Constr Approx (2011) 34:371–391
DOI 10.1007/s00365-010-9125-4

CONSTRUCTIVE APPROXIMATION

Regularity of Tensor Product Approximations to Square Integrable Functions

André Uschmajew

Received: 26 November 2009 / Revised: 25 June 2010 / Accepted: 14 September 2010 /
Published online: 22 December 2010

Abstract We investigate first-order conditions for canonical and optimal subspace (Tucker format) tensor product approximations to square integrable functions. They reveal that the best approximation and all of its factors have the same smoothness as the approximated function itself. This is not obvious, since the approximation is performed in L^2.

Keywords Tensor products · Low-rank approximation · Optimal subspace approximation · Sobolev spaces

Mathematics Subject Classification (2000) 15A69 · 41A46 · 49N60

1 Introduction

The L^2 function spaces, whether real- or complex-valued, possess a natural tensor product structure. For example, if we consider dimensions d_1 and $d_2 \in \mathbb{N}$, then the identity

$$L^2\big(\mathbb{R}^{d_1+d_2}\big) = L^2\big(\mathbb{R}^{d_1}\big) \otimes L^2\big(\mathbb{R}^{d_2}\big) \tag{1.1}$$

expresses the well-known fact that each function $f \in L^2(\mathbb{R}^{d_1+d_2})$ can be written as a (probably infinite) sum of products

$$f(x, y) = \sum_{k=1}^{r} u_k(x) v_k(y)$$

Communicated by: Wolfgang Dahmen.

A. Uschmajew (✉)
Institut für Mathematik, Technische Universität Berlin, Straße des 17. Juni 136, 10623 Berlin, Germany
e-mail: uschmajew@math.tu-berlin.de

of functions $u_k \in L^2(\mathbb{R}^{d_1})$ and $v_k \in L^2(\mathbb{R}^{d_1})$ almost everywhere. But (1.1) means more than that. If we set

$$(u \otimes v)(x, y) := u(x)v(y) \tag{1.2}$$

for $u \in L^2(\mathbb{R}^{d_1})$ and $v \in L^2(\mathbb{R}^{d_2})$, then the right-hand side of (1.1) is defined as the closure of the linear hull of functions $\{u \otimes v\}$ with respect to the norm induced by the inner product, which is generated from

$$(u_1 \otimes v_1, u_2 \otimes v_2)_{L^2(\mathbb{R}^{d_1})\otimes L^2(\mathbb{R}^{d_2})} := (u_1, u_2)_{L^2(\mathbb{R}^{d_1})} \cdot (v_1, v_2)_{L^2(\mathbb{R}^{d_2})} \tag{1.3}$$

via linear expansion. Here $(u_1, u_2)_{L^2(\mathbb{R}^{d_n})}$ denotes the usual inner product in $L^2(\mathbb{R}^{d_n})$. Due to Fubini's theorem, the inner product (1.3) coincides with that of $L^2(\mathbb{R}^{d_1+d_2})$:

$$(u_1 \otimes u_2, v_1 \otimes v_2)_{L^2(\mathbb{R}^{d_1})\otimes L^2(\mathbb{R}^{d_2})} = (u_1 \otimes u_2, v_1 \otimes v_2)_{L^2(\mathbb{R}^{d_1+d_2})}.$$

This is why (1.1) holds in the sense of Hilbert spaces.[1] The observation can be generalized to arbitrary dimensions and orders, that is,

$$L^2\big(\mathbb{R}^{\mathbf{d}}\big) := L^2\big(\mathbb{R}^{d_1+d_2+\cdots+d_N}\big) = \bigotimes_{n=1}^{N} L^2\big(\mathbb{R}^{d_n}\big)$$

for arbitrary positive integers $N, d_1, d_2, \ldots, d_N$. In the following, we will use $\mathbb{R}^{\mathbf{d}}$ with a multi-index $\mathbf{d} = (d_1, d_2, \ldots, d_N)$ as an abbreviation for $\mathbb{R}^{d_1+d_2+\cdots+d_N}$.

Given a function $f \in L^2(\mathbb{R}^{\mathbf{d}})$, there are a lot of theoretical and practical reasons one would like to obtain a decomposition

$$f = \sum_{k=1}^{r} u_k^1 \otimes u_k^2 \otimes \cdots \otimes u_k^N \tag{1.4}$$

with $u_k^n \in L^2(\mathbb{R}^{d_n})$ and small r. For instance, this would allow an integration of f to be performed much more easily and with a significantly reduced numeric cost. In fact, such decompositions seem suitable for breaking the "curse of dimensionality," which arises in the numerical treatment of high-dimensional problems, see, e.g., [8] for a survey. However, it may be quite difficult or impossible to obtain such a decomposition. A simple expansion of f into tensor products of fixed basis functions of $L^2(\mathbb{R}^{d_n})$ does not solve the problem, since it usually leads to $r = \infty$. In fact, the case that an arbitrary chosen function f from $L^2(\mathbb{R}^{\mathbf{d}})$ admits a finite representation (1.4) can be expected to be an exception, since, by definition, $L^2(\mathbb{R}^{\mathbf{d}})$ is the completion of the set of functions with this property. There may be hope for functions with special structure coming from special applications, but this is not the subject of this article.

[1] Of course the Lebesgue measure is constructed in this special way. The point is that an identity like (1.1) is not necessarily true for other Hilbert spaces of functions such as H^1.

The minimal $r \in \mathbb{N} \cup \{\infty\}$ needed in (1.4) for a decomposition of f is called the *rank* of f. If the rank is too large or even infinite, one seeks for an approximation of low rank $r < \infty$, that is, one tries to solve the problem[2]

$$\left\| f - \sum_{k=1}^{r} u_k^1 \otimes u_k^2 \otimes \cdots \otimes u_k^N \right\|_0 = \min \tag{1.5}$$

for $u_k^n \in L^2(\mathbb{R}^{d_n})$. This is called the *best canonical rank-r approximation* and can be seen as a nonlinear m-term approximation (with $m = r$) for which the dictionary consists of all rank-one functions. We ask the following question: Which regularity can we expect the factors u_k^n of a best approximation to have in comparison to the regularity of f? This question is important for instance in quantum chemistry [2, 3, 5, 9, 10], where tensor products are used to approximate wave functions of atoms and molecules, and such wave functions have to lie in the Sobolev space H^1 to be physically meaningful. More generally, the regularity of tensor product approximations is relevant in the treatment of high-dimensional partial differential equations by tensor methods. It may also be useful from an approximation theoretical point of view, e.g., for estimating convergence rates of procedures solving (1.5) or related questions. Furthermore, a high regularity of the factors allows for a further compression and sparse representation of the solution.

Unfortunately, minimization problem (1.5) can be ill-posed[3] and does not admit a solution for every choice of f, see [11] for this issue. To prevent our results from being too vacuous we will focus on local minima of (1.5). Additionally, we analyze in Sect. 4.5 a regularized minimization problem used in practice.

There are other tensor product approximation models, which have the benefit of being well-posed problems because they contain more structure. In this paper we shall also investigate the approximation of f by a tensor from the so-called *optimal subspace*, also known as *Tucker format* or *rank-*$(r_1, r_2, \dots, r_N)$ *approximation*:

$$\left\| f - \sum_{k_1=1}^{r_1} \sum_{k_2=1}^{r_2} \cdots \sum_{k_N=1}^{r_N} \alpha_{k_1 k_2 \dots k_N} u_{k_1}^1 \otimes u_{k_2}^2 \otimes \cdots \otimes u_{k_N}^N \right\|_0 = \min. \tag{1.6}$$

The integers $r_1, r_2, \dots, r_N > 0$ are given, the functions $u_{k_n}^n \in L^2(\mathbb{R}^{d_n})$ and the coefficients $\alpha_{k_1 k_2 \dots k_N}$ are to be determined. The special case $r_1 = r_2 = \cdots = r_N = 1$ is the canonical rank-one approximation. The approximating tensor in (1.6) is an element of the subspace $\bigotimes_{n=1}^N U_n$ with $U_n = \operatorname{span}\{u_1^n, u_2^n, \dots, u_{r_n}^n\}$, and since every tensor $g \in \bigotimes_{n=1}^N U_n$ can be written in this form, we can reformulate (1.6) into

$$\|f - g\|_0 = \min, \quad g \in \bigotimes_{n=1}^{N} U_n, \quad U_n \subseteq L^2(\mathbb{R}^{d_n}), \quad \dim U_n \le r_n,$$

[2]To enhance the readability, we will now use the abbreviations $\|\cdot\|_0$ and $(\cdot,\cdot)_0$ for the norm and inner product of $L^2(\mathbb{R}^d)$. The dimension d will always be clear from the context. In the case of complex spaces, the inner product is supposed to be conjugate linear in the *second* argument.

[3]Exceptions of this statement are the cases $r = 1$ or $N \le 2$.

with the subspaces U_n as unknowns. This explains the name of the model. It was shown in [13] that (1.6) is a well-posed problem, that is, has at least one solution for every choice of f. Again, one might ask for the regularity of a solution.

The answer to this question for both types of approximation is given in Theorems 4.5 and 5.2, respectively: the factors u_k^n in the solution essentially have the same regularity as the function f with regard to the corresponding variable x_n. Since a solution is a sum of products and each factor lies in, say, the Sobolev space H^s, the mixed regularity is even higher in the sense that certain mixed partial derivatives up to order sN are possible. As a consequence, an H^s function of finite rank always possesses certain mixed derivatives up to degree sN, see Corollary 4.9. A similar result holds for finite-dimensional subspace decompositions of functions (Corollary 5.6).

Since we know that (1.6) always has L^2 solutions, Theorem 5.2 implies that for $f \in H^s(\mathbb{R}^{\mathbf{d}})$ and $r_1, r_2, \dots, r_N$ chosen appropriately the problem

$$\left\| f - \sum_{k_1=1}^{r_1} \sum_{k_2=1}^{r_2} \cdots \sum_{k_N=1}^{r_N} \alpha_{k_1 k_2 \dots k_N} u_{k_1}^1 \otimes u_{k_2}^2 \otimes \cdots \otimes u_{k_N}^N \right\|_0 = \min,$$
$$u_{k_n}^n \in H^s(\mathbb{R}^{d_n}),$$

stated in H^s, also admits solutions, which coincide with the ones of (1.6) (Corollary 5.8). From an opposite point of view, this means that switching to the larger space L^2 for approximating f does not affect the regularity of the solution. This is not obvious, since possible minimizers cannot be bounded a priori in H^s (whereas they can in L^2), which is usually necessary to apply any sort of Weierstrass theorem. The proofs of Theorems 4.5 and 5.2 consist of an analysis of necessary first-order conditions which the L^2 solutions of (1.5) and (1.6) have to satisfy.

The results derived in this paper appear valid and expectable at first glance. However, since we did not find them rigorously formulated or proved in the literature, we decided to write this article to show that best tensor product approximations to square integrable functions preserve regularity. This is remarkable because the approximation is performed in L^2. It is interesting to note that in [12] very similar arguments are applied by Tyrtyshnikov in a finite-dimensional setting, to show that best low-rank approximations preserve linear constraints which the target tensor is subject to.

2 The Main Idea

We present the key idea of this article. Recall definition (1.2) of the tensor product of two functions.

Lemma 2.1 *Let $f \in L^2(\mathbb{R}^{d_1+d_2})$ and let $v_1, v_2, \dots, v_r \in L^2(\mathbb{R}^{d_2})$, $r \in \mathbb{N}$, be linearly independent. Let further $G = [g_{kl}]$ with $g_{kl} = (v_k, v_l)_0$ denote the Gram matrix of $v_1, v_2, \dots, v_r$. Then the unique solution $u_1, u_2, \dots, u_r \in L^2(\mathbb{R}^{d_1})$ of*

$$\left\| f - \sum_{k=1}^{r} u_k \otimes v_k \right\|_0^2 = \min \tag{2.1}$$

is almost everywhere given by

$$u_k(x) = \int f(x,y)\overline{w}_k(y)\,\mathrm{d}y, \quad k = 1, 2, \ldots, r, \tag{2.2}$$

where $w_k = \sum_{l=1}^{r} \gamma_{kl} v_l$ *and* $\Gamma = [\gamma_{kl}]$ *denotes the inverse of the* G.

Remark 2.2 It is well known that the Gram matrix of $v_1, v_2, \ldots, v_r$ is invertible if and only if $v_1, v_2, \ldots, v_r$ are linearly independent. As we will see in the proof, formulae (2.2) are simply the normal equations for the inner integral in (2.1). They are an analog of the well-known solution formula for alternating least squares approximation of a discrete tensor in the canonical low-rank format [7].

Proof For almost every $x \in \mathbb{R}^{d_1}$ the function $y \mapsto f_x(y) = f(x,y)$ is square integrable. Since $v_1, v_2, \ldots, v_r$ are linearly independent, for these values of x the inner integral

$$I_u(x) = \int \left| f(x,y) - \sum_{k=1}^{r} u_k(x) v(y) \right|^2 \mathrm{d}y = \left\| f_x - \sum_{k=1}^{r} u_k(x) v_k \right\|_0^2$$

is minimal if and only if the coefficients $u_k(x)$ satisfy the normal equations

$$0 = \left(\sum_{k=1}^{r} u_k(x) v_k - f_x, v_l \right)_0 = \sum_{k=1}^{r} g_{kl} u_k(x) - (f_x, v_l)_0, \quad l = 1, 2, \ldots, r.$$

This is equivalent to

$$u_k(x) = \sum_{l=1}^{r} \overline{\gamma}_{kl} (f_x, v_l)_0 = \left(f_x, \sum_{l=1}^{r} \gamma_{kl} v_l \right)_0,$$

which is (2.2). If chosen this way, $u_1, u_2, \ldots, u_r$ obviously solve (2.1). Let $\widetilde{u}_1, \widetilde{u}_2, \ldots, \widetilde{u}_r$ be another solution. Then for almost every $x \in \mathbb{R}^{d_1}$ we have that f_x is square integrable and $I_{\widetilde{u}}(x) = I_u(x)$ is minimal. As follows from the considerations above, this implies $\widetilde{u}_k(x) = u_k(x)$ for $k = 1, 2, \ldots, r$, which proves the uniqueness in $L^2(\mathbb{R}^{d_1})$. □

Equality (2.2) is the starting point for our regularity analysis. As one can expect, the regularity of $u(x)$ in (2.2) depends on the regularity properties of the target function $f(x,y)$ with regard to the variable x only. We will make this statement more precise in the next section.

3 Spaces of Anisotropic Regularity

In this article we mainly focus on Sobolev spaces H^s of s-times weakly differentiable functions. Let $\mathscr{F}: L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ denote the Fourier transform. For a function $u \in L^2(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$ it is given by

$$(\mathscr{F}u)(\omega) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \int u(x) e^{-i\omega\cdot x}\,\mathrm{d}x. \tag{3.1}$$

Let $s \geq 0$. Then the Sobolev space $H^s(\mathbb{R}^d)$ is defined as the completion of the space of infinitely differentiable functions with compact support on $\mathbb{R}^d$ under the norm

$$\|u\|_s^2 = \int \left(1 + |\omega|^2\right)^s \left|(\mathscr{F}u)(\omega)\right|^2 \mathrm{d}\omega.$$

Since the domain is unbounded, the following definition is equivalent:

$$H^s\left(\mathbb{R}^d\right) = \left\{u \in L^2\left(\mathbb{R}^d\right) \mid \|u\|_s^2 < \infty\right\}.$$

In general, s can be any nonnegative real number. If s is an integer, the definition leads to the classical Sobolev spaces of s-times weakly differentiable functions. For $s = 0$ one obtains L^2.

The spaces H^s contain functions of isotropic regularity, that is, functions which possess the same smoothness with regard to every spatial variable. Assume we have a partition $\mathbf{d} = (d_1, d_2, \ldots, d_N)$ of $\mathbb{R}^{\mathbf{d}} = \mathbb{R}^{d_1+d_2+\cdots+d_N}$. In accordance with the tensor product viewpoint of this article, and in order to obtain more precise results, we want to allow for the possibility that the target function $f \in L^2(\mathbb{R}^{\mathbf{d}})$ in the tensor product approximation has different regularity properties for each variable $x_n \in \mathbb{R}^{d_n}$. For this purpose, we define the norms

$$\|f\|_{s,n}^2 = \int \left(1 + |\omega_n|^2\right)^s \left|(\mathscr{F}f)(\omega)\right|^2 \mathrm{d}\omega,$$

where $\mathscr{F}$ is now the Fourier transform on $L^2(\mathbb{R}^{\mathbf{d}})$ and $\omega \in \mathbb{R}^{\mathbf{d}}$ is partitioned into $\omega_n \in \mathbb{R}^{d_n}$. The corresponding Hilbert space is denoted by

$$H^{s,n}\left(\mathbb{R}^{\mathbf{d}}\right) = \left\{f \in L^2\left(\mathbb{R}^{\mathbf{d}}\right) \mid \|f\|_{s,n} < \infty\right\}.$$

This space contains functions f for which $x_n \mapsto f(x_1, x_2, \ldots, x_N)$ is in $H^s(\mathbb{R}^{d_n})$ for almost every $(x_1, \ldots, x_{n-1}, x_{n+1}, \ldots, x_N)$. The Sobolev space $H^s(\mathbb{R}^{\mathbf{d}})$ is obviously contained in $H^{s,n}(\mathbb{R}^{\mathbf{d}})$ for every choice of n. In fact, it is completely characterized by this property.

The following lemma shows how the image of an integral operator is related to the regularity of its kernel function.

Lemma 3.1 *Let $s \geq 0$, $f \in H^{s,1}(\mathbb{R}^{d_1+d_2})$, and $v \in L^2(\mathbb{R}^{d_2})$. Then u defined almost everywhere by*

$$u(x) = \int f(x, y) v(y)\,\mathrm{d}y \tag{3.2}$$

belongs to $H^s(\mathbb{R}^{d_1})$ *and satisfies the estimate*

$$\|u\|_s \le \|f\|_{s,1}\|v\|_0.$$

Proof We denote by $\mathscr{F}_1$, $\mathscr{F}_2$, and $\mathscr{F}$ the Fourier transforms on $L^2(\mathbb{R}^{d_1})$, $L^2(\mathbb{R}^{d_2})$ and $L^2(\mathbb{R}^{d_1+d_2})$, respectively. It is possible to write the function f as

$$f = \sum_{k=1}^{\infty} \phi_k \otimes \psi_k$$

with $\phi_k \in L^2(\mathbb{R}^{d_1}) \cap L^1(\mathbb{R}^{d_1})$ and $\psi_k \in L^2(\mathbb{R}^{d_2}) \cap L^1(\mathbb{R}^{d_2})$, where the limit is taken in L^2 (take, for instance, constant functions with rectangular support). Since the integral operator (3.2) depends continuously on the kernel function, we have

$$u = \sum_{k=1}^{\infty} \phi_k \int \psi_k(y)v(y)\,\mathrm{d}y$$

in L^2. Consequently,

$$\|u\|_s^2 = \int \left(1+|\omega|^2\right)^s \left| \int \sum_{k=1}^{\infty} (\mathscr{F}_1\phi_k)(\omega)\psi_k(y)v(y)\,\mathrm{d}y \right|^2 \mathrm{d}\omega. \tag{3.3}$$

Note that $\|\sum_{k=1}^{\infty}(\mathscr{F}_1\phi_k) \otimes \psi_k\|_0 = \|f\|_0$ by Plancherel's theorem, which implies that $y \mapsto \sum_{k=1}^{\infty}(\mathscr{F}_1\phi_k)(\omega)\psi_k(y)$ is square integrable for almost every ω. Therefore, we can apply Schwarz's inequality on the inner integral in (3.3). Using once more Plancherel's theorem, we obtain the estimate

$$\|u\|_s^2 \le \iint \left(1+|\omega|^2\right)^s \left| \sum_{k=1}^{\infty} (\mathscr{F}_1\phi_k)(\omega)(\mathscr{F}_2\psi_k)(\xi) \right|^2 \mathrm{d}\xi\,\mathrm{d}\omega \cdot \|v\|_0^2.$$

With the aid of (3.1) one verifies that the double integral equals $\|f\|_{s,1}$. □

4 Regularity of Canonical Low-Rank Approximations

Combining Lemma 3.1 with Lemma 2.1, we obtain a regularity result for tensor product approximations with two factors.

Lemma 4.1 *Consider the situation in Lemma* 2.1, *but with* $f \in H^{s,1}(\mathbb{R}^{d_1+d_2})$. *Then every* u_k *belongs to* $H^s(\mathbb{R}^{d_1})$ *and satisfies* $\|u_k\|_s \le \|f\|_{s,1}\|w_k\|_0$.

We intend to use this lemma to prove the regularity of canonical low-rank approximations with an arbitrary number of factors.

4.1 Notation

We fix $N \geq 2$ and a partition $\mathbf{d} = (d_1, d_2, \dots, d_N)$ of $\mathbb{R}^{\mathbf{d}} = \mathbb{R}^{d_1+d_2+\cdots+d_N}$ and introduce some abbreviations in order to treat the general case clearly. We set

$$T(\mathbf{u}_k) = u_k^1 \otimes u_k^2 \otimes \cdots \otimes u_k^N.$$

Each u_k^n belongs to $L^2(\mathbb{R}^{d_n})$, and their tensor product $T(\mathbf{u}_k)$ of order N is a function in $L^2(\mathbb{R}^{\mathbf{d}})$, namely,

$$T(\mathbf{u}_k)(x) = \prod_{n=1}^{N} u_k^n(x_n),$$

where $x = (x_1, x_2, \dots, x_N)$ is partitioned correspondingly into $x_n \in \mathbb{R}^{d_n}$. Furthermore, for fixed $n \in \{1, 2, \dots, N\}$ the reduced tensor products of order $N-1$,

$$T^n(\mathbf{u}_k) = u_k^1 \otimes \cdots \otimes u_k^{n-1} \otimes u_k^{n+1} \otimes \cdots \otimes u_k^N, \tag{4.1}$$

will be useful. These are functions in $L^2(\mathbb{R}^{d_1+\cdots+d_{n-1}+d_{n+1}+\cdots+d_N})$, namely,

$$T^n(\mathbf{u}_k)(\widetilde{x}) = \prod_{\substack{m=1\\ m\neq n}}^{N} u_k^m(x_m),$$

with $\widetilde{x}$ partitioned correspondingly.

4.2 Rank

Every function $g \in L^2(\mathbb{R}^{\mathbf{d}})$ can be written as a sum of tensor products

$$g = \sum_{k=1}^{r} T(\mathbf{u}_k) \tag{4.2}$$

with $r \in \mathbb{N} \cup \{\infty\}$.

Definition 4.2 The *rank* of g is the minimal $r \in \mathbb{N} \cup \{\infty\}$ such that g can be represented in the form (4.2). One writes $\operatorname{rank} g = r$.

Of course, the rank of g depends on the partition $\mathbf{d}$ of the spatial space $\mathbb{R}^{\mathbf{d}}$ into spaces of lower dimension, that is, on the choice of $d_1, d_2, \dots, d_N$. Throughout this text we assume this partition to be fixed.

The following observation will be of importance. We have found it in [4].

Lemma 4.3 *Let* $g = \sum_{k=1}^{r} T(\mathbf{u}_k)$ *with* $r = \operatorname{rank} g < \infty$. *Then for every* $n \in \{1, 2, \dots, N\}$ *the set*

$$\{T^n(\mathbf{u}_1), T^n(\mathbf{u}_2), \dots, T^n(\mathbf{u}_r)\}$$

with $T^n(\mathbf{u}_k)$ *as in* (4.1) *is linearly independent in* $L^2(\mathbb{R}^{d_1+\cdots+d_{n-1}+d_{n+1}+\cdots+d_N})$.

Proof It is sufficient[4] to prove this for $n = 1$. If we assume $T^1(\mathbf{u}_r) = \sum_{k=1}^{r-1} \beta_k T^1(\mathbf{u}_k)$, then

$$g = \sum_{k=1}^{r} T(\mathbf{u}_k) = \sum_{k=1}^{r-1} \mathbf{u}_k^1 \otimes T^1(\mathbf{u}_k) + \mathbf{u}_r^1 \otimes T^1(\mathbf{u}_r) = \sum_{k=1}^{r-1} \big(\mathbf{u}_k^1 + \beta_k \mathbf{u}_r^1\big) \otimes T^1(\mathbf{u}_k).$$

This contradicts $\operatorname{rank} g = r$. □

4.3 Canonical rank-r Approximation

Let $f \in L^2(\mathbb{R}^{\mathbf{d}})$ be given. The best canonical rank-r approximation of f is defined as a solution g of the problem

$$\|f - g\|_0 = \min, \quad \operatorname{rank} g \le r, \tag{4.3}$$

or, in parametrized form,

$$\left\| f - \sum_{k=1}^{r} T(\mathbf{u}_k) \right\|_0 = \min. \tag{4.4}$$

We emphasize again that for certain choices of f and r the problem can be ill-posed and does not admit a global solution. Moreover, these cases are not rare [11]. In the following, we are concerned with local minima of (4.3), that is, with local minima of $g \mapsto \|f - g\|_0$ with respect to the set of functions of rank at most r.

Lemma 4.4 *If g is a local minimum of (4.3) and $r \le \operatorname{rank} f$, then $\operatorname{rank} g = r$.*

Proof This is more or less clear, since otherwise we could find a nonzero rank-one approximation h to $f - g$ such that

$$\big\| f - (g + \lambda h) \big\|_0 < \|f - g\|_0$$

for all $0 < \lambda \le 1$. This would contradict the local optimality of g. □

Theorem 4.5 *Let $r \le \operatorname{rank} f$, $r < \infty$, and $g = \sum_{k=1}^{r} T(\mathbf{u}_k)$ be a local minimum of (4.3). Then for each $n = 1, 2, \dots, N$ the functions u_k^n inherit the regularity of the function f with regard to the variable $x_n \in \mathbb{R}^{d_n}$, that is, if f belongs to $H^{s_n,n}(\mathbb{R}^{\mathbf{d}})$, then u_k^n is in $H^{s_n}(\mathbb{R}^{d_n})$ for $k = 1, 2, \dots, r$.*

Proof We fix u_k^n for $n \ge 2$ and show that $u_1^1, u_2^1, \dots, u_r^1 \in H^{s_1}(\mathbb{R}^{d_1})$. The other components can be treated in the same way by permuting the order of the factors. The set of functions $u_1^1, u_2^1, \dots, u_r^1$ is a local minimum of the problem

$$\left\| f - \sum_{k=1}^{r} u_k^1 \otimes T^1(\mathbf{u}_k) \right\|_0^2 = \min.$$

[4] A permutation of the factors of a tensor would leave its rank invariant.

Since this functional is nonnegative and quadratic in u_k^1, a local minimum is also the global minimum. According to Lemma 4.4 the solution g of (4.3) has rank r. Hence the functions $T^1(\mathbf{u}_1), T^1(\mathbf{u}_2), \ldots, T^1(\mathbf{u}_r)$ are linearly independent by Lemma 4.3. The theorem is therefore an application of Lemma 4.1. □

Remark 4.6 Of course, the theorem implies that the solution g itself is in $H^{s_n,n}(\mathbb{R}^{\mathbf{d}})$ for $n = 1, 2, \ldots, N$, but as mentioned in the introduction one obtains an even higher mixed regularity. Since g is a sum of tensor products, it is possible to take derivatives of order s_n independently in each direction n, that is, g possesses certain mixed derivatives up to order $s_1 + s_2 + \cdots + s_N$. Function spaces of mixed derivatives can play an important role for approximation in high spatial dimensions. Recent work by Yserentant [15] has shown this in the context of the electronic Schrödinger equation in quantum chemistry.

Remark 4.7 The condition $r \leq \operatorname{rank} f$ appears naturally, for if one solves (4.3) with $r > \operatorname{rank} f$, one gets the exact solution $g = f$ but usually not in a minimal representation. Then additional terms u_k^n could appear that are not in H^{s_n} but would cancel out in a minimal rank-r representation.

Remark 4.8 In the theorem it is assumed that g is a local minimum of (4.3), that is, a local minimum on the admissible set of functions of rank at most r. In practice, one tries to solve the parametrized problem (4.4) as a function of the factors u_k^n. It seems to be a difficult question whether a local minimum in the parameter space is also a local minimum of (4.3). At least for global minima, if existent, this is obviously the case. As one can see in the proof, the theorem is in principle valid for local minima of (4.4) if one additionally assumes that the solution has full rank r. One may conjecture that this is always the case, but a rigorous proof is required.

For the case $r = \operatorname{rank} f < \infty$ the theorem makes the following interesting claim:

Corollary 4.9 *If* rank f *is finite*, *then in every representation*

$$f = \sum_{k=1}^{\operatorname{rank} f} T(\mathbf{u}_k)$$

the functions u_k^n have the same regularity as f with regard to the variable $x_n \in \mathbb{R}^{d_n}$, that is, *if $f \in H^{s_n,n}(\mathbb{R}^{\mathbf{d}})$*, *then $u_k^n \in H^{s_n}(\mathbb{R}^{d_n})$ for $k = 1, 2, \ldots, r$.*

According to Remark 4.6 this means that for instance an H^s function of finite rank always possesses certain mixed derivatives up to order sN. This is an interesting interaction between the algebraic concept of rank and the analytic concept of smoothness.

4.4 Norm Estimates

Theorem 4.5 guarantees that the factors u_k^n in a best low-rank approximation lie in certain Sobolev spaces, but gives no estimate on the corresponding Sobolev norm.

The estimation of the norm of the factors is an intrinsic difficulty of the canonical tensor format. It is closely related to the fact that the problem can be ill-posed.

To obtain a reasonable formula for each factor, one should first eliminate the scaling indeterminacy by writing a solution g of (4.3) in the normalized form

$$g = \sum_{k=1}^{r} \alpha_k T(\mathbf{u}_k), \quad \|u_k^n\|_0 = 1 \tag{4.5}$$

for $k = 1, 2, \ldots, r$ and $n = 1, 2, \ldots, N$. The coefficients α_k carry the L^2 norm of each summand, that is, $|\alpha_k| = \|\alpha_k T(\mathbf{u}_k)\|_0$. For each n, let $G^n = [g_{kl}^n] = [(T^n(\mathbf{u}_k), T^n(\mathbf{u}_l))_0]$ be the Gram matrix of the complementary factors $T^n(\mathbf{u}_1)$, $T^n(\mathbf{u}_2), \ldots, T^n(\mathbf{u}_r)$, and let $\Gamma^n = [\gamma_{kl}^n]$ be its inverse. Under the assumptions of Theorem 4.5, Lemma 4.1 provides the general estimate

$$\|\alpha_k u_k^n\|_{s_n} \le \|f\|_{s_n,n} \|w_k\|_0 \tag{4.6}$$

with

$$w_k = \sum_{l=1}^{r} \gamma_{kl}^n T^n(\mathbf{u}_k).$$

Denoting the kth row vector of Γ^n by γ_k^n and utilizing that the spectral norm $\|G^n\|_2$ dominates the diagonal entries $g_{kk}^n = 1$ of G^n, we have

$$\begin{aligned} \|w_k\|_0^2 &= (\gamma_k^n)^* G^n \gamma_k^n \\ &\le \|G^n\|_2 \|\gamma_k^n\|_2^2 \le \|G^n\|_2 \|\Gamma^n\|_2^2 \le \|G^n\|_2^2 \|\Gamma^n\|_2^2 \le (\kappa_2(G^n))^2, \end{aligned}$$

where $\kappa_2(G^n) = \|G^n\|_2 \|\Gamma^n\|_2 = \lambda_{\max}(G^n)/\lambda_{\min}(G^n)$ is the spectral condition number of G^n. From (4.6) one thus obtains

$$\|u_k^n\|_{s_n} \le \frac{\kappa_2(G^n)}{|\alpha_k|} \|f\|_{s_n,n}. \tag{4.7}$$

This estimate can be useful if lower bounds for $|\alpha_k|$ and upper bounds for $\kappa_2(G^n)$ are available. However, as mentioned above, such bounds are hard to determine a priori.

In the case of the L^2 norm ($s_n = 0$), inequality (4.7) states that

$$1 \le \frac{\kappa_2(G^n)}{|\alpha_k|} \|f\|_0 \tag{4.8}$$

for all k and n. Therefore, the best bound we can hope to get in (4.7) is $\|f\|_{s_n,n}/\|f\|_0$, which would be a perfect estimate. We can also deduce from (4.8) that $\kappa_2(G^n)$ has to be large if $|\alpha_k|$ is large. Unfortunately, the converse does not need to be true, that is, $|\alpha_k|$ can be very small even if all G^n are bad-conditioned. However, in this case the term $T(\mathbf{u}_k)$ is of less importance in the low-rank approximation (4.5) (and may be removed). Assuming that all the matrices G^n have almost the same condition, we hence obtain the best norm estimates (4.7) for the most important rank-one terms $T(\mathbf{u}_k)$. This observation warrants further detailed investigation.

To get some deeper insight into the behavior of $\kappa_2(G^n)$, let $G_m = [(u_k^m, u_l^m)_0]$ denote the Gram matrices of the factors $u_1^m, u_2^m, \dots, u_r^m$. For each n, observe that G^n is the Hadamard (pointwise) product of the matrices G_m for $m \neq n$. All these matrices are positive semidefinite and have ones on the diagonal. By a well-known theorem on the eigenvalues of Hadamard products [6, Theorem 5.3.4], one can hence for each n deduce the estimate[5]

$$\kappa_2\big(G^n\big) \leq \frac{\min_{m\neq n}(\lambda_{\max}(G_m))}{\max_{m\neq n}(\lambda_{\min}(G_m))} \leq \min_{m\neq n}\big(\kappa_2(G_m)\big).$$

In particular, if one of the Gram matrices of the factors, say G_m, is well-conditioned, then so is G^n for all $n \neq m$. In the best case, the set $u_1^m, u_2^m, \dots, u_r^m$ is orthonormal, i.e., $G_m = I$ (identity matrix), and this implies $G^n = I$ for $n \neq m$. For these n, inequality (4.7) then gives a good estimate. However, $\kappa_2(G^m)$ can still be arbitrarily large.

To explain the relation between α_k and $\kappa_2(G^n)$, we introduce still another Gram matrix G, namely the one of the rank-one functions $T(\mathbf{u}_1), T(\mathbf{u}_2), \dots, T(\mathbf{u}_r)$ themselves, that is, $G = [(T(\mathbf{u}_k), T(\mathbf{u}_l))_0]$. The vector $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_r)^T$ is the solution of the normal equation

$$G^T \alpha = \phi,$$

where $\phi_k = (f, T(\mathbf{u}_k))_0$, $k = 1, 2, \dots, r$. For each $n = 1, 2, \dots, N$, we can write G as the Hadamard product of G_n and G^n. By the theorem cited above, this implies

$$\kappa_2(G) \leq \min_n\big(\min\big\{\kappa_2(G_n), \kappa_2\big(G^n\big)\big\}\big).$$

Now, if $\|\alpha\|_2$ is much larger than $\|f\|_0$, this indicates that $\kappa_2(G)$ is large. Consequently, all of the matrices G_n and G^n have to be bad-conditioned or (the former) even singular. This explains inequality (4.8).

4.5 Regularization

It was mentioned above that the canonical rank-r approximation (4.3) can be an ill-posed problem. This is related to the phenomenon of diverging rank-one terms $T(\mathbf{u}_k)$ in a minimizing sequence [11]. In numerical practice one therefore often adds a regularization term to bound the norms of the $T(\mathbf{u}_k)$. We want to show that our regularity results remain valid in this case.

For a given $\lambda > 0$, we consider the problem of finding functions $u_k^n \in L^2(\mathbb{R}^{d_n})$, $k = 1, 2 \dots, r$ and $n = 1, 2, \dots, N$, such that

$$\left\| f - \sum_{k=1}^{r} T(\mathbf{u}_k) \right\|_0^2 + \lambda \sum_{k=1}^{r} \|T(\mathbf{u}_k)\|_0^2 = \min. \tag{4.9}$$

[5] It is possible that $\lambda_{\min}(G_m) = 0$ for all $m \neq n$ or even for all m. In this case, the value of the middle and right term in the inequality is understood as ∞, and the estimate is pointless.

It is possible to prove that this problem has a solution for every choice of f and r. The trick is to expand the u_k^n into an orthonormal basis of $\operatorname{span}\{u_1^n, u_2^n, \dots, u_r^n\}$, that is,

$$u_k^n = \sum_{l=1}^{r} c_{kl}^n \xi_l^n, \tag{4.10}$$

with $(\xi_k^n, \xi_l^n)_0 = \delta_{kl}$. In principle, this is a subspace decomposition as discussed in the next section. One can assume that for each mode n the norms $\|u_k^n\|_0$ of the factors are equal for $k = 1, 2, \dots, r$ (equilibrated). This then implies that in a minimizing sequence for (4.9) all the coefficients c_{kl}^n in (4.10) remain bounded and hence can be assumed to converge. In particular, one may fix them a priori. As can be easily seen, the norms $\|T(\mathbf{u}_k)\|_0$ and $\|\sum_{k=1}^r T(\mathbf{u}_k)\|_0$ of the solution are then also determined in advance. It therefore remains to maximize the real part of the overlap $(f, \sum_{k=1}^r T(\mathbf{u}_k))_0$ with respect to the parametrization (4.10) (with the c_{kl}^n being fixed). In each mode n, this is a convex maximization problem of a weakly sequentially continuous functional on the set of all orthonormal systems $\{\xi_1^n, \xi_2^n, \dots, \xi_r^n\}$, that is, on a Stiefel manifold, and is known to admit a solution. For details we refer to [13, Theorem 2.12] and in particular to Sect 3.8 therein for a very similar problem.

Theorem 4.10 *Let* $r \le \operatorname{rank} f$, $r < \infty$. *Assume for a local minimum of* (4.9) *that* $\operatorname{rank}(\sum_{k=1}^r T(\mathbf{u}_k)) = r$. *Then for each* $n = 1, 2, \dots, N$ *the functions* u_k^n *inherit the regularity of the function* f *with regard to the variable* $x_n \in \mathbb{R}^{d_n}$, *that is, if* f *belongs to* $H^{s_n,n}(\mathbb{R}^{\mathbf{d}})$, *then* u_k^n *is in* $H^{s_n}(\mathbb{R}^{d_n})$ *for* $k = 1, 2, \dots, r$.

Proof Again, we limit ourselves to $n = 1$. The functions $u_1^1, u_2^1, \dots, u_r^1$ locally minimize the functional

$$\left\| f - \sum_{k=1}^{r} u_k^1 \otimes T^1(\mathbf{u}_k) \right\|_0^2 + \lambda \sum_{k=1}^{r} \|u_k^1\|_0^2 \|T^1(\mathbf{u}_k)\|_0^2, \tag{4.11}$$

with the u_k^n being fixed for $n \ge 2$. We now mimic the proof of Lemma 2.1. Let $x = x_1$, $y = (x_2, \dots, x_N)$, $u_k = u_k^1$, and $v_k = T^1(\mathbf{u}_k)$ for abbreviation. Then for almost every $x \in \mathbb{R}^{d_1}$ the function $y \mapsto f_x(y) = f(x, y)$ is square integrable. For almost every of these x the inner integral of (4.11),

$$\begin{aligned} I_u(x) &= \int \left(\left| f_x(y) - \sum_{k=1}^{r} u_k(x) v_k(y) \right|^2 + \lambda \sum_{k=1}^{r} |u_k(x)|^2 |v_k(y)|^2 \right) \mathrm{d}y \\ &= \left\| f_x - \sum_{k=1}^{r} u_k(x) v_k \right\|_0^2 + \lambda \sum_{k=1}^{r} |u_k(x)|^2 \|v_k\|_0^2, \end{aligned}$$

has to be minimal, which is the case if and only if the coefficients $u_1(x), u_2(x), \dots, u_r(x)$ satisfy the first-order condition

$$\sum_{k=1}^{r}(1+\lambda\delta_{kl})(v_k, v_l)_0 u_k(x) = (f_x, v_l)_0, \quad l = 1, 2, \dots, r.$$

Namely, by Lemma 4.3, the diagonally shifted Gram matrix $G^\lambda = [(1+\lambda\delta_{kl})(v_k, v_l)_0]$ is positive definite. Denoting the inverse by $\Gamma^\lambda = [\gamma_{kl}^\lambda]$, we obtain, for almost every $x \in \mathbb{R}^{d_1}$, the solution formula

$$u_k(x) = \sum_{l=1}^{r} \overline{\gamma}_{kl}^\lambda (f_x, v_l)_0 = \left(f_x, \sum_{l=1}^{r} \gamma_{kl}^\lambda v_l \right)_0 = \int f(x,y) \sum_{l=1}^{r} \overline{\gamma}_{kl}^\lambda \overline{v}_l(y)\, \mathrm{d}y$$

for $k = 1, 2, \dots, r$. By Lemma 3.1, this proves the theorem. □

The question whether the condition $\operatorname{rank}(\sum_{k=1}^{r} T(\mathbf{u}_k)) = r$ is really necessary in the above theorem requires further investigation. Similarly to Lemma 4.4, it seems natural to conjecture that it already follows from the assumption $\operatorname{rank} f \geq r$.

5 Regularity of Optimal Subspace Approximations

In this section we will prove the regularity of optimal subspace approximations. The arguments are basically the same as for the canonical rank-r approximation.

5.1 Subspace rank

Again, let $\mathbf{d} = (d_1, d_2, \dots, d_N)$ be a fixed partition of $\mathbb{R}^{\mathbf{d}} = \mathbb{R}^{d_1+d_2+\dots+d_N}$. Every function $g \in L^2(\mathbb{R}^{\mathbf{d}})$ can be written in the form

$$g = \sum_{k_1=1}^{r_1} \sum_{k_2=1}^{r_2} \dots \sum_{k_N=1}^{r_N} \alpha_{k_1 k_2 \dots k_N} u_{k_1}^1 \otimes u_{k_2}^2 \otimes \dots \otimes u_{k_N}^N \tag{5.1}$$

with $u_{k_n}^n \in L^2(\mathbb{R}^{d_n})$ and $r_n \in \mathbb{N} \cup \{\infty\}$. This is called a *subspace representation* or *Tucker decomposition*; the tensor $\boldsymbol{\alpha}$ is called the *core tensor*. If for each n we set $U_n = \operatorname{span}\{u_1^n, u_2^n, \dots, u_{r_n}^n\}$, then (5.1) simply states that $g \in \bigotimes_{n=1}^{N} U_n$.

Definition 5.1 For $n = 1, 2, \dots, N$ the *n-rank* of g is defined as the minimal number r_n for which there is a subspace $U_n \subseteq L^2(\mathbb{R}^{d_n})$ of dimension r_n such that

$$g \in L^2(\mathbb{R}^{d_1}) \otimes \dots \otimes L^2(\mathbb{R}^{d_{n-1}}) \otimes U_n \otimes L^2(\mathbb{R}^{d_{n+1}}) \otimes \dots \otimes L^2(\mathbb{R}^{d_N}).$$

We write $\operatorname{rank}_n g$ for the n-rank and say that g has *subspace rank* (or *Tucker rank*) $(r_1, r_2, \dots, r_N)$ if and only if $r_n = \operatorname{rank}_n g$ for $n = 1, 2, \dots, N$.

The n-ranks generalize the concepts of row and column rank of a matrix. If they are finite, it can be easily seen (for instance by projection arguments) that the subspaces U_n in the definition above are uniquely determined[6] and that $g \in \bigotimes_{n=1}^N U_n$. The U_n are therefore called *minimal supporting subspaces of* g. The n-ranks can equivalently be defined as the (canonical) ranks of the *mode-n matricizations* of g, as it is frequently done for finite-dimensional tensors [7]. For example, $\operatorname{rank}_1 g$ equals the rank of g when treated as an order-2 tensor $g \in L^2(\mathbb{R}^{d_1}) \otimes L^2(\mathbb{R}^{d_2+\cdots+d_N})$.

5.2 Optimal Subspace Approximation

Given $f \in L^2(\mathbb{R}^{\mathbf{d}})$ and $r_1, r_2, \dots, r_N \in \mathbb{N}$, the problem of optimal subspace approximation (also called Tucker approximation) consists of finding a function $g \in L^2(\mathbb{R}^{\mathbf{d}})$ such that

$$\|f - g\|_0 = \min, \quad \operatorname{rank}_n g \le r_n,\ n = 1, 2, \dots, N. \tag{5.2}$$

In the parametrized form we seek for functions $u^n_{k_n} \in L^2(\mathbb{R}^{d_n})$ and a core tensor $\boldsymbol{\alpha}$ such that

$$\left\| f - \sum_{k_1=1}^{r_1} \sum_{k_2=1}^{r_2} \cdots \sum_{k_N=1}^{r_N} \alpha_{k_1 k_2 \dots k_N} u^1_{k_1} \otimes u^2_{k_2} \otimes \cdots \otimes u^N_{k_N} \right\|_0 = \min. \tag{5.3}$$

As mentioned in the introduction, the true unknowns in this problem are the N subspaces $U_n = \operatorname{span}\{u^n_1, u^n_2, \dots, u^n_{r_n}\}$. It can be shown that (5.2) has at least one global solution [13], but it is not clear whether the optimal subspaces U_n are uniquely determined. This is definitely not the case if $\operatorname{rank}_n g < r_n$ for at least one n. Unfortunately, there is no analog to Lemma 4.4. We do not know under which conditions the subspace rank of a solution g will indeed be $(r_1, r_2, \dots, r_N)$. If one considers as a trivial example the case $N = 2$, that is, a function of the form

$$f = \sum_{k=1}^{\infty} u_k \otimes v_k,$$

then every solution g of (5.2) with $r_1 = 1$ and r_2 arbitrarily chosen will have rank 1 and hence subspace rank $(1, 1)$. So one should choose $r_2 = 1$ to get the result with the fewest terms. This is a priori hard to see in general so that usually an uncertainty concerning the initial choice of the r_n will remain. Therefore, the regularity result presented below focuses on the minimal supporting subspaces of a solution. The claim is that they are subspaces of Sobolev spaces corresponding to the smoothness of f.

Theorem 5.2 *Suppose that $r_n < \infty$ for $n = 1, 2, \dots, N$ and g is a local minimum of* (5.2), *that is, a local minimum with respect to the set of functions whose n-rank is at most r_n for all $n = 1, 2, \dots, N$. Let $(q_1, q_2, \dots, q_N)$ be the subspace rank of g,*

[6] In the case $r_n = \infty$ an additional minimality condition has to be imposed to obtain a unique subspace U_n, but this case is not of much interest here.

and let U_n denote the q_n-dimensional supporting subspaces such that $g \in \bigotimes_{n=1}^{N} U_n$. Then $f \in H^{s_n,n}(\mathbb{R}^{\mathbf{d}})$ implies that U_n is a subspace of $H^{s_n}(\mathbb{R}^{d_n})$. In other words, in every representation

$$g = \sum_{k_1=1}^{q_1} \sum_{k_2=1}^{q_2} \cdots \sum_{k_N=1}^{q_N} \alpha_{k_1 k_2 \ldots k_N} u_{k_1}^1 \otimes u_{k_2}^2 \otimes \cdots \otimes u_{k_N}^N$$

the functions $u_{k_n}^n$ belong to $H^{s_n}(\mathbb{R}^{d_n})$ for $k_n = 1, 2, \ldots, q_n$ and $n = 1, 2, \ldots, N$.

Proof We prove this only for $n = 1$. We can find a decomposition $g = \sum_{k=1}^{q_1} u_k^1 \otimes T_k^1$ with $u_k^1 \in U_1$ and $T_k^1 \in L^2(\mathbb{R}^{d_1+\cdots+d_N})$ for $k = 1, 2, \ldots, q_1$. Since q_1 is minimal, the T_k^1 have to be linearly independent (see the proof of Lemma 4.3). Since the set of functions $u_1^1, u_2^1, \ldots, u_{q_1}^1$ is a (global) minimum of the quadratic problem

$$\left\| f - \sum_{k=1}^{q_1} u_k^1 \otimes T_k^1 \right\|_0^2 = \min,$$

the claim of the theorem is again a consequence of Lemma 4.1. □

Remark 5.3 Concerning the mixed regularity of g, Remark 4.6 applies.

Remark 5.4 Obviously, we have $q_n \le r_n$ for $n = 1, 2, \ldots, N$. In view of the fact that the smoothness of the solution g does not depend on its representation, Theorem 5.2 appears quite satisfactory. However, from a constructive point of view it would be useful to know the values of the q_n and derive conditions which ensure $q_n = r_n$. This question requires further research, but since it is purely algebraic in nature we will not deliberate on it in this article.

Remark 5.5 Closely related to the previous remark is the observation that the theorem is also valid for local minima in the parameter space, that is, for local minima of (5.3), provided that they have full subspace rank $(r_1, r_2, \ldots, r_N)$. In particular, to guarantee $u_{k_n}^n \in H^{s_n}(\mathbb{R}^{d_n})$ for a fixed direction n it is sufficient that $\operatorname{rank}_n g = r_n$. For numerical computations one always considers the parametrized problem (5.3).

The theorem implies the regularity of exact subspace decompositions if the subspaces are finite-dimensional.

Corollary 5.6 *If* $\operatorname{rank}_n f$ *is finite and $f \in H^{s_n,n}(\mathbb{R}^{\mathbf{d}})$ for $n = 1, 2, \ldots, N$, then the minimal supporting subspaces of f are subspaces of $H^{s_n}(\mathbb{R}^{d_n})$.*

Remark 5.7 Let U_n be finite-dimensional subspaces of $L^2(\mathbb{R}^{d_n})$ and $V = \bigotimes_{n=1}^{N} U_n$. Then the corollary states that $H^{s_n,n}(\mathbb{R}^{\mathbf{d}}) \cap V \neq \emptyset$ for $n = 1, 2, \ldots, N$ if and only if $U_n \subseteq H^{s_n}(\mathbb{R}^{d_n})$.

Since (5.2) always admits at least one solution, we are allowed to reformulate Theorem 5.2 as an existence result.

Corollary 5.8 *Let $f \in H^{s_n,n}(\mathbb{R}^{\mathbf{d}})$ for $n = 1, 2, \ldots, N$. Then the problem*

$$\left\| f - \sum_{k_1=1}^{r_1} \sum_{k_2=1}^{r_2} \cdots \sum_{k_N=1}^{r_N} \alpha_{k_1 k_2 \ldots k_N} u_{k_1}^1 \otimes u_{k_2}^2 \otimes \cdots \otimes u_{k_N}^N \right\|_0 = \min, \quad u_{k_n}^n \in H^{s_n}\big(\mathbb{R}^{d_n}\big),$$

stated in the spaces H^{s_n}, always admits a solution. The set of solutions coincides with the one of problem (5.2).

5.3 Norm estimates

Under the assumption of Theorem 5.2, let $g \in \bigotimes_{n=1}^N U_n$ be a local minimum of (5.2). Using the singular value decomposition, it is possible to find orthonormal bases $u_1^1, u_2^1, \ldots, u_{q_1}^1$ of U_1 and $T_1^1, T_2^1, \ldots, T_{q_1}^1$ of $\bigotimes_{n=2}^N U_n$ such that

$$g = \sum_{k=1}^{q_1} \sigma_k^1 u_k^1 \otimes T_k^1.$$

Permuting the factors of the tensor product, that is, the spatial variables of the function g, one can find a basis $u_1^n, u_2^n, \ldots, u_{q_n}^n$ and a set $\sigma_1^n, \sigma_2^n, \ldots, \sigma_{q_n}^n$ of "singular values" with this property for each of the subspaces U_n. In this case, Lemma 4.1 provides the estimate

$$\|u_k^n\|_{s_n} \le \frac{\|f\|_{s_n,n}}{|\sigma_k^n|} \le \frac{\|f\|_{s_n,n}}{\min_k |\sigma_k^n|}$$

for each $n = 1, 2, \ldots, N$. Note that $\sigma_k^n = |(f, u_k^n \otimes T_k^n)_0|$, since g is the best approximation of f in $\bigotimes_{n=1}^N U_n$.

6 Classically Differentiable Functions

Up to this point, our notion of regularity focused on the Sobolev spaces H^s of weakly differentiable functions in order to stay within a Hilbert space setting. But of course one may also think of regularity as classical differentiability. We want to catch up on this issue here. The key observation of our analysis, as stated in Lemma 2.1, is that the solutions of least squares tensor product approximations satisfy an integral equation of the form

$$u(x) = \int f(x, y) v(y) \, \mathrm{d}y. \tag{6.1}$$

This opens the door for a lot of possible refinements of the previous results and can be regarded as the central statement of this article. Above, we argued that the function u has the same H^s-regularity as the kernel f with respect to the variable x (Lemma 3.1). Under certain conditions this transfers to differentiable and continuously differentiable functions. For simplicity we consider only the latter. The basic tool is the following standard result from measure theory:

Lemma 6.1 *Let U be an open subset of $\mathbb{R}^{d_1}$ and $\phi: \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \to \mathbb{C}$ be an integrable function which possesses continuous partial derivatives $\frac{\partial \phi}{\partial x_\nu}(x, y)$ for $\nu = 1, 2, \dots, d_1$. Assume there exists an integrable function $\eta \geq 0$ on $\mathbb{R}^{d_2}$ such that*

$$\left| \frac{\partial \phi}{\partial x_\nu}(x, y) \right| \leq \eta(y)$$

for all $(x, y) \in U \times \mathbb{R}^{d_2}$. Then the function

$$u(x) = \int \phi(x, y)\, \mathrm{d}y$$

is integrable and continuously differentiable on U with the partial derivatives

$$\frac{\partial u}{\partial x_\nu}(x) = \int \frac{\partial \phi}{\partial x_\nu}(x, y)\, \mathrm{d}y.$$

Remark 6.2 For the partial differentiability of u it is only necessary that ϕ is partially differentiable. The continuity of the partial derivatives $\frac{\partial u}{\partial x_\nu}$ is a separate assertion which requires the continuity of the partial derivatives of ϕ. A proof of both statements can be found for instance in [1].

We present two examples of how this result can be used to refine Lemma 3.1. Since we are dealing with functions on the whole space, it is clear that, in order to bound the partial derivatives as required in Lemma 6.1, one has to incorporate their decay behavior. As a first example where this works, we consider the spaces $C_0^m(\mathbb{R}^d)$ of m-times continuously differentiable functions with compact support in $\mathbb{R}^d$, including $m = \infty$. If $u \in C_0^m(\mathbb{R}^d)$, then, in usual multi-index notation,

$$D^\alpha u(x) = \frac{\partial^{|\alpha|} u(x)}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \dots \partial x_d^{\alpha_d}}$$

exists and is continuous for all multi-indices α with $|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_d \leq m$. For simplicity we only treat the isotropic case here. A generalization to spaces of anisotropic smoothness is straightforward.

Lemma 6.3 *Let $f \in C_0^m(\mathbb{R}^{d_1+d_2})$, $m \in \mathbb{N} \cup \{\infty\}$, and $v \in L^2(\mathbb{R}^{d_2})$. Then the function u defined by (6.1) is in $C_0^m(\mathbb{R}^{d_1})$ and*[7]

$$D^\alpha u(x) = \int D_x^\alpha f(x, y) v(y)\, \mathrm{d}y \tag{6.2}$$

for all multi-indices α with $|\alpha| \leq m$.

[7] By D_x^α we indicate partial derivatives with regard to $x \in \mathbb{R}^{d_1}$.

Proof We give a proof for $m = 1$; the general case follows by induction. If we set

$$\phi(x, y) = f(x, y)v(y),$$

choose an arbitrary bounded open set $U \in \mathbb{R}^{d_1}$, and assume $v(y) < \infty$ for all $y \in \mathbb{R}^{d_2}$, we find the assumptions of Lemma 6.1 being satisfied. Indeed, as a majorizing function we may choose

$$\eta(y) = \sup_{x \in U} \left| f(x, y) \right| \left| v(y) \right|.$$

This function is integrable, which follows from the fact that

$$y \mapsto \sup_{x \in U} \left| f(x, y) \right|$$

is square integrable as a continuous function with bounded support. Since U is arbitrary, (6.2) follows for all $x \in \mathbb{R}^{d_1}$. □

As a second example we consider Schwartz spaces $\mathscr{S}$ of rapidly decreasing functions. These are infinitely differentiable functions $u(x)$ for which

$$x \mapsto x^{\beta} D^{\alpha} u(x)$$

remains bounded for all multi-indices α and β, see, e.g., [14].

Lemma 6.4 *Let $f \in \mathscr{S}(\mathbb{R}^{d_1+d_2})$ and $v \in L^2(\mathbb{R}^{d_2})$. Then the function u defined by* (6.1) *is in $\mathscr{S}(\mathbb{R}^{d_1})$, and its partial derivatives are given by* (6.2).

Proof Let B denote the unit ball in $\mathbb{R}^{d_2}$. Any Schwartz function $f \in \mathscr{S}(\mathbb{R}^{d_1+d_2})$ can be bounded on $\mathbb{R}^{d_1} \times B$ by a constant and outside of this region by a scalar multiple of $y \mapsto |y^{\beta}|^{-1}$, with β chosen in such a way that this function is square integrable on $\mathbb{R}^{d_2} \setminus B$. Thus, $f(x, y)$ can be bounded by a square integrable function $\eta(y)$ on the whole space $\mathbb{R}^{d_1+d_2}$. Consequently, the product $|f(x, y)v(y)|$ with the square integrable function v can be bounded by the integrable function $|\eta(y)v(y)|$. Since $D^{\alpha} f$ is a Schwartz function for every multi-index α, a repeated application of Lemma 6.1 on the functions $\phi(x, y) = D^{\alpha} f(x, y)v(y)$ therefore shows that u is infinitely differentiable, and that its partial derivatives are given by (6.2). In the same way one shows that

$$x^{\beta} D^{\alpha} u(x) = \int x^{\beta} D_x^{\alpha} f(x, y)v(y)\, \mathrm{d}y$$

remains bounded for every choice of β by bounding the rapidly decreasing function $x^{\beta} D_x^{\alpha} f(x, y)$ by a square integrable function of y. This proves that $u \in \mathscr{S}$. □

Concerning the application to tensor product approximations, we do not want to reformulate every single theorem. Instead, we summarize the result.

Theorem 6.5 *In Theorems 4.5 and 5.2 as well as in Corollaries 4.9, 5.6, and 5.8, the spaces* $H^{s_n,n}$ *may be replaced by the spaces* C_0^m *or* $\mathscr{S}$ *or their anisotropic variants.*

Here we used that C_0^m and $\mathscr{S}$ are both subspaces of L^2.

7 Conclusion

We have shown that first-order conditions for the factors in unconstrained best tensor product approximation problems in L^2 take the form of integral equations

$$u(x) = \int f(x, y)v(y)\,\mathrm{d}y,$$

with the target function f one wants to approximate as kernel. In particular, we considered canonical rank-r and optimal subspace approximations. This observation leads more or less immediately to results on the regularity of the solutions in dependence on the regularity of f itself. We demonstrated this for the spaces H^s, C_0^m, and $\mathscr{S}$ over the whole space. One can interpret the result as follows: if the target function f is smooth enough, then best tensor product approximations of f, even though one will hardly ever be able to calculate them exactly, have as good and even better regularity properties than f itself. This shows, for example, that tensor product approaches in quantum chemistry are valid from a physical point of view. Another conclusion is that a best low-rank approximation of a smooth function can be expected to have a high approximability on the set of admissible low-rank functions, for instance, by infinitely differentiable functions. This lets us hope that methods for calculating tensor product approximations can be constructed which converge fast to almost optimal results, provided that they are sophisticated enough to exploit the regularity of the factors shown in this article.

We have stated and proved the theorems for L^2 spaces over $\mathbb{R}^d$. This enabled us to define the Sobolev regularity via the Fourier transform and to give a simple proof of Lemma 3.1. If $A_n \subseteq \mathbb{R}^{d_n}$ are measurable domains, then

$$\bigotimes_{n=1}^{N} L^2(A_n) = L^2(A_1 \times A_2 \times \cdots \times A_N).$$

The results can be extended to such spaces if one can link the mapping properties of integral operators to the regularity of the kernel function. We suppose that this is possible in many situations.

Acknowledgements The author wishes to thank Heinz-Jürgen Flad, Christian Kamm, Christian Kreusler, Harry Yserentant, and Andreas Zeiser for helpful suggestions, which have been incorporated in the final version of this article. The useful comments of the referees are also gratefully acknowledged.

References

1. Bauer, H.: Measure and Integration Theory. De Gruyter, Berlin/New York (2001)

2. Beylkin, G., Mohlenkamp, M.J., Pérez, F.: Approximating a wavefunction as an unconstrained sum of Slater determinants. J. Math. Phys. **49**, 032107 (2008)
3. Chinnamsetty, S.R., Espig, M., Flad, H.-J., Khoromskij, B.N., Hackbusch, W.: Tensor product approximation with optimal rank in quantum chemistry. J. Chem. Phys. **127**, 084110 (2007)
4. Greub, W.H.: Multilinear Algebra. Springer, Berlin (1967)
5. Helgaker, T., Jørgensen, P., Olsen, J.: Molecular Electronic-Structure Theory. Wiley, New York (1999)
6. Horn, R.A., Johnson, C.R.: Topics in Matrix Analysis. Cambridge University Press, Cambridge (1991)
7. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. SIAM Rev. **51**(3), 455–500 (2009)
8. Khoromskij, B.N.: Tensor-structured numerical methods in scientific computing: Survey on recent advances. Preprint 21/2010 MPI MiS Leipzig (2010)
9. Le Bris, C.: A general approach for multiconfiguration methods in quantum molecular chemistry. Ann. Inst. H. Poincaré **11**, 441–484 (1994)
10. Lewin, M.: Solutions of the multiconfiguration equations in quantum chemistry. Arch. Ration. Mech. Anal. **171**, 83–114 (2004)
11. De Silva, V., Lim, L.-H.: Tensor rank and the ill-posedness of the best low-rank approximation problem. SIAM J. Matrix Anal. Appl. **30**(3), 1084–1127 (2008)
12. Tyrtyshnikov, E.: Preservation of linear constraints in approximation of tensors. Numer. Math. Theory Methods Appl. **2**(4), 421–426 (2009)
13. Uschmajew, A.: Well-posedness of convex maximization problems on Stiefel manifolds and orthogonal tensor product approximations. Numer. Math. **115**(2), 309–331 (2010)
14. Yosida, K.: Functional Analysis. Springer, Berlin (1974)
15. Yserentant, H.: Regularity and Approximability of Electronic Wave Functions. Springer, Berlin/Heidelberg (2010)