^{\varrho_d}(\mathcal{F}_d)$ of $L_2^{\varrho_d}(\Omega_d)$. Now it can be checked that the corresponding covariance operator $C_{\nu_d}^{\varrho_d}: \text{id}_d^{\varrho_d}(\mathcal{F}_d) \rightarrow L_2^{\varrho_d}(\Omega_d)$ of the measure ν_d is given by

$$f \mapsto (C_{\nu_d}^{\varrho_d} f)(\cdot) = \int_{\Omega_d} f(\mathbf{y}) K_d(\cdot, \mathbf{y}) \varrho_d(\mathbf{y}) d\lambda^d(\mathbf{y}).$$

This operator is self-adjoint, compact and has a finite trace due to the integrability assumption on K_d . Consequently, there exists a countable set of non-trivial eigenpairs $(\lambda_{d,\varrho_d,i}, \eta_{d,\varrho_d,i})$ where the eigenfunctions $\eta_{d,\varrho_d,i}$ are mutually orthogonal (and normalized) with respect to the $L_2^{\varrho_d}(\Omega_d)$ -norm; see also Hickernell and Woźniakowski [HW00].

Once more it turns out that the optimal algorithm $A_{n,d}^*$ in this setting is given by the orthogonal projection of the input function onto the subspace spanned by the eigenfunctions $\eta_{d,\varrho_d,i}$ which correspond to the n largest eigenvalues $\lambda_{d,\varrho_d,i}$. In contrast to the worst case setting the n th minimal average case error is

$$e^{\text{avg}}(n, d; \text{id}_d^{\varrho_d} : \mathcal{F}_d \rightarrow L_2^{\varrho_d}(\Omega_d)) = \left(\sum_{i=n+1}^{\infty} \lambda_{d,\varrho_d,i} \right)^{1/2}, \quad n \in \mathbb{N}_0, \quad d \in \mathbb{N},$$

if we assume a non-increasing ordering of the sequence $(\lambda_{d,\varrho_d,i})_{i=1}^{\infty}$.¹⁸ Based on the latter error formula it is possible to obtain characterizations of several types of tractability similar to the assertions given in Section 2.4.3; see, e.g., [NW08, Chapter 6].

We complete the discussion with the observation that the sets of (non-trivial) eigenpairs $(\lambda_{d,\varrho_d,i}, \eta_{d,\varrho_d,i})$ of the operators $C_{\nu_d}^{\varrho_d}$ as defined above and $W_d^{\varrho_d}$ from Example 2.14 coincide, since $C_{\nu_d}^{\varrho_d}$ only takes values in $\mathcal{H}(K_d)$. To be precise, we note that $K_d(\mathbf{x}, \mathbf{y}) = K_d(\mathbf{y}, \mathbf{x})$ equals $(\text{App}_d^{\varrho_d} K_d(\cdot, \mathbf{x}))(\mathbf{y})$ for each fixed \mathbf{x} and λ^d -almost every $\mathbf{y} \in \Omega_d$. Hence the chain of equations

$$\begin{aligned} (C_{\nu_d}^{\varrho_d} f)(\mathbf{x}) &= \int_{\Omega_d} f(\mathbf{y}) (\text{App}_d^{\varrho_d} K_d(\cdot, \mathbf{x}))(\mathbf{y}) \varrho_d(\mathbf{y}) \, d\lambda^d(\mathbf{y}) = \langle f, \text{App}_d^{\varrho_d} K_d(\cdot, \mathbf{x}) \rangle_{L_2^{\varrho_d}(\Omega_d)} \\ &= \langle (\text{App}_d^{\varrho_d})^\dagger f, K_d(\cdot, \mathbf{x}) \rangle_{\mathcal{H}(K_d)} = ((\text{App}_d^{\varrho_d})^\dagger f)(\mathbf{x}) \end{aligned}$$

holds true for every $f \in \text{id}_d^{\varrho_d}(\mathcal{F}_d) \subset L_2^{\varrho_d}(\Omega_d)$ and λ^d -almost all $\mathbf{x} \in \Omega_d$.¹⁹ Clearly $(\text{App}_d^{\varrho_d})^\dagger$ maps into $\mathcal{H}(K_d)$ per definition. Thus, for every eigenfunction $\eta \in \text{id}_d^{\varrho_d}(\mathcal{F})$ of $C_{\nu_d}^{\varrho_d}$, i.e.

$$\lambda \eta = C_{\nu_d}^{\varrho_d} \eta = (\text{App}_d^{\varrho_d})^\dagger \eta, \quad \lambda^d\text{-a.e. on } \Omega_d, \quad (2.23)$$

we can find a representer $\bar{\eta} \in \mathcal{H}(K_d)$ with $\bar{\eta} = \eta$ in the sense of $L_2^{\varrho_d}(\Omega_d)$, such that the equalities in (2.23) hold pointwise on the whole set Ω_d and therefore also in the norm of $\mathcal{H}(K_d)$. Now it is easy to check that $(\lambda, \bar{\eta})$ indeed is an eigenpair of $W_d^{\varrho_d} = (\text{App}_d^{\varrho_d})^\dagger \text{App}_d^{\varrho_d}$, normalized w.r.t. the $L_2^{\varrho_d}(\Omega_d)$ -norm. Conversely every eigenpair $(\lambda, \bar{\eta})$ of the operator $W_d^{\varrho_d}$ obviously fulfills $\lambda \bar{\eta} = C_{\nu_d}^{\varrho_d} \bar{\eta}$ interpreted in $L_2^{\varrho_d}(\Omega_d)$.

In conclusion we see that the knowledge of these eigenpairs implies complexity assertions for both approximation problems in the respective (quite different) settings. \square

¹⁸For the ease of notation we moreover assumed here that all the eigenvalues are strictly positive.

¹⁹Observe that $C_{\nu_d}^{\varrho_d} f \in L_2^{\varrho_d}(\Omega_d)$, i.e. it is uniquely defined on Ω_d up to a set of measure zero.

CHAPTER 3

Problems on Hilbert spaces with scaled norms

The present chapter deals with a generalization of tensor product problems $S = (S_d)_{d \in \mathbb{N}}$ between Hilbert spaces in the sense of [Section 2.4](#). We introduce additional scaling factors s_d to the norm of the source spaces \mathcal{F}_d and analyze their influence on the squared singular values $\lambda_{d,s_d,i}$ of the new problem operators S_{d,s_d} . Using the techniques from [Section 2.4.2](#) we conclude optimal algorithms for these modified problems at the end of [Section 3.1](#). Afterwards, in [Section 3.2](#), we investigate tractability properties of this class of problems w.r.t. the worst case setting. Finally we present some applications of the obtained results in [Section 3.3](#).

3.1 Definitions, eigenpairs and the optimal algorithm

Let H_1 and \mathcal{G}_1 be arbitrary Hilbert spaces with inner products $\langle \cdot, \cdot \rangle_{H_1}$ and $\langle \cdot, \cdot \rangle_{\mathcal{G}_1}$, respectively. Further assume $S_1 \in \mathcal{K}(H_1, \mathcal{G}_1)$ to be a compact linear operator between these spaces. Following the constructions given in [Section 2.4.1](#) for any $d \in \mathbb{N}$ there exist uniquely defined d -fold tensor product spaces of H_1 and \mathcal{G}_1 . Let us denote these spaces by $H_d = H_1 \otimes \dots \otimes H_1$ and $\mathcal{G}_d = \mathcal{G}_1 \otimes \dots \otimes \mathcal{G}_1$, respectively. Finally we define $S = (S_d)_{d \in \mathbb{N}}$ to be the sequence of multivariate tensor product operators constructed out of S_1 .

In contrast to [Section 2.4](#) we now adapt the source spaces of our multivariate problem by introducing an additional positive *sequence of scaling factors* $s = (s_d)_{d \in \mathbb{N}}$. That is, for every $d \in \mathbb{N}$ we define \mathcal{F}_d to be Hilbert space H_d equipped with the inner product

$$\langle \cdot, \cdot \rangle_{\mathcal{F}_d} = \frac{1}{s_d} \langle \cdot, \cdot \rangle_{H_d}, \quad \text{where } s_d > 0. \quad (3.1)$$

Obviously \mathcal{F}_d algebraically coincides with H_d whereas the norms (induced by the respective inner products) are equivalent. Accordingly, the operators S_d are still well-defined for any $d \in \mathbb{N}$ when we replace H_d by \mathcal{F}_d . On the other hand the approximability properties of S crucially depend on the used norms since we need

to consider the whole unit ball $\tilde{\mathcal{F}}_d = \mathcal{B}(\mathcal{F}_d)$ when dealing with the worst case setting. So let us denote the modified problem by $S_{(s)} = (S_{d,s_d} : \mathcal{F}_d \rightarrow \mathcal{G}_d)_{d \in \mathbb{N}}$.

From [Section 2.3](#) we know that for the n th optimal algorithms for $S_{(s)}$ we need to study the eigenpairs of $W_{d,s_d} = S_{d,s_d}^\dagger S_{d,s_d}$. Although S_{d,s_d} equals S_d (as a mapping) we can not claim that $W_{d,s_d} = W_d$ since S_{d,s_d}^\dagger does not necessarily coincide with S_d^\dagger . Nevertheless, there exists a strong relation. The following proposition extends [Proposition 2.10](#) to the case of scaled problems in the mentioned sense. Keep in mind that the eigenpairs of the univariate (unscaled) operator $W_1 = S_1^\dagger S_1$ are given by $\{(\lambda_m, e_m) \mid m \in \mathcal{M}_1\}$, where $\mathcal{M}_1 = \{m \in \mathbb{N} \mid m < v(W_1) + 1\}$ and $0 < \lambda_{m+1} \leq \lambda_m$ for all $m < v(W_1)$.

Proposition 3.1. For $d \in \mathbb{N}$ the non-trivial eigenpairs of the operator $W_{d,s_d} = S_{d,s_d}^\dagger S_{d,s_d}$ are given by $\left\{ \left(\tilde{\lambda}_{d,s_d,\mathbf{m}}, \tilde{\phi}_{d,s_d,\mathbf{m}} \right) \mid \mathbf{m} \in \mathcal{M}_d = (\mathcal{M}_1)^d \right\}$, where

$$\tilde{\lambda}_{d,s_d,\mathbf{m}} = s_d \tilde{\lambda}_{d,\mathbf{m}} = s_d \prod_{k=1}^d \lambda_{m_k} \quad \text{and} \quad \tilde{\phi}_{d,s_d,\mathbf{m}} = \sqrt{s_d} \tilde{\phi}_{d,\mathbf{m}} = \sqrt{s_d} \bigotimes_{k=1}^d \phi_{m_k}. \quad (3.2)$$

Proof. Since $S_{d,s_d} f = S_d f$ for every $f \in \mathcal{F}_d$ (or H_d , respectively) we have

$$\langle S_{d,s_d} f, g \rangle_{\mathcal{G}_d} = \langle S_d f, g \rangle_{\mathcal{G}_d} = \langle f, S_d^\dagger g \rangle_{H_d} = s_d \cdot \langle f, S_d^\dagger g \rangle_{\mathcal{F}_d} = \langle f, (s_d \cdot S_d^\dagger) g \rangle_{\mathcal{F}_d}$$

for all $f \in \mathcal{F}_d$ and $g \in \mathcal{G}_d$. Thus, [\(2.4\)](#) and the uniqueness of the adjoint operator²⁰ yield that $S_{d,s_d}^\dagger = s_d \cdot S_d^\dagger$ holds pointwise and, consequently, W_{d,s_d} equals $s_d \cdot W_d$ as a mapping. Hence, from [Proposition 2.10](#) and the linearity of W_d we conclude that [\(3.2\)](#) indeed are eigenpairs of W_{d,s_d} . Due to the factor $\sqrt{s_d}$ and the relation [\(3.1\)](#) the eigenelements $\tilde{\phi}_{d,s_d,\mathbf{m}}$ are properly normalized in \mathcal{F}_d .

It remains to show that there cannot exist eigenpairs other than [\(3.2\)](#). This can be seen using arguments similar to them in the second part of the proof of [Proposition 2.10](#). To this end, note that due to [\(3.1\)](#) the inner product in \mathcal{F}_d equals zero if and only if the elements under consideration are orthogonal in H_d . \blacksquare

[Proposition 3.1](#) in hand, the rest of this section is straightforward. Namely, we can use the bijections $\psi = \psi_d$ from [Section 2.4.2](#) to define the non-increasing sequences $(\lambda_{d,s_d,i})_{i \in \mathbb{N}}$ by

$$\lambda_{d,s_d,i} = \begin{cases} \tilde{\lambda}_{d,s_d,\psi(i)}, & 1 \leq i < v(W_1)^d + 1, \\ 0, & \text{otherwise,} \end{cases}$$

for every $d \in \mathbb{N}$. The corresponding reordered eigenelements are denoted by $\phi_{d,s_d,i}$, $i < v(W_1)^d + 1$. Moreover, we again use [Corollary 2.7](#) to see that for any $d \in \mathbb{N}$ the n th optimal algorithm for S_{d,s_d} , $n \in \mathbb{N}_0$, is given by

$$A_{n,d,s_d}^* : \mathcal{F}_d \rightarrow \mathcal{G}_d, \quad f \mapsto A_{n,d,s_d}^*(f) = \sum_{i=1}^{\min\{n, v(W_1)^d\}} \langle f, \phi_{d,s_d,i} \rangle_{\mathcal{F}_d} \cdot S_{d,s_d} \phi_{d,s_d,i}.$$

²⁰Note that, clearly, S_{d,s_d} is compact if and only if S_d is compact.

It realizes the n th minimal worst case error in dimension d which equals

$$e^{\text{wor}}(n, d; S_{d,s_d}) = \Delta^{\text{wor}}(A_{n,d,s_d}^*; S_{d,s_d}) = \sqrt{\lambda_{d,s_d,n+1}}. \quad (3.3)$$

In particular, the case $n = 0$, i.e. the initial error

$$\varepsilon_d^{\text{init}} = \sqrt{\lambda_{d,s_d,1}} = \sqrt{s_d \cdot \lambda_1^d},$$

will play an important role in what follows.

3.2 Complexity

Similar to [Section 2.4](#) we proceed with the analysis of the information complexity of scaled tensor product problems $S_{(s)} = (S_{d,s_d})_{d \in \mathbb{N}}$ in the worst case setting. We first take a look at necessary and sufficient conditions for (strong) polynomial tractability with respect to absolute errors. Afterwards, in [Section 3.2.2](#), we complete these assertions and investigate respective conditions for weak tractability and the curse of dimensionality. Finally we will see in [Section 3.2.3](#) that the obtained improvements due to scaling are completely ruled out when we turn to the normalized error criterion.

As usual $\lambda = (\lambda_m)_{m \in \mathbb{N}}$ denotes the (extended) sequence of squared singular values of the underlying operator $S_1: H_1 \rightarrow \mathcal{G}_1$. To avoid triviality we assume that $\lambda_2 > 0$ throughout the rest of this section. The reason for this assumption is explicitly stated in [Section 2.4.3](#).

3.2.1 Polynomial tractability

The next statement is originally based on Theorem 3.1 of Woźniakowski [[Woz94b](#)] which provided the underlying idea for [[NW08](#), Theorem 5.5]. We extend the results stated there to the case of scaled problems.

Theorem 3.2. Let $S_{(s)} = (S_{d,s_d})_{d \in \mathbb{N}}$ denote a scaled tensor product problem in the sense of [Section 3.1](#). Assume $\lambda_2 > 0$ and consider the worst case setting w.r.t. the absolute error criterion. Then the following assertions are equivalent:

- (I) $S_{(s)}$ is strongly polynomially tractable.
- (II) $S_{(s)}$ is polynomially tractable.
- (III) There exists $\tau \in (0, \infty)$ such that $\lambda \in \ell_\tau$ and $\sup_{d \in \mathbb{N}} s_d \|\lambda\|_{\ell_\tau}^d < \infty$.
- (IV) There exists $\varrho \in (0, \infty)$ such that $\lambda \in \ell_\varrho$ and $\limsup_{d \rightarrow \infty} s_d^{1/d} < \frac{1}{\lambda_1}$.

If one of these (and hence all) conditions applies then the exponent of strong polynomial tractability is given by

$$p^* = \inf\{2\tau \mid \tau \text{ fulfills condition (III)}\}.$$

Proof. Step 1. Since (I) clearly implies (II) we start by proving “(II) \Rightarrow (III)”. Therefore let $S_{(s)}$ be polynomially tractable with the constants $C, p > 0$ and $q \geq 0$. Then [Theorem 2.8](#) yields that for all $\varrho > p/2$,

$$0 < C_\varrho = \sup_{d \in \mathbb{N}} \frac{1}{d^{2q/p}} \left(\sum_{i=\lceil(1+C)d^q\rceil}^{\infty} (\lambda_{d,s_d,i})^\varrho \right)^{1/\varrho} < \infty.$$

Because of $\lambda_{d,s_d,i} = s_d \lambda_{d,i}$ for any $d, i \in \mathbb{N}$ due to (3.2), this particularly implies that $s_1 \left(\sum_{m=\lceil 1+C \rceil}^{\infty} \lambda_m^\varrho \right)^{1/\varrho} \leq C_\varrho$ is finite and hence $\lambda = (\lambda_m)_{m \in \mathbb{N}} \in \ell_\varrho$. Moreover, we have

$$\sum_{i=1}^{\infty} (\lambda_{d,i})^\varrho = \|\lambda|_{\ell_\varrho}\|^{e d}, \quad \text{as well as} \quad \sum_{i=1}^{\lceil(1+C)d^q\rceil-1} (\lambda_{d,i})^\varrho \leq \lambda_1^{e d} (1+C) d^q,$$

and therefore

$$\|\lambda|_{\ell_\varrho}\|^{e d} - \lambda_1^{e d} (1+C) d^q \leq \left(\frac{C_\varrho d^{2q/p}}{s_d} \right)^\varrho \quad \text{for all } d \in \mathbb{N}. \quad (3.4)$$

Now let $\tau > \varrho$ and assume that (III) is violated for this τ . Then $\sup_{d \in \mathbb{N}} s_d \|\lambda|_{\ell_\tau}\|^d$ is infinite since $\lambda \in \ell_\varrho$ and $\ell_\varrho \hookrightarrow \ell_\tau$ with $\|\lambda|_{\ell_\tau}\| < \|\lambda|_{\ell_\varrho}\|$. That means, for any $C_0 \in (0, \infty)$ there necessarily exists a sequence $(d_k)_{k \in \mathbb{N}} \subset \mathbb{N}$ such that for every $k \in \mathbb{N}$

$$C_0 \leq s_{d_k} \|\lambda|_{\ell_\tau}\|^{d_k} = s_{d_k} \|\lambda|_{\ell_\varrho}\|^{d_k} / (t_{\varrho,\tau})^{d_k},$$

where we set $t_{\varrho,\tau} = \|\lambda|_{\ell_\varrho}\| / \|\lambda|_{\ell_\tau}\| > 1$. Hence, at least for all k larger than a certain $k_0 \in \mathbb{N}$, we conclude that $C_0 \leq s_{d_k} \|\lambda|_{\ell_\varrho}\|^{d_k} / d_k^{2q/p}$. In particular, we can choose $C_0 > C_\varrho$ such that $0 < C_1 = C_\varrho / C_0 < 1$. Therefore (3.4) implies

$$\|\lambda|_{\ell_\varrho}\|^{e d_k} - \lambda_1^{e d_k} (1+C) d_k^q \leq C_1^\varrho \|\lambda|_{\ell_\varrho}\|^{e d_k}, \quad k \geq k_0,$$

which leads to

$$\left(1 + \left(\frac{\lambda_2}{\lambda_1} \right)^\varrho \right)^{d_k} \leq \left(\sum_{m=1}^{\infty} \left(\frac{\lambda_m}{\lambda_1} \right)^\varrho \right)^{d_k} = \frac{\|\lambda|_{\ell_\varrho}\|^{e d_k}}{\lambda_1^{e d_k}} \leq C_2 d_k^q$$

for all $k \geq k_0$ and some $C_2 = (1+C)/(1-C_1^\varrho) > 0$. Since $\lambda_2 > 0$ and $\varrho > 0$ this is a contradiction and thus we have condition (III) for every $\tau > p/2$. Note that this also shows that

$$\inf\{2\tau \mid \tau \text{ fulfills condition (III)}\} \leq p^*.$$

Step 2. Next we show that (I) follows from (III). Let $\tau > 0$ be given such that (III) holds true and set $p = q = r = 0$, as well as $C = 1$. Then, with $f(d) = \left\lceil C (\min \{s_d \lambda_1^d, 1\})^{-p/2} d^q \right\rceil = 1$, we have

$$C_\tau = \sup_{d \in \mathbb{N}} \frac{1}{d^r} \left(\sum_{i=f(d)}^{\infty} (\lambda_{d,s_d,i})^\tau \right)^{1/\tau} = \sup_{d \in \mathbb{N}} s_d \|\lambda\|_{\ell_\tau}^d < \infty.$$

Once more we apply [Theorem 2.8](#) to obtain

$$n_{\text{abs}}^{\text{wor}}(\varepsilon, d; S_{d,s_d}) \leq (1 + C_\tau^\tau) \varepsilon^{-2\tau} \quad \text{for all } \varepsilon \in (0, 1] \quad \text{and every } d \in \mathbb{N}.$$

Thus $S_{(s)}$ is strongly polynomially tractable and

$$p^* \leq \inf \{2\tau \mid \tau \text{ fulfills condition (III)}\}.$$

Step 3. The implication “(III) \Rightarrow (IV)” can be seen as follows. Assume (III) to be valid for some $0 < \tau < \infty$ and set $\varrho = \tau$. Then, clearly, $\lambda \in \ell_\varrho$. If we now assume (IV) to be violated then for any $\delta > 0$ there needs to exist a sequence $(d_k)_{k \in \mathbb{N}} \subset \mathbb{N}$ such that for all k

$$s_{d_k}^{1/d_k} \geq 1/(\lambda_1 + \delta).$$

Hence, $s_{d_k} \|\lambda\|_{\ell_\tau}^{d_k} \geq (\|\lambda\|_{\ell_\tau} / (\lambda_1 + \delta))^{d_k}$ tends to infinity (as $k \rightarrow \infty$) if we take δ small enough so that $\lambda_1 + \delta < \|\lambda\|_{\ell_\tau}$. Since $\tau \in (0, \infty)$ and

$$\lambda_1 < (\lambda_1^\tau + \lambda_2^\tau)^{1/\tau} \leq \|\lambda\|_{\ell_\tau} < \infty$$

there needs to be some $\alpha_\tau \in (0, \infty)$ such that $\|\lambda\|_{\ell_\tau} = \lambda_1 + \alpha_\tau$. Choosing e.g. $\delta = \alpha_\tau/2$ gives the needed contradiction.

Step 4. Finally we have to show that conversely (IV) also implies (III). Therefore assume that we have (IV) for some $\varrho \in (0, \infty)$. Then there exist constants $d_0 \in \mathbb{N}$ and $\delta > 0$ such that

$$s_d^{1/d} \leq 1/(\lambda_1 + \delta) \quad \text{for all } d \geq d_0.$$

Furthermore note that the function $N(\tau) = \|\lambda\|_{\ell_\tau}$ is strictly decreasing and continuous on the interval $[\varrho, \infty]$ and that $N(\varrho) > \lambda_1 = N(\infty)$ because of the ordering of $\lambda = (\lambda_m)_{m \in \mathbb{N}}$. Hence there necessarily exists some $\tau \in [\varrho, \infty)$ such that $N(\tau) \leq \lambda_1 + \delta/2$, say. Thus, $\lambda \in \ell_\tau$ and for every $d \geq d_0$ we obtain

$$s_d \|\lambda\|_{\ell_\tau}^d \leq \left(\frac{\lambda_1 + \delta/2}{\lambda_1 + \delta} \right)^d \leq 1.$$

Since the term on the left is also finite for any $d = 1, \dots, d_0$ this completes the proof. \blacksquare

Observe that [Theorem 3.2](#) is not very surprising. Indeed, the second assertion in condition (IV) is equivalent to the fact that the d th root of the initial error $\varepsilon_d^{\text{init}}$ is asymptotically strictly less than 1. Hence if $S_1: H_1 \rightarrow \mathcal{G}_1$ (and thus also the sequence λ) is given then we need to select scaling factors s_d such that $\varepsilon_d^{\text{init}} \rightarrow 0$, $d \rightarrow \infty$, in order to obtain polynomial tractability. More advanced illustrations will be given in [Section 3.3](#).

3.2.2 Weak tractability and the curse

To formulate necessary and sufficient conditions for weak tractability w.r.t. the worst case setting and the absolute error criterion we need some additional notation. Therefore let $S_{(s)} = (S_{d,s_d})_{d \in \mathbb{N}}$ denote a scaled tensor product problem between Hilbert spaces as explained in [Section 3.1](#) and assume $\lambda_2 > 0$. Then for fixed $d \in \mathbb{N}$ and $0 < \varepsilon < \varepsilon_d^{\text{init}} = s_d^{1/2} \lambda_1^{d/2}$ formula (3.3) implies

$$\begin{aligned} n(\varepsilon, d) &= \min \{n \in \mathbb{N}_0 \mid \lambda_{d,s_d,n+1} \leq \varepsilon^2\} = \#\{\mathbf{j} \in \mathbb{N}^d \mid s_d \cdot \lambda_{j_1} \cdots \lambda_{j_d} > \varepsilon^2\} \\ &= \#\left\{\mathbf{j} \in \mathbb{N}^d \mid \frac{\lambda_{j_1}}{\lambda_1} \cdots \frac{\lambda_{j_d}}{\lambda_1} > \left(\frac{\varepsilon}{\varepsilon_d^{\text{init}}}\right)^2\right\}. \end{aligned} \quad (3.5)$$

By counting the number of indices equal to one we conclude that

$$n(\varepsilon, d) = 1 + \sum_{k=1}^d \binom{d}{k} \cdot \#\left\{\mathbf{j} = (j_1, \dots, j_k) \in (\mathbb{N} \setminus \{1\})^k \mid \prod_{l=1}^k \frac{\lambda_{j_l}}{\lambda_1} > \left(\frac{\varepsilon}{\varepsilon_d^{\text{init}}}\right)^2\right\}.$$

Now we distinguish two cases. First assume that $\lambda_1 = \lambda_2$. Then, obviously, each the sets in the latter equality contains at least one element. Otherwise, in the case $\lambda_1 > \lambda_2$, some of these k -dimensional sets might be empty if k is larger than some $k_d(\varepsilon)$. The reason is that $\lambda_1 > \lambda_2 \geq \lambda_m$ for $m \geq 2$ implies that every factor in $\prod_{l=1}^k \lambda_{j_l}/\lambda_1$ is strictly smaller than 1. In detail, $(\varepsilon/\varepsilon_d^{\text{init}})^2 \geq (\lambda_2/\lambda_1)^k$ is equivalent to

$$k > k_d(\varepsilon) = \left\lceil \frac{1}{\ln(\lambda_1/\lambda_2)} \cdot \ln(\varepsilon_d^{\text{init}}/\varepsilon)^2 \right\rceil - 1.$$

Hence, denoting $a_d(\varepsilon) = \min\{d, k_d(\varepsilon)\}$ we have

$$n(\varepsilon, d) = 1 + \sum_{k=1}^{a_d(\varepsilon)} \binom{d}{k} \cdot \#\left\{\mathbf{j} \in (\mathbb{N} \setminus \{1\})^k \mid \prod_{l=1}^k \frac{\lambda_{j_l}}{\lambda_1} > \left(\frac{\varepsilon}{\varepsilon_d^{\text{init}}}\right)^2\right\} \quad (3.6)$$

for $d \in \mathbb{N}$ and $0 < \varepsilon < \varepsilon_d^{\text{init}}$. If $\lambda_1 = \lambda_2$ then the same equality remains true when we formally set $k_d(\varepsilon) = \infty$, i.e. $a_d(\varepsilon) = d$. Moreover, for $d \in \mathbb{N}$ we have $a_d(\varepsilon) = 0$ if and only if $\varepsilon \geq (\lambda_2/\lambda_1)^{1/2} \varepsilon_d^{\text{init}}$. If so, then we obtain $n(\varepsilon, d) = 1$ as long as $\varepsilon < \varepsilon_d^{\text{init}}$ and $n(\varepsilon, d) = 0$ otherwise.

Finally the following statement relates the decay properties of the univariate sequence of the squared singular values λ with the growth behavior of the information complexity $n(\varepsilon, d) = n_{\text{abs}}^{\text{wor}}(\varepsilon, d; S_{d, s_d})$. It generalizes an assertion given in Novak and Woźniakowski [NW08, p. 178].

Lemma 3.3. Let $S_{(s)}$ and $\lambda = (\lambda_m)_{m \in \mathbb{N}}$ be given as before. Then, for all $\beta \geq 1$,

$$\lambda_n \in o(\ln^{-2\beta} n), \text{ as } n \rightarrow \infty, \quad \text{if and only if} \quad \ln n(t^\beta, 1) \in o(1/t), \text{ as } t \rightarrow 0.$$

Proof. Assume $\beta \geq 1$ to be fixed and let $t \in (0, (s_1 \lambda_2)^{1/(2\beta)})$. Then (3.5) yields that for $d = 1$

$$n = n(t^\beta, 1) = \min \{n \in \mathbb{N}_0 \mid s_1 \cdot \lambda_{n+1} \leq t^{2\beta}\} \geq 2.$$

Thus we have $s_1 \cdot \lambda_{n(t^\beta, 1)+1} \leq t^{2\beta} < s_1 \cdot \lambda_{n(t^\beta, 1)}$ and $\ln^{2\beta} n \geq 1/4^\beta \cdot \ln^{2\beta}(n+1)$. Combining both these estimates we conclude

$$\frac{s_1}{4^\beta} \cdot \frac{\lambda_{n(t^\beta, 1)+1}}{\ln^{-2\beta}(n(t^\beta, 1)+1)} \leq \left(\frac{\ln n(t^\beta, 1)}{t^{-1}} \right)^{2\beta} < s_1 \cdot \frac{\lambda_{n(t^\beta, 1)}}{\ln^{-2\beta} n(t^\beta, 1)}.$$

Since the one-dimensional information complexity $n(\varepsilon, 1)$ is an increasing function in $1/\varepsilon$ taking the limit for $t \rightarrow 0$ proves the claim. \blacksquare

Now we are well-prepared to present necessary conditions for weak tractability based on the representation of the information complexity given in (3.6).

Proposition 3.4. Weak tractability of $S_{(s)}$ implies

$$\lim_{\varepsilon^{-1}+d \rightarrow \infty} \frac{\ln \sum_{k=0}^{a_d(\varepsilon)} \binom{d}{k}}{\varepsilon^{-1} + d} = 0 \quad \text{and} \quad \lim_{\varepsilon^{-1}+d \rightarrow \infty} \frac{\ln n(\varepsilon_1^{\text{init}} \cdot \varepsilon / \varepsilon_d^{\text{init}}, 1)}{\varepsilon^{-1} + d} = 0. \quad (3.7)$$

If so, then $\lambda_n \in o(\ln^{-2} n)$, as $n \rightarrow \infty$. Furthermore, we have

$$\ln(\varepsilon_d^{\text{init}}) \in o(d), \quad \text{as } d \rightarrow \infty, \quad (3.8)$$

since otherwise $S_{(s)}$ suffers from the curse of dimensionality. If, in addition, $\lambda_1 = \lambda_2$ then we need to claim $\lim_{d \rightarrow \infty} \varepsilon_d^{\text{init}} = 0$ to avoid the curse. Moreover, in this case weak tractability even yields

$$\varepsilon_d^{\text{init}} \in o(1/d), \quad \text{as } d \rightarrow \infty. \quad (3.9)$$

Proof. Step 1. We start by proving the necessity of the first limit condition in (3.7) and study its consequences. To this end, recall that due to the definition of $a_d(\varepsilon)$ we know that all the sets in (3.6) contain at least one element. Consequently, for general $\lambda_1 \geq \lambda_2$ we have

$$n(\varepsilon, d) \geq 1 + \sum_{k=1}^{a_d(\varepsilon)} \binom{d}{k} = \sum_{k=0}^{a_d(\varepsilon)} \binom{d}{k} \quad \text{for } d \in \mathbb{N} \quad \text{and} \quad 0 < \varepsilon < \varepsilon_d^{\text{init}}. \quad (3.10)$$

Now assume the existence of some subsequence $(d_l)_{l \in \mathbb{N}} \subset \mathbb{N}$ such that the initial error $\varepsilon_{d_l}^{\text{init}}$ grows at least exponentially in d_l for l tending to infinity. That is, we assume the condition (3.8) to be violated. Moreover, consider $\varepsilon = \varepsilon_0 \in (0, \inf\{\varepsilon_{d_l}^{\text{init}} \mid l \in \mathbb{N}\})$ to be fixed. Then for any $l \in \mathbb{N}$ and some $\alpha \in (0, 1/2)$ the term $a_{d_l}(\varepsilon_0)$ is bounded from below by $\lfloor \alpha d_l \rfloor$. Accordingly, (3.10) implies $n(\varepsilon_0, d_l) \geq \binom{d_l}{\lfloor \alpha d_l \rfloor}$ for all $l \in \mathbb{N}$. Using similar calculations as in [NW08, p. 178] we see that this lower bound grows exponentially in d_l . This proves the curse of dimensionality for the scaled problem $S_{(s)}$ and thus it contradicts weak tractability.

If we assume in addition that $\lambda_1 = \lambda_2$ then, as already noticed, $a_d(\varepsilon)$ equals d because of $k_d(\varepsilon) = \infty$. Thus we obtain $\sum_{k=0}^{a_d(\varepsilon)} \binom{d}{k} = 2^d$ in this case. Therefore the existence of a sequence $(d_l)_{l \in \mathbb{N}}$ such that $\varepsilon_{d_l}^{\text{init}}$ is larger than some $C > 0$ for all $l \in \mathbb{N}$ would again imply the curse of dimensionality since then we could fix $\varepsilon = \varepsilon_0 = C/2$, say. Moreover $a_d(\varepsilon) = d$ shows that the first part of (3.7) equivalently reads

$$\lim_{\varepsilon^{-1} + d \rightarrow \infty} \frac{d}{\varepsilon^{-1} + d} = 0. \quad (3.11)$$

Observe that in any case the term $d/(\varepsilon^{-1} + d)$ is equivalent to the minimum of 1 and εd (up to some absolute constants). Hence, (3.11) holds true if and only if

$$\lim_{\varepsilon^{-1} + d \rightarrow \infty} \varepsilon d = 0$$

which in turn is equivalent to $\varepsilon_d^{\text{init}} \in o(1/d)$ for $d \rightarrow \infty$. To see this last equivalence, remember that due to Section 1.4 the domain of the sequences $((\varepsilon_k, d_k))_{k \in \mathbb{N}}$ for the limit $\varepsilon^{-1} + d \rightarrow \infty$ is restricted per definition to those for which $\varepsilon_k < \varepsilon_{d_k}^{\text{init}}$.

Step 2. We turn to the proof of the second point in (3.7). Again we distinguish the cases $\lambda_1 = \lambda_2$ and $\lambda_1 > \lambda_2$. For the latter case keep in mind that $a_d(\varepsilon) \geq 1$ if and only if $\varepsilon < \varepsilon_d^{\text{init}}(\lambda_2/\lambda_1)^{1/2}$. If so, then (3.6) shows that

$$\begin{aligned} n(\varepsilon, d) &\geq 1 + \binom{d}{1} \cdot \#\left\{j \geq 2 \mid \frac{\lambda_j}{\lambda_1} > \left(\frac{\varepsilon}{\varepsilon_d^{\text{init}}}\right)^2\right\} \\ &= 1 + d \cdot \#\left\{j \geq 2 \mid s_1 \lambda_j > (\varepsilon_1^{\text{init}} \cdot \varepsilon / \varepsilon_d^{\text{init}})^2\right\} \geq n(\varepsilon_1^{\text{init}} \cdot \varepsilon / \varepsilon_d^{\text{init}}, 1). \end{aligned}$$

On the other hand, if $\varepsilon \in [\varepsilon_d^{\text{init}}(\lambda_2/\lambda_1)^{1/2}, \varepsilon_d^{\text{init}})$ then $n(\varepsilon_1^{\text{init}} \cdot \varepsilon / \varepsilon_d^{\text{init}}, 1)$ is no larger than $n(\varepsilon_1^{\text{init}}(\lambda_2/\lambda_1)^{1/2}, 1)$ which is an absolute, positive constant. Thus, as claimed in (3.7), we conclude

$$0 \leq \frac{\ln n(\varepsilon_1^{\text{init}} \cdot \varepsilon / \varepsilon_d^{\text{init}}, 1)}{\varepsilon^{-1} + d} \leq \max \left\{ \frac{\ln n(\varepsilon, d)}{\varepsilon^{-1} + d}, \frac{\ln n(\varepsilon_1^{\text{init}}(\lambda_2/\lambda_1)^{1/2}, 1)}{\varepsilon^{-1} + d} \right\} \rightarrow 0 \quad (3.12)$$

for $\varepsilon^{-1} + d$ tending to infinity in the above sense. In the case $\lambda_1 = \lambda_2$ we have $a_d(\varepsilon) = d$ which is trivially bounded from below by 1 for any $\varepsilon \in (0, \varepsilon_d^{\text{init}})$. The assertion now follows using the same arguments as in the first part of the previous

case. To complete the proof it finally remains to show that $\lambda_n \in o(\ln^{-2} n)$, as $n \rightarrow \infty$. Let us consider the case $d = d_k \equiv 1$ in (3.12). Then we obtain

$$0 \leq \frac{\ln n(\varepsilon, 1)}{\varepsilon^{-1}} \leq 2 \cdot \frac{\ln n(\varepsilon_1^{\text{init}} \cdot \varepsilon / \varepsilon_1^{\text{init}}, 1)}{\varepsilon^{-1} + 1} \rightarrow 0, \quad \text{as } \varepsilon^{-1} \rightarrow \infty.$$

In other words, weak tractability yields $\ln n(\varepsilon, 1) \in o(\varepsilon^{-1})$ which is equivalent to the claimed assertion due to Lemma 3.3. \blacksquare

Let us add some comments on the latter necessary conditions.

Remark 3.5. First of all note that from (3.7) we concluded (3.8) which is equivalent to the fact that $\limsup_{d \rightarrow \infty} s_d^{1/d} \leq 1/\lambda_1$. Aside from that (3.7) also implies another condition which we will need later on; namely

$$\lim_{\varepsilon^{-1} + d \rightarrow \infty} \frac{\ln n\left(\varepsilon_1^{\text{init}} \cdot (\varepsilon / \varepsilon_d^{\text{init}})^{1/2}, 1\right)}{\varepsilon^{-1} + d} = 0.$$

Conclusively we stress that the second point of (3.7) already indicates a certain trade-off between the decay of the sequence $(\lambda_n)_{n \in \mathbb{N}}$ and the growth of the initial error $\varepsilon_d^{\text{init}}$. Indeed, if $(\lambda_n)_{n \in \mathbb{N}}$ decreases almost logarithmically then $n(t, 1)$ increases subexponentially as t tends to zero. Consequently (3.7) can be fulfilled only if $\varepsilon_d^{\text{init}}$ is polynomially bounded in d . On the other hand, if the (squares of the) singular values tend to zero like the inverse of some polynomial, say, then $n(t, 1)$ grows polynomially in $1/t$ and hence it is enough to assume that the initial error is subexponentially bounded in d to fulfill (3.7). \square

We complement the necessary conditions in Proposition 3.4 by the following sufficient conditions for weak tractability of scaled tensor product problems $S_{(s)} = (S_{d, s_d})_{d \in \mathbb{N}}$. For the proof we essentially follow the arguments of Papageorgiou and Petras [PP09] for the unscaled case which are based on estimates from Woźniakowski [Woz94b].

Proposition 3.6. Let $S_{(s)}$ and $a_d(\varepsilon)$ be defined as before and assume that $\lambda_2 > 0$. If the condition (3.7) from Proposition 3.4 holds true and if we have

$$\lim_{\varepsilon^{-1} + d \rightarrow \infty} \frac{a_d(\varepsilon) \cdot \ln n\left(\varepsilon_1^{\text{init}} \cdot (\varepsilon / \varepsilon_d^{\text{init}})^{1/2}, 1\right)}{\varepsilon^{-1} + d} = 0 \quad (3.13)$$

then $S_{(s)}$ is weakly tractable.

Proof. Given $d \in \mathbb{N}$ and $\varepsilon \in (0, (\lambda_2/\lambda_1)^{1/2} \varepsilon_d^{\text{init}})$ consider the representation (3.6) and keep in mind that for larger ε the information complexity $n(\varepsilon, d)$ is trivially bounded by 1 because then $a_d(\varepsilon) = 0$. For every $k \in \{1, \dots, a_d(\varepsilon)\}$ we have

$$\# \left\{ \mathbf{j} \in (\mathbb{N} \setminus \{1\})^k \left| \prod_{l=1}^k \frac{\lambda_{j_l}}{\lambda_1} > \left(\frac{\varepsilon}{\varepsilon_d^{\text{init}}} \right)^2 \right. \right\} \leq \# \left\{ \mathbf{j} \in \mathbb{N}^{a_d(\varepsilon)} \left| \prod_{l=1}^{a_d(\varepsilon)} \frac{\lambda_{j_l}}{\lambda_1} > \left(\frac{\varepsilon}{\varepsilon_d^{\text{init}}} \right)^2 \right. \right\}$$

since $\lambda_m/\lambda_1 \leq 1$ for all $m \in \mathbb{N}$. Hence we concentrate on all the multi-indices $\mathbf{j} = (j_1, \dots, j_{a_d(\varepsilon)})$ that fulfill

$$\prod_{l=1}^{a_d(\varepsilon)} \frac{\lambda_{j_l}}{\lambda_1} > \left(\frac{\varepsilon}{\varepsilon_d^{\text{init}}} \right)^2. \quad (3.14)$$

Clearly the largest possible index $j_{\max}^{(1)}$ which can appear in those $\mathbf{j} \in \mathbb{N}^{a_d(\varepsilon)}$ is bounded because the sequence $(\lambda_n)_{n \in \mathbb{N}}$ tends to zero as $n \rightarrow \infty$. Indeed, using the arguments given in [PP09] we conclude that

$$j_{\max}^{(1)} \leq \min \left\{ n \in \mathbb{N}_0 \mid s_1 \lambda_{n+1} \leq \left(\varepsilon_1^{\text{init}} \frac{\varepsilon}{\varepsilon_d^{\text{init}}} \right)^2 \right\} = n \left(\varepsilon_1^{\text{init}} \frac{\varepsilon}{\varepsilon_d^{\text{init}}}, 1 \right).$$

More generally, in [PP09] it was noticed that, using the same reasoning, we can bound the i th largest index $j_{\max}^{(i)}$ in (3.14) by

$$j_{\max}^{(i)} \leq n \left(\varepsilon_1^{\text{init}} \cdot (\varepsilon/\varepsilon_d^{\text{init}})^{1/i}, 1 \right).$$

We use this estimate for $i = 1$ and 2 to conclude the upper bound

$$a_d(\varepsilon) \cdot n \left(\varepsilon_1^{\text{init}} \cdot \varepsilon/\varepsilon_d^{\text{init}}, 1 \right) \cdot n \left(\varepsilon_1^{\text{init}} \cdot (\varepsilon/\varepsilon_d^{\text{init}})^{1/2}, 1 \right)^{a_d(\varepsilon)-1}$$

for $\# \{ \mathbf{j} \in \mathbb{N}^{a_d(\varepsilon)} \mid \mathbf{j} \text{ fulfills (3.14)} \}$. Note that due to $\varepsilon < \varepsilon_d^{\text{init}}$ both the univariate complexities in the latter bound need to be at least 1. Therefore we can extend the estimate by adding an additional factor $n \left(\varepsilon_1^{\text{init}} \cdot (\varepsilon/\varepsilon_d^{\text{init}})^{1/2}, 1 \right)$ and replacing $a_d(\varepsilon)$ by d . In summary we have

$$n(\varepsilon, d) \leq d \cdot n \left(\varepsilon_1^{\text{init}} \cdot \varepsilon/\varepsilon_d^{\text{init}}, 1 \right) \cdot n \left(\varepsilon_1^{\text{init}} \cdot (\varepsilon/\varepsilon_d^{\text{init}})^{1/2}, 1 \right)^{a_d(\varepsilon)} \cdot \sum_{k=0}^{a_d(\varepsilon)} \binom{d}{k} \quad (3.15)$$

for each $d \in \mathbb{N}$ and all $\varepsilon \in (0, (\lambda_2/\lambda_1)^{1/2} \varepsilon_d^{\text{init}})$. Because of $n(\varepsilon, d) = 1$ if ε belongs to $[(\lambda_2/\lambda_1)^{1/2} \varepsilon_d^{\text{init}}, \varepsilon_d^{\text{init}})$, the estimate (3.15) remains valid for every $\varepsilon \in (0, \varepsilon_d^{\text{init}})$. Proceeding as in [PP09] we take the logarithm and divide by $\varepsilon^{-1} + d$ to conclude

$$\begin{aligned} \frac{\ln n(\varepsilon, d)}{\varepsilon^{-1} + d} &\leq \frac{\ln(d)}{\varepsilon^{-1} + d} + \frac{\ln \left[n \left(\varepsilon_1^{\text{init}} \cdot \varepsilon/\varepsilon_d^{\text{init}}, 1 \right) \right]}{\varepsilon^{-1} + d} \\ &\quad + \frac{a_d(\varepsilon) \cdot \ln \left[n \left(\varepsilon_1^{\text{init}} \cdot (\varepsilon/\varepsilon_d^{\text{init}})^{1/2}, 1 \right) \right]}{\varepsilon^{-1} + d} + \frac{\sum_{k=0}^{a_d(\varepsilon)} \binom{d}{k}}{\varepsilon^{-1} + d}. \end{aligned}$$

For weak tractability it suffices to show that each of these fractions tends to zero as $\varepsilon^{-1} + d$ approaches infinity. Obviously, for the first one this is true without any further conditions. For the second and fourth fraction the assertion follows from (3.7). Finally the third fraction tends to zero due to the additional condition (3.13) we imposed for this proposition. \blacksquare

To illustrate the obtained results the following theorem considers several cases for the behavior of the initial error $\varepsilon_d^{\text{init}}$.

Theorem 3.7. Let $S_{(s)} = (S_{d,s_d})_{d \in \mathbb{N}}$ denote a scaled tensor product problem in the sense of Section 3.1. Assume that $\lambda_2 > 0$ and consider the worst case setting w.r.t. the absolute error criterion.

- Let $\ln(\varepsilon_d^{\text{init}}) \notin o(d)$, as $d \rightarrow \infty$.
Then $S_{(s)}$ suffers from the curse of dimensionality.
- Let $\varepsilon_d^{\text{init}} \in \Theta(d^\alpha)$, as $d \rightarrow \infty$, for some $\alpha \geq 0$.
 - If $\lambda_1 = \lambda_2$ then $S_{(s)}$ suffers from the curse of dimensionality.
 - In the case $\lambda_1 > \lambda_2$ the problem $S_{(s)}$ is weakly tractable if and only if

$$\lambda_n \in o\left(\ln^{-2(1+\alpha)} n\right), \text{ as } n \rightarrow \infty. \quad (3.16)$$

- Let $\varepsilon_d^{\text{init}} \rightarrow 0$, as d approaches infinity.
Then we never have the curse of dimensionality. Moreover, $S_{(s)}$ is weakly tractable if and only if
 - (i) $\lambda_1 = \lambda_2$ and $\lambda_n \in o(\ln^{-2} n)$, as $n \rightarrow \infty$, and $\varepsilon_d^{\text{init}} \in o(1/d)$, as $d \rightarrow \infty$,
or
 - (ii) $\lambda_1 > \lambda_2$ and $\lambda_n \in o(\ln^{-2} n)$, as $n \rightarrow \infty$.

Proof. Step 1. In this first step we handle the assertions concerning the curse of dimensionality. From the proof of Proposition 3.4 we know that $S_{(s)}$ suffers from the curse if either $\ln(\varepsilon_d^{\text{init}}) \notin o(d)$, or if $\lambda_1 = \lambda_2$ and $\lim_{d \rightarrow \infty} \varepsilon_d^{\text{init}} \neq 0$. Of course the latter condition is fulfilled particularly if the initial error grows polynomially with the dimension d , i.e. if $\varepsilon_d^{\text{init}} \in \Theta(d^\alpha)$ for some $\alpha \geq 0$. Furthermore the fact that we cannot have the curse of dimensionality as long as $\varepsilon_d^{\text{init}}$ tends to zero is clear from the definition.

Step 2. Next we show that weak tractability implies (3.16). Therefore we note that $\varepsilon_d^{\text{init}} \in \Theta(d^\alpha)$ implies the existence of some $c > 0$ such that we have $\varepsilon_d^{\text{init}} \geq c d^\alpha$ for all $d \in \mathbb{N}$. Moreover we see that there is some $d_0 \in \mathbb{N}$ such that $1/c < d^{1+\alpha}$ for every d larger than d_0 . Setting $\varepsilon = 1/d$ now yields

$$\frac{\varepsilon}{\varepsilon_d^{\text{init}}} \leq \frac{1}{c} \cdot \frac{1}{d^{1+\alpha}} < 1 \quad \text{for all } d \geq d_0$$

and $\varepsilon^{-1} + d = 2d \rightarrow \infty$, as $d \rightarrow \infty$. Hence, the sequence $((1/d, d))_{d \geq d_0}$ is admissible for the second limit condition of (3.7) in Proposition 3.4. On the other hand, we have

$$\frac{\ln n(\varepsilon_1^{\text{init}} \cdot \varepsilon / \varepsilon_d^{\text{init}}, 1)}{\varepsilon^{-1} + d} \geq \frac{c'}{2} \cdot \frac{\ln n((c'/d)^{1+\alpha}, 1)}{(c'/d)^{-1}} \geq 0$$

where we set $c' = (\varepsilon_1/c)^{1/(1+\alpha)}$. Thus weak tractability implies $\ln n(t^{1+\alpha}, 1) \in o(1/t)$ for $t \rightarrow 0$. Now the assertion follows from [Lemma 3.3](#).

Step 3. For the case of polynomial initial errors it remains to prove the converse implication, namely that [\(3.16\)](#) is also sufficient for weak tractability provided that $\lambda_1 > \lambda_2$. To this end, we first show that for all $d \in \mathbb{N}$, every $\varepsilon \in (0, \varepsilon_d^{\text{init}})$ and for some $C > 0$,

$$\varepsilon_d^{\text{init}}/\varepsilon \leq C \cdot (\varepsilon^{-1} + d)^{1+\alpha}. \quad (3.17)$$

To see this, we notice that $\varepsilon_d^{\text{init}} \in \Theta(d^\alpha)$ implies the existence of some $C > 0$ such that $\varepsilon_d^{\text{init}} \leq C d^\alpha$ for all $d \in \mathbb{N}$. If $\alpha = 0$ then [\(3.17\)](#) is obvious. For the case $\alpha > 0$ we apply Young's inequality²¹ and obtain $\varepsilon_d^{\text{init}}/\varepsilon \leq C(d^{1+\alpha} + (\varepsilon^{-1})^{1+\alpha})$. Now the inequality [\(3.17\)](#) follows from the relation $\|\cdot\|_{\ell_{1+\alpha}} \leq \|\cdot\|_{\ell_1}$, $\alpha \geq 0$, for (two-dimensional) sequence spaces.

We want to conclude weak tractability from [Proposition 3.6](#). Hence we have to check the limit conditions stated in [\(3.7\)](#) and [\(3.13\)](#). In what follows we abbreviate the notation and set

$$t = t(\varepsilon, d) = \varepsilon^{-1} + d.$$

Given [\(3.17\)](#), as well as the definition of $a_d(\varepsilon)$ in front of formula [\(3.6\)](#), it is easy to see that

$$a_d(\varepsilon) \in \mathcal{O}(\ln(t)), \text{ as } t \rightarrow \infty.$$

In particular, we have $a_d(\varepsilon) < \lfloor [t]/2 \rfloor$, if t is sufficiently large. Moreover note that $d < t$ implies $\binom{d}{k} \leq \binom{\lfloor t \rfloor}{k}$ for all $k \in \{0, 1, \dots, a_d(\varepsilon)\}$ such that $\binom{\lfloor t \rfloor}{a_d(\varepsilon)}$ is an upper bound for each of those binomial coefficients $\binom{d}{k}$. Consequently,

$$\begin{aligned} \ln \sum_{k=0}^{a_d(\varepsilon)} \binom{d}{k} &\leq \ln \left((a_d + 1) \cdot \binom{\lfloor t \rfloor}{a_d(\varepsilon)} \right) \leq \ln(2 a_d(\varepsilon) (e \lfloor t \rfloor)^{a_d(\varepsilon)}) \\ &\leq \ln 2 + \ln(a_d(\varepsilon)) + a_d(\varepsilon) \cdot (1 + \ln(t)) \\ &\in \mathcal{O}(\ln^2(t)) \subseteq o(t), \end{aligned}$$

for $t \rightarrow \infty$. In other words, the first part of condition [\(3.7\)](#) is fulfilled. Also the second limit condition in [\(3.7\)](#) can be shown easily using [\(3.17\)](#). Indeed, due to the assumption in [\(3.16\)](#) (or its equivalent reformulation due to [Lemma 3.3](#), respectively) we conclude that

$$\frac{\ln n(\varepsilon_1^{\text{init}} \cdot \varepsilon/\varepsilon_d^{\text{init}}, 1)}{\varepsilon^{-1} + d} \leq \frac{\ln n(\varepsilon_1^{\text{init}} C (\varepsilon^{-1} + d)^{-(1+\alpha)}, 1)}{\varepsilon^{-1} + d} = \frac{1}{C'} \cdot \frac{\ln n((C'/t)^{1+\alpha}, 1)}{(C'/t)^{-1}}$$

²¹Recall that Young's inequality states that $a, b \geq 0$ and $p, q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$ yields that $ab \leq \frac{1}{p} a^p + \frac{1}{q} b^q$. We use this assertion for $a = d^\alpha$, $b = \varepsilon^{-1}$ and $p = 1 + 1/\alpha$, $q = 1 + \alpha$.

tends to zero as $\varepsilon^{-1} + d$ (and therefore also t/C') approaches infinity. Finally,

$$\begin{aligned} a_d(\varepsilon) \cdot \ln n\left(\varepsilon_1^{\text{init}} \cdot (\varepsilon/\varepsilon_d^{\text{init}})^{1/2}, 1\right) &\leq a_d(\varepsilon) \cdot \ln n\left(\left(\frac{C''}{\varepsilon^{-1} + d}\right)^{(1+\alpha)/2}, 1\right) \\ &\in \mathcal{O}(\ln t) \cdot \mathcal{O}(t^{1/2}) \subseteq o(t) \end{aligned}$$

for $t = \varepsilon^{-1} + d$ tending to infinity. Hence we have shown (3.13). Now the application of Proposition 3.6 completes the proof for the case of polynomial initial errors.

Step 4. In this last step we consider the case of initial errors which tend to zero for d tending to infinity. We already know from Proposition 3.4 that $\lambda_n \in o(\ln^{-2} n)$, $n \rightarrow \infty$, (or equivalently $\ln n(t, 1) \in o(1/t)$, as $t \rightarrow 0$) is necessary for weak tractability, independent of the relation of the two largest (squares of the) singular values λ_1 and λ_2 to each other. Moreover Proposition 3.4 states that $\varepsilon_d^{\text{init}} \in o(1/d)$, $d \rightarrow \infty$, is a necessary condition when we assume $\lambda_1 = \lambda_2$, in addition. It remains to show that these conditions are also sufficient for weak tractability in the particular situations.

If $\lambda_1 > \lambda_2$ then we can exactly follow the lines of Step 3 with $\alpha = 0$ in order to conclude the assertion. Hence we are left with the case $\lambda_1 = \lambda_2$. Similar to the previous step we want to apply Proposition 3.6 and thus we need to check the conditions in (3.7) and (3.13). Setting

$$u = u(\varepsilon, d) = \frac{d}{\varepsilon^{-1} + d} \quad (3.18)$$

we note that (due to (3.11) in the proof of Proposition 3.4) u tends to zero if $\varepsilon^{-1} + d \rightarrow \infty$. This follows from $\varepsilon_d^{\text{init}} \in o(1/d)$, as $d \rightarrow \infty$, and, on the other hand, it implies the first condition in (3.7) because $\sum_{k=0}^{a_d(\varepsilon)} \binom{d}{k}$ equals 2^d . Since, in particular, $\varepsilon_d^{\text{init}} \leq C$ for some $C > 0$, we have $C' = C/\varepsilon_1^{\text{init}} > 0$ and thus we obtain

$$\frac{\ln n(\varepsilon_1^{\text{init}} \cdot \varepsilon/\varepsilon_d^{\text{init}}, 1)}{\varepsilon^{-1} + d} \leq \frac{\ln n(\varepsilon_1^{\text{init}}/C \cdot \varepsilon, 1)}{\varepsilon^{-1} + d} \leq C' \cdot \frac{\ln n((C' \cdot (\varepsilon^{-1} + d))^{-1}, 1)}{C' \cdot (\varepsilon^{-1} + d)} \rightarrow 0,$$

as $\varepsilon^{-1} + d \rightarrow \infty$ due to $\ln n(t, 1) \in o(1/t)$, $t \rightarrow 0$. In other words, we have shown the second condition in (3.7). To see that also (3.13) holds true we once more use $\varepsilon_d^{\text{init}} \in o(1/d) \subseteq \mathcal{O}(1/d)$ as well as Young's inequality to conclude

$$\left(\frac{\varepsilon_d^{\text{init}}}{\varepsilon}\right)^{1/2} \leq C_1 \cdot \frac{1}{d} \cdot d^{1/2} \cdot \left(\frac{1}{\varepsilon}\right)^{1/2} \leq \frac{C_1}{2} \cdot \frac{\varepsilon^{-1} + d}{d}$$

with some $C_1 > 0$. Hence, using (3.18) we have $\varepsilon_1^{\text{init}} \cdot (\varepsilon/\varepsilon_d^{\text{init}})^{1/2} \geq C_2 u$ and therefore

$$\frac{d \cdot \ln n(\varepsilon_1^{\text{init}} \cdot (\varepsilon/\varepsilon_d^{\text{init}})^{1/2}, 1)}{\varepsilon^{-1} + d} \leq \frac{1}{C_2} \cdot (C_2 u) \cdot \ln n(C_2 u, 1) \rightarrow 0 \quad \text{if } \varepsilon^{-1} + d \rightarrow \infty,$$

because then $C_2 u = C_2 u(\varepsilon, d)$ tends to zero. Since $a_d(\varepsilon) = d$ this yields (3.13) and we are allowed to conclude weak tractability from Proposition 3.6. \blacksquare

Before we turn to normalized errors we want to stress the point that [Theorem 3.7](#) contains at least two surprising results. At first, we can have weak tractability even if the initial error of $S_{(s)}$ grows with increasing dimension. Hence, although the performance of the zero algorithm gets steadily worse for $d \rightarrow \infty$ we are not necessarily faced with the curse of dimensionality. In contrast, remember that we need decreasing initial errors in order to conclude polynomial tractability. Secondly, it seems to be quite surprising that also in the case $\lambda_1 = \lambda_2$ we can break the curse by imposing only moderate additional conditions on the scaling sequence s . Indeed, it is enough to guarantee that $\varepsilon_d^{\text{init}} = \sqrt{s_d \lambda_1^d} \in o(1/d)$ for $d \rightarrow \infty$.

3.2.3 Normalized errors

We complete our studies of the complexity of scaled problems $S_{(s)} = (S_{d,s_d})_{d \in \mathbb{N}}$ by investigating tractability properties with respect to the normalized error criterion. This can be done by analyzing the information complexity of a related problem w.r.t. absolute errors.

Let $\lambda = (\lambda_n)_{n \in \mathbb{N}}$ and $s = (s_d)_{d \in \mathbb{N}}$ be fixed and define a tensor product problem $T = (T_d: H_d \rightarrow \mathcal{G}_d)_{d \in \mathbb{N}}$ out of the building blocks $T_1 = (1/\sqrt{\lambda_1} S_1): H_1 \rightarrow \mathcal{G}_1$ as described in the proof of [Theorem 2.12](#). Then the extended sequence of squared singular values of T_d , based on the univariate sequence $\mu = (\mu_m)_{m \in \mathbb{N}} = (\lambda_m/\lambda_1)_{m \in \mathbb{N}}$, reads

$$(\mu_{d,1,i})_{i \in \mathbb{N}} = (\mu_{d,i})_{i \in \mathbb{N}} = \left(\frac{\lambda_{d,i}}{\lambda_{d,1}} \right)_{i \in \mathbb{N}} = \left(\frac{\lambda_{d,s_d,i}}{s_d \lambda_{d,1}} \right)_{i \in \mathbb{N}} = \left(\frac{\lambda_{d,s_d,i}}{(\varepsilon_d^{\text{init}})^2} \right)_{i \in \mathbb{N}}.$$

Here the second subscript in $\mu_{d,1,i}$ indicates that T can be seen as a trivially scaled tensor product problem. Furthermore, $\varepsilon_d^{\text{init}} = \sqrt{s_d \lambda_1^d}$ denotes the initial error of S_{d,s_d} . Thus, from [\(3.5\)](#) applied to T and $S_{(s)}$ we conclude

$$\begin{aligned} n_{\text{abs}}^{\text{wor}}(\varepsilon, d; T_d) &= \min \{n \in \mathbb{N}_0 \mid \mu_{d,1,n+1} \leq \varepsilon^2\} \\ &= \min \{n \in \mathbb{N}_0 \mid \lambda_{d,s_d,n+1} \leq (\varepsilon \cdot \varepsilon_d^{\text{init}})^2\} = n_{\text{abs}}^{\text{wor}}(\varepsilon \cdot \varepsilon_d^{\text{init}}, d; S_{d,s_d}). \end{aligned} \quad (3.19)$$

By definition this also equals $n_{\text{norm}}^{\text{wor}}(\varepsilon, d; S_{d,s_d})$, i.e. the information complexity of S_{d,s_d} w.r.t. to the normalized error criterion. This relation in hand, we can use our results from the previous subsections to prove the following assertion.

Theorem 3.8. Let $S_{(s)} = (S_{d,s_d})_{d \in \mathbb{N}}$ denote a scaled tensor product problem in the sense of [Section 3.1](#). Assume $\lambda_2 > 0$ and consider the worst case setting w.r.t. the normalized error criterion.

- Let $\lambda_1 = \lambda_2$. Then $S_{(s)}$ suffers from the curse of dimensionality.
- Let $\lambda_1 > \lambda_2$. Then $S_{(s)}$ is not polynomially tractable. Moreover, in this case $S_{(s)}$ is weakly tractable if and only if $\lambda_n \in o(\ln^{-2} n)$, as $n \rightarrow \infty$.

Proof. Note that for all $d \in \mathbb{N}$ the initial error of T_d is 1 since μ_1 , as well as the scaling parameters, equal 1. Thus, obviously, condition (IV) in [Theorem 3.2](#) is violated and therefore T is polynomially intractable w.r.t. to the absolute error criterion. Moreover, the second point of [Theorem 3.7](#) with $\alpha = 0$ shows that T suffers from the curse of dimensionality if $\mu_1 = \mu_2$. Otherwise, i.e. if $\mu_1 > \mu_2$, the problem T is weakly tractable if and only if $\mu_n \in o(\ln^{-2} n)$, as $n \rightarrow \infty$. Since we set $\mu_m = \lambda_m/\lambda_1$, $m \in \mathbb{N}$, all these conditions on $\mu = (\mu_m)_{m \in \mathbb{N}}$ are fulfilled if and only if the corresponding assertions holds true for the sequence $\lambda = (\lambda_m)_{m \in \mathbb{N}}$. Equation (3.19) finally shows that every complexity assertion for T w.r.t. absolute errors is equivalent to the corresponding statement for $S_{(s)}$ and the normalized error criterion. This simple observation completes the proof. \blacksquare

In conclusion the scaling sequence $s = (s_d)_{d \in \mathbb{N}}$ does not have any influence on the complexity of $S_{(s)}$, as long as we consider normalized errors. So the advantages of scaling are completely ruled out in this setting.

3.3 Examples

In this last part of [Chapter 3](#) we briefly discuss two applications of the complexity results obtained in the previous section. We start by proving that our assertions reproduce the known facts for unscaled tensor product problems studied in [Theorem 2.11](#) and [Theorem 2.12](#).

Example 3.9 (Unscaled problems). Let $S_{(s)} = (S_{d,s_d})_{d \in \mathbb{N}}$ denote a tensor product problem between Hilbert spaces in the sense of [Section 3.1](#) where all the scaling factors s_d equal 1. As usual we assume $\lambda_2 > 0$ and consider the worst case setting. Then for every $d \in \mathbb{N}$ the operators S_{d,s_d} coincide with S_d as defined in [Section 2.4.1](#). Since we already saw that for the normalized error criterion the conditions stated in [Theorem 3.8](#) exactly match the assertions of [Theorem 2.12](#), it remains to consider the absolute error criterion. Here $\varepsilon_d^{\text{init}}$ is given by $\lambda_1^{d/2}$. Hence there are three scenarios for the behavior of the initial error depending on the largest squared singular value λ_1 of the underlying operator S_1 .

From [Theorem 3.2](#) we know that strong polynomial tractability and polynomial tractability are equivalent; see (I) and (II), respectively. Moreover, condition (IV) shows that this holds if and only if $\lambda_1 < 1$ and $\lambda = (\lambda_m)_{m \in \mathbb{N}} \in \ell_\tau$ for some $\tau \in (0, \infty)$. In this case the exponent of strong polynomial tractability is given by

$$p^* = \inf \left\{ 2\tau \left| \sup_{d \in \mathbb{N}} \|\lambda\|_{\ell_\tau}^d < \infty \right. \right\} = \inf \left\{ 2\tau \left| \left(\sum_{m=1}^{\infty} (\lambda_m)^\tau \right)^{1/\tau} \leq 1 \right. \right\}.$$

In turn, $\lambda_1 \geq 1$ yields polynomial intractability. More precisely, if $\lambda_1 > 1$ then the initial error grows exponentially in d and $S = S_{(s)}$ suffers from the curse of dimensionality due to the first point of [Theorem 3.7](#). Setting $\alpha = 0$ the second point

of the latter theorem describes the case of constant initial errors which corresponds to the case $\lambda_1 = 1$ in the unscaled situation. In detail, if $\lambda_2 = \lambda_1 = 1$ then we are faced with the curse again. In contrast, if $\lambda_2 < 1$ then we have weak tractability if and only if

$$\lambda_n \in o(\ln^{-2} n), \quad \text{as } n \rightarrow \infty. \quad (3.20)$$

Finally the initial error tends to zero exponentially fast if $\lambda_1 < 1$. The last point of [Theorem 3.7](#) thus shows that in this case the curse of dimensionality is not possible and that (3.20) is necessary and sufficient for weak tractability.

Altogether these results exactly match the conditions stated in [Theorem 2.11](#). Hence, scaled tensor product problems indeed yield a generalization. \square

So let us turn to a more advanced application. To this end, recall the definition of $S_{d,s_d} : \mathcal{F}_d \rightarrow \mathcal{G}_d$ in [Section 3.1](#). There we constructed the source spaces \mathcal{F}_d by scaling the norm in the tensor product space $H_d = H_1 \otimes \dots \otimes H_1$. Alternatively we can think of \mathcal{F}_d as the successively taken tensor product of some building blocks $H^{(k)}$, $k = 1, \dots, d$, in the sense of [Section 2.4.1](#), where we define $H^{(k)}$ to be the univariate space H_1 scaled by some factor $s^{(k)} > 0$. That is, let

$$\langle \cdot, \cdot \rangle_{H^{(k)}} = \frac{1}{s^{(k)}} \langle \cdot, \cdot \rangle_{H_1}.$$

Then the scaling factor s_d in dimension d is given by $\prod_{k=1}^d s^{(k)} > 0$. The following example illustrates how the behavior of the *generator sequence* $(s^{(k)})_{k \in \mathbb{N}}$ effects the complexity of $S_{(s)}$.

Example 3.10. Because scaling has no influence on assertions for normalized errors we restrict ourselves to the absolute error criterion in what follows. For simplicity we further assume that $\lambda_1 > \lambda_2 > 0$ and that the generator sequence is non-increasing, i.e.

$$s^{(1)} \geq s^{(2)} \geq \dots \geq s^{(k)} \geq \dots > 0, \quad k \in \mathbb{N}.$$

Then [Theorem 3.2](#) states that $S_{(s)}$ is strongly polynomially tractable if and only if the geometric mean of the first d elements $s^{(k)}$ is asymptotically strictly smaller than $1/\lambda_1$, provided that $\lambda \in \ell_\tau$ for some $\tau > 0$. This holds iff at most finitely many of these generators are bounded from below by $1/\lambda_1$. Moreover, from [Remark 3.5](#) we know that we need

$$\limsup_{d \rightarrow \infty} s_d^{1/d} = \limsup_{d \rightarrow \infty} \left(\prod_{k=1}^d s^{(k)} \right)^{1/d} \leq \frac{1}{\lambda_1}$$

in order to obtain weak tractability. Therefore let the generators be given by

$$s^{(k)} = \frac{1}{\lambda_1} \cdot (1 + \delta_k), \quad k \in \mathbb{N},$$

with a non-increasing null sequence $(\delta_k)_{k \in \mathbb{N}}$ and note that then some elementary calculations yield

$$\exp\left(\frac{c}{2} \cdot \sum_{k=1}^d \delta_k\right) \leq \varepsilon_d^{\text{init}} \leq \exp\left(\frac{1}{2} \cdot \sum_{k=1}^d \delta_k\right), \quad d \in \mathbb{N},$$

where $c = \ln(1 + \delta_1)/\delta_1 \leq 1$. Furthermore this observation shows that $\varepsilon_d^{\text{init}} \in \Theta(d^\alpha)$, as $d \rightarrow \infty$, for some $\alpha \geq 0$ implies that

$$L = \lim_{d \rightarrow \infty} \frac{1}{\ln d} \sum_{k=1}^d \delta_k \in 2\alpha \cdot \left[1, \frac{1}{c}\right].$$

Conversely, from the existence of L it follows that for any $\delta > 0$ there is some $d_0 = d_0(\delta)$ such that

$$\varepsilon_d^{\text{init}} \in [d^{\alpha_1}, d^{\alpha_2}] \quad \text{for all } d \geq d_0,$$

where $\alpha_1 = c \cdot L/2 - \delta$ and $\alpha_2 = L/2 + \delta$. Hence, if L is sufficiently small then the initial error $\varepsilon_d^{\text{init}}$ behaves like a polynomial of small degree and thus a quite slow decay of the sequence $(\lambda_n)_{n \in \mathbb{N}}$ is enough to conclude weak tractability using [Theorem 3.7](#). \square

CHAPTER 4

Problems on function spaces with weighted norms

In [NW09] it is shown that the approximation problem defined on $C^\infty([0, 1]^d)$ is intractable. In fact, Novak and Woźniakowski considered the linear space F_d of all real-valued infinitely differentiable functions f defined on the unit cube $[0, 1]^d$ in d dimensions for which the norm

$$\|f\|_{F_d} = \sup_{\alpha \in \mathbb{N}_0^d} \|D^\alpha f\|_{L_\infty([0, 1]^d)} \quad (4.1)$$

of $f \in F_d$ is finite. In this case the (uniform) approximation problem is given by the sequence of solution operators $S = (S_d)_{d \in \mathbb{N}}$,

$$S_d = \text{id}_d: \tilde{F}_d \rightarrow L_\infty([0, 1]^d), \quad f \mapsto \text{id}_d(f) = f, \quad d \in \mathbb{N}, \quad (4.2)$$

defined on the unit ball $\tilde{F}_d = \mathcal{B}(F_d)$ of F_d . The authors studied this problem in the worst case setting using algorithms from the classes $\mathcal{A}_d^{n, \text{cont}}$ and $\mathcal{A}_d^{n, \text{adapt}}$ as defined in Section 1.3.

The *initial error* of this problem is given by $\varepsilon_d^{\text{init}} = e^{\text{wor}}(0, d; \text{id}_d) = 1$, the norm of the embedding $F_d \hookrightarrow L_\infty$, since $A_{0,d} \equiv 0$ is a valid choice of an algorithm which does not use any information of f ; see Proposition 2.4. This means that the problem is well-scaled such that there is no difference in studying the absolute or the normalized error criterion.

Now [NW09, Theorem 1] yields that the n th minimal worst case error of L_∞ -approximation defined on F_d satisfies

$$e^{\text{wor}}(n, d; \text{id}_d) = 1 \quad \text{for all } n = 0, 1, \dots, 2^{\lfloor d/2 \rfloor} - 1. \quad (4.3)$$

Therefore, for all $d \in \mathbb{N}$ and every $\varepsilon \in (0, 1)$, the information complexity is bounded from below by

$$n^{\text{wor}}(\varepsilon, d; \text{id}_d) \geq 2^{\lfloor d/2 \rfloor}.$$

Hence the problem suffers from the curse of dimensionality; in particular it is intractable. One possibility to avoid this exponential dependence on d , i.e. to break the curse, is to shrink the function space F_d by introducing *weights*.

In the present chapter we follow this idea. We show that turning to spaces equipped with *product weights* can dramatically improve the tractability behavior of certain problems such as uniform approximation. In [Section 4.1](#) we formally introduce the concept of weighted spaces by considering the examples of weighted Banach spaces of smooth functions and of weighted reproducing kernel Hilbert spaces. Uniform approximation in the latter class of spaces then is studied in [Section 4.2](#). Afterwards, in [Section 4.3](#), we show how to use the obtained upper error bounds for the L_∞ -approximation problem defined on scales of smooth functions. Moreover we prove corresponding lower bounds on the information complexity which enable us to give necessary and sufficient conditions for several kinds of tractabilities in terms of the used weights. Most of the results stated in this chapter are published in the article [\[Wei12b\]](#).

4.1 The concept of weighted spaces

The idea to introduce weights directly into the norm of the function space appeared for the first time in a paper of Sloan and Woźniakowski in 1998; see [\[SW98\]](#). They studied the integration problem defined over some Sobolev Hilbert space, equipped with so-called *product weights*, to explain the overwhelming success of QMC integration rules. Thenceforth weighted problems attracted a lot of attention.

For example it turned out that tractability of approximation of linear compact operators between Hilbert spaces can be fully characterized in terms of the weights and the singular values of the operators if we use information operations from the class Λ^{all} . The proof of this kind of assertions is once again based on the singular value decomposition; see [Section 2.3](#). One such result is given in [Section 4.4](#) below.

But first let us illustrate the concept of weighted spaces by modifying the space F_d we introduced before.

4.1.1 Weighted Banach spaces of smooth functions

A closer look at the norm given in [\(4.1\)](#) yields that for $f \in \mathcal{B}(F_d)$ we have

$$\|D^\alpha f | L_\infty([0, 1]^d)\| \leq 1 \quad \text{for all } \alpha \in \mathbb{N}_0^d. \quad (4.4)$$

Hence every derivative is equally important. In order to shrink the space, for each $\alpha \in \mathbb{N}_0^d$ we replace the right-hand side of inequality [\(4.4\)](#) by a non-negative weight γ_α . For α with $|\alpha| = 1$ this means that we control the importance of every single variable. So, the norm in the weighted space F_d^γ is now given by

$$\|f | F_d^\gamma\| = \sup_{\alpha \in \mathbb{N}_0^d} \frac{1}{\gamma_\alpha} \|D^\alpha f | L_\infty([0, 1]^d)\|, \quad (4.5)$$

where we demand $D^\alpha f$ to be equal to zero if $\gamma_\alpha = 0$. It is clear from the construction that we indeed shrink the space if all γ_α are chosen strictly less than one.

Since this approach is quite general we restrict ourselves to so-called *product weights* (with uniformly bounded generators) in what follows. Thus we assume that for every $d \in \mathbb{N}$ there exists an ordered and uniformly bounded sequence

$$C_\gamma \geq \gamma_{d,1} \geq \gamma_{d,2} \geq \dots \geq \gamma_{d,d} \geq 0.$$

Then for $d \in \mathbb{N}$ the product weight sequence $\gamma = (\gamma_\alpha)_{\alpha \in \mathbb{N}_0^d}$ is given by

$$\gamma_\alpha = \prod_{j=1}^d (\gamma_{d,j})^{\alpha_j}, \quad \alpha \in \mathbb{N}_0^d. \quad (4.6)$$

Note that the dependence of x_j on f is now controlled by the so-called *generator weight* $\gamma_{d,j}$. Since $\gamma_{d,j} = 0$ for some $j \in \{1, \dots, d\}$ implies that f does not depend on x_j, \dots, x_d we assume that $\gamma_{d,d} > 0$ in the rest of this chapter. Moreover observe that the ordering of $\gamma_{d,j}$ is without loss of generality. Later on we will see that tractability of our problem will only depend on summability properties of the generator weights.

Among other things, we show in [Section 4.3.3](#) that for the L_∞ -approximation problem defined on the Banach spaces F_d^γ with the norm given above and generator weights $\gamma_{d,j} \equiv \gamma^{(j)} \in \Theta(j^{-\beta})$ we have

- intractability for $\beta = 0$,
- weak tractability but no polynomial tractability for $0 < \beta < 1$,
- strong polynomial tractability if $1 < \beta$.

Furthermore, we prove that for $\beta = 1$ the problem is not strongly polynomially tractable.

4.1.2 Weighted Hilbert spaces and weighted RKHS

Let us briefly discuss the idea of weighted norms in the case of Hilbert (function) spaces, before we turn to weighted RKHSs. Our approach is based on a generalization of the so-called ANOVA²² decomposition of d -variate functions f , where d is an arbitrary large integer. For the ease of presentation we follow the lines of [\[NW08, Section 5.3.1\]](#). Thus, we focus our attention on Hilbert function spaces constructed out of tensor products and equipped with some assumptions that can be significantly relaxed. For further information on more general settings the interested reader is referred to [\[KSWW10b\]](#) and the references therein.

Given a d -fold tensor product space $H_d = H_1 \otimes \dots \otimes H_1$, $d \in \mathbb{N}$, as well as an orthonormal basis $\{e_i \mid i \in \mathbb{N}\}$ of the underlying univariate Hilbert space²³ H_1 that

²²analysis of variance.

²³We assume H_1 to be separable and infinite-dimensional to keep the notation as short as possible.

contains the constant function $e_1 \equiv 1$, it is easy to see that every $f \in H_d$ can be represented as

$$f = \sum_{\mathbf{u} \subseteq \{1, \dots, d\}} f_{\mathbf{u}}.$$

In this decomposition the (formally d -variate) functions $f_{\mathbf{u}}$ solely depend on the variables x_j with index $j \in \mathbf{u}$. The main advantage of this kind of representation is that for fixed f the collection of all $f_{\mathbf{u}}$, $\mathbf{u} \subseteq \{1, \dots, d\}$, can be taken mutually orthogonal w.r.t. the inner product $\langle \cdot, \cdot \rangle_{H_d}$ in H_d . Therefore the norm of $f \in H_d$ can be expressed by

$$\|f\|_{H_d}^2 = \sum_{\mathbf{u} \subseteq \{1, \dots, d\}} \|f_{\mathbf{u}}\|_{H_d}^2 = \sum_{\mathbf{u} \subseteq \{1, \dots, d\}} \|f_{\mathbf{u},1}\|_{H_{|\mathbf{u}|}}^2,$$

where $f_{\mathbf{u},1}$ equals $f_{\mathbf{u}}$ interpreted as an element of the $|\mathbf{u}|$ -fold tensor product space $H_{|\mathbf{u}|}$ of the closed subspace

$$H_1' = \{h \in H_1 \mid \langle h, e_1 \rangle_{H_1} = 0\} \subset H_1$$

with itself. That is, in the unweighted situation the contribution of each $f_{\mathbf{u}}$ to the norm of $f \in H_d$ is the same.

Now suppose that we have some additional, a priori knowledge about the importance of some (groups of) variables in dimension d . This can be modeled by assigning positive²⁴ weights $\gamma_{d,\mathbf{u}}$ to each of the 2^d subsets \mathbf{u} of $\{1, \dots, d\}$. We denote the collection of these weights by $\gamma_{(d)} = \{\gamma_{d,\mathbf{u}} \mid \mathbf{u} \subseteq \{1, \dots, d\}\}$. Then it can be verified that

$$\langle f, g \rangle_{\gamma_{(d)}} = \sum_{\mathbf{u} \subseteq \{1, \dots, d\}} \frac{1}{\gamma_{d,\mathbf{u}}} \langle f_{\mathbf{u}}, g_{\mathbf{u}} \rangle_{H_d} = \sum_{\mathbf{u} \subseteq \{1, \dots, d\}} \frac{1}{\gamma_{d,\mathbf{u}}} \langle f_{\mathbf{u},1}, g_{\mathbf{u},1} \rangle_{H_{|\mathbf{u}|}} \quad (4.7)$$

defines an inner product on the tensor product space H_d which implies an equivalent norm depending on $\gamma_{(d)}$. The Hilbert space H_d endowed with this new inner product will be denoted by $H_d^{\gamma_{(d)}}$. At this point we need to stress the fact that for general weights $\gamma_{(d)}$ these spaces are no longer tensor product spaces, although their construction is based on $H_d = H_1 \otimes \dots \otimes H_1$ and $H_{|\mathbf{u}|}$, respectively. To overcome this problem we restrict ourselves to the case of *product weights* in the following. Thus we assume

$$\gamma_{d,\mathbf{u}} = \prod_{k \in \mathbf{u}} \gamma_{d,k} \quad (4.8)$$

for some positive $\gamma_{d,k}$, $k = 1, \dots, d$, and every $\mathbf{u} \subseteq \{1, \dots, d\}$. Then it can be checked that indeed $H_d^{\gamma_{(d)}}$ is again a tensor product space. For the study of other

²⁴Also zero weights are possible but for reasons of simplification we do not discuss this more complicated situation in the present brief introduction to weighted Hilbert spaces.

types of weights such as *finite-order*, *finite-diameter*, *order-dependent* or the recently developed *POD²⁵ weights* we refer to Novak and Woźniakowski [NW08, Section 5.3.2] and to Kuo, Schwab and Sloan [KSS11].

In the last decade it turned out that weighted norms provide a powerful tool to vanquish the curse of dimensionality that we are often faced with. Since the $H_d^{\gamma(d)}$'s are still Hilbert spaces the complexity analysis of weighted problems $S^{\gamma(d)} = (S_d^{\gamma(d)} : \mathcal{B}(H_d^{\gamma(d)}) \rightarrow \mathcal{G}_d)_{d \in \mathbb{N}}$ again is based on the singular value decomposition presented in Section 2.3.1; at least in the cases where the target spaces \mathcal{G}_d are also Hilbert spaces. Fortunately, the introduced weights enter the spectrum of the operator $W_d^{\gamma(d)} = (S_d^{\gamma(d)})^\dagger S_d^{\gamma(d)}$ in a straightforward way. Therefore in many cases tractability properties of S can be fully characterized in terms of the singular values and the introduced weights.

For our purposes weighted Hilbert spaces that possess a reproducing kernel are of particular interest. Typical examples of such weighted RKHSs are the following unanchored Sobolev spaces endowed with product weights which will play an important role in our further argumentation; see also Sloan and Woźniakowski [SW02]. Instead of applying the presented approach which is based on decompositions we use the common procedure and define them directly.

Example 4.1 (Unanchored Sobolev spaces \mathcal{H}_d^γ). As usual we start with the definition for $d = 1$ and $\gamma > 0$. Then the space \mathcal{H}_1^γ is nothing but the Sobolev space of all absolutely continuous real-valued functions f defined on the unit interval $[0, 1]$ whose first derivative²⁶ f' belongs to the space $L_2([0, 1])$. The difference to the classical Sobolev space is the inner product which here depends on the parameter γ :

$$\begin{aligned} \langle f, g \rangle_{\mathcal{H}_1^\gamma} &= \langle f, g \rangle_{L_2([0,1])} + \gamma^{-1} \langle f', g' \rangle_{L_2([0,1])} \\ &= \int_0^1 f(x) g(x) \, d\lambda^1(x) + \gamma^{-1} \int_0^1 f'(x) g'(x) \, d\lambda^1(x), \quad f, g \in \mathcal{H}_1^\gamma. \end{aligned} \quad (4.9)$$

For the sake of completeness we define the space \mathcal{H}_1^0 as the limit of \mathcal{H}_1^γ for $\gamma \rightarrow 0$. Consequently the derivatives of $f \in \mathcal{H}_1^0$ need to vanish λ^1 -almost everywhere on $[0, 1]$ which implies that the space \mathcal{H}_1^0 only consists of constant functions. This coincides with the common convention $0/0 = 0$.

Note that the univariate space \mathcal{H}_1^γ algebraically coincides with its anchored analogue $\tilde{\mathcal{H}}_1^\gamma$ where the term $\langle f, g \rangle_{L_2([0,1])}$ in (4.9) is replaced by $f(a) \cdot g(a)$ for some *anchor point* $a \in [0, 1]$. For details we refer to [SW02] and [Wei12b]. Finally we mention that for positive parameters γ all these definitions imply equivalent norms on the classical Sobolev space $W_2^1([0, 1])$.

Once more the d -variate spaces \mathcal{H}_d^γ for $d > 1$ are defined by a tensor product construction similar to Section 2.4.1. We set $\mathcal{H}_d^\gamma = \bigotimes_{k=1}^d \mathcal{H}_1^{\gamma_{d,k}}$, where now γ

²⁵product and order-dependent.

²⁶in the weak or distributional sense

denotes a (subset of a) product weight sequence $(\gamma_\alpha)_{\alpha \in \{0,1\}^d}$ induced by some generator weights $\gamma_{d,k}$, $k = 1, \dots, d$; see (4.6). Remember that at the beginning of this chapter we assumed $\gamma_{d,d} > 0$ for all $d \in \mathbb{N}$. That is, we avoid to take the trivial spaces \mathcal{H}_1^0 as factors in the definition of \mathcal{H}_d^γ .

How does the inner product of \mathcal{H}_d^γ looks like? Following the lines of Section 2.4.1 it is uniquely determined by the coordinate-wise inner products of the factors of simple tensors $f = \bigotimes_{k=1}^d f_k$ and $g = \bigotimes_{k=1}^d g_k$, where $f_k, g_k \in \mathcal{H}_1^{\gamma_{d,k}}$ for $k = 1, \dots, d$. Consequently,

$$\begin{aligned} \langle f, g \rangle_{\mathcal{H}_d^\gamma} &= \prod_{k=1}^d \langle f_k, g_k \rangle_{\mathcal{H}_1^{\gamma_{d,k}}} = \prod_{k=1}^d \left(\langle f_k, g_k \rangle_{L_2([0,1])} + \frac{1}{\gamma_{d,k}} \langle f'_k, g'_k \rangle_{L_2([0,1])} \right) \\ &= \sum_{\mathbf{u} \subseteq \{1, \dots, d\}} \prod_{k \in \mathbf{u}} \frac{1}{\gamma_{d,k}} \cdot \prod_{k \in \mathbf{u}} \langle f'_k, g'_k \rangle_{L_2([0,1])} \cdot \prod_{j \in \{1, \dots, d\} \setminus \mathbf{u}} \langle f_j, g_j \rangle_{L_2([0,1])} \\ &= \sum_{\mathbf{u} \subseteq \{1, \dots, d\}} \prod_{k \in \mathbf{u}} \frac{1}{\gamma_{d,k}} \cdot \int_{[0,1]^d} \prod_{k \in \mathbf{u}} f'_k(x_k) g'_k(x_k) \prod_{j \in \{1, \dots, d\} \setminus \mathbf{u}} f_j(x_j) g_j(x_j) d\lambda^d(\mathbf{x}) \\ &= \sum_{\mathbf{u} \subseteq \{1, \dots, d\}} \frac{1}{\gamma_{d,\mathbf{u}}} \cdot \int_{[0,1]^d} \frac{\partial^{|\mathbf{u}|} f}{\partial x_{\mathbf{u}}}(\mathbf{x}) \frac{\partial^{|\mathbf{u}|} g}{\partial x_{\mathbf{u}}}(\mathbf{x}) d\lambda^d(\mathbf{x}), \end{aligned}$$

where we used (4.8) and the shorthand notation $\partial^{|\mathbf{u}|}/(\partial x_{\mathbf{u}})$ for $\prod_{k \in \mathbf{u}} \partial/(\partial x_k)$. Note that this representation resembles (4.7) from the general approach to weighted Hilbert spaces introduced at the beginning of this subsection. For our purposes it is more convenient to rewrite the subsets $\mathbf{u} \subseteq \{1, \dots, d\}$ in terms of multi-indices $\alpha = (\alpha_1, \dots, \alpha_d) \in \{0, 1\}^d$. In detail, we set $\alpha_k = 1$ if and only if $k \in \mathbf{u}$ and $\alpha_k = 0$ otherwise. Then we can express the norm of any $f \in \mathcal{H}_d^\gamma$ by

$$\|f\|_{\mathcal{H}_d^\gamma}^2 = \sum_{\alpha \in \{0,1\}^d} \frac{1}{\gamma_\alpha} \cdot \int_{[0,1]^d} |D^\alpha f(\mathbf{x})|^2 d\lambda^d(\mathbf{x}) \quad (4.10)$$

since then $\gamma_{\mathbf{u}} = \gamma_\alpha$. The inner products of the multivariate anchored spaces, $\tilde{\mathcal{H}}_d^\gamma$, can be found by a similar reasoning; see [Wei12b, p. 67] for the final result.

It is known (cf. Micchelli and Wahba [MW81]) that the univariate spaces \mathcal{H}_1^γ are reproducing kernel Hilbert spaces for any $\gamma > 0$. Consequently, this property is transferred to the multivariate tensor product space. To stress this fact we write $\mathcal{H}(K_d^\gamma)$ for \mathcal{H}_d^γ in what follows. Equation (5) in [WW09] now states that the reproducing kernel $K_d^\gamma: [0, 1]^d \times [0, 1]^d \rightarrow \mathbb{R}$ in dimension $d \geq 1$ is given by²⁷

$$\begin{aligned} K_d^\gamma(\mathbf{x}, \mathbf{y}) &= \prod_{k=1}^d \frac{\sqrt{\gamma_{d,k}}}{\sinh(\sqrt{\gamma_{d,k}})} \cosh(\sqrt{\gamma_{d,k}}(1 - \max\{x_k, y_k\})) \cosh(\sqrt{\gamma_{d,k}} \min\{x_k, y_k\}), \end{aligned}$$

²⁷Here \sinh and \cosh denote the hyperbolic sine and cosine functions, respectively.

$\mathbf{x}, \mathbf{y} \in [0, 1]^d$. For $d = 1$ this kernel formula follows from Thomas-Agnan [Tho96, Corollary 2] whereas the higher-dimensional generalization for product weights γ results from the tensor product structure; see (2.20) in Section 2.5. In particular we note that K_d^γ is continuous (and thus also bounded) along its diagonal

$$\{(\mathbf{x}, \mathbf{y}) \in [0, 1]^{2d} \mid \mathbf{x} = \mathbf{y}\}.$$

Moreover, from [WW09, Lemma 4.1] we know that for $\gamma > 0$ the set

$$E_1(\gamma) = \{e_{1,\gamma,i}: [0, 1] \rightarrow \mathbb{R} \mid i \in \mathbb{N}\}$$

with $e_{1,\gamma,1} \equiv 1$ and

$$e_{1,\gamma,i}(x) = \cos(\pi(i-1)x) \cdot \sqrt{\frac{2\gamma}{\gamma + \pi^2(i-1)^2}}, \quad x \in [0, 1], \quad i \geq 2,$$

builds an orthonormal basis in the univariate space $\mathcal{H}(K_1^\gamma)$. Applying the arguments from Section 2.4.1 this leads to an ONB $E_d(\gamma)$ of $\mathcal{H}(K_d^\gamma) = \bigotimes_{k=1}^d \mathcal{H}(K_1^{\gamma_{d,k}})$ that consists of tensor product functions

$$\tilde{e}_{d,\gamma,\mathbf{m}} = \bigotimes_{k=1}^d e_{1,\gamma_{d,k},m_k}, \quad \mathbf{m} = (m_1, \dots, m_d) \in \mathbb{N}^d. \quad (4.11)$$

For a direct proof of this result we refer to [NW08, Appendix A.2.1]²⁸ and to [WW09, Lemma 4.2]. Actually, these proofs show a little bit more; namely that the functions $\tilde{e}_{d,\gamma,\mathbf{m}}$ together with

$$\tilde{\lambda}_{d,\gamma,\mathbf{m}} = \prod_{k=1}^d \lambda_{1,\gamma_{d,k},m_k} = \prod_{k=1}^d \frac{\gamma_{d,k}}{\gamma_{d,k} + \pi^2(m_k - 1)^2}, \quad \mathbf{m} \in \mathbb{N}^d, \quad (4.12)$$

describe the full set of eigenpairs $\{(\tilde{\lambda}_{d,\gamma,\mathbf{m}}, \tilde{e}_{d,\gamma,\mathbf{m}}) \mid \mathbf{m} \in \mathbb{N}^d\}$ of the operator $W_d^\gamma = (S_d^\gamma)^\dagger S_d^\gamma$ where $S_d^\gamma: \mathcal{H}_d^\gamma \hookrightarrow L_2([0, 1]^d)$ denotes the solution operator of the L_2 -approximation problem on $\mathcal{H}_d^\gamma = \mathcal{H}(K_d^\gamma)$. \square

4.2 Uniform approximation in reproducing kernel Hilbert spaces

The main result of this section is based on a paper of Kuo, Wasilkowski and Woźniakowski [KWW08]. In contrast to the presentation given in [Wei12b] we decided to apply this result to the case of the unanchored Sobolev Space introduced in Section 4.1.2 instead of the anchored analogue studied in [KWW08]. This opens

²⁸Note the missing factor 1/2 in [NW08, p. 351, line 5].

up the opportunity to explain the underlying ideas without literally repeating the proof given in [KWW08] while obtaining a result which is (according to our knowledge) not published elsewhere so far.

We start with an upper error bound which remains valid for any reproducing kernel Hilbert space $\mathcal{H}(K_d)$ of real-valued functions f on $[0, 1]^d$ with

$$\operatorname{ess-sup}_{\mathbf{x} \in [0, 1]^d} K_d(\mathbf{x}, \mathbf{x}) < \infty. \quad (4.13)$$

This condition guarantees that $\mathcal{H}(K_d)$ is continuously embedded into $L_\infty([0, 1]^d)$ since the reproducing property (2.17), together with the Hahn-Banach theorem (cf. [Yos80, IV.6 Cor.2]), yields that $\|\operatorname{id}_d\|_{\mathcal{L}(\mathcal{H}(K_d), L_\infty([0, 1]^d))}$ is given by

$$\begin{aligned} \sup_{f \in \mathcal{B}(\mathcal{H}(K_d))} \|f\|_{L_\infty([0, 1]^d)} &= \operatorname{ess-sup}_{\mathbf{x} \in [0, 1]^d} \sup_{f \in \mathcal{B}(\mathcal{H}(K_d))} |f(\mathbf{x})| \\ &= \operatorname{ess-sup}_{\mathbf{x} \in [0, 1]^d} \sup_{f \in \mathcal{B}(\mathcal{H}(K_d))} \left| \langle f, K_d(\cdot, \mathbf{x}) \rangle_{\mathcal{H}(K_d)} \right| \\ &= \operatorname{ess-sup}_{\mathbf{x} \in [0, 1]^d} K_d(\mathbf{x}, \mathbf{x})^{1/2}. \end{aligned}$$

Now the mentioned upper bound reads as follows:

Proposition 4.2. For $d \in \mathbb{N}$ consider a RKHS $\mathcal{H}(K_d)$, where K_d fulfills (4.13), i.e. $\mathcal{H}(K_d) \hookrightarrow L_\infty([0, 1]^d)$. Furthermore, suppose $\Xi = \{\xi_j : [0, 1]^d \rightarrow \mathbb{R} \mid j \in \mathbb{N}\}$ to be some orthonormal basis of $\mathcal{H}(K_d)$ and let $n \in \mathbb{N}_0$. Then the algorithm $A_{n,d}^\Xi \in \mathcal{A}_d^{n, \operatorname{lin}}(\Lambda^{\operatorname{all}})$, given by

$$f \mapsto A_{n,d}^\Xi f = \sum_{j=1}^n \langle f, \xi_j \rangle_{\mathcal{H}(K_d)} \xi_j(\cdot),$$

for uniform approximation on $\mathcal{H}(K_d)$ fulfills

$$\Delta^{\operatorname{wor}}(A_{n,d}^\Xi; \operatorname{id}_d : \mathcal{B}(\mathcal{H}(K_d)) \rightarrow L_\infty([0, 1]^d)) \leq \left\| \sum_{j=n+1}^{\infty} \xi_j(\cdot)^2 \right\|_{L_\infty([0, 1]^d)}^{1/2}. \quad (4.14)$$

Proof. Since Ξ builds an ONB we may represent any $f \in \mathcal{H}(K_d)$ by its basis expansion, $f = \sum_{j=1}^{\infty} \langle f, \xi_j \rangle_{\mathcal{H}(K_d)} \xi_j$. Therefore Parseval's identity implies

$$\begin{aligned} |f(\mathbf{x}) - A_{n,d}^\Xi f(\mathbf{x})| &= |(f - A_{n,d}^\Xi f)(\mathbf{x})| = \left| \sum_{j=n+1}^{\infty} \langle f, \xi_j \rangle_{\mathcal{H}(K_d)} \xi_j(\mathbf{x}) \right| \\ &= \left| \left\langle f, \sum_{j=n+1}^{\infty} \xi_j(\mathbf{x}) \xi_j \right\rangle_{\mathcal{H}(K_d)} \right| \end{aligned}$$

which can be estimated from above using the inequality of Cauchy and Schwarz. Thus we obtain

$$\begin{aligned} |f(\mathbf{x}) - A_{n,d}^{\Xi} f(\mathbf{x})| &\leq \|f\|_{\mathcal{H}(K_d)} \cdot \left\| \sum_{j=n+1}^{\infty} \xi_j(\mathbf{x}) \xi_j \Big|_{\mathcal{H}(K_d)} \right\| \\ &= \|f\|_{\mathcal{H}(K_d)} \cdot \left(\sum_{j=n+1}^{\infty} \xi_j(\mathbf{x})^2 \right)^{1/2} \end{aligned} \quad (4.15)$$

for every $f \in \mathcal{H}(K_d)$ and all fixed $\mathbf{x} \in [0, 1]^d$. Taking the (essential) supremum with respect to \mathbf{x} in the d -dimensional unit cube and the supremum over all $f \in \mathcal{B}(\mathcal{H}(K_d))$ gives the desired result. \blacksquare

We note in passing that we can easily prove more than we stated in the latter assertion. In what follows we only need the given upper error bound such that we restrict ourselves to some brief comments on further results in the next remark.

Remark 4.3. For fixed $\mathbf{x} \in [0, 1]^d$ we see that the function $f^* = C \cdot \sum_{j=n+1}^{\infty} \xi_j(\mathbf{x}) \xi_j$ with $C > 0$ gives equality in (4.15). Of course, we can choose the constant C such that $\|f^*\|_{\mathcal{H}(K_d)} = 1$ provided that \mathbf{x} is not a common root of ξ_j for all $j > n$. Hence, the upper bound in (4.14) is sharp.

Moreover, [KWW08, Theorem 2] shows that the n th minimal worst case error for L_{∞} -approximation on $\mathcal{H}(K_d)$ is given by

$$e^{\text{wor}}(n, d; \text{id}_d: \mathcal{B}(\mathcal{H}(K_d)) \rightarrow L_{\infty}([0, 1]^d)) = \inf_{\Xi = \{\xi_j\}_{j \in \mathbb{N}}} \left\| \sum_{j=n+1}^{\infty} \xi_j(\cdot)^2 \Big|_{L_{\infty}([0, 1]^d)} \right\|^{1/2},$$

where the infimum is taken w.r.t. all orthonormal bases $\Xi \subset \mathcal{H}(K_d)$. Thus, any clever choice of the basis Ξ in Proposition 4.2 leads to algorithms $A_{n,d}^{\Xi}$ with almost optimal worst case errors. \square

Next we apply Proposition 4.2 to the weighted unanchored Sobolev spaces \mathcal{H}_d^{γ} introduced in Section 4.1.2 using the basis $\Xi = E_d(\gamma)$ given in (4.11). Since the ordering of the basis functions $\xi \in \Xi$ is essential for our application we rearrange them non-increasingly with respect to their L_{∞} -norm:

$$\|\xi_j\|_{L_{\infty}([0, 1]^d)} \geq \|\xi_{j+1}\|_{L_{\infty}([0, 1]^d)} \quad \text{for all } j \in \mathbb{N}. \quad (4.16)$$

We obtain an estimate which resembles the corresponding result for the anchored case studied in [Wei12b, Proposition 2].

Corollary 4.4. For $n \in \mathbb{N}_0$ and $d \in \mathbb{N}$ there exists an algorithm $A_{n,d}^* \in \mathcal{A}_d^{n, \text{lin}}(\Lambda^{\text{all}})$ for uniform approximation on \mathcal{H}_d^{γ} such that for every $\tau \in (1/2, 1)$

$$\Delta^{\text{wor}}(A_{n,d}^*; \text{id}_d: \mathcal{B}(\mathcal{H}_d^{\gamma}) \rightarrow L_{\infty}([0, 1]^d)) < a_{\tau} \exp \left(b_{\tau} \sum_{k=1}^d (\gamma_{d,k})^{\tau} \right) \cdot n^{-(1-\tau)/(2\tau)},$$

where the constants $a_{\tau}, b_{\tau} > 0$ are independent of γ, n , and d .

Proof. To keep the notation as short as possible we abbreviate the L_∞ -norm in d dimensions, $\|\cdot\|_{L_\infty([0,1]^d)}$, by $\|\cdot\|_d$ within this proof.

Following our plan we fix $n \in \mathbb{N}_0$, as well as $d \in \mathbb{N}$, and take $A_{n,d}^* = A_{n,d}^\Xi$ defined in Proposition 4.2 with $\Xi = E_d(\gamma)$ as above. From (4.11) we conclude for $d = 1$ and any $\gamma > 0$ that

$$\|e_{1,\gamma,1}^2\|_1 = 1 \quad \text{and} \quad \|e_{1,\gamma,i}^2\|_1 = \frac{2\gamma}{\gamma + \pi^2(i-1)^2} < \frac{2\gamma}{\pi^2} \cdot (i-1)^{-2}, \quad i \geq 2.$$

Moreover, for every simple tensor $f = \bigotimes_{k=1}^d f_k \in \mathcal{H}(K_d^\gamma)$ we clearly have

$$\|f\|_d = \prod_{k=1}^d \|f_k\|_1 \quad \text{and} \quad f(\mathbf{x})^2 = \prod_{k=1}^d f_k(x_k)^2, \quad \mathbf{x} \in [0,1]^d.$$

Consequently, for any $j \in \mathbb{N}$ and all $\tau \in (1/2, \infty)$ the ordering of Ξ given in (4.16) implies

$$\begin{aligned} j \cdot \|\xi_j^2\|_d^\tau &\leq \sum_{m=1}^{\infty} \|\xi_m^2\|_d^\tau = \sum_{\mathbf{m} \in \mathbb{N}^d} \|\tilde{e}_{d,\gamma,\mathbf{m}}^2\|_d^\tau = \prod_{k=1}^d \sum_{i=1}^{\infty} \|e_{1,\gamma_{d,k},i}^2\|_1^\tau \\ &= \prod_{k=1}^d \left(1 + \sum_{i=2}^{\infty} \|e_{1,\gamma_{d,k},i}^2\|_1^\tau \right) < \prod_{k=1}^d \left(1 + \left(\frac{2\gamma_{d,k}}{\pi^2} \right)^\tau \sum_{i=2}^{\infty} (i-1)^{-2\tau} \right) \\ &= \prod_{k=1}^d (1 + c_\tau \gamma_{d,k}^\tau), \end{aligned}$$

where we set $c_\tau = (2/\pi^2)^\tau \zeta(2\tau)$. Hence, if $\tau \in (1/2, 1)$ then

$$\left\| \sum_{j=n+1}^{\infty} \xi_j^2 \right\|_d \leq \sum_{j=n+1}^{\infty} \|\xi_j^2\|_d < \sum_{j=n+1}^{\infty} j^{-1/\tau} \cdot \left(\prod_{k=1}^d (1 + c_\tau \gamma_{d,k}^\tau) \right)^{1/\tau} < \infty.$$

Since the first factor is no larger than $\int_n^\infty x^{-1/\tau} d\lambda^1(x) = \tau/(1-\tau) \cdot n^{-(1-\tau)/\tau}$ and the second factor can be bounded by $\exp\left(c_\tau/\tau \cdot \sum_{k=1}^d (\gamma_{d,k})^\tau\right)$ we conclude

$$\left\| \sum_{j=n+1}^{\infty} \xi_j(\cdot)^2 \right\|_{L_\infty([0,1]^d)}^{1/2} < a_\tau \exp\left(b_\tau \sum_{k=1}^d (\gamma_{d,k})^\tau\right) \cdot n^{-(1-\tau)/(2\tau)}$$

with $a_\tau = \sqrt{\tau/(1-\tau)}$ and $b_\tau = c_\tau/(2\tau) = (2/\pi^2)^\tau \zeta(2\tau)/(2\tau)$. Now the claim follows from (4.14) in Proposition 4.2. \blacksquare

4.3 Uniform approximation in Banach spaces of smooth functions

Our derivation of necessary and sufficient conditions for various kinds of tractability for the L_∞ -approximation problem defined on the weighted spaces F_d^γ introduced in Section 4.1.1 is based on simple embedding arguments. To this end, we consider a whole scale of Banach spaces \mathcal{F}_d^γ (where F_d^γ is a special case of). Then we first study lower bounds on the n th minimal error on a space $\mathcal{P}_d^\gamma \hookrightarrow \mathcal{F}_d^\gamma$ which consists of d -variate polynomials of low degree. Afterwards, in Section 4.3.2, we use the results for $\mathcal{H}_d^\gamma \hookrightarrow \mathcal{F}_d^\gamma$ from Section 4.2 to conclude corresponding upper bounds. Finally we discuss a couple of concrete examples in Section 4.3.3.

4.3.1 Lower bounds for spaces of low-degree polynomials

Following the lines of [Wei12b, Section 4] we use Proposition 2.2 to obtain a lower bound for the L_∞ -approximation error for the space

$$\mathcal{P}_d^\gamma = \text{span} \left\{ p_{\mathbf{i}} : [0, 1]^d \rightarrow \mathbb{R}, p_{\mathbf{i}}(\mathbf{x}) = \mathbf{x}^{\mathbf{i}} = \prod_{j=1}^d (x_j)^{i_j} \mid \mathbf{i} = (i_1, \dots, i_d) \in \{0, 1\}^d \right\}$$

of all real-valued d -variate polynomials of degree at most one in each coordinate direction, defined on the unit cube $[0, 1]^d$. We equip this linear space with the weighted norm

$$\|f\|_{\mathcal{P}_d^\gamma} = \max_{\alpha \in \{0, 1\}^d} \frac{1}{\gamma_\alpha} \|D^\alpha f\|_{L_\infty([0, 1]^d)}, \quad f \in \mathcal{P}_d^\gamma, \quad (4.17)$$

similar to (4.5), where γ is a product weight sequence as described in (4.6), and study the worst case setting.

Theorem 4.5. For $d \in \mathbb{N}$ and $n \in \mathbb{N}_0$ assume $A_{n,d} \in \mathcal{A}_d^{n,\text{cont}} \cup \mathcal{A}_d^{n,\text{adapt}}$ to be an arbitrary algorithm for the uniform approximation problem defined on \mathcal{P}_d^γ . Then we have

$$\Delta^{\text{wor}}(A_{n,d}; \text{id}_d : B_r(\mathcal{P}_d^\gamma) \rightarrow L_\infty([0, 1]^d)) \geq r \quad \text{for all } r \geq 0$$

provided that $n < 2^s$, where $s = s(\gamma, d) \in \{0, 1, \dots, d\}$ is some integer such that

$$s > \frac{1}{2 + C_\gamma} \cdot \left(\sum_{j=1}^d \gamma_{d,j} - 2 \right). \quad (4.18)$$

Proof. The proof of this lower error bound consists of several steps. First we fix $d \in \mathbb{N}$ and construct a partition of the set of coordinates $\{1, \dots, d\}$ into $s + 1$ parts which we will need later and with $s = s(\gamma, d)$ satisfying (4.18). In a second step we

define a special linear subspace $V \subseteq \mathcal{P}_d^\gamma$ with $\dim V = 2^s$. Step 3 then shows that V satisfies the assumptions of [Proposition 2.2](#). The proof is completed in Step 4.

Step 1. For $k \in \{0, \dots, d\}$ let us define inductively $m_0 = 0$ and

$$m_k = \inf \left\{ t \in \mathbb{N} \mid m_{k-1} < t \leq d, \text{ with } 2 \leq \sum_{j=m_{k-1}+1}^t \gamma_{d,j} \right\}$$

with the usual convention $\inf \emptyset = \infty$. Note that the infimum coincides with the minimum in the finite case, since then $m_k \in \mathbb{N}$. Moreover we set

$$s = \max \{k \in \{0, \dots, d\} \mid m_k < \infty\}.$$

We denote $I_k = \{m_{k-1} + 1, m_{k-1} + 2, \dots, m_k\}$ for $k = 1, \dots, s$. Thus, this gives a uniquely defined disjoint partition of the set

$$\{1, \dots, d\} = \left(\bigcup_{k=1}^s I_k \right) \cup \{m_s + 1, \dots, d\},$$

and m_k denotes the last element of the block I_k . For all $k = 1, \dots, s$ we conclude

$$2 \leq \sum_{j \in I_k} \gamma_{d,j} < 2 + \gamma_{d,m_k} \leq 2 + C_\gamma,$$

where C_γ is the uniform upper bound for $\gamma_{d,j}$; see [Section 4.1.1](#). Finally, summation of these inequalities gives

$$\sum_{j=1}^d \gamma_{d,j} < \sum_{k=1}^s \sum_{j \in I_k} \gamma_{d,j} + 2 < (2 + C_\gamma)s + 2,$$

and [\(4.18\)](#) follows immediately.

If $s = 0$ then we can stop at this point since the initial error is 1 as the norm of the embedding $\mathcal{P}_d^\gamma \hookrightarrow L_\infty$ (cf. [Proposition 2.4](#)) and the remaining assertion is trivial. Hence, from now on we can assume that $s > 0$ and thus $m_s \geq 1$.

Step 2. To apply [Proposition 2.2](#) we have to construct a linear subspace V of $\mathcal{F} = \mathcal{P}_d^\gamma$ such that the condition [\(2.1\)](#) holds for the target space $\mathcal{G} = L_\infty([0, 1]^d)$, the embedding operator $S = \text{id}_d$, and $a = 1$. Note that we restrict ourselves to the set

$$\widehat{\mathcal{F}} = \{f \in \mathcal{F} \mid f \text{ depends only on } x_1, \dots, x_{m_s}\},$$

since we can interpret $\widehat{\mathcal{F}}$ as the space $\mathcal{P}_{m_s}^\gamma$ by a simple isometric isomorphism.

We are ready to construct a suitable space V using the partition from Step 1. We define V as the span of all functions $g_{\mathbf{i}}: [0, 1]^{m_s} \rightarrow \mathbb{R}$, $\mathbf{i} = (i_1, \dots, i_s) \in \{0, 1\}^s$, of the form

$$g_{\mathbf{i}}(\mathbf{x}) = \prod_{k=1}^s \left(\sum_{j \in I_k} \gamma_{d,j} \cdot x_j \right)^{i_k}, \quad \mathbf{x} \in X = [0, 1]^{m_s}.$$

Clearly, V is a linear subspace of $\mathcal{P}_{m_s}^\gamma$ and with the interpretation above it is also a linear subspace of \mathcal{F} . Moreover it is easy to see that we have by construction

$$\|g|_{\mathcal{F}}\| = \|g|_{\mathcal{P}_{m_s}^\gamma}\| \quad \text{and} \quad \|g|_{L_\infty(X)}\| = \|g|_{L_\infty([0,1]^d)}\| \quad \text{for } g \in V.$$

Finally we note that $\dim V = \#\{0,1\}^s = 2^s$. It remains to show that this subspace is the right choice to prove the claim using [Proposition 2.2](#).

Step 3. The proof of the needed condition (2.1),

$$\|g|_{\mathcal{P}_{m_s}^\gamma}\| \leq \|g|_{L_\infty(X)}\| \quad \text{for all } g \in V,$$

is a little bit technical. Due to the special structure of the functions $g \in V$, the left-hand side reduces to $\max \{\gamma_\alpha^{-1} \|D^\alpha g|_{L_\infty(X)}\| \mid \alpha \in \mathbb{M}\}$, where the maximum is taken over all multi-indices α in the set

$$\mathbb{M} = \left\{ \alpha \in \{0,1\}^{m_s} \mid \sum_{j \in I_k} \alpha_j \leq 1 \text{ for all } k = 1, \dots, s \right\}.$$

This is simply because for $\alpha \notin \mathbb{M}$ we have $D^\alpha g \equiv 0$ and then the inequality is trivial. To simplify the notation let us define

$$T: \{0,1\}^{m_s} \rightarrow \mathbb{N}_0^s, \quad \alpha \mapsto T(\alpha) = \sigma = (\sigma_1, \dots, \sigma_s),$$

where

$$\sigma_k = \sum_{j \in I_k} \alpha_j \quad \text{for } k = 1, \dots, s.$$

Note that $T(\mathbb{M}) = \{0,1\}^s$. Moreover, for every $g = \sum_{i \in \{0,1\}^s} c_i g_i(\cdot) \in V$ we define a function

$$h_g: Z = \prod_{k=1}^s \left[0, \sum_{j \in I_k} \gamma_{d,j} \right] \rightarrow \mathbb{R}, \quad \mathbf{z} \mapsto h_g(\mathbf{z}) = \sum_{i \in \{0,1\}^s} c_i \prod_{k=1}^s z_k^{i_k} = \sum_{i \in \{0,1\}^s} c_i \mathbf{z}^i.$$

Hence, $h_g(\mathbf{z}) = g(\mathbf{x})$ under the transformation $\mathbf{x} \mapsto \mathbf{z}$ such that

$$z_k = \sum_{j \in I_k} \gamma_{d,j} x_j \quad \text{for every } k = 1, \dots, s \quad \text{and every } \mathbf{x} \in X.$$

The span, W , of all functions $h: Z \rightarrow \mathbb{R}$ with this structure is a linear space, too. Furthermore, easy calculus yields that

$$(D_{\mathbf{x}}^\alpha g)(\mathbf{x}) = \left(\prod_{j=1}^{m_s} (\gamma_{d,j})^{\alpha_j} \right) (D_{\mathbf{z}}^{T(\alpha)} h_g)(\mathbf{z}) \quad (4.19)$$

for all $g \in V$, $\alpha \in \mathbb{M}$ and $\mathbf{x} \in X$. Here the \mathbf{x} and \mathbf{z} in $D_{\mathbf{x}}^{\alpha}$ and $D_{\mathbf{z}}^{T(\alpha)}$ indicate differentiation with respect to \mathbf{x} and \mathbf{z} , respectively. Since the mapping $\mathbf{x} \mapsto \mathbf{z}$ is surjective we obtain $\|D^{\alpha}g|L_{\infty}(X)\| = \gamma_{\alpha} \|D^{T(\alpha)}h_g|L_{\infty}(Z)\|$ by the form of γ given by (4.6). Thus,

$$\max_{\alpha \in \mathbb{M}} \frac{1}{\gamma_{\alpha}} \|D^{\alpha}g|L_{\infty}(X)\| = \max_{\sigma \in \{0,1\}^s} \|D^{\sigma}h_g|L_{\infty}(Z)\|.$$

Observe that (4.19) with $\alpha = 0$ particularly yields that $\|g|L_{\infty}(X)\| = \|h_g|L_{\infty}(Z)\|$. Therefore the claim reduces to

$$\max_{\sigma \in \{0,1\}^s} \|D^{\sigma}h_g|L_{\infty}(Z)\| \leq \|h_g|L_{\infty}(Z)\| \quad \text{for every } g \in V.$$

We show this estimate for every $h \in W$, i.e.,

$$\|D^{\sigma}h|L_{\infty}(Z)\| \leq \|h|L_{\infty}(Z)\| \quad \text{for all } \sigma \in \{0,1\}^s. \quad (4.20)$$

We start with the special case of one derivative. That is, we first consider $\sigma = \mathbf{e}_k$ for a certain $k \in \{1, \dots, s\}$. Since h is affine in each coordinate we can represent it as

$$h(\mathbf{z}) = a(\mathbf{z}_{(\mathbf{k})}) \cdot z_k + b(\mathbf{z}_{(\mathbf{k})})$$

with functions a and b which only depend on $\mathbf{z}_{(\mathbf{k})} = (z_1, \dots, z_{k-1}, z_{k+1}, \dots, z_s)$. Hence we have $(D^{\mathbf{e}_k}h)(\mathbf{z}) = a(\mathbf{z}_{(\mathbf{k})})$ and we need to show that

$$|a(\mathbf{z}_{(\mathbf{k})})| \leq \max \left\{ |b(\mathbf{z}_{(\mathbf{k})})|, \left| a(\mathbf{z}_{(\mathbf{k})}) \cdot \sum_{j \in I_k} \gamma_{d,j} + b(\mathbf{z}_{(\mathbf{k})}) \right| \right\}. \quad (4.21)$$

This is obviously true for every $\mathbf{z} \in Z$ with $a(\mathbf{z}_{(\mathbf{k})}) = 0$. For $a(\mathbf{z}_{(\mathbf{k})}) \neq 0$ we can divide by $|a(\mathbf{z}_{(\mathbf{k})})|$ to get

$$1 \leq \max \left\{ |t|, \left| \sum_{j \in I_k} \gamma_{d,j} - t \right| \right\}$$

if we set $t = -b(\mathbf{z}_{(\mathbf{k})})/a(\mathbf{z}_{(\mathbf{k})})$. The last maximum is minimal if both of its entries coincide. This is for $t = \frac{1}{2} \sum_{j \in I_k} \gamma_{d,j}$. Consequently, we need to ensure that

$$2 \leq \sum_{j \in I_k} \gamma_{d,j}$$

to conclude (4.21) for all admissible $\mathbf{z} \in Z$. But this is true for every $k \in \{1, \dots, s\}$ by definition of the sets I_k in Step 1. Thus we have shown (4.20) for the special case $\sigma = \mathbf{e}_k$ for all $k \in \{1, \dots, s\}$.

The inequality (4.20) also holds true for every $\sigma \in \{0, 1\}^s$ by an easy inductive argument on the cardinality of $|\sigma|$. Indeed, if $|\sigma| \geq 2$ then $\sigma = \sigma' + e_k$ with $|\sigma'| = |\sigma| - 1$. We now need to estimate $\|D^{\sigma'+e_k}h\|_{L_\infty(Z)}$. Since $(D^{e_k}h)(z) = a(z_{(k)})$ has the same structure as the function h itself, we see that $\|D^{\sigma'+e_k}h\|_{L_\infty(Z)}$ equals $\|D^{\sigma'}a(z_{(k)})\|_{L_\infty(Z)}$ and the proof of (4.20) then is completed by the inductive step.

Step 4. Collecting the previous equalities and estimates we obtain

$$\begin{aligned} \|g|_{\mathcal{P}_d^\gamma}\| &= \|g|_{\mathcal{P}_{m_s}^\gamma}\| = \max_{\substack{\alpha \in \{0,1\}^{m_s} \\ T(\alpha) \in \{0,1\}^s}} \frac{1}{\gamma_\alpha} \|D^\alpha g\|_{L_\infty(X)} = \max_{\sigma \in \{0,1\}^s} \|D^\sigma h_g\|_{L_\infty(Z)} \\ &\leq \|h_g\|_{L_\infty(Z)} = \|g\|_{L_\infty(X)} = \|g|_{L_\infty([0,1]^d)}\| \end{aligned}$$

for every $g \in V$, where V is a linear subspace of $\mathcal{F} = \mathcal{P}_d^\gamma$ with $\dim V = 2^s$. Therefore Proposition 2.2 with $a = 1$ yields that for $n < \dim V$ the worst case error

$$\Delta^{\text{wor}}(A_{n,d}; \text{id}_d: B_r(\mathcal{P}_d^\gamma) \rightarrow L_\infty([0,1]^d))$$

of any algorithm $A_{n,d}$ from the class $\mathcal{A}_d^{n,\text{cont}} \cup \mathcal{A}_d^{n,\text{adapt}}$ is lower bounded by r , the radius of the centered ball $B_r(\mathcal{P}_d^\gamma)$. \blacksquare

4.3.2 Complexity results via embeddings

Keeping in mind the assertions shown in the previous sections, we are ready to give conditions for tractability of the uniform approximation problem

$$\text{App} = (\text{App}_d)_{d \in \mathbb{N}}, \quad \text{App}_d: \mathcal{B}(\mathcal{F}_d^\gamma) \rightarrow L_\infty([0,1]^d), \quad \text{App}_d(f) = \text{id}_d(f) = f.$$

We suppose $(\mathcal{F}_d^\gamma)_{d \in \mathbb{N}}$ to be a sequence of Banach spaces of real-valued functions f defined on the unit cube $[0,1]^d$. We further assume that this sequence depends on product weights $\gamma = (\gamma_\alpha)_{\alpha \in \mathbb{N}_0^d}$ and fulfills one of the following simple assumptions:

(A4.1) $\mathcal{P}_d^\gamma \hookrightarrow \mathcal{F}_d^\gamma$ with norm

$$C_{1,d} \leq c \cdot d^{q_1} \quad \text{for all } d \in \mathbb{N}$$

and some absolute constants $c, q_1 \geq 0$,

(A4.2) $\mathcal{F}_d^\gamma \hookrightarrow \mathcal{H}_d^\gamma$ with norm

$$C_{2,d} \leq a \cdot \exp\left(b \cdot \sum_{j=1}^d (\gamma_{d,j})^t\right) \quad \text{for all } d \in \mathbb{N} \quad (4.22)$$

and some absolute constants $a > 0, b \geq 0$, as well as a parameter $t \in (0, 1]$ independent of d and γ .

Here the spaces \mathcal{P}_d^γ and $\mathcal{H}_d^\gamma = \mathcal{H}(K_d^\gamma)$ are defined as in Section 4.3.1 and Section 4.1.2, respectively.

To simplify the notation we use the commonly known definitions of the so-called *sum exponents*²⁹ for the product weight sequence $\gamma = (\gamma_\alpha)_{\alpha \in \mathbb{N}_0^d}$, $d \in \mathbb{N}$, induced by

²⁹Note that some authors use the name *decay* for $1/p(\cdot)$.

uniformly bounded generator weights $0 < \gamma_{d,j} \leq C_\gamma$, $j = 1, \dots, d$; see (4.6). We set

$$p(\gamma) = \inf \left\{ \kappa \geq 0 \left| P_\kappa(\gamma) = \limsup_{d \rightarrow \infty} \sum_{j=1}^d (\gamma_{d,j})^\kappa < \infty \right. \right\},$$

as well as

$$q(\gamma) = \inf \left\{ \kappa \geq 0 \left| Q_\kappa(\gamma) = \limsup_{d \rightarrow \infty} \frac{\sum_{j=1}^d (\gamma_{d,j})^\kappa}{\ln(d+1)} < \infty \right. \right\},$$

with the usual convention that $\inf \emptyset = \infty$.

The following necessary conditions for (strong) polynomial tractability slightly generalize Theorem 2 of [Wei12b].

Proposition 4.6 (Necessary conditions). Assume that (A4.1) holds true with some $q_1 \geq 0$. Consider L_∞ -approximation over $(\mathcal{F}_d^\gamma)_{d \in \mathbb{N}}$ in the worst case setting with respect to the class of algorithms $\mathcal{A}_d^{n,\text{cont}} \cup \mathcal{A}_d^{n,\text{adapt}}$ and the absolute error criterion. Then

$$n^{\text{wor}}(\varepsilon, d; \text{App}_d) > \frac{1}{2} \cdot 2^\wedge \left(\frac{1}{2 + C_\gamma} \sum_{j=1}^d \gamma_{d,j} \right) \quad (4.23)$$

for all $d \in \mathbb{N}$ and every $\varepsilon \in (0, C_{1,d}^{-1})$. Hence,

- if the problem App is polynomially tractable then $q(\gamma) \leq 1$,
- if $q_1 = 0$ and the problem is strongly polynomially tractable then $p(\gamma) \leq 1$.

Proof. Let $d \in \mathbb{N}$. Due to (A4.1), every algorithm $A_{n,d} \in \mathcal{A}_d^{n,\text{cont}} \cup \mathcal{A}_d^{n,\text{adapt}}$ for L_∞ -approximation defined on \mathcal{F}_d^γ also applies to the embedded space \mathcal{P}_d^γ . Furthermore the embedding constant $C_{1,d}$ implies that the ball $B_r(\mathcal{P}_d^\gamma)$ of radius $r = C_{1,d}^{-1}$ in \mathcal{P}_d^γ is completely contained in the unit ball $\mathcal{B}(\mathcal{F}_d^\gamma)$ of \mathcal{F}_d^γ . Therefore,

$$\begin{aligned} \Delta^{\text{wor}}(A_{n,d}; \text{App}_d: \mathcal{B}(\mathcal{F}_d^\gamma) \rightarrow L_\infty([0, 1]^d)) \\ \geq \Delta^{\text{wor}} \left(A_{n,d}|_{\mathcal{P}_d^\gamma}; \text{id}_d: B_r(\mathcal{P}_d^\gamma) \rightarrow L_\infty([0, 1]^d) \right). \end{aligned}$$

From Theorem 4.5 we have that the latter quantity is lower bounded by $r = C_{1,d}^{-1}$ provided that $n < 2^s$, where $s = s(\gamma, d) \in \{0, \dots, d\}$ satisfies (4.18). Since this lower bound holds for any such $A_{n,d}$ it remains valid for the n th minimal error, i.e.

$$e^{\text{wor}}(n, d; \text{App}_d) \geq C_{1,d}^{-1} \quad \text{for all } n < 2^s.$$

Hence we obtain $n^{\text{wor}}(\varepsilon, d; \text{App}_d) \geq 2^s$ for all $d \in \mathbb{N}$ and every $\varepsilon \in (0, C_{1,d}^{-1})$ which implies (4.23) using (4.18).

Now suppose the problem $\text{App} = (\text{App}_d)_{d \in \mathbb{N}}$ to be polynomially tractable. Then there are constants $C, p > 0$ and $q_2 \geq 0$ such that

$$n^{\text{wor}}(\varepsilon, d; \text{App}_d) \leq C \varepsilon^{-p} d^{q_2} \quad \text{for all } d \in \mathbb{N} \quad \text{and } \varepsilon \in (0, 1].$$

For any given $d \in \mathbb{N}$ we can take, say, $\varepsilon = \varepsilon(d) = \frac{1}{2} \cdot \min \{1, C_{1,d}^{-1}\}$ to conclude

$$2^\wedge \left(\frac{1}{2 + C_\gamma} \sum_{j=1}^d \gamma_{d,j} \right) < C' \max \{1, C_{1,d}^p\} d^{q_2} \quad (4.24)$$

for some $C' > 0$ independent of d . If we now assume that $C_{1,d} \in \mathcal{O}(d^{q_1})$ then the right-hand side of the last inequality belongs to $\mathcal{O}(d^{p q_1 + q_2})$, as $d \rightarrow \infty$. Provided that $\max \{q_1, q_2\} > 0$ this is equivalent to the boundedness of $\sum_{j=1}^d \gamma_{d,j} / \ln(d+1)$ such that we arrive at $q(\gamma) \leq 1$, as claimed.

Finally, the case of strong polynomial tractability can be treated similarly by setting $q_1 = q_2 = 0$ in the latter bounds. Then we obtain that $\sum_{j=1}^d \gamma_{d,j}$ is uniformly bounded in d which implies $p(\gamma) \leq 1$. \blacksquare

Of course, the conditions $q(\gamma) \leq 1$ and $p(\gamma) \leq 1$ are also necessary for polynomial and strong polynomial tractability with respect to smaller classes of algorithms such as, e.g., $\mathcal{A}_d^{n,\text{lin}}(\Lambda^{\text{all}})$.

Observe that one of the improvements compared to [Wei12b, Theorem 2] is the possibility to choose the uniform upper bound for the generator weights, C_γ , different than 1. Moreover, now we have weaker conditions on the embedding constant $C_{1,d}$. For the application we have in mind we will see that there still $C_{1,d} = 1$. But we note in passing that the stated conclusions for (strong) polynomial tractability are only special instances of the more general bound (4.24) obtained in the latter proof which we will not investigate further.

We next assume (A4.2) and show that slightly stronger conditions on the product weights γ than in Proposition 4.6 are sufficient for polynomial and strong polynomial tractability, respectively. This is stated in the next assertion which can be found as Theorem 3 in [Wei12b].

Proposition 4.7 (Sufficient conditions). Suppose that (A4.2) holds true with some $t \in (0, 1]$. Consider L_∞ -approximation over $(\mathcal{F}_d^\gamma)_{d \in \mathbb{N}}$ in the worst case setting with respect to the class of linear algorithms $\mathcal{A}_d^{n,\text{lin}}(\Lambda^{\text{all}})$ and the absolute error criterion. Then

- $q(\gamma) < t$ implies polynomial tractability,
- $p(\gamma) < t$ implies strong polynomial tractability.

Proof. Due to (A4.2), the restriction of the algorithm $A_{n,d}^*$ in Corollary 4.4 from \mathcal{H}_d^γ to \mathcal{F}_d^γ is admissible for L_∞ -approximation over \mathcal{F}_d^γ . Furthermore, due to the

linearity of $A_{n,d}^*$, we have

$$\begin{aligned} \|f - A_{n,d}^* f\|_{L_\infty([0,1]^d)} &\leq \Delta^{\text{wor}}(A_{n,d}^*; \text{id}_d: \mathcal{B}(\mathcal{H}_d^\gamma) \rightarrow L_\infty([0,1]^d)) \cdot \|f\|_{\mathcal{H}_d^\gamma} \\ &\leq \Delta^{\text{wor}}(A_{n,d}^*; \text{id}_d: \mathcal{B}(\mathcal{H}_d^\gamma) \rightarrow L_\infty([0,1]^d)) \cdot C_{2,d} \cdot \|f\|_{\mathcal{F}_d^\gamma} \end{aligned}$$

for all $f \in \mathcal{F}_d^\gamma$. Therefore we can estimate the n th minimal error by

$$\begin{aligned} e^{\text{wor}}(n, d; \text{App}_d) &\leq \Delta^{\text{wor}}\left(A_{n,d}^*|_{\mathcal{F}_d^\gamma}; \text{App}_d: \mathcal{B}(\mathcal{F}_d^\gamma) \rightarrow L_\infty([0,1]^d)\right) \\ &\leq C_{2,d} \cdot \Delta^{\text{wor}}(A_{n,d}^*; \text{id}_d: \mathcal{B}(\mathcal{H}_d^\gamma) \rightarrow L_\infty([0,1]^d)) \\ &\leq a \cdot a_\tau \cdot \exp\left(b \sum_{j=1}^d (\gamma_{d,j})^t + b_\tau \sum_{j=1}^d (\gamma_{d,j})^\tau\right) \cdot n^{-(1-\tau)/(2\tau)}, \end{aligned}$$

where τ is an arbitrary number from $(1/2, 1)$. Choosing n such that the right-hand side is not greater than a given $\varepsilon \in (0, 1]$, we obtain an estimate for the information complexity with respect to the class of linear algorithms,

$$n^{\text{wor}}(\varepsilon, d; \text{App}_d) \leq c_1 \cdot \varepsilon^{-2\tau/(1-\tau)} \cdot \exp\left(c_2 \sum_{j=1}^d (\gamma_{d,j})^t + c_3 \sum_{j=1}^d (\gamma_{d,j})^\tau\right), \quad (4.25)$$

where the non-negative constants c_1 , c_2 and c_3 only depend on τ , a and b .

Suppose that $q(\gamma) < t$. Then $Q_\kappa(\gamma)$ is finite for every $\kappa > q(\gamma)$. Taking $\kappa = t$ we obtain

$$\frac{\sum_{j=1}^d (\gamma_{d,j})^t}{\ln(d+1)} \cdot \ln(d+1) \leq (Q_t(\gamma) + \delta) \cdot \ln(d+1) = \ln(d+1)^{Q_t(\gamma)+\delta}$$

for every $\delta > 0$ whenever d is larger than a certain $d_\delta \in \mathbb{N}$. This means that the factor $\exp\left(c_2 \sum_{j=1}^d (\gamma_{d,j})^t\right)$ in (4.25) is polynomially dependent on d . On the other hand, we can choose $\tau \in (\max\{q(\gamma), 1/2\}, 1)$ such that $Q_\tau(\gamma)$ is finite and thus the factor $\exp\left(c_3 \sum_{j=1}^d (\gamma_{d,j})^\tau\right)$ in (4.25) is also polynomially dependent on d . So, for this value of τ we can rewrite (4.25) as

$$n^{\text{wor}}(\varepsilon, d; \text{App}_d) \in \mathcal{O}\left(\varepsilon^{-2\tau/(1-\tau)} \cdot (d+1)^{c_4}\right),$$

with c_4 , as well as the implied factor in the \mathcal{O} -notation, independent of d and ε which means that the problem is polynomially tractable, as claimed.

Suppose finally that $p(\gamma) < t$. Then the sums $\sum_{j=1}^d (\gamma_{d,j})^t$ and $\sum_{j=1}^d (\gamma_{d,j})^\tau$ for $\tau \in (\max\{p(\gamma), 1/2\}, 1)$ are both uniformly bounded in d . Consequently (4.25) yields strong polynomial tractability, and completes the proof. \blacksquare

The conditions in Proposition 4.7 are obviously also sufficient if we consider larger classes of algorithms such as, e.g., $\mathcal{A}_d^{n,\text{cont}} \cup \mathcal{A}_d^{n,\text{adapt}}$. Moreover note that the given proof also provides explicit upper bounds for the exponents of tractability.

Let us briefly discuss the different roles of the assumptions (A4.1) and (A4.2) in the following remark.

Remark 4.8. Assumption (A4.1) is used to find a lower bound on the information complexity for the space \mathcal{F}_d^γ as long the space \mathcal{P}_d^γ is continuously embedded in \mathcal{F}_d^γ with an embedding constant which grows at most polynomially with the dimension d . Such an embedding can be shown for several different classes of functions.

On the other hand, assumption (A4.2) is used to find an upper bound on the information complexity for the space \mathcal{F}_d^γ as long as it is continuously embedded in the unanchored weighted Sobolev space $\mathcal{H}_d^\gamma = \mathcal{H}(K_d^\gamma)$ with an embedding constant depending exponentially on the sum of some power of the generators $\gamma_{d,j}$ of the product weights γ . This considerably restricts the choice of \mathcal{F}_d^γ . We need this assumption in order to use the linear algorithm $A_{n,d}^*$ defined on the space \mathcal{H}_d^γ and the error bound given in Corollary 4.4.

Obviously, we can replace the space \mathcal{H}_d^γ in (A4.2) by any other space which contains at least \mathcal{P}_d^γ and for which we know a linear algorithm using n linear functionals whose worst case error is polynomial in n^{-1} with an explicit dependence on the product weights γ . \square

We now show that the assumptions (A4.1) and (A4.2) allow us to characterize weak tractability and the curse of dimensionality.

Theorem 4.9 (Weak tractability and the curse of dimensionality). Suppose that for a sequence of Banach spaces $(\mathcal{F}_d^\gamma)_{d \in \mathbb{N}}$ equipped with product weights γ the assumptions (A4.1) and (A4.2) hold true with some parameter $t \in (0, 1]$. Consider the L_∞ -approximation problem App in the worst case setting and with respect to the absolute error criterion. Then the following statements are equivalent:

- (i) The problem is weakly tractable with respect to the class $\mathcal{A}_d^{n,\text{lin}}(\Lambda^{\text{all}})$.
- (ii) The problem is weakly tractable with respect to the class $\mathcal{A}_d^{n,\text{cont}} \cup \mathcal{A}_d^{n,\text{adapt}}$.
- (iii) There is no curse of dimensionality for the class $\mathcal{A}_d^{n,\text{lin}}(\Lambda^{\text{all}})$.
- (iv) There is no curse of dimensionality for the class $\mathcal{A}_d^{n,\text{cont}} \cup \mathcal{A}_d^{n,\text{adapt}}$.
- (v) For all $\kappa > 0$ we have $\lim_{d \rightarrow \infty} \frac{1}{d} \sum_{j=1}^d (\gamma_{d,j})^\kappa = 0$.
- (vi) There exists $\kappa \in (0, t)$ such that $\lim_{d \rightarrow \infty} \frac{1}{d} \sum_{j=1}^d (\gamma_{d,j})^\kappa = 0$.

Proof. We start by showing that (vi) implies (i), i.e.,

$$\lim_{\varepsilon^{-1} + d \rightarrow \infty} \frac{\ln(n^{\text{wor}}(\varepsilon, d; \text{App}_d))}{\varepsilon^{-1} + d} = 0,$$

where the information complexity is taken with respect to the class $\mathcal{A}_d^{n,\text{lin}}(\Lambda^{\text{all}})$ of linear algorithms that use continuous linear functionals. By the arguments used in the proof of Proposition 4.7 we obtain estimate (4.25) for all ε in $(0, 1]$, as well as for every $d \in \mathbb{N}$, and all $\tau \in (1/2, 1)$, due to assumption (A4.2). Clearly, for

$\kappa \in (0, t)$ as in the hypothesis and $t \in (0, 1]$ as in the embedding condition, we find $\tau \in (1/2, 1)$ such that $\kappa < \min \{t, \tau\}$. So, since $\gamma_{d,j} \leq C_\gamma$, we can estimate

$$\sum_{j=1}^d (\gamma_{d,j})^s = C_\gamma^s \cdot \sum_{j=1}^d \left(\frac{\gamma_{d,j}}{C_\gamma} \right)^s \leq C_\gamma^{s-\kappa} \cdot \sum_{j=1}^d (\gamma_{d,j})^\kappa \leq C \cdot \sum_{j=1}^d (\gamma_{d,j})^\kappa,$$

where s either equals t or τ and $C = \max \{1, C_\gamma\}$. Therefore the right-hand side of (4.25) can be estimated from above and thus

$$\frac{\ln(n^{\text{wor}}(\varepsilon, d; \text{App}_d))}{\varepsilon^{-1} + d} \leq \frac{\ln(c_1)}{\varepsilon^{-1} + d} + \frac{2\tau}{1-\tau} \cdot \frac{\ln(\varepsilon^{-1})}{\varepsilon^{-1} + d} + C \cdot \max \{c_2, c_3\} \cdot \frac{\sum_{j=1}^d (\gamma_{d,j})^\kappa}{\varepsilon^{-1} + d}$$

tends to zero when $\varepsilon^{-1} + d$ approaches infinity, as claimed.

Clearly, (i) \Rightarrow (ii) \Rightarrow (iv) and (i) \Rightarrow (iii) \Rightarrow (iv). Moreover the implication from (v) to (vi) is obvious. Hence, it only remains to show that (iv) \Rightarrow (v).

From (A4.1) we have estimate (4.23). Then the absence of the curse of dimensionality implies

$$\lim_{d \rightarrow \infty} \frac{1}{d} \sum_{j=1}^d \gamma_{d,j} = 0.$$

Now Jensen's inequality yields that

$$\frac{1}{d} \sum_{j=1}^d \gamma_{d,j} \geq \left(\frac{1}{d} \sum_{j=1}^d (\gamma_{d,j})^\kappa \right)^{1/\kappa} \quad \text{for } 0 < \kappa \leq 1,$$

because $f(y) = y^\kappa$ is a concave function for $y > 0$. This shows

$$\lim_{d \rightarrow \infty} \frac{1}{d} \sum_{j=1}^d (\gamma_{d,j})^\kappa = 0 \quad \text{for all } 0 < \kappa \leq 1.$$

Finally, for every $\kappa \geq 1$ we can estimate $\gamma_{d,j} \geq C_\gamma^{1-\kappa} (\gamma_{d,j})^\kappa$ since $\gamma_{d,j} \leq C_\gamma$ for $j = 1, \dots, d$. Therefore $\lim_{d \rightarrow \infty} d^{-1} \sum_{j=1}^d (\gamma_{d,j})^\kappa = 0$ also holds true for $\kappa > 1$, and the proof is complete. \blacksquare

4.3.3 Conclusions and applications

In this last part of the current section we give some examples to illustrate the obtained complexity results. To this end, we only have to prove the corresponding embeddings, i.e. we need to verify assumption (A4.1) and/or (A4.2) from the beginning of Section 4.3.2.

Example 4.10 (Limiting cases \mathcal{P}_d^γ and \mathcal{H}_d^γ). To begin with, we check the case where $\mathcal{F}_d^\gamma = \mathcal{P}_d^\gamma$ for every $d \in \mathbb{N}$. Then (A4.1) obviously holds with $C_{1,d} = 1$, i.e. $c = 1$ and $q_1 = 0$. To prove (A4.2), note that the algebraical inclusion $\mathcal{F}_d^\gamma \subset \mathcal{H}_d^\gamma$ is trivial by the definition of $\mathcal{H}_d^\gamma = \mathcal{H}(K_d^\gamma)$ given in Section 4.1.2. For $f \in \mathcal{F}_d^\gamma = \mathcal{P}_d^\gamma$ we calculate

$$\|f|_{\mathcal{H}_d^\gamma}\|^2 \leq \sum_{\alpha \in \{0,1\}^d} \frac{1}{\gamma_\alpha} \int_{[0,1]^d} \|D^\alpha f|_{L_\infty([0,1]^d)}\|^2 d\lambda^d(\mathbf{x}) \leq \|f|_{\mathcal{F}_d^\gamma}\|^2 \cdot \sum_{\alpha \in \{0,1\}^d} \gamma_\alpha$$

using (4.10), as well as (4.17). Hence the norm of the embedding $\mathcal{F}_d^\gamma \hookrightarrow \mathcal{H}_d^\gamma$ is bounded by

$$\left(\sum_{\alpha \in \{0,1\}^d} \gamma_\alpha \right)^{1/2} = \left(\prod_{j=1}^d (1 + \gamma_{d,j}) \right)^{1/2} \leq \exp \left(\frac{1}{2} \sum_{j=1}^d \gamma_{d,j} \right).$$

So, with $a = 1$, $b = 1/2$, and $t = 1$ the assumption (A4.2) is also fulfilled and we can apply the stated assertions from Section 4.3.2 for the spaces $\mathcal{F}_d^\gamma = \mathcal{P}_d^\gamma$, $d \in \mathbb{N}$.

We now turn to the case $\mathcal{F}_d^\gamma = \mathcal{H}_d^\gamma$. Unfortunately, the estimate above indicates that (A4.1) may not hold for $\mathcal{F}_d^\gamma = \mathcal{H}_d^\gamma$ with $C_{1,d} \in \mathcal{O}(d^{q_1})$ without imposing additional conditions on the product weights γ . Nevertheless, in this case assumption (A4.2) is trivially true with $C_{2,d} = 1$, i.e., $a = 1$, $b = 0$, and $t = 1$. Therefore we can apply Proposition 4.7 for this space. Thus the problem is polynomially tractable if $q(\gamma) < 1$ and we have strong polynomial tractability if $p(\gamma) < 1$. It can be shown that these conditions are also necessary; see Section 4.4. \square

Next we discuss a more advanced sequence of Banach function spaces.

Example 4.11 ($C^{(1,\dots,1)}$). For every $d \in \mathbb{N}$ consider the space

$$\mathcal{F}_d^\gamma = \{f: [0,1]^d \rightarrow \mathbb{R} \mid f \in C^{(1,\dots,1)}([0,1]^d), \text{ where } \|f|_{\mathcal{F}_d^\gamma}\| < \infty\}$$

of functions which are once continuously differentiable in every coordinate direction, where

$$\|f|_{\mathcal{F}_d^\gamma}\| = \max_{\alpha \in \{0,1\}^d} \frac{1}{\gamma_\alpha} \|D^\alpha f|_{L_\infty([0,1]^d)}\|.$$

Since \mathcal{P}_d^γ is a linear subset of \mathcal{F}_d^γ and, due to (4.17), the norm $\|\cdot|_{\mathcal{P}_d^\gamma}\|$ is simply the restriction of $\|\cdot|_{\mathcal{F}_d^\gamma}\|$ we have $\mathcal{P}_d^\gamma \hookrightarrow \mathcal{F}_d^\gamma$ with an embedding factor $C_{1,d} = 1$ and hence (A4.1) holds true. For the norm $C_{2,d}$ of the embedding $\mathcal{F}_d^\gamma \hookrightarrow \mathcal{H}_d^\gamma$, the same estimates hold exactly as in the previous example and, moreover, the set inclusion is obvious. Therefore also assumption (A4.2) is fulfilled and we can apply the propositions and theorems of Section 4.3.2 to the sequence $(\mathcal{F}_d^\gamma)_{d \in \mathbb{N}}$. \square

Our last example $\mathcal{F}_d^\gamma = F_d^\gamma$, for all $d \in \mathbb{N}$, finally shows that even very high smoothness does not improve the conditions for tractability.

Example 4.12 (C^∞). For $d \in \mathbb{N}$ and product weights γ let

$$\mathcal{F}_d^\gamma = F_d^\gamma = \{f: [0, 1]^d \rightarrow \mathbb{R} \mid f \in C^\infty([0, 1]^d) \text{ with } \|f\|_{F_d^\gamma} < \infty\},$$

where the norm is given by (4.5). Obviously, $\mathcal{P}_d^\gamma \subset C^\infty$, because functions from \mathcal{P}_d^γ are at most linear in each coordinate. This moreover implies that $D^\alpha f \equiv 0$ for all $\alpha \in \mathbb{N}_0^d \setminus \{0, 1\}^d$. Therefore, once again we have

$$\|f\|_{\mathcal{P}_d^\gamma} = \max_{\alpha \in \{0, 1\}^d} \frac{1}{\gamma_\alpha} \|D^\alpha f\|_{L_\infty([0, 1]^d)} = \|f\|_{\mathcal{F}_d^\gamma} \quad \text{for all } f \in \mathcal{P}_d^\gamma.$$

Together this yields $\mathcal{P}_d^\gamma \hookrightarrow \mathcal{F}_d^\gamma$ with an embedding constant $C_{1,d} = 1$ for all $d \in \mathbb{N}$. In addition, also (A4.2) can be concluded as in the examples above. So, even infinite smoothness leads to the the same conditions for tractability and the curse of dimensionality as before. \square

Note that in the latter example we do not need to claim a product structure for the weights according to multi-indices $\alpha \in \mathbb{N}_0^d \setminus \{0, 1\}^d$. Furthermore, this example is a generalization of the space F_d studied in [NW09]. For $\gamma_\alpha \equiv 1$ we reproduce the intractability result stated there because then F_d^γ equals F_d for each $d \in \mathbb{N}$.

In conclusion we discuss the tractability behavior of uniform approximation defined on one of the spaces \mathcal{F}_d^γ above using a special class of product weights γ which are *independent of the dimension d* . That is, for the generator weights we claim that

$$\gamma_{d,j} \equiv \gamma^{(j)} \in \Theta(j^{-\beta}) \quad \text{for some } \beta \geq 0, \quad (4.26)$$

and all j and $d \in \mathbb{N}$. The imposed polynomial behavior of $\gamma^{(j)}$ is a typical example in the theory of product weights. Clearly, $p(\gamma)$ is finite if and only if $\beta > 0$, and if so then $p(\gamma) = 1/\beta$. For details see [NW08, Section 5.3.4].

If $\beta = 0$ then the L_∞ -approximation problem $\text{App} = (\text{App}_d)_{d \in \mathbb{N}}$ is intractable (more precisely it suffers from the curse of dimensionality) due to [Theorem 4.9](#), assertion (v), since then $d^{-1} \sum_{j=1}^d \gamma_{d,j}$ does not tend to zero. For $\beta \in (0, 1)$, easy calculus yields $q(\gamma) > 1$. So, using [Proposition 4.6](#) we conclude polynomial intractability in this case. On the other hand, for all δ and κ with $0 < \delta < \kappa \leq 1$, we have

$$\frac{\sum_{j=1}^d j^{-\kappa}}{d} = \frac{\sum_{j=1}^d j^{-\kappa} d^{\kappa-(1+\delta)}}{d^{\kappa-\delta}} \leq \frac{\sum_{j=1}^d j^{-(1+\delta)}}{d^{\kappa-\delta}} \rightarrow 0 \quad \text{for } d \rightarrow \infty$$

and if $\kappa > 1$ then the most left fraction obviously tends to zero, too. Hence condition (vi) of [Theorem 4.9](#) holds and the problem is weakly tractable for all $\beta > 0$.

For $\beta = 1$ we use inequality (4.23) from [Proposition 4.6](#) and estimate

$$\sum_{j=1}^d \gamma_{d,j} \geq c \cdot \ln(d+1)$$

for some positive c . Therefore, for every $\varepsilon \in (0, 1)$ the information complexity $n(\varepsilon, d; \text{App}_d)$ is lower bounded polynomially in $d \in \mathbb{N}$. This proves that strong polynomial tractability does not hold for $\beta = 1$. Moreover, it is easy to show that in this case the sufficient condition $q(\gamma) < 1$ for polynomial tractability is not fulfilled. So, we do not know whether polynomial tractability holds or not.

Finally, consider $\beta > 1$ in (4.26). Then we easily see that $p(\gamma) = \frac{1}{\beta} < 1 = t$. Thus Proposition 4.7 provides strong polynomial tractability in this situation.

In summary, we proved all the assertions we claimed at the end of Section 4.1.1.

4.4 Possible extensions and further results

Note that the main result of this chapter, the lower bound given in Theorem 4.5, can be easily transferred from $[0, 1]^d$ to more general domains $\Omega_d \subset \mathbb{R}^d$. Indeed, the case $\Omega_d = [c_1, c_2]^d$, where $c_1 < c_2$, can be immediately obtained using the presented techniques. It turns out that in this case we have to modify estimate (4.18) by a constant which depends only on the length of the interval $[c_1, c_2]$. Consequently, the general tractability behavior does not change.

Another extension of the obtained results is possible if we consider L_p -norms ($1 \leq p < \infty$) instead of the L_∞ -norm. In Section 4.4.1 we briefly discuss these norms for the unweighted case. Then the modifications for the weighted case are obvious and thus we leave it for the interested reader. In passing we correct a small mistake stated in [NW09].

Finally, in Section 4.4.2, we show that the algorithm studied in Corollary 4.4 is essentially optimal for the uniform approximation problem on the unanchored weighted Sobolev space $\mathcal{H}(K_d^\gamma)$ defined in Example 4.1.

4.4.1 L_p -approximation

As in [Wei12b, Section 7] we follow Novak and Woźniakowski [NW09] and define the spaces

$$F_{d,p} = \left\{ f \in C^\infty([c_1, c_2]^d) \mid \|f\|_{F_{d,p}} = \sup_{\alpha \in \mathbb{N}_0^d} \|D^\alpha f\|_{L_p([c_1, c_2]^d)} < \infty \right\}$$

for $1 \leq p < \infty$ and $d \in \mathbb{N}$, where we assume that $l = c_2 - c_1 > 0$. In what follows we want to approximate $f \in F_{d,p}$ in the norm of L_p . That is, we modify (4.2) and consider the problem $S^p = (S_d^p)_{d \in \mathbb{N}}$ given by

$$S_d^p = \text{id}_d^p: \mathcal{B}(F_{d,p}) \rightarrow L_p([c_1, c_2]^d), \quad f \mapsto \text{id}_d^p(f) = f.$$

Hence we try to minimize the n th minimal worst case error

$$e_p^{\text{wor}}(n, d; \text{id}_d^p) = \inf_{A_{n,d}} \sup_{f \in \mathcal{B}(F_{d,p})} \|f - A_{n,d}(f)\|_{L_p([c_1, c_2]^d)}$$

which now depends on the additional integrability parameter p . Observe that, without loss of generality, we can restrict ourselves to the case $[c_1, c_2] = [0, l]$.

In order to conclude a lower bound analogue to (4.3) and Theorem 4.5, i.e., $e_p^{\text{wor}}(n, d; \text{id}_d^p) \geq 1$ for $n < 2^s$, we once again use Proposition 2.2 with $\mathcal{F} = F_{d,p}$ and $\mathcal{G} = L_p([0, l]^d)$.³⁰ The authors of [NW09] suggest to use the subspace $V_d^{(k)} \subset F_{d,p}$ defined as

$$V_d^{(k)} = \text{span} \left\{ g_i : [0, l]^d \rightarrow \mathbb{R}, \mathbf{x} \mapsto g_i(\mathbf{x}) = \prod_{j=1}^s \left(\sum_{m=(j-1)k+1}^{jk} x_m \right)^{i_j} \mid \mathbf{i} \in \{0, 1\}^s \right\},$$

where $s = \lfloor d/k \rfloor$ and $k \in \mathbb{N}$ such that $kl \geq 2(p+1)^{1/p}$. Hence, if $l < 2(p+1)^{1/p}$ then we have to use blocks of variables with size $k > 1$, in order to guarantee (2.1). That is, to fulfill the condition

$$\|g\|_{F_{d,p}} \leq \|g\|_{L_p([0, l]^d)} \quad \text{for all } g \in V_d^{(k)}. \quad (4.27)$$

Therefore Novak and Woźniakowski defined $k = \lceil 2(p+1)^{1/p}/l \rceil$, but this is too small as the following example shows.

Example 4.13. For $d \geq 4$ take $l = 1$, i.e. $[c_1, c_2]^d = [0, 1]^d$, and $p = 1$. Then $k = 4$ should be a proper choice, but for $g^*(\mathbf{x}) = (x_1 + x_2 + x_3 + x_4) - 2$ it can be checked (using a computer algebra system) that

$$\|g^*\|_{L_1([0, 1]^d)} = \frac{7}{15} < 1 = \left\| \frac{\partial g^*}{\partial x_1} \right\|_{L_1([0, 1]^d)}.$$

This obviously contradicts (4.27). □

For an exhaustive proof of an assertion which states that a slightly larger choice of $k \in \mathbb{N}$ suffices to conclude the desired intractability result we need to show the following technical lemma first. Its proof is based on some well-known arguments from Banach space geometry.

Lemma 4.14. Let $p \in [1, \infty)$ and $k \in \mathbb{N}$. Then

$$\mathfrak{J}_{k,p} = \int_{[-1/2, 1/2]^k} \left| \sum_{m=1}^k z_m \right|^p d\lambda^k(\mathbf{z}) \geq C_p \cdot k^{p/2} \quad (4.28)$$

with some $C_p \geq 1/[(2\sqrt{2})^p(1+p)]$ independent of k .

³⁰Note that it is sufficient to restrict ourselves to the case $r = 1$ since now we do not need to take care of embedding constants as in the proof of Proposition 4.6.

Proof. For $k = 1$ we easily calculate $\mathfrak{I}_{1,p} = 1/[2^p(1+p)]$. Hence, without loss of generality we can assume $k \geq 2$ in what follows.

To abbreviate the notation, let us define

$$f = f_k: \mathbb{R}^k \rightarrow \mathbb{R}, \quad \mathbf{z} = (z_1, \dots, z_k) \mapsto f(\mathbf{z}) = \sum_{m=1}^k z_m \quad (4.29)$$

for any fixed $k \geq 2$. Moreover, for given vectors $\mathbf{z}, \mathbf{y} \in \mathbb{R}^k$, let $\langle \mathbf{z}, \mathbf{y} \rangle$ denote the inner product $\sum_{m=1}^k z_m y_m$ in \mathbb{R}^k . In the special case $\mathbf{y} = \boldsymbol{\xi} = 1/\sqrt{k} \cdot (\mathbf{1}, \dots, \mathbf{1}) \in \mathbb{S}^{k-1}$ it is $\langle \mathbf{z}, \boldsymbol{\xi} \rangle = t$ for a given $t \in \mathbb{R}$ if and only if $f(\mathbf{z}) = t\sqrt{k}$. Furthermore note that every \mathbf{y} in the k -dimensional unit sphere $\mathbb{S}^{k-1} \subset \mathbb{R}^k$ uniquely defines a hyperplane $\mathbf{y}^\perp = \{\mathbf{z} \in \mathbb{R}^k \mid \langle \mathbf{z}, \mathbf{y} \rangle = 0\}$ perpendicular to \mathbf{y} which contains zero. Therefore, for $\mathbf{y} = \boldsymbol{\xi}$ and every $t \in [0, \infty)$, the set

$$\mathfrak{H}_t = \boldsymbol{\xi}^\perp + t \cdot \boldsymbol{\xi} = \{\mathbf{z} \in \mathbb{R}^k \mid \langle \mathbf{z}, \boldsymbol{\xi} \rangle = t\}$$

describes a parallel shifted hyperplane in \mathbb{R}^k with distance t to the origin. Using Fubini's theorem, this leads to the following representation:

$$\begin{aligned} \mathfrak{I}_{k,p} &= \int_{[-1/2, 1/2]^k} |f(\mathbf{z})|^p \, d\lambda^k(\mathbf{z}) = 2 \cdot \int_{\substack{[-1/2, 1/2]^k \\ \langle \mathbf{z}, \boldsymbol{\xi} \rangle \geq 0}} f(\mathbf{z})^p \, d\lambda^k(\mathbf{z}) \\ &= 2 \cdot k^{p/2} \cdot \int_0^\infty t^p \left(\int_{[-1/2, 1/2]^k \cap \mathfrak{H}_t} 1 \, d\lambda^k(\mathbf{z}) \right) \, d\lambda^1(t). \end{aligned}$$

Now we see that the inner integral describes the $(k-1)$ -dimensional volume

$$v(t) = \lambda^{k-1}([-1/2, 1/2]^k \cap \mathfrak{H}_t)$$

of the parallel section of the unit cube with the hyperplanes defined above. Because of Ball's famous theorem we know that $v(0) \leq \sqrt{2}$ holds independently of k ; see, e.g., Chapter 7 in the monograph of Koldobsky [Kol05]. Moreover taking $\mathfrak{H}_0 = \boldsymbol{\xi}^\perp$ provides a central hyperplane section of the unit cube. From this observation we conclude that

$$\int_0^\infty v(t) \, d\lambda^1(t) = \frac{1}{2} \cdot \lambda^k([-1/2, 1/2]^k) = \frac{1}{2}$$

because of the symmetry of $[-1/2, 1/2]^k$ w.r.t. \mathfrak{H}_0 . In addition, by Brunn's theorem (cf. [Kol05, Theorem 2.3]), the function v is non-negative and non-increasing on the interval $[0, \infty)$. Thus v is related to the distribution function of a certain non-negative real-valued random variable X , up to some normalizing factor, i.e. $v(t) = v(0) \cdot \mathbb{P}(\{X \geq t\})$. Using Hölder's inequality³¹ we obtain $\mathbb{E}(X^{1+p}) \geq (\mathbb{E}X)^{1+p}$ and, respectively,

$$\mathfrak{I}_{k,p} = k^{p/2} \cdot 2 \int_0^\infty t^p v(t) \, d\lambda^1(t) \geq k^{p/2} \cdot \frac{2}{v(0)^p (1+p)} \left(\int_0^\infty v(t) \, d\lambda^1(t) \right)^{1+p}$$

³¹See also [Kol05, Lemma 7.5].

by integration by parts.

In summary we have shown (4.28) and hence the proof is complete. \blacksquare

Now the mentioned intractability result reads as follows:

Proposition 4.15. Let $1 \leq p < \infty$ and $l > 0$. Moreover, choose $k \in \mathbb{N}$ such that

$$k \geq \kappa_{p,l} = \lceil 8(p+1)^{2/p}/l^2 \rceil. \quad (4.30)$$

Then condition (4.27) holds for $V_d^{(k)} \subset F_{d,p}$. Hence the L_p -approximation problem $S^p = (\text{id}_d^p: F_{d,p} \rightarrow L_p([0, l]^d))_{d \in \mathbb{N}}$ suffers from the curse of dimensionality since

$$e_p^{\text{wor}}(n, d; \text{id}_d^p) \geq 1 \quad \text{for all } n < 2^{\lfloor d/k \rfloor}$$

and every $d \in \mathbb{N}$.

Proof. Due to the structure of the functions g from $V_d^{(k)}$, it suffices to show that

$$\|D^\alpha g|_{L_p([0, l]^{ks})}\| \leq \|g|_{L_p([0, l]^{ks})}\| \quad \text{for all } g \in V_d^{(k)} \quad \text{and every } \alpha \in \mathbb{M}_d^{(k)},$$

where the set of multi-indices $\mathbb{M}_d^{(k)}$ is defined by

$$\mathbb{M}_d^{(k)} = \left\{ \alpha = (\alpha_1, \dots, \alpha_{ks}) \in \{0, 1\}^{ks} \left| \sum_{m \in I_j} \alpha_m \leq 1 \text{ for all } j = 1, \dots, s \right. \right\}$$

and $I_j = \{(j-1)k+1, \dots, jk\}$. Observe that $\mathbb{M}_d^{(k)}$ depends on d via $s = \lfloor d/k \rfloor$. Similar to the proof of Theorem 4.5, we only need to consider the case $\alpha = \mathbf{e}_t \in \{0, 1\}^{ks}$ with $t \in I_j$. The rest then follows by induction.

Given $t \in I_j$ for some $j \in \{1, \dots, s\}$ we can represent every fixed $g \in V_d^{(k)}$, as well as its partial derivative $D^{\mathbf{e}_t} g$, by some functions $a, b: [0, l]^{k(s-1)} \rightarrow \mathbb{R}$ (depending on g and j) such that

$$g(\mathbf{x}) = a(\tilde{\mathbf{x}}) \sum_{m=1}^k y_m + b(\tilde{\mathbf{x}}) \quad \text{and} \quad (D^{\mathbf{e}_t} g)(\mathbf{x}) = a(\tilde{\mathbf{x}}), \quad \mathbf{x} \in [0, l]^{ks}.$$

Here we split the ks -dimensional vector $\mathbf{x} = (\mathbf{x}_{I_1}, \dots, \mathbf{x}_{I_{j-1}}, \mathbf{y}, \mathbf{x}_{I_{j+1}}, \dots, \mathbf{x}_{I_s})$ into $\tilde{\mathbf{x}} = (\mathbf{x}_{I_1}, \dots, \mathbf{x}_{I_{j-1}}, \mathbf{x}_{I_{j+1}}, \dots, \mathbf{x}_{I_s}) \in [0, l]^{k(s-1)}$ and $\mathbf{y} = (y_1, \dots, y_k) \in [0, l]^k$, where \mathbf{x}_{I_j} denotes the k -dimensional block of components x_m in \mathbf{x} with coordinates $m \in I_j$. Using this representation we can rewrite the inequality $\|D^{\mathbf{e}_t} g|_{L_p([0, l]^{ks})}\| \leq \|g|_{L_p([0, l]^{ks})}\|$ as

$$\begin{aligned} & \int_{[0, l]^{k(s-1)}} \int_{[0, l]^k} |a(\tilde{\mathbf{x}})|^p \, d\lambda^k(\mathbf{y}) \, d\lambda^{k(s-1)}(\tilde{\mathbf{x}}) \\ & \leq \int_{[0, l]^{k(s-1)}} \int_{[0, l]^k} \left| a(\tilde{\mathbf{x}}) \sum_{m=1}^k y_m + b(\tilde{\mathbf{x}}) \right|^p \, d\lambda^k(\mathbf{y}) \, d\lambda^{k(s-1)}(\tilde{\mathbf{x}}) \end{aligned}$$

such that it is enough to prove a pointwise estimate of the inner integrals for ($\lambda^{k(s-1)}$ -almost every) fixed $\tilde{\mathbf{x}} \in [0, l]^{k(s-1)}$ with $a = a(\tilde{\mathbf{x}}) \neq 0$. Easy calculus yields

$$\int_{[0, l]^k} \left| a \sum_{m=1}^k y_m + b \right|^p d\lambda^k(\mathbf{y}) = l^{p+k} \cdot \int_{[-1/2, 1/2]^k} \left| a \sum_{m=1}^k z_m + b \right|^p d\lambda^k(\mathbf{z})$$

for some constant $b' \in \mathbb{R}$ that depends on $b = b(\tilde{\mathbf{x}})$. Note that the right-hand side of the latter equality is minimized for $b' = 0$. Therefore we can estimate the left-hand side from below by

$$\begin{aligned} \int_{[0, l]^k} \left| a \sum_{m=1}^k y_m + b \right|^p d\lambda^k(\mathbf{y}) &\geq |a|^p \cdot l^{p+k} \cdot \int_{[-1/2, 1/2]^k} \left| \sum_{m=1}^k z_m \right|^p d\lambda^k(\mathbf{z}) \\ &= \int_{[0, l]^k} |a|^p d\lambda^k(\mathbf{y}) \cdot l^p \cdot \int_{[-1/2, 1/2]^k} \left| \sum_{m=1}^k z_m \right|^p d\lambda^k(\mathbf{z}). \end{aligned}$$

To complete the proof it remains to show that our choice of $k \geq \kappa_{p, l}$, with $\kappa_{p, l}$ given in (4.30), implies that

$$\int_{[-1/2, 1/2]^k} \left| \sum_{m=1}^k z_m \right|^p d\lambda^k(\mathbf{z}) \geq l^{-p} \quad (4.31)$$

but this easily follows from Lemma 4.14 above. \blacksquare

Actually, using other proof methods we can slightly improve the lower bound for C_p in Lemma 4.14 and thus also $\kappa_{p, l}$ in formula (4.30) of Proposition 4.15. This is the subject of our final remark within this subsection:

Remark 4.16. Let $\mathbf{Y} = (Y_1, \dots, Y_k)$ denote a random vector of $k \in \mathbb{N}$ independent copies of some uniformly $[-1/2, 1/2]$ -distributed random variable Y_0 . Then $\mathfrak{J}_{k, p}$ can be interpreted as the p th absolute moment $\mathbb{E}(|f_k(\mathbf{Y})|^p)$ of $f_k(\mathbf{Y})$, where f_k again is given by (4.29). In the case of even $p = 2N$, $N \in \mathbb{N}$, this can be calculated exactly using the multinomial theorem. For $k, N \in \mathbb{N}$ we obtain

$$\mathfrak{J}_{k, 2N} = \mathbb{E}(|f_k(\mathbf{Y})|^{2N}) = 2^{-2N} \sum_{\substack{j=(j_1, \dots, j_k) \in \mathbb{N}_0^k \\ j_1 + \dots + j_k = N}} \binom{2N}{2j_1, \dots, 2j_k} \prod_{m=1}^k \frac{1}{2j_m + 1},$$

where we used the independence of the Y_m 's and fact that

$$\mathbb{E}(Y_0^n) = \int_{-1/2}^{1/2} y^n d\lambda^1(y) = \begin{cases} 0, & \text{if } n = 2j + 1, \\ (2j + 1)^{-1} \cdot 2^{-2j}, & \text{if } n = 2j, \end{cases}$$

and $j \in \mathbb{N}_0$. In particular, we conclude

$$\mathfrak{J}_{k, 2} = \frac{1}{2^2 \cdot 3} \cdot k \quad \text{and} \quad \mathfrak{J}_{k, 4} = \frac{1}{48} \cdot k \left(k - \frac{2}{5} \right) \geq \frac{1}{2^4 \cdot 5} \cdot k^2.$$

Since $\mathfrak{J}_{k,p} = \left\| |f_k| \mathbf{L}_p([-1/2, 1/2]^k) \right\|^p$ we can use the monotonicity of the Lebesgue spaces in order to estimate C_p for the remaining powers p . For $k \in \mathbb{N}$ and $1 \leq q \leq p < \infty$ we obtain $\mathfrak{J}_{k,p} \geq (\mathfrak{J}_{k,q})^{p/q} \geq (C_q)^{p/q} k^{p/2}$, i.e. $C_p \geq (C_q)^{p/q}$, provided that $\mathfrak{J}_{k,q} \geq C_q k^{q/2}$. Consequently, we can take

$$k \geq \begin{cases} \lceil 12/l^2 \rceil, & \text{if } 2 \leq p < 4, \\ \lceil 4\sqrt{5}/l^2 \rceil, & \text{if } 4 \leq p \end{cases}$$

to fulfill (4.31) in the proof of Proposition 4.15. This clearly improves the bound $k \geq \kappa_{p,l}$ in (4.30). \square

Nevertheless, we want to stress the point that also with these improvements the lower bounds on k are not sharp since we know from [NW09] that in the limit case $p = \infty$ we can take $k = \lceil 2/l \rceil$. On the other hand, we note that Hoeffding's inequality implies the existence of some universal constants C'_p such that $\mathfrak{J}_{k,p} \leq C'_p k^{p/2}$ for all $p \in [1, \infty)$ and every $k \in \mathbb{N}$. Thus the estimates on the integrals $\mathfrak{J}_{k,p}$ are of the right order in k such that we need other proof techniques to obtain a better dependence of k on $l = c_2 - c_1$.

4.4.2 Uniform approximation in the weighted Sobolev space

To show that the linear algorithm $A_{n,d}^*$ studied in Corollary 4.4 is essentially optimal for L_∞ -approximation on the unanchored Sobolev space $\mathcal{H}_d^\gamma = \mathcal{H}(K_d^\gamma)$ in the worst case setting we study (weighted) L_2 -approximation on a related Banach space \mathcal{F}_d in the average case setting; see Example 2.15 for details. The relation of these two problems is given by the assertion below which follows from [KWW08, Theorem 1].

Proposition 4.17. For $d \in \mathbb{N}$ let $\mathcal{H}(K_d)$ denote a RKHS induced by a kernel $K_d: [0, 1]^d \times [0, 1]^d \rightarrow \mathbb{R}$ that satisfies (4.13).³² Moreover, define the set of non-vanishing probability density functions ϱ on the unit cube by

$$\mathcal{D}_d = \left\{ \varrho: [0, 1]^d \rightarrow [0, \infty) \left| \int_{[0,1]^d} \varrho(\mathbf{x}) \, d\lambda^d(\mathbf{x}) = 1 \quad \text{and} \quad \varrho > 0 \text{ } (\lambda^d\text{-a.e.}) \right. \right\}.$$

Then, for every $n \in \mathbb{N}_0$ and all $d \in \mathbb{N}$,

$$\begin{aligned} e^{\text{wor}}(n, d; \text{id}_d: \mathcal{B}(\mathcal{H}(K_d)) \rightarrow L_\infty([0, 1]^d)) \\ \geq \sup_{\varrho \in \mathcal{D}_d} e^{\text{avg}}(n, d; \text{id}_d^\varrho: \mathcal{F}_d \rightarrow L_2^\varrho([0, 1]^d)). \end{aligned}$$

Here the n th minimal errors are taken with respect to all algorithms from the class $\mathcal{A}_d^{n, \text{lin}}(\Lambda^{\text{all}})$.

³²Note that (4.13) clearly implies that $\int_{[0,1]^d} K_d(\mathbf{x}, \mathbf{x}) \varrho(\mathbf{x}) \, d\lambda^d(\mathbf{x})$ is finite for every probability density function ϱ on $[0, 1]^d$.

In particular, it follows that the (n th minimal) worst case error for L_∞ -approximation on the unit ball of the Sobolev space \mathcal{H}_d^γ is lower bounded by the average case error of *unweighted* L_2 -approximation on the corresponding Banach space. That is, we set $K_d = K_d^\gamma$ and $\varrho = \chi_{[0,1]^d} \in \mathcal{D}_d$ in the following.

In turn we have (strong) polynomial tractability for the uniform approximation problem w.r.t. the worst case setting only if average case L_2 -approximation is polynomially tractable, as long as we consider the absolute error criterion. Due to [NW08, Theorem 6.1] we know that the latter holds true if and only if there exist a positive constant c_1 , non-negative q_1, q_2 and $\tau \in (0, 1)$ such that

$$c_2 = \sup_{d \in \mathbb{N}} \frac{1}{d^{q_2}} \left(\sum_{i=\lceil c_1 d^{q_1} \rceil}^{\infty} (\lambda_{d,i})^\tau \right)^{1/\tau} < \infty,$$

where $(\lambda_{d,i})_{i=1}^\infty$ denotes the sequence of eigenvalues of the correlation operator C_{ν_d} with respect to a non-increasing ordering. Moreover we have strong polynomial tractability if and only if this holds with $q_1 = q_2 = 0$.

Because of the observation at the end of Example 2.15 it suffices to consider the eigenvalues of $W_d^\gamma = (S_d^\gamma)^\dagger S_d^\gamma: \mathcal{H}_d^\gamma \rightarrow \mathcal{H}_d^\gamma$, where S_d^γ describes the embedding $\mathcal{H}_d^\gamma \hookrightarrow L_2([0, 1]^d)$. Recall that these eigenvalues are given by

$$\left\{ \tilde{\lambda}_{d,\gamma,\mathbf{m}} = \prod_{k=1}^d \lambda_{1,\gamma_{d,k},m_k} = \prod_{k=1}^d \frac{\gamma_{d,k}}{\gamma_{d,k} + \pi^2(m_k - 1)^2} \mid \mathbf{m} = (m_1, \dots, m_d) \in \mathbb{N}^d \right\};$$

see (4.12) at the end of Example 4.1. Thus we only need to reorder this set appropriately using a rearrangement $\psi_d: \mathbb{N} \rightarrow \mathbb{N}^d$ such that

$$\lambda_{d,i} = \tilde{\lambda}_{d,\gamma,\psi_d(i)} \geq \tilde{\lambda}_{d,\gamma,\psi_d(i+1)} \quad \text{for all } i \in \mathbb{N}.$$

Given $d \in \mathbb{N}$, $\tau \in (0, 1)$, as well as $c_1 > 0$, and $q_1 \geq 0$ we estimate

$$\begin{aligned} \sum_{i=\lceil c_1 d^{q_1} \rceil}^{\infty} (\lambda_{d,i})^\tau &= \sum_{\mathbf{m} \in \mathbb{N}^d} \left(\tilde{\lambda}_{d,\gamma,\mathbf{m}} \right)^\tau - \sum_{i=1}^{\lceil c_1 d^{q_1} \rceil - 1} (\lambda_{d,i})^\tau \\ &\geq \prod_{k=1}^d \sum_{m \in \mathbb{N}} (\lambda_{1,\gamma_{d,k},m})^\tau - (\lambda_{d,1})^\tau (\lceil c_1 d^{q_1} \rceil - 1) \\ &\geq \prod_{k=1}^d \left(1 + \sum_{m=2}^{\infty} \left(\frac{\gamma_{d,k}}{\gamma_{d,k} + \pi^2(m-1)^2} \right)^\tau \right) - c_1 d^{q_1}, \end{aligned}$$

since $\lambda_{d,1} = \tilde{\lambda}_{d,\gamma,(1,\dots,1)} = \prod_{k=1}^d \lambda_{d,\gamma_{d,k},1} = 1$. Due to the boundedness of the generator weights $\gamma_{d,k} \leq C_\gamma$ for every $k \in \{1, \dots, d\}$, we can further estimate the sum by

$$\sum_{m=2}^{\infty} \left(\frac{\gamma_{d,k}}{\gamma_{d,k} + \pi^2(m-1)^2} \right)^\tau \geq \gamma_{d,k}^\tau \sum_{i=1}^{\infty} \frac{(C_\gamma')^\tau}{i^{2\tau}} = \gamma_{d,k}^\tau (C_\gamma')^\tau \zeta(2\tau),$$

where we set $C'_\gamma = (C_\gamma + \pi^2)^{-1}$. Because of $\ln(1 + y) \geq y/(1 + y)$ for all $y \geq 0$ we conclude that for $k = 1, \dots, d$ and some positive C depending on C_γ and τ

$$\ln \left(1 + \gamma_{d,k}^\tau (C'_\gamma)^\tau \zeta(2\tau) \right) \geq \frac{(C'_\gamma)^\tau \zeta(2\tau)}{1 + \gamma_{d,k}^\tau (C'_\gamma)^\tau \zeta(2\tau)} \cdot \gamma_{d,k}^\tau \geq C \cdot \gamma_{d,k}^\tau.$$

Consequently, this yields

$$\sum_{i=\lceil c_1 d^{q_1} \rceil}^{\infty} (\lambda_{d,i})^\tau \geq \prod_{k=1}^d \exp(C \cdot \gamma_{d,k}^\tau) - c_1 d^{q_1} = \exp\left(C \sum_{k=1}^d \gamma_{d,k}^\tau\right) - c_1 d^{q_1}.$$

Therefore, polynomial tractability implies $q(\gamma) < 1$ and strong polynomial tractability is possible only if $p(\gamma) < 1$. Here p and q describe the sum exponents of the product weight sequence $\gamma = (\gamma_\alpha)_{\alpha \in \mathbb{N}_0^d}$, $d \in \mathbb{N}$, defined at the beginning of [Section 4.3.2](#).

Together with [Proposition 4.7](#) this finally proves

Theorem 4.18. Consider the uniform approximation problem defined on the sequence of unanchored Sobolev spaces $(\mathcal{H}_d^\gamma)_{d \in \mathbb{N}}$, where the product weight sequence γ is constructed out of a uniformly bounded generator sequence $C_\gamma \geq \gamma_{d,1} \geq \dots \geq \gamma_{d,d}$, $d \in \mathbb{N}$. We study this problem in the worst case setting and with respect to the absolute error criterion. Then we have

- polynomial tractability if and only if $q(\gamma) < 1$ and
- strong polynomial tractability if and only if $p(\gamma) < 1$.

CHAPTER 5

Problems on Hilbert spaces with (anti)symmetry conditions

In this last chapter we describe an essentially new kind of a priori knowledge which can help to overcome the curse of dimensionality. As in [Section 2.4](#), we study compact linear problems $S = (S_d)_{d \in \mathbb{N}}$ defined between tensor products of Hilbert spaces but now we restrict our attention to problem elements which fulfill certain (anti)symmetry conditions. After investigating some basic properties of the related subspaces of (anti)symmetric problem elements in [Section 5.1](#) we construct a linear algorithm that uses finitely many continuous linear functionals and show an explicit formula for its worst case error in terms of the eigenvalues $\lambda = (\lambda_m)_{m \in \mathbb{N}}$ of the operator $W_1 = S_1^\dagger S_1$. Moreover, in [Section 5.2](#) we show that this algorithm is optimal w.r.t. a wide class of algorithms. Next we clarify the influence of different (anti)symmetry conditions on the complexity, compared to the case for the classical unrestricted problem studied in [Section 2.4.3](#). In particular, we give necessary and sufficient conditions for (strong) polynomial tractability of (anti)symmetric problems in [Section 5.3](#). Apart from the absolute error criterion we also deal with normalized errors. Finally, in [Section 5.4](#), we discuss several applications. [Section 5.4.2](#) particularly indicates how to apply our results to the approximation problem for wavefunctions.

Most of the results stated in this chapter are already published in the articles [\[Wei11\]](#) and [\[Wei12a\]](#). At some points we improve the known results and/or proof techniques slightly. In particular, the presented results also hold for problems defined on finite-dimensional or on non-separable source spaces.

5.1 Basic definitions related to (anti)symmetry

The aim of this section is to introduce the notion of (anti)symmetry in Hilbert spaces. In order to illustrate this concept we mainly deal with *function spaces*. For this purpose in [Section 5.1.1](#) we start by defining (anti)symmetry properties for functions which will lead us to orthogonal projections, mapping the whole space onto its subspace of (anti)symmetric functions. In [Section 5.1.2](#) it will turn out that these projections applied to a given basis of a tensor product Hilbert function

space lead us to handsome formulas for orthonormal bases of the subspaces. Finally we generalize our approach and define (anti)symmetry conditions for *arbitrary* tensor product Hilbert spaces based on the deduced results for function spaces. Section 5.1.3 is devoted to this generalization.

5.1.1 Hilbert function spaces

Following Hamaekers [Ham09, Section 2.5] we use a general approach to (anti)symmetric functions which also can be found in [Wei12a]. Consider H to be a (possibly non-separable) Hilbert space of real-valued multivariate functions f defined on some domain Ω in \mathbb{R}^d , where we assume $d \geq 2$ to be fixed. Furthermore, take an arbitrary non-empty subset of coordinates $I \subseteq \{1, \dots, d\}$. For every such subset we define the set

$$\mathcal{S}_I = \left\{ \pi: \{1, \dots, d\} \rightarrow \{1, \dots, d\} \mid \pi \text{ bijective and } \pi|_{\{1, \dots, d\} \setminus I} = \text{id} \right\} \quad (5.1)$$

of all permutations on $\{1, \dots, d\}$ that leave the complement of I fixed. To abbreviate the notation we identify $\pi \in \mathcal{S}_I$ with the corresponding permutation π' on \mathbb{R}^d ,

$$\pi': \mathbb{R}^d \rightarrow \mathbb{R}^d, \quad \mathbf{x} = (x_1, \dots, x_d) \mapsto \pi'(\mathbf{x}) = (x_{\pi(1)}, \dots, x_{\pi(d)}).$$

For an appropriate definition of partial (anti)symmetry of functions $f \in H$ we need the following simple assumptions. Given any $\pi \in \mathcal{S}_I$ we assume that

$$(A5.1) \quad \mathbf{x} \in \Omega \text{ implies } \pi(\mathbf{x}) \in \Omega,$$

$$(A5.2) \quad f \in H \text{ implies } f(\pi(\cdot)) \in H \text{ and}$$

$$(A5.3) \quad \text{there ex. } c_\pi \geq 0 \text{ (independent of } f) \text{ such that } \|f(\pi(\cdot))\|_H \leq c_\pi \|f\|_H.$$

A function $f \in H$ is called *partially symmetric w.r.t. I* (or *I -symmetric* for short) if any permutation $\pi \in \mathcal{S}_I$ applied to the argument \mathbf{x} does not affect the value of f . Hence,

$$f(\mathbf{x}) = f(\pi(\mathbf{x})) \quad \text{for all } \mathbf{x} \in \Omega \quad \text{and every } \pi \in \mathcal{S}_I. \quad (5.2)$$

Moreover, we call a function $f \in H$ *partially antisymmetric w.r.t. I* (or *I -antisymmetric*, respectively) if f changes its sign by exchanging the variables x_i and x_j with each other, where $i, j \in I$. That is, we have

$$f(\mathbf{x}) = (-1)^{|\pi|} f(\pi(\mathbf{x})) \quad \text{for all } \mathbf{x} \in \Omega \quad \text{and every } \pi \in \mathcal{S}_I, \quad (5.3)$$

where $|\pi|$ denotes the *inversion number* of the permutation π . The term $(-1)^{|\pi|}$ therefore coincides with the *sign*, or *parity* of π and is equal to the determinant of the associated permutation matrix. In the case $\#I = 1$ we do not claim any (anti)symmetry, since then the set $\mathcal{S}_I = \{\text{id}\}$ is trivial. For $I = \{1, \dots, d\}$ functions f which satisfy (5.2) or (5.3), respectively, are called *fully (anti)symmetric*.

Note that, in particular, formula (5.3) yields that the value $f(\mathbf{x})$ of I -antisymmetric functions f equals zero if $x_i = x_j$ with $i \neq j$ and $i, j \in I$. For I -symmetric functions such an implication does not hold. Therefore the (partial) antisymmetry property is a somewhat more restrictive condition than the (partial) symmetry property with respect to the same subset I . As we will see in Section 5.3 this will also affect our complexity estimates.

Next we define the so-called *symmetrizer* \mathfrak{S}_I^H and *antisymmetrizer* \mathfrak{A}_I^H on H with respect to the subset I by

$$\mathfrak{S}_I^H : H \rightarrow H, \quad f \mapsto \mathfrak{S}_I^H(f) = \frac{1}{\#\mathcal{S}_I} \sum_{\pi \in \mathcal{S}_I} f(\pi(\cdot))$$

and

$$\mathfrak{A}_I^H : H \rightarrow H, \quad f \mapsto \mathfrak{A}_I^H(f) = \frac{1}{\#\mathcal{S}_I} \sum_{\pi \in \mathcal{S}_I} (-1)^{|\pi|} f(\pi(\cdot)).$$

If there is no danger of confusion we use the notation \mathfrak{S}_I and \mathfrak{A}_I instead of \mathfrak{S}_I^H and \mathfrak{A}_I^H , respectively. The following lemma collects some basic properties. It generalizes Lemma 10.1 in Zeiser [Zei10].

Lemma 5.1. For $\emptyset \neq I \subseteq \{1, \dots, d\}$ both the mappings $P_I \in \{\mathfrak{S}_I, \mathfrak{A}_I\}$ define bounded linear operators on the Hilbert space H with $P_I^2 = P_I$. Thus, \mathfrak{S}_I and \mathfrak{A}_I provide projections of H onto the closed linear subspaces

$$\mathfrak{S}_I(H) = \{f \in H \mid f \text{ satisfies (5.2)}\} \quad \text{and} \quad \mathfrak{A}_I(H) = \{f \in H \mid f \text{ satisfies (5.3)}\} \quad (5.4)$$

of all partially (anti)symmetric functions w.r.t. I in H , respectively. If, in addition,

$$\langle f(\pi(\cdot)), g(\pi(\cdot)) \rangle_H = \langle f, g \rangle_H \quad \text{for all } f, g \in H \quad \text{and every } \pi \in \mathcal{S}_I \quad (5.5)$$

then the operators are self-adjoint and hence the projections are orthogonal. Consequently,

$$H = \mathfrak{S}_I(H) \oplus (\mathfrak{S}_I(H))^\perp = \mathfrak{A}_I(H) \oplus (\mathfrak{A}_I(H))^\perp. \quad (5.6)$$

Proof. Obviously $P_I \in \{\mathfrak{S}_I, \mathfrak{A}_I\}$ is well-defined due to the assumptions (A5.1) and (A5.2). The linearity directly follows from the definition and, using (A5.3), we see that the operator norm of P_I is bounded by $\max\{c_\pi \mid \pi \in \mathcal{S}_I\}$.

To show that the operators are idempotent, i.e. that $P_I^2 = P_I$, we first prove that $\mathfrak{A}_I(f)$ satisfies (5.3) for every $f \in H$. Therefore, we use the representation

$$\begin{aligned} (\mathfrak{A}_I(f))(\pi(\cdot)) &= \frac{1}{\#\mathcal{S}_I} \sum_{\sigma \in \mathcal{S}_I} (-1)^{|\sigma|} f(\sigma(\pi(\cdot))) = \frac{1}{\#\mathcal{S}_I} \sum_{\lambda \in \mathcal{S}_I} (-1)^{|\lambda|+|\pi|} f(\lambda(\cdot)) \\ &= (-1)^{|\pi|} (\mathfrak{A}_I(f))(\cdot) \end{aligned}$$

for every fixed $\pi \in \mathcal{S}_I$. Here we imposed $\lambda = \sigma \circ \pi \in \mathcal{S}_I$ and used that

$$|\lambda \circ \pi^{-1}| = |\lambda| + |\pi^{-1}| = |\lambda| + |\pi|.$$

Hence we have shown $\mathfrak{A}_I(H) \subseteq \{f \in H \mid f \text{ satisfies (5.3)}\}$. In a second step, it is easy to check that for every function $g \in H$ which satisfies (5.3) it is $\mathfrak{A}_I(g) = g$. Thus, $\{f \in H \mid f \text{ satisfies (5.3)}\} \subseteq \mathfrak{A}_I(H)$ and \mathfrak{A}_I is a projector onto $\mathfrak{A}_I(H)$. Since the same arguments also apply for the symmetrizer \mathfrak{S}_I this shows (5.4), as well as $P_I^2 = P_I$ for $P_I \in \{\mathfrak{S}_I, \mathfrak{A}_I\}$.

To prove the self-adjointness of P_I we need to show that for f and g in H we have $\langle P_I f, g \rangle_H = \langle f, P_I g \rangle_H$. To this end, note that (5.5) is equivalent to the fact that

$$\langle f(\pi(\cdot)), g \rangle_H = \langle f, g(\sigma(\cdot)) \rangle_H, \quad f, g \in H, \quad \pi \in \mathcal{S}_I,$$

where we set $\sigma = \pi^{-1}$ and used (A5.2). Now the claimed assertion follows from the bilinearity of the inner product $\langle \cdot, \cdot \rangle_H$. Moreover, orthogonality and the decompositions stated in (5.6) are simple consequences. \blacksquare

We note in passing that (5.5) already implies (A5.3). Furthermore, the notion of partially (anti)symmetric functions can be easily extended to more than one subset I . Therefore, consider two non-empty subsets of coordinates $I, J \subset \{1, \dots, d\}$ with $I \cap J = \emptyset$. Then we call a function $f \in H$ *multiple partially (anti)symmetric w.r.t. I and J* if f satisfies (5.2), or (5.3), respectively, for I and J . Since I and J are disjoint we observe that $\pi \circ \sigma = \sigma \circ \pi$ for all $\pi \in \mathcal{S}_I$ and $\sigma \in \mathcal{S}_J$. Hence the linear projections $P_I \in \{\mathfrak{S}_I, \mathfrak{A}_I\}$ and $P_J \in \{\mathfrak{S}_J, \mathfrak{A}_J\}$ commute on H . That is, we have $P_I \circ P_J = P_J \circ P_I$. Further extensions to more than two disjoint subsets of coordinates are possible. We will restrict ourselves to the case of at most two coordinate subsets, because in particular wavefunctions can be modeled as functions which are antisymmetric w.r.t. I and $J = I^c$, where I^c denotes the complement of I in $\{1, \dots, d\}$; see, e.g., Section 5.4.2.

5.1.2 Tensor products of Hilbert function spaces

In the previous subsection the function space H was a somewhat abstract Hilbert space of d -variate real-valued functions. Indeed, for the definition of (anti)symmetry we do not need to claim any product structure. On the other hand, it is also motivated by applications to consider tensor product function spaces; see, e.g., Section 3.6 in Yserentant [Yse10]. In detail, it is well-known that so-called spaces of dominated mixed smoothness, e.g. $W_2^{(1, \dots, 1)}(\mathbb{R}^{3d})$, can be represented as certain tensor products; see Section 1.4.2 in Hansen [Han10].

Anyway, let us take into account such a structure, i.e. let us assume that

$$H = H_d = H_1 \otimes \dots \otimes H_1 \quad (d \geq 2 \text{ times}),$$

where H_1 is a suitable Hilbert space of functions $f: D \rightarrow \mathbb{R}$; see also the constructions given in [Section 2.4.1](#). There it is stated that we can construct an orthonormal basis E_d of H_d out of a given ONB E_1 of H_1 ; see [\(2.13\)](#). Since now we deal with function spaces, the d -fold simple tensors in E_d are d -variate functions $e_{d,\mathbf{j}}: D^d \rightarrow \mathbb{R}$. More precisely, they are given by

$$e_{d,\mathbf{j}}(\mathbf{x}) = \prod_{l=1}^d e_{j_l}(x_l), \quad \text{where } \mathbf{x} = (x_1, \dots, x_d) \in D^d \quad \text{and} \quad \mathbf{j} \in \mathcal{J}_d = (\mathcal{J}_1)^d,$$

provided that $E_1 = \{e_m: D \rightarrow \mathbb{R} \mid m \in \mathcal{J}_1\}$ denotes the underlying ONB in H_1 . To exploit this representation we start with a simple observation.

Let $d \in \mathbb{N}$. Moreover assume $\mathbf{j} \in \mathcal{J}_d$ and $\mathbf{x} \in D^d$, as well as a non-empty subset I of $\{1, \dots, d\}$, to be arbitrarily fixed. If we define $\sigma = \pi^{-1} \in \mathcal{S}_I$ then

$$e_{d,\mathbf{j}}(\boldsymbol{\pi}(\mathbf{x})) = \prod_{l=1}^d e_{j_l}(x_{\pi(l)}) = \prod_{l=1}^d e_{j_{\sigma(l)}}(x_l) = e_{d,\boldsymbol{\sigma}(\mathbf{j})}(\mathbf{x}). \quad (5.7)$$

For simplicity, once again we identified $\boldsymbol{\pi}(\mathbf{j}) = \pi(j_1, \dots, j_d)$ with $(j_{\pi(1)}, \dots, j_{\pi(d)})$ for $\mathbf{j} \in \mathcal{J}_d = (\mathcal{J}_1)^d$. Since $\mathbf{x} \in D^d$ was arbitrary and $|\boldsymbol{\pi}| = |\pi^{-1}| = |\sigma|$ we obtain

$$\mathfrak{S}_I e_{d,\mathbf{j}} = \frac{1}{\#\mathcal{S}_I} \sum_{\sigma \in \mathcal{S}_I} e_{d,\boldsymbol{\sigma}(\mathbf{j})} \quad \text{and} \quad \mathfrak{A}_I e_{d,\mathbf{j}} = \frac{1}{\#\mathcal{S}_I} \sum_{\sigma \in \mathcal{S}_I} (-1)^{|\sigma|} e_{d,\boldsymbol{\sigma}(\mathbf{j})} \quad (5.8)$$

for all $\mathbf{j} \in \mathcal{J}_d$. Besides this, [\(5.7\)](#) can be used to verify that [\(5.5\)](#) in [Lemma 5.1](#) always holds true for (unweighted) tensor products of Hilbert function spaces.

Note that in general, i.e. for arbitrary $\mathbf{j} \in \mathcal{J}_d$ and $\sigma \in \mathcal{S}_I$, the tensor products $e_{d,\boldsymbol{\sigma}(\mathbf{j})}$ and $e_{d,\mathbf{j}}$ do not coincide, because taking the tensor product is not commutative in general. Therefore \mathfrak{S}_I is not simply the identity on the set of basis functions $E_d = \{e_{d,\mathbf{j}} \mid \mathbf{j} \in \mathcal{J}_d\}$. On the other hand, we see that for different $\mathbf{j} \in \mathcal{J}_d$ many of the functions $\mathfrak{S}_I e_{d,\mathbf{j}}$ coincide. Of course the same holds true for $\mathfrak{A}_I e_{d,\mathbf{j}}$, at least up to a factor of (-1) .

We will see in the following that for $P_I \in \{\mathfrak{S}_I, \mathfrak{A}_I\}$ a linearly independent subset of all projections $\{P_I e_{d,\mathbf{j}} \mid \mathbf{j} \in \mathcal{J}_d\}$ equipped with suitable normalizing constants can be used as an ONB of the linear subspace $P_I(H_d)$ of I -(anti)symmetric functions in H_d . For the application we have in mind, we need this result only in the case where the underlying space H_1 is separable. Without loss of generality, we can thus assume that³³

$$\mathcal{J}_1 = \mathcal{M}_1 = \{m \in \mathbb{N} \mid m < \dim H_1 + 1\}$$

and consequently $\mathcal{J}_d = \mathcal{M}_d = (\mathcal{M}_1)^d \subseteq \mathbb{N}^d$. Clearly, in the most interesting case the set \mathcal{J}_d equals \mathbb{N}^d .

³³Note that also the case of abstract, countable index sets \mathcal{J}_1 can be reduced to this form by the application of some simple isomorphism.

To state the claimed assertion, we need a further definition. For fixed $d \geq 2$ and $I \subseteq \{1, \dots, d\}$, let us introduce a function

$$M_I = M_{I,d}: \mathbb{N}^d \rightarrow \{0, \dots, \#I\}^{\#I}$$

which counts how often different indices occur in a given multi-index $\mathbf{j} \in \mathbb{N}^d$ among the subset I of coordinates, ordered with respect to their rate. To give an example let $d = 7$ and $I = \{1, \dots, 6\}$. Then $M_{I,7}$ applied to $\mathbf{j} = (12, 4, 4, 12, 6, 4, 4) \in \mathbb{N}^7$ gives the $\#I = 6$ dimensional vector $\mathbf{M}_{I,7}(\mathbf{j}) = (3, 2, 1, 0, 0, 0)$, because \mathbf{j} contains the number “4” three times among the coordinates j_1, \dots, j_6 , “12” two times, and so on. Since in this example there are only three different numbers involved, the fourth to sixth coordinates of $\mathbf{M}_{I,7}(\mathbf{j})$ equal zero. Obviously, M_I is invariant under all permutations $\pi \in \mathcal{S}_I$ of the argument. Thus,

$$\mathbf{M}_I(\mathbf{j}) = \mathbf{M}_I(\pi(\mathbf{j})) \quad \text{for all } \mathbf{j} \in \mathbb{N}^d \quad \text{and} \quad \pi \in \mathcal{S}_I.$$

In addition, since $\mathbf{M}_I(\mathbf{j})$ again is a multi-index, we see that $|\mathbf{M}_I(\mathbf{j})| = \#I$ and $\mathbf{M}_I(\mathbf{j})!$ are well-defined for every $\mathbf{j} \in \mathbb{N}^d$. Prepared with this tool, we are ready to prove the following

Lemma 5.2. Assume $E_d = \{e_{d,\mathbf{j}} \mid \mathbf{j} \in \mathcal{M}_d\}$ to be a given orthonormal tensor product basis in the space H_d and let $\emptyset \neq I = \{i_1, \dots, i_{\#I}\} \subseteq \{1, \dots, d\}$. Moreover, for $P_I \in \{\mathfrak{S}_I, \mathfrak{A}_I\}$ define the functions $\xi_{\mathbf{j}}: D^d \rightarrow \mathbb{R}$ by

$$\xi_{\mathbf{j}} = \sqrt{\frac{\#\mathcal{S}_I}{\mathbf{M}_I(\mathbf{j})!}} \cdot P_I(e_{d,\mathbf{j}}) \quad \text{for } \mathbf{j} \in \mathcal{M}_d.$$

Then the set $\Xi_d = \{\xi_{\mathbf{k}} \mid \mathbf{k} \in \nabla_d\}$ builds an orthonormal basis of the partially (anti)symmetric subspace $P_I(H_d)$, where ∇_d is given by

$$\nabla_d = \begin{cases} \{\mathbf{k} \in \mathcal{M}_d \mid k_{i_1} \leq k_{i_2} \leq \dots \leq k_{i_{\#I}}\}, & \text{if } P_I = \mathfrak{S}_I, \\ \{\mathbf{k} \in \mathcal{M}_d \mid k_{i_1} < k_{i_2} < \dots < k_{i_{\#I}}\}, & \text{if } P_I = \mathfrak{A}_I. \end{cases} \quad (5.9)$$

Proof. To abbreviate the notation, we suppress the index H_d at the inner products $\langle \cdot, \cdot \rangle_{H_d}$ in this proof.

Step 1. We start by proving orthonormality. Therefore let us recall (5.8) and remember that now $\mathcal{J}_d = \mathcal{M}_d$. For $P_I = \mathfrak{A}_I$ and $\mathbf{j}, \mathbf{k} \in \nabla_d$ easy calculations yield

$$\begin{aligned} \langle \xi_{\mathbf{j}}, \xi_{\mathbf{k}} \rangle &= \frac{\#\mathcal{S}_I}{\sqrt{\mathbf{M}_I(\mathbf{j})! \cdot \mathbf{M}_I(\mathbf{k})!}} \langle \mathfrak{A}_I(e_{d,\mathbf{j}}), \mathfrak{A}_I(e_{d,\mathbf{k}}) \rangle \\ &= \frac{1}{\#\mathcal{S}_I \sqrt{\mathbf{M}_I(\mathbf{j})! \cdot \mathbf{M}_I(\mathbf{k})!}} \sum_{\pi, \sigma \in \mathcal{S}_I} (-1)^{|\pi|+|\sigma|} \langle e_{d,\pi(\mathbf{j})}, e_{d,\sigma(\mathbf{k})} \rangle. \end{aligned}$$

Of course, up to the factor controlling the sign, the same is true for the case $P_I = \mathfrak{S}_I$. Now assume that there exists $l \in \{1, \dots, d\}$ such that $j_l \neq k_l$. Then the

ordering of $\mathbf{j}, \mathbf{k} \in \nabla_d$ implies that $\boldsymbol{\pi}(\mathbf{j}) \neq \boldsymbol{\sigma}(\mathbf{k})$ for all $\sigma, \pi \in \mathcal{S}_I$, since π and σ leave the coordinates $l \in I^c$ fixed. Hence, we conclude that we have $\boldsymbol{\pi}(\mathbf{j}) = \boldsymbol{\sigma}(\mathbf{k})$ only if $\mathbf{j} = \mathbf{k}$.

At this point we have to distinguish the antisymmetric and the symmetric case. For $P_I = \mathfrak{A}_I$ the only way to conclude $\boldsymbol{\pi}(\mathbf{j}) = \boldsymbol{\sigma}(\mathbf{k})$ is to claim $\mathbf{j} = \mathbf{k}$ and $\pi = \sigma$. Furthermore we see that in the antisymmetric case we have $M_I(\mathbf{j})! = 1$ for all $\mathbf{j} \in \nabla_d$, because then all coordinates j_l , where $l \in I$, differ. Therefore, in this case the last inner product coincides with $\delta_{\mathbf{j}, \mathbf{k}} \cdot \delta_{\pi, \sigma}$ because of the mutual orthonormality of the elements from $E_d = \{e_{d, \mathbf{j}} \mid \mathbf{j} \in \mathcal{M}_d\}$. Hence we arrive at

$$\langle \xi_{\mathbf{j}}, \xi_{\mathbf{k}} \rangle = \frac{1}{\#\mathcal{S}_I} \sum_{\pi \in \mathcal{S}_I} (-1)^{2|\pi|} \delta_{\mathbf{j}, \mathbf{k}} = \delta_{\mathbf{j}, \mathbf{k}} \quad \text{for all } \mathbf{j}, \mathbf{k} \in \nabla_d,$$

as claimed.

So, let us consider the case $P_I = \mathfrak{S}_I$ and $\mathbf{j} = \mathbf{k} \in \nabla_d$, since we already saw that otherwise $\langle \xi_{\mathbf{j}}, \xi_{\mathbf{k}} \rangle$ equals zero. Then for fixed $\sigma \in \mathcal{S}_I$ there are $M_I(\mathbf{j})!$ different permutations $\pi \in \mathcal{S}_I$ such that $\boldsymbol{\pi}(\mathbf{j}) = \boldsymbol{\sigma}(\mathbf{j})$. This leads to

$$\langle \xi_{\mathbf{j}}, \xi_{\mathbf{j}} \rangle = \frac{1}{\#\mathcal{S}_I \cdot M_I(\mathbf{j})!} \sum_{\sigma \in \mathcal{S}_I} M_I(\mathbf{j})! = 1$$

and completes the proof of orthonormality.

Step 2. It remains to show that the span of $\Xi_d = \{\xi_{\mathbf{k}} \mid \mathbf{k} \in \nabla_d\}$ is dense in $P_I(H_d)$ for $P_I \in \{\mathfrak{S}_I, \mathfrak{A}_I\}$. Note that every multi-index $\mathbf{j} \in \mathcal{M}_d$ can be represented by a uniquely defined multi-index $\mathbf{k} \in \nabla_d$ and exactly $M_I(\mathbf{k})!$ different permutations $\pi \in \mathcal{S}_I$ such that $\mathbf{j} = \boldsymbol{\pi}(\mathbf{k})$. Assume that $f \in \mathfrak{A}_I(H_d)$, i.e. $f \in H_d$ satisfies (5.3). Then (5.7) together with (5.5) yields

$$\langle f, e_{d, \mathbf{j}} \rangle = (-1)^{|\pi|} \cdot \langle f, e_{d, \boldsymbol{\pi}(\mathbf{k})} \rangle \quad \text{for all } \mathbf{j} \in \mathcal{M}_d \quad \text{and} \quad \pi \in \mathcal{S}_I. \quad (5.10)$$

Now expanding f with respect to the basis functions in $E_d \subset H_d$ gives

$$\begin{aligned} f &= \sum_{\mathbf{j} \in \mathcal{M}_d} \langle f, e_{d, \mathbf{j}} \rangle e_{d, \mathbf{j}} = \sum_{\mathbf{k} \in \nabla_d} \sum_{\pi \in \mathcal{S}_I} \frac{\langle f, e_{d, \boldsymbol{\pi}(\mathbf{k})} \rangle e_{d, \boldsymbol{\pi}(\mathbf{k})}}{M_I(\mathbf{k})!} \\ &= \sum_{\mathbf{k} \in \nabla_d} \frac{1}{M_I(\mathbf{k})!} \sum_{\pi \in \mathcal{S}_I} (-1)^{|\pi|} \langle f, e_{d, \mathbf{k}} \rangle e_{d, \boldsymbol{\pi}(\mathbf{k})} \\ &= \sum_{\mathbf{k} \in \nabla_d} \sqrt{\frac{\#\mathcal{S}_I}{M_I(\mathbf{k})!}} \cdot \langle f, e_{d, \mathbf{k}} \rangle \cdot \sqrt{\frac{\#\mathcal{S}_I}{M_I(\mathbf{k})!}} \cdot \mathfrak{A}_I(e_{d, \mathbf{k}}), \end{aligned}$$

where we used (5.8) for the last equality. Furthermore, due to the self-adjointness of \mathfrak{A}_I , we have $\langle f, e_{d, \mathbf{k}} \rangle = \langle \mathfrak{A}_I f, e_{d, \mathbf{k}} \rangle = \langle f, \mathfrak{A}_I e_{d, \mathbf{k}} \rangle$, such that finally $f \in \mathfrak{A}_I(H_d)$ possesses the representation

$$f = \sum_{\mathbf{k} \in \nabla_d} \langle f, \xi_{\mathbf{k}} \rangle \cdot \xi_{\mathbf{k}}$$

since $\xi_{\mathbf{k}} = \sqrt{\#\mathcal{S}_I/\mathbf{M}_I(\mathbf{k})!} \cdot \mathfrak{A}_I(e_{d,\mathbf{k}})$ per definition. This proves the assertion for the case $P_I = \mathfrak{A}_I$. The remaining case $P_I = \mathfrak{S}_I$ can be treated in the same way. ■

Observe that in the antisymmetric case the definition of $\xi_{\mathbf{j}}$ for $\mathbf{j} \in \nabla_d$ simplifies, since then $\mathbf{M}_I(\mathbf{j})! = 1$ for all $\mathbf{j} \in \nabla_d$. Moreover we see that in this case ∇_d is trivial if $d > \#\mathcal{M}_1$. Hence we should assume that $\dim H_1$ is infinite in order to work with antisymmetric tensor products for arbitrarily many building blocks. We note in passing that the square of the normalizing factor, $\#\mathcal{S}_I/\mathbf{M}_I(\mathbf{j})!$, coincides with the multinomial coefficient $\binom{|\mathbf{M}_I(\mathbf{j})|}{\mathbf{M}_I(\mathbf{j})}$ which is quite natural due to combinatorial issues. Furthermore, in the special case $I = \{1, 2, \dots, \#I\}$ we have

$$P_I(H_d) = P_I\left(\bigotimes_{m \in I} H_1\right) \otimes \left(\bigotimes_{m \notin I} H_1\right).$$

That is, we can consider the subspace of I -(anti)symmetric functions $f \in H_d$ as the tensor product of the set of all fully (anti)symmetric $\#I$ -variate functions with the $(d - \#I)$ -fold tensor product of H_1 . If $\#I = 1$, i.e. if we do not claim any (anti)symmetry, then $P_I(H_d) = H_d$ and thus we have $\nabla_d = \mathcal{M}_d$, as well as $\Xi_d = E_d$. Modifications in connection with multiple partially (anti)symmetric functions are obvious.

5.1.3 Arbitrary tensor product Hilbert spaces

Up to now we exclusively dealt with Hilbert *function* spaces. However, the proofs of Lemma 5.1 and Lemma 5.2 yield that there are only a few key arguments in connection with (anti)symmetry such that we can cut out this restriction. We briefly sketch the points which need to be changed.

Starting from the very beginning we have to adapt the definition of I -(anti)symmetry due to (5.2) and (5.3) in Section 5.1.1. Of course it is sufficient to define this property at first only for basis elements. Therefore, if $E_d = \{e_{d,\mathbf{k}} \mid \mathbf{k} \in (\mathcal{J}_1)^d = \mathcal{J}_d\}$ denotes a tensor product ONB of H_d and $\emptyset \neq I \subseteq \{1, \dots, d\}$ is given then we call an element $e_{d,\mathbf{k}} = \bigotimes_{l=1}^d e_{k_l}$ *partially symmetric with respect to I (I -symmetric)*, if

$$e_{d,\mathbf{k}} = e_{d,\pi(\mathbf{k})} \quad \text{for all } \pi \in \mathcal{S}_I,$$

where \mathcal{S}_I and $\pi(\mathbf{k}) = (k_{\pi(1)}, \dots, k_{\pi(d)}) \in \mathcal{J}_d$ are defined as before. Analogously, we define I -*antisymmetry* with an additional factor $(-1)^{|\pi|}$. Finally, an arbitrary element in H_d is called I -(anti)symmetric if in its basis expansion every element with non-vanishing coefficient possesses this property.³⁴

Next, the *antisymmetrizer* \mathfrak{A}_I is given as the uniquely defined continuous extension of the linear mapping

$$\tilde{\mathfrak{A}}_I: E_d \rightarrow H_d, \quad e_{d,\mathbf{k}} \mapsto \frac{1}{\#\mathcal{S}_I} \sum_{\pi \in \mathcal{S}_I} (-1)^{|\pi|} e_{d,\pi(\mathbf{k})} \quad (5.11)$$

³⁴Note that even in the non-separable case any such expansion only has countably many terms.

from E_d to H_d . Again the *symmetrizer* \mathfrak{S}_I is given in a similar way. Hence, in the general setting we define the mappings using formula (5.8) which we derived for the special case of function spaces. Note that the triangle inequality yields $\|P_I\| \leq 1$, for $P_I \in \{\mathfrak{S}_I, \mathfrak{A}_I\}$.

Once more we denote the sets of all I -(anti)symmetric elements of H_d by $P_I(H_d)$, where $P_I \in \{\mathfrak{S}_I, \mathfrak{A}_I\}$. Observe that this can be justified since the operators P_I again provide orthogonal projections onto closed linear subspaces. That is, a generalization of Lemma 5.1 remains valid also in the more general case of tensor products of arbitrary Hilbert space which we consider here. This can be shown using (5.11) and its analogue for \mathfrak{S}_I , as well as with the help of some simple extension arguments. Moreover, also the proof of Lemma 5.2 can be adapted to the generalized setting. Indeed, the only difference is the conclusion of formula (5.10) in Step 2. Now, for arbitrary Hilbert spaces, this simply follows from our definitions. Then the rest of the proof transfers literally.

Finally and without going into details, we stress the point that further generalizations are possible. Here we can think of tensor products of arbitrary Hilbert spaces with multiple partial (anti)symmetry conditions or of scaled tensor products in the sense of Chapter 3. Since the corresponding calculations are straightforward we leave them to the reader.

5.2 Optimal algorithms for (anti)symmetric problems

Keeping the definitions and assertions from the previous Section 5.1 in mind, we are ready to study algorithms for linear problems defined on (anti)symmetric subsets of tensor product Hilbert spaces.

Let $S_d: H_d \rightarrow \mathcal{G}_d$ denote a tensor product problem in the sense of Section 2.4. It is constructed out of a compact linear operator $S_1: H_1 \rightarrow \mathcal{G}_1$ between arbitrary Hilbert spaces H_1 and \mathcal{G}_1 via a tensor product construction; see Section 2.4.1. Hence, let $H_d = H_1 \otimes \dots \otimes H_1$ in what follows and refer to the problem of approximating $S = (S_d)_{d \in \mathbb{N}}$ as the *entire d -variate problem*. Note that we completely solved this problem in Section 2.4. In detail, the n th optimal algorithm $A_{n,d}^*$, given by (2.16), was related to a certain subset $\{e_{d,j} = \tilde{\phi}_{d,j} \mid j \in \mathcal{M}_d\}$ of a tensor product ONB.

In contrast, now we are interested in the approximation of the restriction

$$S_{d,I_d} = S_d|_{P_{I_d}(H_d)}: P_{I_d}(H_d) \rightarrow \mathcal{G}_d$$

of S_d to some (anti)symmetric subspace $P_{I_d}(H_d)$ as defined in Section 5.1.3, where $P_{I_d} \in \{\mathfrak{S}_{I_d}, \mathfrak{A}_{I_d}\}$ and $\emptyset \neq I_d \subseteq \{1, \dots, d\}$ for $d \in \mathbb{N}$. We refer to $S_I = (S_{d,I_d})_{d \in \mathbb{N}}$ as the *I -(anti)symmetric problem*. Using the notation from Section 1.2 we thus have $\mathcal{F}_d = P_{I_d}(H_d)$ and, consequently, $\tilde{\mathcal{F}}_d = \mathcal{B}(P_{I_d}(H_d))$.

Due to (5.10) it is quite clear that $A_{n,d}^*$ cannot be optimal in this restricted setting since it calculates redundant pieces of information. Hence we need to go

beyond this naive attempt to solve I -(anti)symmetric problems efficiently. On the other hand, $P_{I_d}(H_d)$ equipped with the inner product of H_d , $\langle \cdot, \cdot \rangle_{H_d}$, again is a Hilbert space. Therefore we basically know how to construct an optimal algorithm; see Section 2.3.2. If $\#I_d = 1$ then our new algorithm should resemble $A_{n,d}^*$, because then we do not claim any (anti)symmetry and thus we deal with the entire tensor product problem.

Before we state the main assertion of this section we present an auxiliary result which shows that any optimal algorithm A^* for S_{d,I_d} needs to preserve the (anti)symmetry properties of its domain of definition. The following proposition generalizes Lemma 10.2 in Zeiser [Zei10] where this assertion was shown for the approximation problem, that is for $S_{d,I_d} = \text{id}: P_{I_d}(H_d) \rightarrow \mathcal{G}_d$.

Proposition 5.3. Let $d > 1$ and $\emptyset \neq I \subseteq \{1, \dots, d\}$ be arbitrarily fixed. Furthermore, for $X \in \{H, \mathcal{G}\}$ let P_I^X denote the (anti)symmetrizer $P_I \in \{\mathfrak{S}_I, \mathfrak{A}_I\}$ on X_d with respect to I . Then we have

$$(S_d \circ P_I^H)(g) = (P_I^{\mathcal{G}} \circ S_d)(g) \quad \text{for any } g \in H_d. \quad (5.12)$$

Moreover, for all $A: P_I^H(H_d) \rightarrow \mathcal{G}_d$ and every $f \in P_I^H(H_d)$,

$$\|S_{d,I}f - Af\|_{\mathcal{G}_d}^2 = \|S_{d,I}f - P_I^{\mathcal{G}}(Af)\|_{\mathcal{G}_d}^2 + \|Af - P_I^{\mathcal{G}}(Af)\|_{\mathcal{G}_d}^2. \quad (5.13)$$

Hence an optimal algorithm A^* for $S_{d,I}$ preserves (anti)symmetry, i.e.

$$A^*f \in P_I^{\mathcal{G}}(\mathcal{G}_d) \quad \text{for all } f \in P_I^H(H_d).$$

Proof. The proof is organized as follows. First we show that the tensor product operator S_d and the (anti)symmetrizer P_I commute on H_d , i.e. it holds (5.12). In a second step we conclude (5.13) out of this. The (anti)symmetry of A^*f for an optimal algorithm A^* then follows immediately.

Step 1. Assume $E_d = \{e_{d,j} \mid \mathbf{j} \in \mathcal{J}_d\}$ to be an arbitrary tensor product ONB of H_d , as defined in (2.13). Then, for fixed $\mathbf{j} \in \mathcal{J}_d$, formula (5.11) and the structure of $S_d = S_1 \otimes \dots \otimes S_1$ yields in the case $P_I = \mathfrak{A}_I$

$$\begin{aligned} S_d(\mathfrak{A}_I^H(e_{d,\mathbf{j}})) &= S_d \left(\frac{1}{\#\mathfrak{S}_I} \sum_{\pi \in \mathfrak{S}_I} (-1)^{|\pi|} \bigotimes_{l=1}^d e_{j_{\pi(l)}} \right) \\ &= \frac{1}{\#\mathfrak{S}_I} \sum_{\pi \in \mathfrak{S}_I} (-1)^{|\pi|} \bigotimes_{l=1}^d S_1(e_{j_{\pi(l)}}) = \mathfrak{A}_I^{\mathcal{G}}(S_d(e_{d,\mathbf{j}})). \end{aligned}$$

Obviously the same is true for $P_I = \mathfrak{S}_I$. Hence, (5.12) holds at least on the set of basis elements E_d of H_d . Because of the representation $g = \sum_{\mathbf{j} \in \mathcal{J}_d} \langle g, e_{d,\mathbf{j}} \rangle_{H_d} \cdot e_{d,\mathbf{j}}$ of $g \in H_d$, as well as the linearity and boundedness of the operators $P_I^H, P_I^{\mathcal{G}}$ and S_d , we can extend the relation (5.12) from E_d to the whole space H_d .

Step 2. Now let $f \in P_I^H(H_d)$ and let Af denote an arbitrary approximation to $S_{d,I}f$. Then $S_{d,I}f = S_d(P_I^H f) = P_I^{\mathfrak{G}}(S_d f)$, due to Step 1. Using the fact that $P_I^{\mathfrak{G}}$ provides an orthogonal projection onto $P_I^{\mathfrak{G}}(\mathfrak{G}_d)$, see (5.6) in Lemma 5.1, we obtain (5.13), i.e.

$$\begin{aligned} \|S_{d,I}f - Af | \mathfrak{G}_d\|^2 &= \|P_I^{\mathfrak{G}}(S_d f) - [P_I^{\mathfrak{G}}(Af) + (\text{id}^{\mathfrak{G}} - P_I^{\mathfrak{G}})(Af)] | \mathfrak{G}_d\|^2 \\ &= \|P_I^{\mathfrak{G}}(S_d f - Af) | \mathfrak{G}_d\|^2 + \|(\text{id}^{\mathfrak{G}} - P_I^{\mathfrak{G}})(Af) | \mathfrak{G}_d\|^2 \\ &= \|S_{d,I}f - P_I^{\mathfrak{G}}(Af) | \mathfrak{G}_d\|^2 + \|Af - P_I^{\mathfrak{G}}(Af) | \mathfrak{G}_d\|^2, \end{aligned}$$

as claimed. ■

Apart from this qualitative assertion, we are interested in an explicit formula for the optimal algorithm, as well as in sharp error bounds. To this end, let $d \in \mathbb{N}$ and $\emptyset \neq I_d = \{i_1, \dots, i_{\#I}\} \subseteq \{1, \dots, d\}$, as well as $P \in \{\mathfrak{S}, \mathfrak{A}\}$. Furthermore, consider the singular value decomposition of $S_1: H_1 \rightarrow \mathfrak{G}_1$. That is, let $\{(\lambda_m, \phi_m) \mid m \in \mathcal{M}_1\}$ denote the non-trivial eigenpairs of $W_1 = S_1^\dagger S_1$; see Section 2.3.1. Due to Proposition 2.10 in Section 2.4.2 we know that for $d > 1$ the (tensor) product eigenpairs $\{(\tilde{\lambda}_{d,\mathbf{m}}, \tilde{\phi}_{d,\mathbf{m}}) \mid \mathbf{m} \in \mathcal{M}_d\}$ of $W_d = S_d^\dagger S_d$ are given by (2.15). Moreover, $E_d = \Phi_d = \{\tilde{\phi}_{d,\mathbf{m}} \mid \mathbf{m} \in \mathcal{M}_d\}$ builds an tensor product ONB in H_d . Hence, we can apply Lemma 5.2 to $e_{d,\mathbf{j}} = \tilde{\phi}_{d,\mathbf{j}}$, $\mathbf{j} \in \mathcal{M}_d$, in order to obtain an orthonormal basis $\Xi_d = \{\tilde{\xi}_{\mathbf{k}} \mid \mathbf{k} \in \nabla_d\}$ of the partially (anti)symmetric subspaces $P_{I_d}(H_d)$. More precisely, for $\mathbf{k} \in \nabla_d$ we define

$$\tilde{\xi}_{\mathbf{k}} = \sqrt{\frac{\#S_I}{\mathbf{M}_I(\mathbf{k})!}} \cdot P_{I_d} \left(\bigotimes_{l=1}^d \phi_{k_l} \right) \in P_I(H_d) \quad \text{and} \quad \tilde{\lambda}_{d,\mathbf{k}} = \prod_{l=1}^d \lambda_{k_l} > 0, \quad (5.14)$$

where ∇_d is given by (5.9). Similar to the approach in Section 2.4.2, let

$$\psi = \psi_d: \{i \in \mathbb{N} \mid i < \#\nabla_d + 1\} \rightarrow \nabla_d$$

denote a bijection which provides a non-increasing ordering of $\{\tilde{\lambda}_{d,\mathbf{k}} \mid \mathbf{k} \in \nabla_d\}$ and set $\lambda_{d,i} = \tilde{\lambda}_{d,\psi(i)}$, as well as $\xi_{d,i} = \tilde{\xi}_{\psi(i)}$ for $i < \#\nabla_d + 1$. Finally, if $\#\nabla_d$ is finite then we extend the sequence of λ 's by setting $\lambda_{d,i} = 0$ for $i > \#\nabla_d$.

Given this bunch of notations we are well-prepared to prove our main theorem of this section. For every $d \in \mathbb{N}$ it provides a linear algorithm $A'_{n,d}$ which uses at most n continuous linear functionals on the input to approximate the solution operator S_{d,I_d} of a given I_d -(anti)symmetric tensor product problem between Hilbert spaces. Since the worst case error of this algorithm coincides with the n th minimal error of the problem, $A'_{n,d}$ is optimal in this setting; thus it cannot be improved by any other algorithm from the class $\mathcal{A}_d^{n,\text{cont}} \cup \mathcal{A}_d^{n,\text{adapt}}$; see Section 1.3. The assertion reads as follows.

Theorem 5.4. Assume $S_I = (S_{d,I_d})_{d \in \mathbb{N}}$ to be the linear tensor product problem S restricted to the I_d -(anti)symmetric subspaces $P_{I_d}(H_d)$ of the d -fold tensor product spaces H_d . Then for every $d \in \mathbb{N}$ the set

$$\{(\lambda_{d,i}, \xi_{d,i}) \mid 1 \leq i < \#\nabla_d + 1\} = \left\{ \left(\tilde{\lambda}_{d,\mathbf{k}}, \tilde{\xi}_{\mathbf{k}} \right) \mid \mathbf{k} \in \nabla_d \right\} \quad (5.15)$$

denotes the eigenpairs of $W_{d,I_d} = S_{d,I_d}^\dagger S_{d,I_d} : P_{I_d}(H_d) \rightarrow P_{I_d}(H_d)$. Thus, for every $n \in \mathbb{N}_0$ and all $d \in \mathbb{N}$ the linear algorithm $A'_{n,d} : P_{I_d}(H_d) \rightarrow P_{I_d}(\mathcal{G}_d)$ given by

$$A'_{n,d} f = \sum_{i=1}^{\min\{n, \#\nabla_d\}} \langle f, \xi_{d,i} \rangle_{H_d} \cdot S_d \xi_{d,i}, \quad (5.16)$$

is n th optimal for S_{d,I_d} w.r.t. the worst case setting. Furthermore we have

$$e^{\text{wor}}(n, d; P_{I_d}(H_d)) = \Delta^{\text{wor}}(A'_{n,d}; P_{I_d}(H_d)) = \sqrt{\lambda_{d,n+1}}. \quad (5.17)$$

Proof. Since S_I is a compact problem between Hilbert spaces it is enough to prove that for $d \in \mathbb{N}$ the eigenpairs of $W_{d,I_d} = S_{d,I_d}^\dagger S_{d,I_d}$ are given by (5.15). The remaining assertions then follow from Corollary 2.7. Indeed, we only need to show that $W_{d,I_d} \tilde{\xi}_{\mathbf{k}} = \tilde{\lambda}_{d,\mathbf{k}} \cdot \tilde{\xi}_{\mathbf{k}}$ for every $\mathbf{k} \in \nabla_d$ because we already know that the set $\Xi_d = \{\tilde{\xi}_{\mathbf{k}} \mid \mathbf{k} \in \nabla_d\}$ builds an ONB in $P_{I_d}(H_d)$. Hence there cannot be more than these eigenpairs.

To prove the claim, observe that from the first part of Proposition 5.3 it follows

$$S_{d,I_d} = S_d \circ P_{I_d}^H = P_{I_d}^{\mathcal{G}} \circ S_d \quad \text{which implies that} \quad S_{d,I_d} : P_{I_d}^H(H_d) \rightarrow P_{I_d}^{\mathcal{G}}(\mathcal{G}_d).$$

Moreover, due to the self-adjointness of the projectors (see Lemma 5.1), it is easily seen that this yields

$$S_{d,I_d}^\dagger = P_{I_d}^H \circ S_d^\dagger = S_d^\dagger \circ P_{I_d}^{\mathcal{G}} \quad \text{such that} \quad S_{d,I_d}^\dagger : P_{I_d}^{\mathcal{G}}(\mathcal{G}_d) \rightarrow P_{I_d}^H(H_d).$$

Consequently, we have

$$W_{d,I_d} P_{I_d}^H = (P_{I_d}^H S_d^\dagger) (S_d P_{I_d}^H) P_{I_d}^H = P_{I_d}^H (S_d^\dagger P_{I_d}^{\mathcal{G}}) S_d = P_{I_d}^H (S_d^\dagger S_d) = P_{I_d}^H W_d,$$

because of $(P_{I_d}^X)^2 = P_{I_d}^X$, where $X \in \{H, \mathcal{G}\}$. Since for every $\mathbf{j} \in \mathcal{M}_d$ the simple tensor $\tilde{\phi}_{d,\mathbf{j}}$ is an eigenelement of W_d with respect to the eigenvalue $\tilde{\lambda}_{d,\mathbf{j}}$, we conclude

$$W_{d,I_d} \left(P_{I_d}^H \tilde{\phi}_{d,\mathbf{j}} \right) = \tilde{\lambda}_{d,\mathbf{j}} \cdot \left(P_{I_d}^H \tilde{\phi}_{d,\mathbf{j}} \right)$$

from the linearity of $P_{I_d}^H$. In particular, this is true for every $\mathbf{j} = \mathbf{k} \in \nabla_d \subseteq \mathcal{M}_d$. But now we note that $\tilde{\xi}_{\mathbf{k}}$ equals $P_{I_d}^H \tilde{\phi}_{d,\mathbf{k}}$, at least up to some normalizing constant. Hence, using linearity once again, we have proven the claimed assertion. ■

We conclude this section by adding some final remarks on the above theorem.

Remark 5.5. Obviously, our former result for the entire tensor product problem $S = (S_d: H_d \rightarrow \mathcal{G}_d)_{d \in \mathbb{N}}$ in Section 2.4.2 is also covered by Theorem 5.4. We simply have to choose I_d such that $\#I_d = 1$ for every $d \in \mathbb{N}$ and obtain $A'_{n,d} = A^*_{n,d}$. As in this case, the worst case error can be attained by the element $\xi_{d,n+1}$ provided that $n < \#\nabla_d$. Otherwise it trivially equals zero.

It should be clear to the reader how to generalize the results of this section to the case of multiple partially (anti)symmetric problems where we claim (anti)symmetry w.r.t. more than one subset of coordinates I . Recall that this definition is given at the end of Section 5.1.1.

Finally we want to mention that we decided to give a different proof of Theorem 5.4 than in [Wei11] and [Wei12a], respectively. The reason is that the usage of the self-adjointness of the projections P_{I_d} seems to be more elegant than again repeating the arguments used for Corollary 2.7 in Section 2.3.2. Furthermore, now we can handle also problems defined on non-separable or on finite-dimensional source spaces H_d . Thus we slightly generalized our old results. \square

5.3 Complexity of (anti)symmetric problems

Encouraged by the exact formula for the n th minimal worst case error in Theorem 5.4 the intention of the present section is to investigate the information complexity of (anti)symmetric tensor product problems. We restrict our attention to the study of polynomial and strong polynomial tractability in what follows. The aim is to find necessary and sufficient conditions for these properties in terms of the univariate sequence $\lambda = (\lambda_m)_{m \in \mathbb{N}}$ and the number of (anti)symmetry conditions we impose. From the definition of ∇_d in (5.9) it is quite clear that antisymmetric problems are significantly easier than their symmetric counterparts. Therefore, after proving some general assertions, we handle these cases separately in order to conclude sharp conditions. Moreover, we distinguish between the absolute and the normalized error criterion.

Let us fix the basic notation for this section. As before, assume $S_I = (S_{d,I_d})_{d \in \mathbb{N}}$ to denote a tensor product problem $S = (S_d: H_d \rightarrow \mathcal{G}_d)_{d \in \mathbb{N}}$, restricted to some sequence of (anti)symmetric subspaces $P_{I_d}(H_d)$, where $P \in \{\mathfrak{S}, \mathfrak{A}\}$, of the tensor product Hilbert spaces $H_d = H_1 \otimes \dots \otimes H_d$, $d \in \mathbb{N}$. Here for every $d \in \mathbb{N}$ the elements are (anti)symmetric w.r.t. the non-empty subset $I_d \subseteq \{1, \dots, d\}$ of coordinates. The cardinality of these subsets will be denoted by $a_d = \#I_d$ and we set $b_d = d - a_d$ for the number of coordinates without (anti)symmetry conditions. Finally, for $d \in \mathbb{N}$ the non-increasingly ordered eigenvalues $\lambda_{d,i} = \tilde{\lambda}_{d,\psi(i)}$, $i \in \mathbb{N}$, are given by (5.14) and (5.15), respectively. They are constructed out of the squared singular values $\lambda = (\lambda_m)_{m \in \mathbb{N}}$ of the underlying solution operator $S_1: H_1 \rightarrow \mathcal{G}_1$.

As an immediate consequence of (5.17) we see that the initial error of approxi-

mating S_{d,I_d} on the unit ball $\tilde{\mathcal{F}}_d = \mathcal{B}(P_{I_d}(H_d))$ is given by

$$\varepsilon_d^{\text{init}} = e^{\text{wor}}(0, d; P_I(H_d)) = \sqrt{\lambda_{d,1}} = \begin{cases} \sqrt{\lambda_1^d}, & \text{if } P = \mathfrak{S}, \\ \sqrt{\lambda_1^{b_d} \cdot \lambda_1 \cdot \dots \cdot \lambda_{a_d}}, & \text{if } P = \mathfrak{A}. \end{cases}$$

Clearly, we need to assume that this initial error is strictly positive for any reasonably large $d \in \mathbb{N}$ because otherwise we have (strong) polynomial tractability by default. In particular, if the number of antisymmetric coordinates a_d grows with the dimension then this condition implies that the whole sequence of univariate eigenvalues λ need to be strictly positive. Moreover, similar to the entire tensor product problems studied in [Section 2.4.3](#), we always assume that $\lambda_2 > 0$ in order to avoid triviality. Consequently, we have $\#\mathcal{M}_1 \geq 2$.

Now we are ready to conclude a first general condition which is necessary for (strong) polynomial tractability of both symmetric and antisymmetric problems as long as we deal with the absolute error criterion. It is independent of the concrete choice of the (anti)symmetry conditions we impose.

Lemma 5.6 (General necessary conditions, absolute errors). Let $P \in \{\mathfrak{S}, \mathfrak{A}\}$ and consider $S_I = (S_{d,I_d})_{d \in \mathbb{N}}$ as defined above, where I_d is arbitrarily fixed for every $d \in \mathbb{N}$. Then the fact that S_I is polynomially tractable with the constants $C, p > 0$ and $q \geq 0$ implies that $\lambda = (\lambda_m)_{m \in \mathbb{N}} \in \ell_\tau$ for all $\tau > p/2$. Moreover, for any such τ and all $d \in \mathbb{N}$ the following estimate holds:

$$\frac{1}{(\lambda_{d,1})^\tau} \sum_{\mathbf{k} \in \nabla_d} (\tilde{\lambda}_{d,\mathbf{k}})^\tau \leq (1 + C) d^q + C^{2\tau/p} \zeta\left(\frac{2\tau}{p}\right) \left(\frac{d^{2q/p}}{\lambda_{d,1}}\right)^\tau.$$

Proof. From [Theorem 2.8](#) we know that for any $\tau > p/2$ and $r = 2q/p$ polynomial tractability yields

$$\sup_{d \in \mathbb{N}} \frac{1}{d^r} \left(\sum_{i=f(d)}^{\infty} (\lambda_{d,i})^\tau \right)^{1/\tau} < \infty, \quad (5.18)$$

where the function $f: \mathbb{N} \rightarrow \mathbb{N}$ is given by $f(d) = \lceil (1 + C) d^q \rceil$. This particularly implies that the sum in the brackets converges for every fixed $d \in \mathbb{N}$. Therefore, especially for $d = 1$ the tail series $\sum_{i=f(1)}^{\infty} (\lambda_{1,i})^\tau = \sum_{m=\lceil 1+C \rceil}^{\infty} (\lambda_m)^\tau$ needs to be finite which is possible only if $\lambda = (\lambda_m)_{m \in \mathbb{N}} \in \ell_\tau$.

So, let us turn to the second assertion. Obviously (5.18) implies the existence of some constant $C_1 > 0$ such that

$$\sum_{i=f(d)}^{\infty} (\lambda_{d,i})^\tau \leq C_1 d^{r\tau} \quad \text{for all } d \in \mathbb{N}.$$

Indeed, [Theorem 2.8](#) yields that we can take $C_1 = C^{2\tau/p}\zeta(2\tau/p)$. Due to the ordering of $(\lambda_{d,i})_{i \in \mathbb{N}}$ the rest of the sum can also be bounded easily for any $d \in \mathbb{N}$ by

$$\sum_{i=1}^{f(d)-1} (\lambda_{d,i})^\tau \leq (\lambda_{d,1})^\tau \cdot (f(d) - 1).$$

Since $\sum_{\mathbf{k} \in \nabla_d} (\tilde{\lambda}_{d,\mathbf{k}})^\tau = \sum_{i=1}^{\infty} (\lambda_{d,i})^\tau$, it remains to show that $f(d) - 1 \leq (1 + C)d^q$ for every $d \in \mathbb{N}$ which is also obvious due to the definition of f . \blacksquare

5.3.1 Symmetric problems (absolute errors)

Apart from the general assertion $\lambda \in \ell_\tau$, we focus our attention on further necessary conditions for (strong) polynomial tractability in the symmetric setting. The following proposition yields a slight improvement compared to the corresponding assertion stated in [\[Wei12a\]](#) which can be obtained without using essential new ideas.

Proposition 5.7 (Necessary conditions, symmetric case). Let $S_I = (S_{d,I_d})_{d \in \mathbb{N}}$ be the problem considered in [Lemma 5.6](#) and set $P = \mathfrak{S}$. Moreover, assume $\lambda_1 \geq 1$.

- If S_I is polynomially tractable then $b_d \in \mathcal{O}(\ln d)$, as $d \rightarrow \infty$.
- If S_I is strongly polynomially tractable then $b_d \in \mathcal{O}(1)$, as $d \rightarrow \infty$, and $\lambda_1 = 1 > \lambda_2$.

Proof. Assume $\lambda_1 \geq 1$ and let τ be given by [Lemma 5.6](#). Then, independent of the amount of symmetry conditions, we have $\lambda_{d,1} = \lambda_1^d \geq 1$ and there exist absolute constants $r \geq 0$ and $C > 1$ such that

$$\frac{1}{(\lambda_1)^{\tau d}} \sum_{\mathbf{k} \in \nabla_d} (\tilde{\lambda}_{d,\mathbf{k}})^\tau \leq C d^r, \quad d \in \mathbb{N}, \quad (5.19)$$

due to [Lemma 5.6](#). In the case of strong polynomial tractability we even have $r = 0$. For $d \geq 2$ we use the product structure of $\tilde{\lambda}_{d,\mathbf{k}}$, $\mathbf{k} \in \nabla_d$, provided by [\(5.14\)](#). That is, we split the sum w.r.t. the coordinates with and without symmetry conditions. Hence, we conclude

$$\sum_{\mathbf{k}=(\mathbf{h},\mathbf{j}) \in \nabla_d} \tilde{\lambda}_{d,\mathbf{k}}^\tau = \sum_{\mathbf{j} \in (\mathcal{M}_1)^{b_d}} \tilde{\lambda}_{b_d,\mathbf{j}}^\tau \sum_{\substack{\mathbf{h} \in (\mathcal{M}_1)^{a_d}, \\ h_1 \leq \dots \leq h_{a_d}}} \tilde{\lambda}_{a_d,\mathbf{h}}^\tau = \left(\sum_{m=1}^{\#\mathcal{M}_1} \lambda_m^\tau \right)^{b_d} \sum_{\substack{\mathbf{h} \in \mathcal{M}_{a_d}, \\ h_1 \leq \dots \leq h_{a_d}}} \tilde{\lambda}_{a_d,\mathbf{h}}^\tau \quad (5.20)$$

for $d = a_d + b_d \geq 2$ which leads to

$$\left(\sum_{m=1}^{\#\mathcal{M}_1} \left(\frac{\lambda_m}{\lambda_1} \right)^\tau \right)^{b_d} \sum_{\substack{\mathbf{h} \in \mathcal{M}_{a_d}, \\ h_1 \leq \dots \leq h_{a_d}}} \prod_{l=1}^{a_d} \left(\frac{\lambda_{h_l}}{\lambda_1} \right)^\tau \leq C d^r.$$

In any case the second sum in the above inequality is bounded from below by 1. Thus, using $\#\mathcal{M}_1 \geq 2$ we conclude that $(1 + \lambda_2^\tau/\lambda_1^\tau)^{b_d} \leq \left(\sum_{m=1}^{\#\mathcal{M}_1} \lambda_m^\tau/\lambda_1^\tau\right)^{b_d}$ needs to be polynomially bounded from above. Since we always assume $\lambda_2 > 0$ this leads to the claimed bounds on b_d .

It remains to show the assertions on the two largest univariate eigenvalues in the case of strong polynomial tractability. To this end, assume for a moment that $\lambda_1 > 1$. Then, because of $\lambda_2 > 0$, there need to exist some $K \in \mathbb{N}_0$ such that $\lambda_2 \geq (1/\lambda_1)^K$. Now it is easy to see that (independent of the number of symmetry conditions) there are at least $1 + \lfloor d/(K+1) \rfloor$ different $\mathbf{k} \in \nabla_d$ such that $\tilde{\lambda}_{d,\mathbf{k}} \geq 1$. Namely, for $l = 0, \dots, \lfloor d/(K+1) \rfloor$ we can take the first $d-l$ coordinates of $\mathbf{k} \in \nabla_d$ equal to one. To the remaining coordinates we assign the value two and obtain

$$\tilde{\lambda}_{d,\mathbf{k}} = \lambda_1^{d-l} \lambda_2^l \geq \lambda_1^{Kl} \lambda_2^l \geq 1.$$

In other words, we have $\lambda_{d,1+\lfloor d/(K+1) \rfloor} \geq 1$. On the other hand, strong polynomial tractability implies $\sum_{i=\lceil 1+C \rceil}^{\infty} \lambda_{d,i}^\tau \leq C_1$ for some absolute constants $\tau, C, C_1 > 0$ and all $d \in \mathbb{N}$; see (5.18). Consequently, for every $d \geq d_0 = (2+C)(K+1)$ we obtain $1 + \lfloor d/(K+1) \rfloor \geq \lceil 1+C \rceil$ and thus

$$\begin{aligned} C_1 &\geq \sum_{i=\lceil 1+C \rceil}^{\infty} \lambda_{d,i}^\tau \geq \sum_{i=\lceil 1+C \rceil}^{1+\lfloor d/(K+1) \rfloor} \lambda_{d,i}^\tau \\ &\geq \lambda_{d,1+\lfloor d/(K+1) \rfloor}^\tau (2 + \lfloor d/(K+1) \rfloor - \lceil 1+C \rceil) \\ &\geq \frac{d}{K+1} - (1+C). \end{aligned}$$

Obviously this is a contradiction and we conclude $\lambda_1 = 1$. Finally, we need to show that we necessarily have $\lambda_2 < 1$. Assuming that $\lambda_1 = \lambda_2 = 1$ leads to $K = 0$ in the discussion above and hence we obtain the same contradiction as before. Therefore the proof is complete. \blacksquare

Note in passing that independent of the number of symmetry conditions the information complexity $n_{\text{abs}}^{\text{wor}}(\varepsilon, d; S_{d,I_d} : \mathcal{B}(\mathfrak{S}_{I_d}(H_d)) \rightarrow \mathfrak{G}_d)$ needs to grow at least linearly in d if we assume $\lambda_1 \geq 1$ and $\lambda_2 > 0$.

We continue the analysis of I -symmetric problems with respect to the absolute error criterion by proving that the stated necessary conditions are also sufficient for (strong) polynomial tractability. For this purpose we need a rather technical preliminary lemma. For the convenience of the reader we include a full proof that uses only elementary induction arguments.

Lemma 5.8. Let $(\mu_m)_{m \in \mathbb{N}}$ be a non-increasing sequence of non-negative real numbers with $\mu_1 > 0$ and set $\mu_{s,\mathbf{k}} = \prod_{l=1}^s \mu_{k_l}$ for $\mathbf{k} \in \mathbb{N}^s$ and $s \in \mathbb{N}$.

Then, for all $V \in \mathbb{N}_0$ and every $d \in \mathbb{N}$, it holds

$$\sum_{\substack{\mathbf{k} \in \mathbb{N}^d, \\ 1 \leq k_1 \leq \dots \leq k_d}} \mu_{d,\mathbf{k}} \leq (\mu_1)^d d^V \left(1 + V + \sum_{L=1}^d (\mu_1)^{-L} \sum_{\substack{\mathbf{j}^{(L)} \in \mathbb{N}^L, \\ V+2 \leq j_1^{(L)} \leq \dots \leq j_L^{(L)}}} \mu_{L,\mathbf{j}^{(L)}} \right). \quad (5.21)$$

Proof. Step 1. By induction on s we first prove that for every fixed $m \in \mathbb{N}$

$$\sum_{\substack{\mathbf{k} \in \mathbb{N}^s, \\ m \leq k_1 \leq \dots \leq k_s}} \mu_{s,\mathbf{k}} = (\mu_m)^s + \sum_{l=1}^s (\mu_m)^{s-l} \sum_{\substack{\mathbf{j}^{(l)} \in \mathbb{N}^l, \\ m+1 \leq j_1^{(l)} \leq \dots \leq j_l^{(l)}}} \mu_{l,\mathbf{j}^{(l)}} \quad \text{for all } s \in \mathbb{N}. \quad (5.22)$$

Easy calculations show that this holds at least for the initial step $s = 1$. Therefore, assume the assertion (5.22) to be true for some $s \in \mathbb{N}$. Then

$$\begin{aligned} \sum_{\substack{\mathbf{k} \in \mathbb{N}^{s+1}, \\ m \leq k_1 \leq \dots \leq k_{s+1}}} \mu_{s+1,\mathbf{k}} &= \sum_{k_1=m}^{\infty} \mu_{k_1} \sum_{\substack{\mathbf{h} \in \mathbb{N}^s, \\ k_1 \leq h_1 \leq \dots \leq h_s}} \mu_{s,\mathbf{h}} \\ &= \mu_m \sum_{\substack{\mathbf{h} \in \mathbb{N}^s, \\ m \leq h_1 \leq \dots \leq h_s}} \mu_{s,\mathbf{h}} + \sum_{\substack{\mathbf{k} \in \mathbb{N}^{s+1}, \\ m+1 \leq k_1 \leq \dots \leq k_{s+1}}} \mu_{s+1,\mathbf{k}}. \end{aligned}$$

Now, by inserting the induction hypothesis for the first sum and renaming \mathbf{k} to $\mathbf{j}^{(s+1)}$ in the remaining sum, we conclude that $\sum_{\substack{\mathbf{k} \in \mathbb{N}^{s+1}, \\ m \leq k_1 \leq \dots \leq k_{s+1}}} \mu_{s+1,\mathbf{k}}$ equals

$$(\mu_m)^{s+1} + \sum_{l=1}^s (\mu_m)^{s+1-l} \sum_{\substack{\mathbf{j}^{(l)} \in \mathbb{N}^l, \\ m+1 \leq j_1^{(l)} \leq \dots \leq j_l^{(l)}}} \mu_{l,\mathbf{j}^{(l)}} + \sum_{\substack{\mathbf{j}^{(s+1)} \in \mathbb{N}^{s+1}, \\ m+1 \leq j_1^{(s+1)} \leq \dots \leq j_{s+1}^{(s+1)}}} \mu_{s+1,\mathbf{j}^{(s+1)}}.$$

Hence (5.22) also holds for $s + 1$ and the induction is complete.

Step 2. Here we prove (5.21) via another induction on $V \in \mathbb{N}_0$. Therefore, let $d \in \mathbb{N}$ be arbitrarily fixed. The initial step, $V = 0$, corresponds to (5.22) for $s = d$ and $m = 1$. Thus assume (5.21) to be valid for some fixed $V \in \mathbb{N}_0$. Then, by using (5.22) for $s = L$ and $m = V + 2$, we see that the right-hand side of (5.21) equals

$$(\mu_1)^d d^V \left(1 + V + \sum_{L=1}^d (\mu_1)^{-L} \left((\mu_{V+2})^L + \sum_{l=1}^L (\mu_{V+2})^{L-l} \sum_{\substack{\mathbf{j}^{(l)} \in \mathbb{N}^l, \\ (V+2)+1 \leq j_1^{(l)} \leq \dots \leq j_l^{(l)}}} \mu_{l,\mathbf{j}^{(l)}} \right) \right).$$

Now we estimate $1 + V$ by $d(1 + V)$, take advantage of the non-increasing ordering of $(\mu_m)_{m \in \mathbb{N}}$, and extend the inner sum from L to d in order to obtain

$$\sum_{\substack{\mathbf{k} \in \mathbb{N}^d, \\ 1 \leq k_1 \leq \dots \leq k_d}} \mu_{d, \mathbf{k}} \leq (\mu_1)^d d^{V+1} \left(1 + (V+1) + \sum_{l=1}^d (\mu_1)^{-l} \sum_{\substack{\mathbf{j}^{(l)} \in \mathbb{N}^l, \\ (V+1)+2 \leq j_1^{(l)} \leq \dots \leq j_l^{(l)}}} \mu_{l, \mathbf{j}^{(l)}} \right).$$

Since this estimate corresponds to (5.21) for $V+1$ the claim is proven. \blacksquare

Now the sufficient conditions read as follows.

Proposition 5.9 (Sufficient conditions, symmetric case). Let $P = \mathfrak{S}$, assume S_I to be the problem considered in Lemma 5.6, and let $\lambda = (\lambda_m)_{m \in \mathbb{N}} \in \ell_{\tau_0}$ for some $\tau_0 \in (0, \infty)$.

- If $\lambda_1 < 1$ then S_I is strongly polynomially tractable.
- If $\lambda_1 = 1 > \lambda_2$ and $b_d \in \mathcal{O}(1)$ then S_I is strongly polynomially tractable.
- If $\lambda_1 = 1$ and $b_d \in \mathcal{O}(\ln d)$, as $d \rightarrow \infty$, then S_I is polynomially tractable.

Proof. Step 1. We start the proof by exploiting the property $\lambda \in \ell_{\tau_0}$; namely we use the ordering of $(\lambda_m)_{m \in \mathbb{N}}$ to conclude that

$$m \lambda_m^{\tau_0} \leq \lambda_1^{\tau_0} + \dots + \lambda_m^{\tau_0} < \sum_{i=1}^{\infty} \lambda_i^{\tau_0} = \|\lambda\|_{\ell_{\tau_0}}^{\tau_0} < \infty \quad \text{for any } m \in \mathbb{N}.$$

Hence, there exists some $C_{\tau_0} > 0$ such that λ_m is bounded from above by $C_{\tau_0} \cdot m^{-r}$ for every $r \leq 1/\tau_0$. Therefore there is some index such that for every larger $m \in \mathbb{N}$ we have $\lambda_m < 1$. We denote the smallest of these indices by m_0 . Similar to the calculations of Novak and Woźniakowski [NW08, p. 180] this leads to

$$\sum_{m=m_0}^{\infty} \lambda_m^{\tau} \leq (p+1) \lambda_{m_0}^{\tau} + C_{\tau_0}^{\tau} \int_{m_0+p}^{\infty} x^{-\tau r} d\lambda^1(x) = (p+1) \lambda_{m_0}^{\tau} + \frac{C_{\tau_0}^{\tau}}{\tau r - 1} \frac{1}{(m_0 + p)^{\tau r - 1}}$$

for every $p \in \mathbb{N}_0$ and all τ such that $\tau r > 1$. In particular, with $r = 1/\tau_0$ we obtain for all $\tau > \tau_0$ and any $p \in \mathbb{N}_0$ the estimate

$$\sum_{m=m_0}^{\infty} (\lambda_m)^{\tau} \leq (p+1) (\lambda_{m_0})^{\tau} + \frac{1/\tau}{1/\tau_0 - 1/\tau} \left(\frac{C_{\tau_0}^{1/(1/\tau_0 - 1/\tau)}}{m_0 + p} \right)^{\tau(1/\tau_0 - 1/\tau)}.$$

Note that for a given $\delta > 0$ there exists some constant $\tau_1 \geq \tau_0$ such that for all $\tau > \tau_1$ it is $1/(1/\tau_0 - 1/\tau) \in (\tau_0, \tau_0 + \delta)$. Hence, if $p \in \mathbb{N}_0$ is sufficiently large then we conclude that for all $\tau > \tau_1$

$$\sum_{m=m_0}^{\infty} (\lambda_m)^{\tau} \leq (p+1) (\lambda_{m_0})^{\tau} + \frac{\tau_0 + \delta}{\tau_1} \left(\frac{C_1}{m_0 + p} \right)^{\tau/(\tau_0 + \delta)},$$

where we set $C_1 = \max\{1, C_{\tau_0}^{\tau_0+\delta}\} < m_0 + p$. Finally, since $\lambda_{m_0} < 1$, both the summands tend to zero as τ approaches infinity. In particular, there need to exist some $\tau > \tau_1 \geq \tau_0$ such that

$$\sum_{m=m_0}^{\infty} (\lambda_m)^\tau \leq \frac{1}{2}.$$

Step 2. Now all the stated assertions can be seen using the second point of [Theorem 2.8](#). Indeed, for polynomial tractability it is sufficient to show that

$$\sum_{\mathbf{k} \in \nabla_d} \left(\tilde{\lambda}_{d,\mathbf{k}}\right)^\tau = \sum_{i=1}^{\infty} (\lambda_{d,i})^\tau \leq C d^{r\tau} \quad \text{for all } d \in \mathbb{N} \quad (5.23)$$

and some $C, \tau > 0$, as well as some $r \geq 0$. If this even holds for $r = 0$ we obtain strong polynomial tractability.

In the case $\lambda_1 < 1$ we can estimate the sum on the left of (5.23) from above by $(\sum_{m=1}^{\infty} \lambda_m)^\tau$ since clearly $\nabla_d \subseteq \mathcal{M}_d \subseteq \mathbb{N}^d$. Using Step 1 with $m_0 = 1$ we conclude that $\sum_{\mathbf{k} \in \nabla_d} (\tilde{\lambda}_{d,\mathbf{k}})^\tau \leq 2^{-d}$ for some large $\tau > \tau_0$. Hence the problem is strongly polynomially tractable in this case.

For the proof of the remaining points we assume that $\lambda_1 = 1$. In any case we have

$$\sum_{\mathbf{k} \in \nabla_1} \left(\tilde{\lambda}_{1,\mathbf{k}}\right)^\tau \leq \sum_{m=1}^{\infty} (\lambda_m)^{\tau_0} = \|\lambda\|_{\ell_{\tau_0}}^{\tau_0} < \infty$$

for all $\tau \geq \tau_0$ because of $\lambda \in \ell_{\tau_0}$. Therefore we can assume $d \geq 2$ in the following. Recall that we can split the first sum in (5.23) w.r.t. the coordinates with and without symmetry conditions. That is, for $d = a_d + b_d \geq 2$ we use (5.20).

If $\lambda_2 < 1$ and b_d is universally bounded then the first factor in this splitting can be bounded by a constant and the second factor can be estimated using [Lemma 5.8](#) with $V = 0$, d replaced by a_d and μ replaced by λ^τ .³⁵ Consequently, for any $\tau \geq \tau_0$,

$$\sum_{\substack{\mathbf{h} \in \mathcal{M}_{a_d}, \\ h_1 \leq \dots \leq h_{a_d}}} \left(\tilde{\lambda}_{a_d,\mathbf{h}}\right)^\tau \leq 1 + \sum_{L=1}^{a_d} \sum_{\substack{\mathbf{j}^{(L)} \in \mathbb{N}^L, \\ 2 \leq j_1^{(L)} \leq \dots \leq j_L^{(L)}}} \left(\tilde{\lambda}_{L,\mathbf{j}^{(L)}}\right)^\tau \leq 1 + \sum_{L=1}^{a_d} \left(\sum_{m=2}^{\infty} (\lambda_m)^\tau\right)^L. \quad (5.24)$$

Now, with the help of Step 1 and the properties of geometric series, we see that if τ is large enough then (5.24) can be estimated further by $1 + \sum_{L=1}^{\infty} 2^{-L} = 2$. In summary also $\sum_{\mathbf{k} \in \nabla_d} (\tilde{\lambda}_{d,\mathbf{k}})^\tau$ is universally bounded in this case and therefore the problem S_I is strongly polynomially tractable.

To prove the last point we argue in the same manner. Here the assumption $b_d \in \mathcal{O}(\ln d)$, as $d \rightarrow \infty$, yields that the first factor in the splitting (5.20) is

³⁵Observe that this choice particularly implies that $\mu_m = 0$ for any $m > \#\mathcal{M}_1$.

polynomially bounded in d . For the second factor we again apply [Lemma 5.8](#), but in this case we set $V = m_0 - 2$, where m_0 denotes the first index $m \in \mathbb{N}$ such that $\lambda_m < 1$. Keep in mind that this index is at least two because of $\lambda_1 = 1$. On the other hand, it needs to be finite, since $\lambda \in \ell_{\tau_0}$. Therefore, due to the same arguments as above, the second factor in the splitting [\(5.20\)](#) is polynomially bounded in d , too. All in all, this proves [\(5.23\)](#) and thus S_I is polynomially tractable in this case. ■

We summarize the results obtained for I -symmetric tensor product problems $S_I = (S_{d,I_d})_{d \in \mathbb{N}}$ in the following theorem.

Theorem 5.10 (Polynomial tractability of sym. problems, absolute errors). Let $S_1: H_1 \rightarrow \mathcal{G}_1$ denote a compact linear operator between two Hilbert spaces and let $\lambda = (\lambda_m)_{m \in \mathbb{N}}$ be the sequence of eigenvalues of $W_1 = S_1^\dagger S_1$ w.r.t. a non-increasing ordering. Moreover, for $d > 1$ let $\emptyset \neq I_d \subseteq \{1, \dots, d\}$ and assume $S_I = (S_{d,I_d})_{d \in \mathbb{N}}$ to be the linear tensor product problem $S = (S_d)_{d \in \mathbb{N}}$ restricted to the I_d -symmetric subspaces $\mathfrak{S}_{I_d}(H_d)$ of the d -fold tensor product spaces H_d . Consider the worst case setting with respect to the absolute error criterion and let $\lambda_2 > 0$. Then S_I is strongly polynomially tractable if and only if $\lambda \in \ell_\tau$ for some $\tau \in (0, \infty)$ and

- $\lambda_1 < 1$, or
- $1 = \lambda_1 > \lambda_2$ and $(d - \#I_d) \in \mathcal{O}(1)$, as $d \rightarrow \infty$.

Moreover, provided that $\lambda_1 \leq 1$ the problem is polynomially tractable if and only if $\lambda \in \ell_\tau$ for some $\tau \in (0, \infty)$ and

- $\lambda_1 < 1$, or
- $\lambda_1 = 1$ and $(d - \#I_d) \in \mathcal{O}(\ln d)$, as $d \rightarrow \infty$.

Note that we do not have sufficient conditions for polynomial tractability in the case when $\lambda_1 > 1$. We only know that $(d - \#I_d) \in \mathcal{O}(\ln d)$, as $d \rightarrow \infty$, is necessary in this situation. Anyway, we completely characterized strong polynomial tractability of symmetric problems. In this respect we improved the results known from [\[Wei12a\]](#). Moreover, we have shown that the stated results also hold for finite-dimensional and for non-separable source spaces H_1 .

Before we turn to the complexity of antisymmetric problems we briefly focus on the normalized error criterion for the I -symmetric setting in the next subsection.

5.3.2 Symmetric problems (normalized errors)

Due to [\(2.10\)](#) and [\(5.14\)](#) the information complexity of I -symmetric problems $S_I = (S_{d,I_d})_{d \in \mathbb{N}}$ in the worst case setting w.r.t. the normalized error criterion is given by

$$n_{\text{norm}}^{\text{wor}}(\varepsilon', d; \mathfrak{S}_{I_d}(H_d)) = \# \left\{ \mathbf{k} \in \nabla_d \left| \frac{\tilde{\lambda}_{d,\mathbf{k}}}{\lambda_{d,1}} = \prod_{l=1}^d \left(\frac{\lambda_{k_l}}{\lambda_1} \right) > (\varepsilon')^2 \right. \right\}$$

for $\varepsilon' \in (0, 1)$ and $d \in \mathbb{N}$, since we have $(\varepsilon_d^{\text{init}})^2 = \lambda_{d,1} = \lambda_1^d$ for any kind of symmetric problem. In contrast, for the absolute error criterion (2.9) yields that $n_{\text{abs}}^{\text{wor}}(\varepsilon, d; \mathfrak{S}_{I_d}(H_d)) = \#\left\{\mathbf{k} \in \nabla_d \mid \tilde{\lambda}_{d,\mathbf{k}} = \prod_{l=1}^d \lambda_{k_l} > \varepsilon^2\right\}$, where $\varepsilon > 0$ and $d \in \mathbb{N}$. Hence, using the ideas in the proof of Theorem 2.12 it suffices to study a scaled tensor product problem $T_d: \mathfrak{S}_{I_d}(H_d) \rightarrow \mathfrak{G}_d$ w.r.t. the absolute error criterion in order to obtain tractability results for S_I in the normalized situation. To this end, recall that the squared singular values of T_1 equal $\mu = (\mu_m)_{m \in \mathbb{N}}$ with $\mu_m = \lambda_m / \lambda_1$ such that we always have $\mu_1 = 1$. Furthermore, we obviously have $\mu \in \ell_\tau$ if and only if $\lambda \in \ell_\tau$. This leads to the following theorem.

Theorem 5.11 (Polynomial tractability of symmetric problems, normalized errors). Consider the situation of Theorem 5.10. We study the the worst case setting with respect to the normalized error criterion. Then $S_I = (S_{d,I_d})_{d \in \mathbb{N}}$ is strongly polynomially tractable if and only if

$$\lambda \in \ell_\tau \text{ for some } \tau \in (0, \infty) \quad \text{and} \quad \lambda_1 > \lambda_2 \quad \text{and} \quad (d - \#I_d) \in \mathcal{O}(1), \text{ as } d \rightarrow \infty.$$

Moreover, the problem S_I is polynomially tractable if and only if

$$\lambda \in \ell_\tau \text{ for some } \tau \in (0, \infty) \quad \text{and} \quad (d - \#I_d) \in \mathcal{O}(\ln d), \text{ as } d \rightarrow \infty.$$

5.3.3 Antisymmetric problems (absolute errors)

We start this subsection with sufficient conditions for (strong) polynomial tractability which slightly improve the results stated in [Wei12a, Proposition 5].

Proposition 5.12 (Sufficient conditions, antisymmetric case). Let $P = \mathfrak{A}$, suppose $S_I = (S_{d,I_d})_{d \in \mathbb{N}}$ to be the problem considered in Lemma 5.6, and let $\lambda = (\lambda_m)_{m \in \mathbb{N}} \in \ell_{\tau_0}$ for some $\tau_0 \in (0, \infty)$.

- If $\lambda_1 < 1$ then S_I is strongly polynomially tractable, independent of the number of antisymmetry conditions.
- If $\lambda_1 \geq 1$ and if there exist constants $\tau \geq \tau_0$, $d_0 \in \mathbb{N}$, as well as $C \geq 1$, and $q \geq 0$ such that for the number of antisymmetric coordinates a_d in dimension d it holds that

$$\frac{\ln(a_d!)}{d} + \frac{\ln(C d^q)}{d} \geq \ln(\|\lambda\|_{\ell_\tau}^\tau) \quad \text{for all } d \geq d_0 \quad (5.25)$$

then the problem S_I is polynomially tractable. If this even holds for $q = 0$ then we obtain strong polynomial tractability.

Proof. Like for the symmetric setting, the proof of these sufficient conditions is based on the second point of Theorem 2.8. We show that under the given assumptions for some $\tau \geq \tau_0$ the whole sum of the eigenvalues

$$\sum_{\mathbf{k} \in \nabla_d} \left(\tilde{\lambda}_{d,\mathbf{k}}\right)^\tau = \sum_{i=1}^{\infty} (\lambda_{d,i})^\tau \quad (5.26)$$

is universally bounded, or polynomially bounded in d , respectively. Note that since we deal with the case $P = \mathfrak{A}$ now, the set ∇_d is given by the second line in (5.9). Moreover observe that for $d = 1$ there is no antisymmetry condition at all. That is, we have $\nabla_1 = \mathcal{M}_1 \subseteq \mathbb{N}$ and the sums in (5.26) equal $\|\lambda|_{\ell_\tau}\|^\tau \leq \|\lambda|_{\ell_{\tau_0}}\|^\tau$ in this case. Therefore, due to the hypothesis $\lambda \in \ell_{\tau_0}$, the term for $d = 1$ is finite.

Hence, let $d \geq 2$ be arbitrarily fixed. Without loss of generality we may reorder the set of coordinates such that $I_d = \{i_1, \dots, i_{a_d}\} = \{1, \dots, a_d\}$. That means, we assume partial antisymmetry with respect to the first a_d coordinates. For $s \in \mathbb{N}$ with $s \geq d$ let us define cubes of multi-indices

$$Q_{d,s} = \{1, \dots, s\}^d.$$

Furthermore, let $U_{a_d,s} = \{\mathbf{j} \in Q_{a_d,s} \mid j_1 < j_2 < \dots < j_{a_d}\}$ denote the a_d -dimensional projection of $Q_{d,s}$ which reflects the assumed antisymmetry conditions. With this notation we obtain

$$\sum_{\mathbf{k} \in \nabla_d} (\tilde{\lambda}_{d,\mathbf{k}})^\tau = \lim_{s \rightarrow \infty} \sum_{\mathbf{k} \in \nabla_d \cap Q_{d,s}} (\tilde{\lambda}_{d,\mathbf{k}})^\tau,$$

where the set of multi-indices under consideration $\nabla_d \cap Q_{d,s}$ can be represented as a subset of $U_{a_d,s} \times Q_{b_d,s}$. We will assume $b_d = d - a_d > 0$ in what follows to ensure this splitting to be non-trivial. Because of the product structure of $\tilde{\lambda}_{d,\mathbf{k}}$, $\mathbf{k} \in \nabla_d$, this implies

$$\sum_{\mathbf{k}=(\mathbf{j},\mathbf{i}) \in \nabla_d \cap Q_{d,s}} (\tilde{\lambda}_{d,\mathbf{k}})^\tau \leq \left(\sum_{\mathbf{j} \in U_{a_d,s}} \prod_{l=1}^{a_d} \lambda_{j_l}^\tau \right) \left(\sum_{\mathbf{i} \in Q_{b_d,s}} \prod_{l=1}^{b_d} \lambda_{i_l}^\tau \right). \quad (5.27)$$

Since the sequence $\lambda = (\lambda_m)_{m \in \mathbb{N}}$ is an element of $\ell_{\tau_0} \hookrightarrow \ell_\tau$ we can easily estimate the second factor for every $s \geq d$ from above by

$$\sum_{\mathbf{i} \in Q_{b_d,s}} \prod_{l=1}^{b_d} \lambda_{i_l}^\tau = \prod_{l=1}^{b_d} \sum_{m=1}^s \lambda_m^\tau = \left(\sum_{m=1}^s \lambda_m^\tau \right)^{b_d} \leq \left(\sum_{m=1}^{\infty} \lambda_m^\tau \right)^{1/\tau \cdot b_d \cdot \tau} = \|\lambda|_{\ell_\tau}\|^{b_d \cdot \tau}. \quad (5.28)$$

To handle the first term we need an additional argument. Note that due to the structure of $U_{a_d,s}$ we have

$$\sum_{\mathbf{j} \in Q_{a_d,s}} \prod_{l=1}^{a_d} \lambda_{j_l}^\tau = \sum_{\substack{\mathbf{j} \in Q_{a_d,s} \\ \exists k,m: j_k=j_m}} \prod_{l=1}^{a_d} \lambda_{j_l}^\tau + a_d! \sum_{\mathbf{j} \in U_{a_d,s}} \prod_{l=1}^{a_d} \lambda_{j_l}^\tau \geq a_d! \sum_{\mathbf{j} \in U_{a_d,s}} \prod_{l=1}^{a_d} \lambda_{j_l}^\tau.$$

Consequently, using the same arguments as in (5.28), this yields the upper bound $\|\lambda|_{\ell_\tau}\|^{a_d \cdot \tau} / (a_d!)$ for the first factor in (5.27). Once again this bound does not depend on $s \geq d$. Hence, due to $d = a_d + b_d$, we conclude that

$$\sum_{\mathbf{k} \in \nabla_d} (\tilde{\lambda}_{d,\mathbf{k}})^\tau = \lim_{s \rightarrow \infty} \sum_{\mathbf{k} \in \nabla_d \cap Q_{d,s}} (\tilde{\lambda}_{d,\mathbf{k}})^\tau \leq \frac{1}{a_d!} \|\lambda|_{\ell_\tau}\|^{\tau d} \quad \text{for every } d \in \mathbb{N}$$

and any choice of \mathfrak{A}_{I_d} . Of course, for every $d < d_0$ this upper bound is trivially less than an absolute constant. Thus, to prove the second assertion of this [Proposition 5.12](#) it is enough to show that

$$\frac{1}{a_d!} \|\lambda | \ell_\tau\|^{\tau d} \leq C d^q \quad \text{for all } d \geq d_0,$$

as well as for some $C \geq 1$ and some $q \geq 0$. But this is equivalent to our hypothesis stated in [\(5.25\)](#). Hence the condition [\(5.25\)](#) implies (strong) polynomial tractability of S_I , independently of the value of λ_1 .

Note that now it suffices to show that $\lambda_1 < 1$ already yields [\(5.25\)](#) with $q = 0$ and $C = 1$ in order to complete the proof. To see this, observe that (due to Step 1 in the proof of [Proposition 5.9](#)) we know that there exists some $\tau > \tau_0$ such that $\|\lambda | \ell_\tau\|^\tau = \sum_{m=1}^{\infty} \lambda_m^\tau$ is strictly less than 1. Thus the right-hand side of [\(5.25\)](#) is negative in this case, whereas the left-hand side is non-negative for every choice of a_d . \blacksquare

Let us briefly comment the latter result. Clearly, for any $q \geq 0$ the term $\ln(C d^q)/d$ in [\(5.25\)](#) tends to zero as d approaches infinity. Hence there is not much difference in the stated sufficient condition for strong polynomial and for polynomial tractability. Moreover, we need to mention that [Theorem 2.8](#) allows us to omit the largest $f(d) - 1$ eigenvalues $\lambda_{d,i}$, where $f(d)$ may grow polynomially in $(\varepsilon_d^{\text{init}})^{-1}$ with d , but we did not use this fact in the above proof.

The next example investigates how fast a_d needs to grow with the dimension d in order to fulfill the condition [\(5.25\)](#).

Example 5.13. For any $d \in \mathbb{N}$ and some $\gamma > 0$ let

$$a_d = \left\lceil \frac{d}{\ln d^\gamma} \right\rceil. \quad (5.29)$$

Then Stirling's formula provides that $a_d \ln(a_d/e) \leq \ln(a_d!) < \ln(a_d) - a_d + a_d \ln(a_d)$ if d (and hence also a_d) is sufficiently large. Consequently,

$$\frac{\ln(a_d!)}{d} \geq \frac{a_d \ln(a_d/e)}{d} \geq \frac{1}{\gamma} \cdot \frac{\ln\left(d \cdot \frac{1}{e\gamma \ln d}\right)}{\ln d} = \frac{1}{\gamma} \left(1 - \frac{\ln(e\gamma \ln d)}{\ln d}\right) \nearrow \frac{1}{\gamma},$$

as $d \rightarrow \infty$. On the other hand, we have $a_d/d \leq 1/(\gamma \ln d) + 1/d$ and thus

$$\frac{\ln(a_d!)}{d} < \frac{\ln(a_d) - a_d}{d} + \left(\frac{1}{\gamma \ln d} + \frac{1}{d}\right) \ln(a_d) = \frac{2 \ln(a_d) - a_d}{d} + \frac{1}{\gamma} \cdot \frac{\ln(a_d)}{\ln d} \leq \frac{1}{\gamma}.$$

So we see that γ in [\(5.29\)](#) needs to be strictly smaller than $\ln^{-1}(\|\lambda | \ell_\tau\|^\tau)$ in order to fulfill [\(5.25\)](#) with $q = 0$. In particular, it follows that assumptions like $a_d = \lceil d^\beta \rceil$ with $\beta < 1$ are not sufficient to conclude tractability using the second point of [Proposition 5.12](#). \square

Now we turn to necessary conditions. As in the symmetric setting [Lemma 5.6](#) yields that $\lambda \in \ell_\tau$ is needed for polynomial tractability. In addition, we will see that we need a condition similar to [\(5.25\)](#), particularly if we deal with slowly decreasing eigenvalues λ .

Proposition 5.14 (Necessary conditions, antisymmetric case). Let $P = \mathfrak{A}$ and assume $S_I = (S_{d,I_d})_{d \in \mathbb{N}}$ to denote the problem considered in [Lemma 5.6](#). Furthermore, let S_I be polynomially tractable with the constants $C, p > 0$ and $q \geq 0$. Then, for d tending to infinity, the initial error $\varepsilon_d^{\text{init}}$ tends to zero faster than the inverse of any polynomial. Moreover, $\lambda = (\lambda_m)_{m \in \mathbb{N}} \in \ell_\tau$ for every $\tau > p/2$ and there exists some $d^* \in \mathbb{N}$, as well as $C_2 \geq 1$, such that

$$\frac{1}{d} \sum_{m=1}^{a_d} \ln \left(\frac{\|\lambda| \ell_\tau\|^\tau}{\lambda_m^\tau} \right) + \frac{\ln(C_2 d^{2q\tau/p})}{d} \geq \ln(\|\lambda| \ell_\tau\|^\tau) \quad \text{for all } d \geq d^*. \quad (5.30)$$

Thus we either have $\lambda_1 < 1$, or $\lim_{d \rightarrow \infty} a_d = \infty$.

Proof. Step 1. For the whole proof assume $\tau > p/2$ to be fixed. Then [Lemma 5.6](#) shows that $\lambda \in \ell_\tau$. Like in [\(5.20\)](#) for the symmetric case, we can split the sum of the eigenvalues such that for all $d \in \mathbb{N}$

$$\sum_{\mathbf{k} \in \nabla_d} \left(\tilde{\lambda}_{d,\mathbf{k}} \right)^\tau = \left(\sum_{m=1}^{\#\mathcal{M}_1} \lambda_m^\tau \right)^{b_d} \sum_{\substack{j \in \mathcal{M}_{a_d}, \\ j_1 < \dots < j_{a_d}}} \left(\tilde{\lambda}_{a_d,j} \right)^\tau \geq \|\lambda| \ell_\tau\|^{\tau b_d} \cdot \lambda_1^\tau \cdot \dots \cdot \lambda_{a_d}^\tau.$$

Hence [Lemma 5.6](#) together with the fact that $\lambda_{d,1} = \lambda_1^{b_d} \cdot \lambda_1 \cdot \dots \cdot \lambda_{a_d}$ gives

$$\left(\frac{\|\lambda| \ell_\tau\|^\tau}{\lambda_1^\tau} \right)^{b_d} \leq (1 + C) d^q + C^{2\tau/p} \zeta \left(\frac{2\tau}{p} \right) \left(\frac{d^{2q/p}}{\lambda_{d,1}} \right)^\tau. \quad (5.31)$$

In what follows we will use this inequality to conclude all the stated assertions.

Step 2. Here we prove the limit property for the initial error $\varepsilon_d^{\text{init}} = \sqrt{\lambda_{d,1}}$, i.e. we need to show that for every fixed polynomial $\mathcal{P} > 0$

$$\lambda_{d,1} \mathcal{P}(d) \longrightarrow 0, \quad \text{as } d \rightarrow \infty. \quad (5.32)$$

Since $\lambda_{d,1} \leq \lambda_1^{b_d} \cdot \lambda_1^{a_d} = \lambda_1^d$ we can restrict ourselves to the non-trivial case $\lambda_1 \geq 1$ in the following. Assume that there exists a subsequence $(d_k)_{k \in \mathbb{N}}$ of natural numbers, as well as some constant $C_0 > 0$, such that $\lambda_{d_k,1} \mathcal{P}(d_k)$ is bounded from below by C_0 for every $k \in \mathbb{N}$. Then for every $d = d_k$ the right-hand side of [\(5.31\)](#) is bounded from above by some other polynomial $\mathcal{P}_1(d_k) > 0$. On the other hand, due to the general condition $\lambda_2 > 0$, the term $\|\lambda| \ell_\tau\|^\tau / \lambda_1^\tau$ is strictly larger than one. Thus it follows that there exists some $C_1 > 0$ such that

$$b_{d_k} \leq C_1 \ln(d_k) \quad \text{for every } k \in \mathbb{N}.$$

Therefore we obtain that $a_{d_k} = d_k - b_{d_k} \rightarrow \infty$, as $k \rightarrow \infty$. Moreover, the assumed boundedness of $\lambda_{d_k,1} \mathcal{P}(d_k)$ leads to

$$C_0 \mathcal{P}(d_k)^{-1} \leq \lambda_{d_k,1} \leq \lambda_1^{C_1 \ln(d_k)} \cdot \lambda_1 \cdot \dots \cdot \lambda_{a_{d_k}} = d_k^{C_1 \ln(\lambda_1)} \cdot \lambda_1 \cdot \dots \cdot \lambda_{a_{d_k}}$$

since $\lambda_1 \geq 1$. In the first step of the proof of [Proposition 5.9](#) we saw that $\lambda \in \ell_\tau$ yields the existence of some $C_\tau > 0$ such that $\lambda_m \leq C_\tau m^{-1/\tau}$ for every $m \in \mathbb{N}$. Indeed, this holds for $C_\tau = \|\lambda\|_{\ell_\tau} > 1$. Hence $\lambda_1^\tau \cdot \dots \cdot \lambda_{a_{d_k}}^\tau \leq C_\tau^{\tau a_{d_k}} (a_{d_k}!)^{-1}$ which gives

$$\left(\frac{a_{d_k}}{e}\right)^{a_{d_k}} \leq a_{d_k}! \leq (C_\tau^\tau)^{a_{d_k}} \mathcal{P}_2(d_k) \quad \text{for all } k \in \mathbb{N}$$

and some other polynomial $\mathcal{P}_2 > 0$. If k is sufficiently large then we conclude that

$$a_{d_k} \leq a_{d_k} \ln\left(\frac{a_{d_k}}{e C_\tau^\tau}\right) \leq \ln(\mathcal{P}_2(d_k)),$$

since $a_{d_k} \rightarrow \infty$ implies $a_{d_k}/(e C_\tau^\tau) \geq e$ for $k \geq k_0$. Therefore the number of antisymmetric coordinates a_d needs to be logarithmically bounded from above for every d out of the sequence $(d_k)_{k \geq k_0}$. Because also b_{d_k} was found to be logarithmically bounded this is a contradiction to the fact $d_k = a_{d_k} + b_{d_k}$. Consequently, the hypothesis $\lambda_{d_k,1} \mathcal{P}(d_k) \geq C_0 > 0$ can not be true for any subsequence $(d_k)_k$. In other words, it holds [\(5.32\)](#).

Step 3. Next we show [\(5.30\)](#). From the former step we know that there needs to exist some $d^* \in \mathbb{N}$ such that $1/\lambda_{d,1} \geq 1$ for all $d \geq d^*$. Hence, [\(5.31\)](#) together with $\tau > p/2$ implies

$$\left(\frac{\|\lambda\|_{\ell_\tau}^\tau}{\lambda_1^\tau}\right)^{b_d} \leq C_2 \left(\frac{d^{2q/p}}{\lambda_{d,1}}\right)^\tau = \frac{C_2 d^{2q\tau/p}}{\lambda_1^{\tau b_d} \cdot \lambda_1^\tau \cdot \dots \cdot \lambda_{a_d}^\tau} \quad \text{for } d \geq d^*,$$

where we set $C_2 = 1 + C + C^{2\tau/p} \zeta(2\tau/p)$. Therefore we obtain

$$C_2 d^{2q\tau/p} \prod_{k=1}^{a_d} \frac{\|\lambda\|_{\ell_\tau}^\tau}{\lambda_k^\tau} \geq \|\lambda\|_{\ell_\tau}^{\tau d}$$

for all $d \geq d^*$, which is equivalent to the claimed estimate [\(5.30\)](#).

Step 4. It remains to show that $\lambda_1 \geq 1$ implies that $\lim_{d \rightarrow \infty} a_d$ is infinite. To this end, note that every summand in [\(5.30\)](#) is strictly positive. If we assume for a moment the existence of a subsequence $(d_k)_{k \in \mathbb{N}}$ such that a_{d_k} is bounded for every $k \in \mathbb{N}$ then the left-hand side of [\(5.30\)](#) is less than some positive constant divided by d_k . Hence it tends to zero if k approaches infinity. On the other hand, the right-hand side of [\(5.30\)](#) is strictly larger than some positive constant, because of $\lambda_1 \geq 1$ and $\lambda_2 > 0$. This contradiction completes the proof. \blacksquare

As mentioned before there are examples such that the sufficient condition (5.25) from Proposition 5.12 is also necessary (up to some constant factor) in order to conclude polynomial tractability in the antisymmetric setting. Now we are ready to give such an example.

Example 5.15. Consider the situation of Lemma 5.6 for $P = \mathfrak{A}$ and assume the problem S_I to be polynomially tractable. In addition, for a fixed $\tau \in (0, \infty)$, let $\lambda = (\lambda_m)_{m \in \mathbb{N}} \in \ell_\tau$ be given such that $\lambda_1 \geq 1$ and assume the existence of some $m_0 \in \mathbb{N}$ such that

$$\lambda_m \geq \frac{\|\lambda | \ell_\tau\|}{m^{\alpha/\tau}} \quad \text{for all } m > m_0 \quad \text{and some } \alpha > 1. \quad (5.33)$$

Then we claim that there exist constants $\bar{d} \in \mathbb{N}$, $C \geq 1$, and $r \geq 0$ such that

$$\alpha \cdot \frac{\ln(a_d!)}{d} + \frac{\ln(C d^r)}{d} \geq \ln(\|\lambda | \ell_\tau\|^\tau) \quad \text{for all } d \geq \bar{d}. \quad (5.34)$$

Recall that due to Proposition 5.12, for the amount of antisymmetry a_d , it was sufficient to assume (5.34) with $\alpha = 1$ in order to conclude (strong) polynomial tractability; see (5.25). Moreover keep in mind that we know from Example 5.13 that $\ln(a_d!)/d$ tends to $1/\gamma$ if we assume a_d to be given by (5.29). Hence in the present example we have strong polynomial tractability if $\gamma < \ln^{-1}(\|\lambda | \ell_\tau\|^\tau)$, whereas the problem is polynomially intractable if $\gamma > \alpha/\ln(\|\lambda | \ell_\tau\|^\tau)$.

Before we prove the assertion it might be useful to give a concrete example where (5.33) holds true. Therefore set $\lambda_m = 1/m^2$, $\tau = m_0 = 1$, and $\alpha = 3$. Then it is easy to check that $\|\lambda | \ell_\tau\| = \zeta(2) = \pi^2/6$ and we obviously have $\lambda_1 = 1$.

To see that (5.34) holds true we can use Proposition 5.14 and, in particular, inequality (5.30). Since $\lambda_1 \geq 1$ we know that $\lim_d a_d = \infty$, i.e. $a_d > m_0$ for every d larger than some $d_1 \in \mathbb{N}$. Furthermore, note that (5.33) is equivalent to

$$\ln\left(\frac{\|\lambda | \ell_\tau\|^\tau}{\lambda_m^\tau}\right) \leq \alpha \ln(m) \quad \text{for all } m > m_0.$$

Hence if $d \geq d_1$ then we can estimate the sum in (5.30) from above by

$$\frac{1}{d} \sum_{m=1}^{a_d} \ln\left(\frac{\|\lambda | \ell_\tau\|^\tau}{\lambda_m^\tau}\right) \leq \frac{m_0}{d} \cdot \ln\left(\frac{\|\lambda | \ell_\tau\|^\tau}{\lambda_{m_0}^\tau}\right) + \frac{\alpha}{d} \sum_{m=m_0+1}^{a_d} \ln(m) \leq \frac{C_\lambda}{d} + \alpha \cdot \frac{\ln(a_d!)}{d}.$$

Obviously, for d larger than some $d_2 \in \mathbb{N}$ the term $C_\lambda + \ln(C_2 d^{2q\tau/p})$ is less than $\ln(C d^r)$, where $C \geq 1$ and $r \geq 0$. Here $r = 0$ if and only if $q = 0$ in (5.14), i.e. if the problem is strongly polynomially tractable. Consequently we can conclude (5.34) from (5.30) by choosing $\bar{d} = \max\{d_1, d_2, d^*\}$. \square

Although there remains a small gap between the necessary and the sufficient conditions for the absolute error criterion, the most important cases of antisymmetric tensor product problems are covered by our results. Let us summarize the main facts.

Theorem 5.16 (Tractability of antisymmetric problems, absolute errors). Let $S_1: H_1 \rightarrow \mathcal{G}_1$ denote a compact linear operator between two Hilbert spaces and let $\lambda = (\lambda_m)_{m \in \mathbb{N}}$ be the sequence of eigenvalues of $W_1 = S_1^\dagger S_1$ w.r.t. a non-increasing ordering. Moreover, for $d > 1$ let $\emptyset \neq I_d \subseteq \{1, \dots, d\}$ and assume $S_I = (S_{d, I_d})_{d \in \mathbb{N}}$ to be the linear tensor product problem $S = (S_d)_{d \in \mathbb{N}}$ restricted to the I_d -antisymmetric subspaces $\mathfrak{A}_{I_d}(H_d)$ of the d -fold tensor product spaces H_d . Consider the worst case setting with respect to the absolute error criterion and let $\lambda_2 > 0$. Then for the case $\lambda_1 < 1$ the following statements are equivalent:

- S_I is strongly polynomially tractable.
- S_I is polynomially tractable.
- There exists a constant $\tau \in (0, \infty)$ such that $\lambda \in \ell_\tau$.

Moreover, the same equivalences hold true if $\lambda_1 \geq 1$ and $\#I_d$ grows linearly with the dimension d .

At this point we mention that for the case of fully antisymmetric problems, i.e. for $\#I_d = a_d = d$, an explicit formula for the information complexity w.r.t. the absolute error criterion is known. Furthermore, simple examples can be constructed which show that we cannot expect the same nice tractability behavior if we deal with normalized errors. For further details the interested reader is referred to [Wei11, Proposition 8].

5.3.4 Antisymmetric problems (normalized errors)

Up to now every complexity assertion in this chapter was mainly based on [Theorem 2.8](#) which dealt with the general situation of arbitrary compact linear operators between Hilbert spaces and with the absolute error criterion. While investigating tractability properties of I -symmetric problems with respect to the normalized error criterion, we were able to use assertions from the absolute error setting. Since for I -antisymmetric problems the structure of the initial error is more complicated, this approach will not work again. Therefore we recall [Theorem 2.9](#) as a replacement of [Theorem 2.8](#) for the normalized setting. This in hand, we can give the following necessary conditions for (strong) polynomial tractability.

Proposition 5.17 (Necessary conditions, antisymmetric case). Let $S_I = (S_{d, I_d})_{d \in \mathbb{N}}$ denote an I -antisymmetric problem as defined at the beginning of [Section 5.3](#) and consider the worst case setting w.r.t. to normalized errors. Then the fact that S_I is polynomially tractable with the constants $C, p > 0$ and $q \geq 0$ implies that $\lambda = (\lambda_m)_{m \in \mathbb{N}} \in \ell_\tau$ for all $\tau > p/2$. Moreover, for d tending to infinity, $\varepsilon_d^{\text{init}}$ tends to zero faster than the inverse of any polynomial and $b_d \in \mathcal{O}(\ln d)$, as $d \rightarrow \infty$. Thus we have $\lim_{d \rightarrow \infty} a_d/d = 1$. In addition, if S_I is strongly polynomially tractable then $b_d \in \mathcal{O}(1)$, as $d \rightarrow \infty$.

Proof. From [Theorem 2.9](#) it follows that there is some $C_1 > 0$ such that

$$\frac{1}{(\lambda_{d,1})^\tau} \sum_{\mathbf{k} \in \nabla_d} \left(\tilde{\lambda}_{d,\mathbf{k}} \right)^\tau = \sum_{i=1}^{\infty} \left(\frac{\lambda_{d,i}}{\lambda_{d,1}} \right)^\tau \leq C_1 d^{2\tau q/p} \quad \text{for every } d \in \mathbb{N} \quad (5.35)$$

and all $\tau > p/2$. Once more the index set ∇_d is given as in [\(5.9\)](#). Indeed, [Theorem 2.9](#) yields that it is sufficient to take $C_1 = 2(1+C)^{2\tau/p} \zeta(2\tau/p)$. As in the proof of [Lemma 5.6](#) it suffices to consider the case $d = 1$ in [\(5.35\)](#) to see that $\lambda \in \ell_\tau$ is necessary for polynomial tractability. Moreover, like with the arguments of Step 1 in the proof of [Proposition 5.14](#), it follows that

$$\left(\frac{\|\lambda\| \ell_\tau^\tau}{\lambda_1^\tau} \right)^{b_d} \leq C_1 d^{2\tau q/p}, \quad d \in \mathbb{N}, \quad (5.36)$$

since $\lambda_{d,1} = \lambda_1^{b_d} \cdot \lambda_1 \cdot \dots \cdot \lambda_{a_d}$. Due to the general assertion $\lambda_2 > 0$ we have $\|\lambda\| \ell_\tau^\tau > \lambda_1^\tau$ and thus polynomial tractability of S_I implies the bound $b_d \leq C_2 \ln(d)$ for some $C_2 \geq 0$, i.e. $b_d \in \mathcal{O}(\ln d)$, as $d \rightarrow \infty$. Therefore we obviously have

$$1 \geq \frac{a_d}{d} = 1 - \frac{b_d}{\ln d} \cdot \frac{\ln d}{d} \geq 1 - C_2 \cdot \frac{\ln d}{d} \rightarrow 1, \quad d \rightarrow \infty.$$

The proof that strong polynomial tractability leads to $b_d \in \mathcal{O}(1)$, as $d \rightarrow \infty$, can be obtained using [\(5.36\)](#) with the same arguments as before and $q = 0$. Finally we need to show the assertion concerning $\varepsilon_d^{\text{init}}$. Here we refer to Step 2 in the proof of [Proposition 5.14](#). ■

5.4 Applications

This last section of the present chapter is devoted to applications of the theory developed previously. In [Section 5.4.1](#) we follow the lines of the introduction of [\[Wei12a\]](#) and illustrate the power of imposing additional (anti)symmetry conditions to linear tensor product problems by using simple toy examples. Afterwards, in [Section 5.4.2](#), we focus our attention to more advanced problems which we are faced with in practice. There we briefly introduce wavefunctions and show how our results allow it to handle the approximation problem for such classes of functions.

5.4.1 Toy examples

The aim of the following simple examples is to show that exploiting an a priori knowledge about (anti)symmetries of a given tensor product problem can help to obtain tractability, but it does not make the problem trivial in general.

Let $S = (S_d: H_d \rightarrow \mathcal{G}_d)_{d \in \mathbb{N}}$ denote a tensor product problem between Hilbert spaces. Remember that due to [Section 2.4](#) for complexity studies it suffices to specify the singular values of the univariate operator S_1 . To simplify the presentation we

slightly abuse the notation and denote the information complexity of the entire problem S by $n^{\text{ent}}(\varepsilon, d)$. We want to compare this quantity with the respective information complexities of the restriction of S to the fully symmetric and the fully antisymmetric subspaces of $(H_d)_{d \in \mathbb{N}}$. These numbers will be denoted by $n^{\text{sym}}(\varepsilon, d)$ and $n^{\text{asy}}(\varepsilon, d)$, respectively.

Clearly, our results yield that in any case (as long as we deal with the worst case setting and the absolute error criterion)

$$n^{\text{asy}}(\varepsilon, d) \leq n^{\text{sym}}(\varepsilon, d) \leq n^{\text{ent}}(\varepsilon, d) \quad \text{for every } \varepsilon > 0 \quad \text{and all } d \in \mathbb{N},$$

where for $d = 1$ the terms coincide, since then we do not claim any (anti)symmetry. To see that additional (anti)symmetry conditions may reduce the information complexity dramatically consider the following three examples.

Example 5.18. Let us have a look at the simple case of a linear operator S_1 with singular values σ such that $\lambda_1 = \lambda_2 = 1$ and $\lambda_j = 0$ for $j \geq 3$. Then the information complexity of the entire tensor product problem can be shown to be

$$n^{\text{ent}}(\varepsilon, d) = 2^d \quad \text{for all } d \in \mathbb{N} \quad \text{and } \varepsilon < 1.$$

Hence the problem suffers from the curse of dimensionality and is therefore intractable.

On the other hand, our results show that in the fully symmetric setting we have polynomial tractability, because

$$n^{\text{sym}}(\varepsilon, d) = d + 1 \quad \text{for all } d \in \mathbb{N} \quad \text{and } \varepsilon < 1.$$

Moreover, it can be proved that in this case the complexity of the fully antisymmetric problem decreases with increasing dimension d and, finally, the problem even gets trivial. In detail, we have

$$n^{\text{asy}}(\varepsilon, d) = \max \{3 - d, 0\} \quad \text{for all } d \in \mathbb{N} \quad \text{and } \varepsilon < 1$$

which yields strong polynomial tractability. □

Example 5.19. Next let us consider a more challenging problem, where $\lambda_1 = \lambda_2 = \dots = \lambda_m = 1$ and $\lambda_j = 0$ for every $j > m \geq 2$. For $m = 2$ this obviously coincides with the example studied above, but letting m increase may tell us more about the structure of (anti)symmetric tensor product problems. In this situation it is easy to check that for every $d \in \mathbb{N}$ and all $\varepsilon < 1$

$$n^{\text{ent}}(\varepsilon, d) = m^d \quad \text{and} \quad n^{\text{asy}}(\varepsilon, d) = \begin{cases} \binom{m}{d}, & d \leq m, \\ 0, & d > m. \end{cases}$$

Since $\binom{m}{d} \geq 2^{d-1}$ for $d \leq \lfloor m/2 \rfloor$, this means that for large m the complexity in the antisymmetric case increases exponentially fast with d up to a certain maximum. Beyond this point it falls back to zero.

The information complexity in the symmetric setting is much harder to calculate for this case. However, it can be seen that we have polynomial tractability, but $n^{\text{sym}}(\varepsilon, d)$ needs to grow at least linearly with d such that the symmetric problem cannot be strongly polynomially tractable, whereas this holds in the antisymmetric setting. The entire problem again suffers from the curse of dimensionality. \square

Example 5.20. For a last illustrating example consider the case $\lambda_1 = 1$ and $\lambda_{j+1} = j^{-\beta}$ for some $\beta \geq 0$ and all $j \in \mathbb{N}$. That means, we have the two largest singular values $\sigma_1 = \sigma_2$ of S_1 equal to one. The remaining series decays like the inverse of some polynomial. If $\beta = 0$ then the operator S_1 is not compact, since the sequence $\lambda = (\lambda_m)_{m \in \mathbb{N}}$ does not tend to zero; hence all the information complexities are infinite in this case.

For $\beta > 0$, any $\delta > 0$, and some $C > 0$ we have

$$n^{\text{ent}}(\varepsilon, d) \geq 2^d, \quad n^{\text{sym}}(\varepsilon, d) \geq d + 1, \quad \text{and} \quad n^{\text{asy}}(\varepsilon, d) \leq C\varepsilon^{-(2/\beta+\delta)},$$

for all $\varepsilon < 1$, as well as every $d \in \mathbb{N}$. Thus, again for the entire problem we observe the curse, whereas the antisymmetric problem is strongly polynomially tractable. Once more, the symmetric problem can be shown to be polynomially tractable. Note that in this example the antisymmetric case is not trivial, because all λ_j are strictly positive. If we replace $j^{-\beta}$ by $\log^{-1}(j+1)$ in this example we obtain (polynomial) intractability even in the antisymmetric setting. \square

5.4.2 Wavefunctions

During the few last decades there has been considerable interest in finding approximations of so-called *wavefunctions*, e.g., solutions of the electronic Schrödinger equation. Due to the *Pauli principle* of quantum physics only functions with certain (anti)symmetry properties are of physical interest. For a more detailed view see, e.g. Hamaekers [Ham09], Yserentant [Yse10], or Zeiser [Zei10]. Furthermore, for a comprehensive introduction to the topic, as well as a historical survey, we refer the reader to Hunziker and Sigal [HS00] and Reed and Simon [RS78].

In particular, the notion of multiple partial antisymmetry w.r.t. two sets of coordinates is useful for describing wavefunctions Ψ . In computational chemistry such functions occur as models which describe quantum states of certain physical d -particle systems. Formally, these functions depend on d blocks of variables $\mathbf{y}_i = (\mathbf{x}^{(i)}, s^{(i)})$, for $i = 1, \dots, d$, which represent the spacial coordinates $\mathbf{x}^{(i)} = (x_1^{(i)}, x_2^{(i)}, x_3^{(i)}) \in \mathbb{R}^3$ and certain additional intrinsic parameters $s^{(i)} \in C$ of each particle \mathbf{y}_i within the system. Hence, rearranging the arguments such that $\mathbf{x} = (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(d)})$ and $\mathbf{s} = (s^{(1)}, \dots, s^{(d)})$ yields that

$$\Psi: (\mathbb{R}^3)^d \times C^d \rightarrow \mathbb{R}, \quad (\mathbf{x}, \mathbf{s}) \mapsto \Psi(\mathbf{x}, \mathbf{s}).$$

In the case of systems of electrons one of the most important parameters is called *spin* and it can take only two values, i.e., $s^{(i)} \in C = \{-\frac{1}{2}, +\frac{1}{2}\}$. Due to the Pauli

principle the only wavefunctions Ψ that are physically admissible are those which are antisymmetric in the sense that for $I \subseteq \{1, \dots, d\}$ and $I^c = \{1, \dots, d\} \setminus I$

$$\Psi(\boldsymbol{\pi}(\mathbf{x}), \boldsymbol{\pi}(\mathbf{s})) = (-1)^{|\pi|} \Psi(\mathbf{x}, \mathbf{s}) \quad \text{for all } \pi \in \mathcal{S}_I \cup \mathcal{S}_{I^c}.$$

Thus Ψ changes its sign if we replace any particles \mathbf{y}_i and \mathbf{y}_j by each other which possess the same spin, i.e. $s^{(i)} = s^{(j)}$. So the set of particles, and therefore also the set of spacial coordinates, naturally split into two groups I_+ and I_- . In detail, for wavefunctions of d particles \mathbf{y}_i we can (without loss of generality) assume that the first $\#I_+$ indices i belong to the group of positive spin, whereas the rest of the particles possess negative spin, i.e. $I_+ = \{1, \dots, \#I_+\}$ and $I_- = I_+^c = \{\#I_+ + 1, \dots, d\}$.

In physics it is well-known that some problems, e.g., the electronic Schrödinger equation, which involve (general) wavefunctions can be reduced to a bunch of similar problems, where each of them only acts on functions $\Psi_{\mathbf{s}}$ out of a certain Hilbert space $\mathcal{F}_d = \mathcal{F}_d(\mathbf{s})$. That is,

$$\Psi_{\mathbf{s}} = \Psi(\cdot, \mathbf{s}) \in \mathcal{F}_d \subset \{f: (\mathbb{R}^3)^d \rightarrow \mathbb{R}\}$$

with a given fixed spin configuration $\mathbf{s} \in C^d$. Of course every possible spin configuration \mathbf{s} corresponds to exactly one choice $I_+ \subseteq \{1, \dots, d\}$ of indices. Moreover, it is known that \mathcal{F}_d is a Hilbert space which possesses a tensor product structure. Therefore we can model wavefunctions as elements of certain classes of smoothness, e.g., $\mathcal{F}_d \subset H_d = H_1 \otimes \dots \otimes H_1 = W_2^{(1, \dots, 1)}((\mathbb{R}^3)^d)$, as Yserentant [Yse10] recently did, and incorporate spin properties by using projections of the type $\mathfrak{A} = \mathfrak{A}_{I_+} \circ \mathfrak{A}_{I_-}$, as defined in Section 5.1.1. In particular, Lemma 5.2 then yields that

$$\mathcal{F}_d = \mathfrak{A}(H_d) = \mathfrak{A}_{I_+}(H_{\#I_+}) \otimes \mathfrak{A}_{I_-}(H_{\#I_-})$$

and the system of all

$$\bar{\xi}_{\mathbf{k}} = \sqrt{\#\mathcal{S}_{I_+} \cdot \#\mathcal{S}_{I_-}} \cdot \mathfrak{A}(e_{\mathbf{k}}), \quad \mathbf{k} \in \bar{\nabla}_d,$$

with

$$\bar{\nabla}_d = \{\mathbf{k} = (\mathbf{i}, \mathbf{j}) \in \mathbb{N}^{\#I_+} \times \mathbb{N}^{\#I_-} \mid i_1 < i_2 < \dots < i_{\#I_+} \text{ and } j_1 < \dots < j_{\#I_-}\}$$

builds an orthonormal basis of $\mathcal{F}_d = \mathfrak{A}(H_d)$, where the set $\{e_{\mathbf{m}} \mid \mathbf{m} \in \mathbb{N}^d\}$ is once again assumed to be an orthonormal tensor product basis of $H_d = H_1 \otimes \dots \otimes H_1$ constructed with the help of $\{e_m \mid m \in \mathbb{N}\}$, an arbitrary orthonormal basis of H_1 .

Note that in the former sections the underlying Hilbert space H_1 always consists of univariate functions. In contrast, wavefunctions of one particle depend on at least three (spacial) variables, but we want to stress the point that this is just a formal issue. Anyway, our approach radically decreases the degrees of freedom and improves the solvability of certain problems $S = (S_d)_{d \in \mathbb{N}}$ like the approximation

problem, i.e. $S_d = \text{id}: H_d \rightarrow \mathcal{G}_d$ for every $d \in \mathbb{N}$, considered in connection with the electronic Schrödinger equation.

[Theorem 5.4](#) provides an algorithm which is optimal for the \mathcal{G}_d -approximation of d -particle wavefunctions in \mathcal{F}_d with respect to all linear algorithms that use at most n continuous linear functionals. Therefore we only need to choose the right ONB $\{e_m = \phi_m \mid m \in \mathbb{N}\}$ of H_1 which coincides with the eigenfunctions of the univariate operator $W_1 = S_1^\dagger S_1$. Moreover, the error can be calculated exactly in terms of the eigenvalues $\lambda = (\lambda_m)_{m \in \mathbb{N}}$ of W_1 .

Furthermore it is possible to prove a modification of [Theorem 5.16](#) for problems dealing with wavefunctions. In fact, for the mentioned approximation problem polynomial tractability as well as strong polynomial tractability are equivalent to the fact that the sequence λ of the squared singular values of the univariate problem belong to some ℓ_τ -space if we consider the absolute error criterion. The reason is that all the assertions in [Section 5.3.3](#) can be easily extended to the multiple partially antisymmetric case. In detail, if we denote the number of antisymmetric coordinates $\mathbf{x}^{(i)}$ within each antisymmetry group $I_d^{(m)} \subseteq \{1, \dots, d\}$ by $a_{d,m}$ with $m = 1, \dots, M$ then the constraint $a_d + b_d = d$ extends to

$$a_{d,1} + \dots + a_{d,M} + b_d = d.$$

Here b_d again denotes the number of coordinates without any antisymmetry condition. In conclusion, the sufficient condition (5.25) in [Proposition 5.12](#) transfers to

$$\frac{1}{d} \sum_{m=1}^M \ln(a_{d,m}!) \geq \|\lambda\|_{\ell_\tau}^\tau, \quad \text{for all } d \geq d_0,$$

which is always satisfied in the case of wavefunctions, since then $M = 2$ and the cardinality $a_{d,m}$ of at least one of the groups of the same spin needs to grow linearly with the dimension d .

Bibliography

- [Aro50] N. ARONSZAJN - *Theory of reproducing kernels*. Trans. Amer. Math. Soc. **68(3)**, 1950, pp. 337–404. [[↑ 29](#) and [30](#)]
- [Bau96] H. BAUER - *Probability Theory*. de Gruyter Studies in Math. 23 (Transl. from the 4th German ed.). Walter de Gruyter & Co., Berlin. 1996. [[↑ 2](#)]
- [Bau01] H. BAUER - *Measure and Integration Theory*. de Gruyter Studies in Math. 26 (Transl. from the German). Walter de Gruyter & Co., Berlin. 2001. [[↑ 2](#)]
- [Bel57] R.E. BELLMAN - *Dynamic Programming*. Princeton Univ. Press, Princeton, NJ. 1957. [[↑ 8](#)]
- [Dei85] K. DEIMLING - *Nonlinear Functional Analysis*. Springer, Berlin. 1985. [[↑ 12](#)]
- [DKS13] J. DICK, F.Y. KUO AND I.H. SLOAN - *High-dimensional integration: the Quasi-Monte Carlo way*. Acta Numerica **22**, 2013, pp. 133 - 288. [[↑ 32](#)]
- [DP10] J. DICK AND F. PILLICHSHAMMER - *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*. Cambridge Univ. Press, Cambridge. 2010. [[↑ 32](#)]
- [Enf73] P. ENFLO - *A counterexample to the approximation problem in Banach spaces*. Acta Math. **130**, 1973, pp. 309–317. [[↑ 14](#)]
- [GW07] M. GNEWUCH AND H. WOŹNIAKOWSKI - *Generalized tractability for multivariate problems, part I: Linear tensor product problems and linear information*. J. Complexity **23(2)**, 2007, pp. 262–295. [[↑ 9](#)]
- [GW08] M. GNEWUCH AND H. WOŹNIAKOWSKI - *Generalized tractability for linear functionals*, in: A. Keller, S. Heinrich and H. Niederreiter (Eds.) - Monte Carlo and Quasi-Monte Carlo Methods 2006. Springer, Berlin. 2008, pp. 359–381. [[↑ 9](#)]
- [GW09] M. GNEWUCH AND H. WOŹNIAKOWSKI - *Generalized tractability for*

- multivariate problems, part II: Linear tensor product problems, linear information, and unrestricted tractability.* Found. Comput. Math. **9(4)**, 2009, pp. 431–460. [↑ 9]
- [GW11] M. GNEWUCH AND H. WOŹNIAKOWSKI - *Quasi-polynomial tractability.* J. Complexity **27(3–4)**, 2011, pp. 312–330. [↑ 9]
- [Ham09] J. HAMAËKERS - *Tensor Product Multiscale Many-Particle Spaces with Finite-Order Weights for the Electronic Schrödinger Equation.* Ph.D. thesis, Uni. Bonn, 2009. *urn:nbn:de:hbz:5N-18339*. [↑ 84 and 112]
- [Han10] M. HANSEN - *Nonlinear Approximation and Function Spaces of Dominating Mixed Smoothness.* Ph.D. thesis, FSU Jena, 2010. *urn:nbn:de:gbv:27-20110121-105128-4*. [↑ 22 and 86]
- [HS00] W. HUNZIKER AND I.M. SIGAL - *The quantum N -body problem.* J. Math. Phys. **41**, 2000, pp. 3448–3510. [↑ 112]
- [HW00] F.J. HICKERNELL AND H. WOŹNIAKOWSKI - *Integration and approximation in arbitrary dimensions.* Advances in Comp. Math. **12**, 2000, pp. 25–58. [↑ 34]
- [Kol05] A. KOLDOBSKY - *Fourier Analysis in Convex Geometry.* Amer. Math. Soc., Providence, RI. 2005. [↑ 77]
- [Kön86] H. KÖNIG - *Eigenvalue Distribution of Compact Operators.* Operator Theory: Advances and Applications 16. Birkhäuser, Basel. 1986. [↑ 15 and 17]
- [KR83] R.V. KADISON AND J.R. RINGROSE - *Fundamentals of the Theory of Operator Algebras. Vol. I: Elementary Theory.* Pure and Applied Mathematics 100. Academic Press Inc., New York. 1983. [↑ 22, 23, and 24]
- [KSS11] F.Y. KUO, C. SCHWAB AND I.H. SLOAN - *Quasi-Monte Carlo methods for high-dimensional integration: The standard (weighted Hilbert space) setting and beyond.* ANZIAM J. **53(1)**, 2011, pp. 1–37. [↑ 57]
- [KSWW10a] F.Y. KUO, I.H. SLOAN, G.W. WASILKOWSKI AND H. WOŹNIAKOWSKI - *Liberating the dimension.* J. Complexity **26(5)**, 2010, pp. 422–454. [↑ 7]
- [KSWW10b] F.Y. KUO, I.H. SLOAN, G.W. WASILKOWSKI AND H. WOŹNIAKOWSKI - *On decompositions of multivariate functions.* Math. Comp. **79**, 2010, pp. 953–966. [↑ 55]
- [KWW08] F.Y. KUO, G.W. WASILKOWSKI AND H. WOŹNIAKOWSKI - *Multi-*

-
- variate L_∞ approximation in the worst case setting over reproducing kernel Hilbert spaces. *J. Approx. Theory* **152(2)**, 2008, pp. 135–160. [[↑ ix](#), [xvii](#), [30](#), [59](#), [60](#), [61](#), and [80](#)]
- [LC85] W.A. LIGHT AND E.W. CHENEY - *Approximation Theory in Tensor Product Spaces*. Lecture Notes in Mathematics 1169. Springer, Berlin. 1985. [[↑ 22](#)]
- [Mat90] P. MATHÉ - *s-Numbers in information-based complexity*. *J. Complexity* **6(1)**, 1990, pp. 41–66. [[↑ 11](#)]
- [MW81] C.A. MICCHELLI AND G. WAHBA - *Design problems for optimal surface interpolation*, in: Z. Ziegler (Eds.) - *Approximation Theory and Applications (Proc. Workshop, Technion—Israel Inst. Tech., Haifa, 1980)*. Academic Press, New York. 1981, pp. 329–348. [[↑ 58](#)]
- [NW08] E. NOVAK AND H. WOŹNIAKOWSKI - *Tractability of Multivariate Problems. Vol. I: Linear Information*. EMS Tracts in Mathematics 6. European Mathematical Society (EMS), Zürich. 2008. [[↑ 4](#), [5](#), [7](#), [9](#), [11](#), [14](#), [17](#), [19](#), [20](#), [21](#), [24](#), [26](#), [27](#), [29](#), [33](#), [34](#), [37](#), [41](#), [42](#), [55](#), [57](#), [59](#), [74](#), [81](#), and [100](#)]
- [NW09] E. NOVAK AND H. WOŹNIAKOWSKI - *Approximation of infinitely differentiable multivariate functions is intractable*. *J. Complexity* **25(4)**, 2009, pp. 398–404. [[↑ iii](#), [vi](#), [xi](#), [xiv](#), [53](#), [74](#), [75](#), [76](#), and [80](#)]
- [NW10] E. NOVAK AND H. WOŹNIAKOWSKI - *Tractability of Multivariate Problems. Vol. II: Standard Information for Functionals*. EMS Tracts in Mathematics 12. European Mathematical Society (EMS), Zürich. 2010. [[↑ 11](#), [32](#), and [33](#)]
- [NW12] E. NOVAK AND H. WOŹNIAKOWSKI - *Tractability of Multivariate Problems. Vol. III: Standard Information for Linear Operators*. EMS Tracts in Mathematics 18. European Mathematical Society (EMS), Zürich. 2012. [[↑ 9](#), [11](#), [14](#), and [33](#)]
- [Pie87] A. PIETSCH - *Eigenvalues and s-Numbers*. Cambridge Studies in Advanced Math. 13. Cambridge Univ. Press, Cambridge. 1987. [[↑ 11](#)]
- [Pie07] A. PIETSCH - *History of Banach Spaces and Linear Operators*. Birkhäuser, Boston, MA. 2007. [[↑ 11](#) and [15](#)]
- [Pin85] A. PINKUS - *n-Widths in Approximation Theory*. Results in Math. and Related Areas 3(7). Springer, Berlin. 1985. [[↑ 11](#), [15](#), and [16](#)]
- [PP09] A. PAPAGEORGIU AND I. PETRAS - *On the tractability of linear tensor product problems in the worst case*. *J. Complexity* **25(5)**, 2009, pp. 415–419. [[↑ 26](#), [27](#), [43](#), and [44](#)]

- [RS78] M. REED AND B. SIMON - *Methods of Modern Mathematical Physics. Vol. IV: Analysis of Operators*. Academic Press, New York. 1978. [[↑ 112](#)]
- [Sie13] P. SIEDLECKI - *Uniform weak tractability*. To appear in: *J. Complexity*, 2013. [[↑ 9](#)]
- [SJ94] I.H. SLOAN AND S. JOE - *Lattice Methods for Multiple Integration*. Oxford Science Publ. Oxford Univ. Press, New York. 1994. [[↑ 32](#)]
- [SW98] I.H. SLOAN AND H. WOŹNIAKOWSKI - *When are quasi-Monte Carlo algorithms efficient for high-dimensional integrals?*. *J. Complexity* **14(1)**, 1998, pp. 1–33. [[↑ 54](#)]
- [SW02] I.H. SLOAN AND H. WOŹNIAKOWSKI - *Tractability of integration in non-periodic and periodic weighted tensor product Hilbert spaces*. *J. Complexity* **18(2)**, 2002, pp. 479–499. [[↑ 57](#)]
- [Tho96] C. THOMAS-AGNAN - *Computing a family of reproducing kernels for statistical applications*. *Numer. Algo.* **13(1)**, 1996, pp. 21–32. [[↑ 59](#)]
- [Tri92] H. TRIEBEL - *Higher Analysis*. Univ. Books for Math. (Transl. from the German). Johann Ambrosius Barth Verlag GmbH, Leipzig. 1992. [[↑ 2](#) and [29](#)]
- [TWW88] J.F. TRAUB, G.W. WASILKOWSKI AND H. WOŹNIAKOWSKI - *Information-based Complexity*. Academic Press Inc., Boston, MA. 1988. [[↑ 14](#)]
- [Wah90] G. WAHBA - *Spline Models for Observational Data*. Soc. Indust. Appl. Math. (SIAM), Philadelphia, PA. 1990. [[↑ 29](#)]
- [Wei11] M. WEIMAR - *The complexity of linear tensor product problems in (anti-) symmetric Hilbert spaces*. Manuscript, available at: <http://arxiv.org/abs/1111.0057>, 2011. [[↑ 19](#), [83](#), [95](#), and [109](#)]
- [Wei12a] M. WEIMAR - *The complexity of linear tensor product problems in (anti)symmetric Hilbert spaces*. *J. Approx. Theory* **164(10)**, 2012, pp. 1345–1368. [[↑ x](#), [xviii](#), [83](#), [84](#), [95](#), [97](#), [102](#), [103](#), and [110](#)]
- [Wei12b] M. WEIMAR - *Tractability results for weighted Banach spaces of smooth functions*. *J. Complexity* **28(1)**, 2012, pp. 59–75. [[↑ x](#), [xvii](#), [12](#), [30](#), [54](#), [57](#), [58](#), [59](#), [61](#), [63](#), [68](#), [69](#), and [75](#)]
- [Woź94a] H. WOŹNIAKOWSKI - *Tractability and strong tractability of linear multivariate problems*. *J. Complexity* **10(1)**, 1994, pp. 96–128. [[↑ 8](#)]
- [Woź94b] H. WOŹNIAKOWSKI - *Tractability and strong tractability of multivariate tensor product problems*. *J. of Computing and Information* **4**, 1994, pp. 1–19. [[↑ 8](#), [37](#), and [43](#)]

-
- [WW07] A.G. WERSCHULZ AND H. WOŹNIAKOWSKI - *Tractability of quasi-linear problems I: General results*. J. Approx. Theory **145(2)**, 2007, pp. 266–285. [[↑ 14](#)]
- [WW09] A.G. WERSCHULZ AND H. WOŹNIAKOWSKI - *Tractability of multivariate approximation over a weighted unanchored Sobolev space*. Constr. Approx. **30**, 2009, pp. 395–421. [[↑ 12](#), [58](#), and [59](#)]
- [Yos80] K. YOSIDA - *Functional Analysis*. Fund. Principles of Math. Sci. 123 (6th ed.). Springer-Verlag, Berlin. 1980. [[↑ 2](#), [16](#), [29](#), and [60](#)]
- [Yse10] H. YSERENTANT - *Regularity and Approximability of Electronic Wave Functions*. Lecture Notes in Mathematics. Springer-Verlag, Berlin. 2010. [[↑ 86](#), [112](#), and [113](#)]
- [Zei10] A. ZEISER - *Direkte Diskretisierung der Schrödingergleichung auf Dünnen Gittern*. Ph.D. thesis, TU Berlin, 2010. *urn:nbn:de:kobv:83-opus-27910*. [[↑ 85](#), [92](#), and [112](#)]
- [ZZ13] H. ZHANG AND J. ZHANG - *Vector-valued reproducing kernel Banach spaces with applications to multi-task learning*. J. Complexity **29(2)**, 2013, pp. 195–215. [[↑ 30](#)]

Lebenslauf

Persönliche Daten

Name: Markus Weimar
Geburtsdatum: 28.02.1986
Geburtsort: Weimar
Nationalität: deutsch

Akademischer Werdegang

05/2004 Abitur am Staatlichen Gymnasium “Bergschule” Apolda
10/2004–09/2009 Studium der Mathematik an der Friedrich-Schiller-Universität (FSU) Jena, partiell gefördert durch die *Studienstiftung des deutschen Volkes*
Diplomarbeit: *Das Haar’sche System in Funktionenräumen vom Typ $A_{p,q}^s(\mathbb{R})$*
10/2009–03/2010 Wissenschaftliche Hilfskraft am Mathematischen Institut der FSU Jena
12/2009–05/2013 Doktorand am Lehrstuhl für Theoretische Numerik der FSU Jena bei Prof. Dr. Erich Novak
04/2010–04/2013 Wissenschaftlicher Mitarbeiter der FSU Jena
10/2012 Vierwöchiger Forschungsaufenthalt bei der *Numerical Integration, Nonlinear Equations & Software (NINES)* Gruppe der *Katholieke Universiteit Leuven* (Belgien)
03/2012 Zweiwöchiger Forschungsaufenthalt bei der *NINES* Gruppe der *Katholieke Universiteit Leuven* (Belgien)

Seit 05/2013 Wissenschaftlicher Mitarbeiter der Philipps-Universität Marburg

Publikationen

Tractability results for weighted Banach spaces of smooth functions.

J. Complexity **28(1)**, 2012, pp. 59–75.

[doi:10.1016/j.jco.2011.03.004](https://doi.org/10.1016/j.jco.2011.03.004)

The complexity of linear tensor product problems in (anti)symmetric Hilbert spaces.

J. Approx. Theory **164(10)**, 2012, pp. 1345–1368.

[doi:10.1016/j.jat.2012.05.016](https://doi.org/10.1016/j.jat.2012.05.016)

Probabilistic star discrepancy bounds for double infinite random matrices (gemeinsam mit C. Aistleitner).

Erscheint in: J. Dick, F.Y. Kuo, G.W. Peters, and I.H. Sloan (Eds.) - *Monte Carlo and Quasi-Monte Carlo Methods 2012*. Springer, Berlin.

Konferenzen, Seminare und Workshops

2009–2013 Oberseminare “Theoretische Numerik” und “Function Spaces”, Jena
Verschiedene Vorträge

06/2010 Workshop Sparsity & Computation, Bonn

07/2010 Workshop Analysis & Numerics Siegmundsburg (WANS’10), Siegmundsburg
Vortrag: *Komplexität hochdimensionaler Probleme - Eine Einführung*

07/2010 54. Workshop Approximationstheorie, Jena

08/2010 9th International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing (MCQMC2010), Warschau (Polen)

10/2010	International Workshop on Smoothness, Approximation and Function Spaces (SAFS2010), Oppurg Vortrag: <i>On error bounds for L_∞-approximation of smooth functions</i>
06/2011	4th Workshop on High-Dimensional Approximation (HDA2011), Bonn Vortrag: <i>Tractability of L_∞-approximation on weighted Banach spaces of smooth functions</i>
07/2011	Foundations of Comput. Mathematics Conference (FoCM2011), Budapest (Ungarn)
02/2012	Tenth International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing (MCQMC2012), Sydney (Australien) Vortrag: <i>Linear tensor product problems in (anti-) symmetric Hilbert spaces</i>
06/2012	3rd International Conference on Uniform Distribution Theory (UDT2012), Smolenice (Slowakei) Vortrag: <i>Probabilistic star discrepancy bounds for double infinite random matrices</i>
09/2012	Dagstuhl Seminar 12391: <i>Algorithms and Complexity for Continuous Problems</i> , Wadern Vortrag: <i>Probabilistic star discrepancy bounds for double infinite random matrices</i>

Lehre

Übungen zu Vorlesungen am Mathematischen Institut der FSU Jena:

WS 2009/2010	2x Analysis 3 (Lehramt Gymnasium), Dr. Haberland
SS 2010	Analysis 2 (Mathe / WiMa B.Sc), Prof. Carl
WS 2010/2011	Analysis 1 (Physik B.Sc), Prof. Hinrichs
SS 2011	Analysis 2 (Physik B.Sc), Prof. Hinrichs
WS 2011/2012	2x Einführung in die Numerische Mathematik und das Wissenschaftliche Rechnen, Prof. Novak
WS 2012/2013	Analysis 3 (Physik / Mathe / WiMa B.Sc), Prof. Schmeißer

Ehrenwörtliche Erklärung

Ich erkläre hiermit, dass mir die Promotionsordnung der Fakultät bekannt ist.

Ferner erkläre ich, dass ich die vorliegende Arbeit selbst und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Weitere Personen waren nicht an der inhaltlich-materiellen Erstellung der Arbeit beteiligt. Insbesondere habe ich hierfür nicht die Hilfe eines Promotionsberaters in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Arbeit stehen.

Die Arbeit wurde bisher weder im In- noch Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt. Insbesondere wurde keine in wesentlichen Teilen ähnliche oder andere Abhandlung bei einer anderen Hochschule als Dissertation eingereicht.

Ich versichere, dass ich nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen habe.

Jena, den 20.02.2013

Markus Weimar