The combination technique and some generalisations

Markus Hegland, Jochen Garcke and Vivien Challis
Mathematical Sciences Institute
Australian National University
Canberra ACT 0200 Australia
{jochen.garcke,markus.hegland}@anu.edu.au
Department of Mathematics
University of Queensland
Brisbane QLD 4072 Australia
vchallis@maths.uq.edu.au

August 18, 2006

Abstract

The combination technique has repeatedly been shown to be an effective tool for the approximation with sparse grid spaces. Little is known about the reasons of this effectiveness and in some cases the combination technique can even break down. It is known, however, that the combination technique produces an exact result in the case of a projection into a sparse grid space if the involved partial projections commute.

The performance of the combination technique is analysed using a projection framework and the C/S decomposition. Error bounds are given in terms of angles between the spanning subspaces or the projections onto these subspaces. Based on this analysis modified combination coefficients are derived which are optimal in a certain sense and which can substantially extend the applicability and performance of the combination technique.

1 Introduction

Standard finite element approaches for applications with more than four variables suffer under the *curse of dimensionality*, their numerical solution is infeasible on the computational equipment available nowadays. Zenger [28] has introduced *sparse grid approximations* into finite elements which substantially reduce the computational complexity at a moderate cost to the accuracy allowing the numerical treatment of problems with ten variables and more. Note that the approximation theory for sparse grids requires slightly stronger smoothness

conditions in comparison to ordinary finite elements, in particular, so called *mixed Sobolev norms* are used instead of the standard Sobolev norms. The underlying idea of a constrained *tensor product decomposition* has been first suggested by Smolyak [25] for numerical integration. The ansatz closely relates to ANOVA *decompositions* [26] used in statistics. In fact, sparse grids have been used for regression problems as well [11, 12, 20]. Note that dimension adaptive sparse grid techniques, which adapt to the particular smoothness of the solution in different dimensions, were introduced in [15, 19].

The elements of the sparse grid space can be represented in a *hierarchical basis* [27] and many algorithms for hierarchical basis methods including wavelets can be used for the solution [5, 20]. Compared to the commonly used nodal basis, a hierarchical basis of, e.g. multilinear functions has its disadvantages, as the corresponding matrices have *reduced sparsity* and a less regular structure. This is due to the fact that the supports of the lower level basis functions are large and intersect nontrivially with many higher level basis functions. These difficulties increase with dimension.

An efficient way to avoid the problem of reduced sparsity is given by the *combination technique* which is mentioned in the original paper by Smolyak [25]. Its introduction in a modern computational framework was given in [18]. Basically, the combination technique uses the fact that sparse grid spaces can be seen as the sum of ordinary finite element spaces. The variational problems can be solved for each of the component spaces independently and the solution in the sparse grid space is then approximated by a linear combination of these partial solutions. For elliptic model problems this approach does introduce an additional error which, however, is of the same order as the sparse grid approximation error [4, 17]. If certain error expansions for the component approximations exist one can show that the combination technique achieves the same approximation properties as sparse grids due to the cancellation of some higher order error terms [4], an effect which is well known from extrapolation techniques. Further advantages of the combination technique include the inherent parallelism and the possibility to utilise the structure, and indeed, even software for standard finite element approximations.

Many computational problems, including Galerkin finite element methods for elliptic partial differential equations, can be viewed as projections into finite dimensional spaces with respect to a suitably chosen scalar product. The combination technique does provide the exact sparse grid solution when the partial projection operators commute. This is the case for interpolation with tensor product spaces [18], but for example is not the case for the Poisson problem or for regression. As mentioned, for the Poisson problem the combination technique nevertheless does provide a very good approximation [4]. However, the approximation can be poor for some large scale regression problems, in particular when strongly correlated features are used – which is not uncommon for data mining applications. One of the authors has thus introduced a variant of the combination technique in [21] where the combination coefficients are chosen adaptively to improve performance.

In the following we look at tools to analyse approximation problems in terms

of partial projections and their commutators. In the general setting we are given a function u and wish to find Pu, where $P : H \to V$ is the orthogonal projection from the (usually infinite dimensional) space H onto our constructed solution space V. We do not know P directly, but have subspaces $V_1, V_2, \ldots, V_m$ which together make up V, i.e. $V = V_1 + V_2 + \cdots + V_m$. We wish to approximate P via a linear combination of the (known or easily computable) orthogonal projections $P_{V_i} : H \to V_i$ onto each of the V_i. Thus we seek combination coefficients c_i such that

$$\sum_{i=1}^{m} c_i P_{V_i} u \approx Pu. \tag{1}$$

Several (new) variants of the combination technique will be explored.

In section 2, we consider the case of two spaces (and possibly their intersection). Conditions for when combinations of projections form again a projection are shown, errors for a general combination technique are derived and combination techniques based on worst case and average case analysis are given. These methods are also combined with a multiplicative approximation. The optimal combination technique ("opticom") is introduced and the error is compared with the error of the classical combination technique. Finally we show some numerical experiments. In section 3, the case of n partial spaces is examined. It is shown how the (classical) combination coefficients can be derived from the "inclusion/exclusion" principle from combinatorics. The requirements for this to hold is that the partial spaces need to form a distributed lattice of vector spaces and, furthermore, the projections onto the elements of this lattice should be additive functions on this lattice. If the projections onto the subspaces do commute these properties are automatically fulfilled. For this case the combination coefficients are defined recursively using the Hasse diagram of the lattice. Two types of new "generalised" combination coefficients are introduced, one which provides a best approximation "on average", i.e. with respect to the Frobenius norm. A second type selects the combination coefficients adaptively so that an optimal combination approximation is obtained. The corresponding method is the "opticom" method. Again we provide some numerical examples. Section 3 concludes with a discussion of both the worst case error bound and an error bound which depends on the vector which is projected. In particular, several combination techniques (including opticom and the classical one) are compared. The final section 4 puts this work in the broader context of high dimensional approximation and shows possible future research projects.

2 Combination techniques for two spaces

In this section the problem of approximating the orthogonal projection into the sum $V = V_1 + V_2$ of two closed subspaces $V_i \subset H$ of a Hilbert space is considered. The approximations are either *additive*, where

$$T^a = c_1 P_{V_1} + c_2 P_{V_2} + c_{12} P_{V_1 \cap V_2}$$

or *multiplicative*, where

$$T^{mc} = c_1 P_{V_1} + c_2 P_{V_2} + c_{12} P_{V_1} P_{V_2}.$$

The following result is based on [1, 2] and shows when an additive combination technique T^a can be a projection into V.

Theorem 1. *Let $V_1, V_2 \subset V$, $V_1 \neq V_2$, and $V_{12} := V_1 \cap V_2 \neq 0$ with orthogonal projections P_{V_1}, P_{V_2}, and $P_{V_{12}}$, respectively. $P := c_1 P_{V_1} + c_2 P_{V_2} + c_{12} P_{V_{12}}$ with $(c_1, c_2, c_{12}) \neq (0, 0, 0)$ is idempotent if and only if*

$$P_{V_1} P_{V_2} = P_{V_2} P_{V_1} = P_{V_{12}},\ c_1 = c_2 = 1 \text{ and } c_{12} \in \{-1, -2\}.$$

Proof.

(i) Idempotency of P is equivalent to

$$c_1(1-c_1)P_{V_1}+c_2(1-c_2)P_{V_2}+c_{12}(1-c_{12}-2c_1-2c_2)P_{V_{12}} = c_1c_2(P_{V_1}P_{V_2}+P_{V_2}P_{V_1}) \tag{2}$$

and it follows from $P_{V_1} P_{V_2} = P_{V_2} P_{V_1} = P_{V_{12}}$, $c_1 = c_2 = 1$ and $c_{12} = -1$ or $c_{12} = -2$ that P is idempotent.

(ii) We now multiply (2) with $P_{V_{12}}$ to get

$$\left((c_1 + c_2 + c_{12}) - (c_1 + c_2 + c_{12})^2\right) P_{V_{12}} = 0$$

and as $P_{V_{12}} \neq 0$ one derives $c_1 + c_2 + c_{12} = 1$ or $c_1 + c_2 + c_{12} = 0$ as necessary conditions for idempotency.

Multiply (2) with P_{V_1} and substitute $c_{12} = 1-(c_1+c_2)$ or $c_{12} = -(c_1+c_2)$:

$$c_1(1-c_1)P_{V_1}+c_2(1-c_2)P_{V_1}P_{V_2}-(c_1+c_2)(1-(c_1+c_2))P_{V_{12}} = c_1c_2(P_{V_1}P_{V_2}+P_{V_1}P_{V_2}P_{V_1})$$

and subtract the transpose of this expression to get

$$c_2(1 - c_1 - c_2)(P_{V_1} P_{V_2} - P_{V_2} P_{V_1}) = 0.$$

If $P_{V_1} P_{V_2} \neq P_{V_2} P_{V_1}$ we have $c_1 + c_2 = 1$ and, as $c_{12} \neq 0$, get $c_{12} = -1$. Inserting this into (2) gives $c_1(1 - c_1)(P_{V_1} - P_{V_2})^2 = 0$ and so $c_1 = 1$, and therefore $c_2 = 0$, or $P_{V_1} = P_{V_2}$, both cases are excluded. Thus $P_{V_1} P_{V_2} = P_{V_2} P_{V_1} = P_{V_{12}}$ and from (2) one gets

$$c_1(1 - c_1)P_{V_1} + c_2(1 - c_2)P_{V_2} - (c_1 + c_2 - c_1^2 - c_2^2)P_{V_{12}} = 0$$

or

$$c_1(1 - c_1)(P_{V_1} - P_{V_{12}}) + c_2(1 - c_2)(P_{V_2} - P_{V_{12}}) = 0$$

from which it follows that $c_1 = 1$ and $c_2 = 1$.

□

The next subsection provides some background, in particular the C/S decomposition and the second subsection introduces several approximations and provides some error bounds.

2.1 Angles, commutators and the CS decomposition

We will characterise the error of an approximation to P_V in terms of geometric quantities relating the two spaces V_1 and V_2. Let us introduce the concept of the *angle* $\alpha(V_1, V_2) \in [0, \pi/2]$ between two closed subspaces of a Hilbert space H. According to [9] the cosine of the angle between two spaces $\cos(\alpha(V_1, V_2)) := c(V_1, V_2)$ is defined as

$$c(V_1, V_2) := \sup\left\{(f_1, f_2) \mid f_i \in V_i \cap (V_1 \cap V_2)^\perp, \|f_i\| \leq 1, i = 1, 2\right\}.$$

It follows directly that $c(V_1, V_2) = c(V_1 \cap (V_1 \cap V_2)^\perp, V_2 \cap (V_1 \cap V_2)^\perp)$. The angle is (essentially) defined for (sub)spaces with intersection 0 and is not changed by adding any orthogonal intersection, i.e. one has $c(V_1 + V_3, V_2 + V_3) = c(V_1, V_2)$ if $V_3 \perp (V_1 + V_2)$. In the case of two one dimensional spaces $V_i = \langle v_i \rangle$ one has $c(V_1, V_2) = (v_1, v_2)/(\|v_1\|\|v_2\|)$, i.e. the ordinary angle. For any two spaces with a null intersection the angle is the *minimal angle* between any two vectors $v_1 \in V_1$ and $v_2 \in V_2$. In particular, if two spaces are orthogonal, their angle is $\pi/2$ and $c(V_1, V_2) = 0$.

The angle can be characterised in terms of the orthogonal projections P_{V_i} into the closed subspaces V_i and the corresponding operator norm, it holds [7]

$$c(V_1, V_2) = \|P_{V_1} P_{V_2} P_{(V_1 \cap V_2)^\perp}\|. \tag{3}$$

For two *commuting operators* P_{V_i}, i.e.

$$P_{V_1} P_{V_2} = P_{V_2} P_{V_1} = P_{V_1 \cap V_2}$$

we immediately observe that $c(V_1, V_2) = 0$ and so $\alpha(V_1, V_2) = \pi/2$. In particular, this is the case where two spaces are orthogonal to each other. Note that one also gets an angle of $\pi/2$ for $V_1 \subset V_2$ (which is different from what one might expect by intuition).

We now use the following result, see, e.g. [3, 16]:

Lemma 1 (C/S decomposition). *Let H be a finite dimensional space and P_{V_i} be the orthogonal projection into the subspace $V_i \subset H, i = 1, 2$. Then there exists an orthogonal basis of H with respect to which the matrices of the projections are of the form*

$$P_{V_1} = \begin{bmatrix} I & & & & & \\ & 0 & & & & \\ & & I & & & \\ & & & I & & \\ & & & & 0 & \\ & & & & & 0 \end{bmatrix}, \quad P_{V_2} = \begin{bmatrix} C^2 & CS & & & & \\ CS & S^2 & & & & \\ & & I & & & \\ & & & 0 & & \\ & & & & I & \\ & & & & & 0 \end{bmatrix},$$

where C, S are positive diagonal real matrices such that $C^2 + S^2 = I$, the symbol I denotes identity matrices of various sizes and the corresponding blocks in the two projection matrices are of the same size.

With Lemma 1 we get a particular representation for the following related projections:

$$P_{V_1 \cap V_2} = \begin{bmatrix} 0 & & & & & \\ & 0 & & & & \\ & & I & & & \\ & & & 0 & & \\ & & & & 0 & \\ & & & & & 0 \end{bmatrix}, \quad P_{V_1 + V_2} = \begin{bmatrix} I & & & & & \\ & I & & & & \\ & & I & & & \\ & & & I & & \\ & & & & I & \\ & & & & & 0 \end{bmatrix},$$

as well as

$$P_{V_1^\perp} = I - P_{V_1} = \begin{bmatrix} 0 & & & & & \\ & I & & & & \\ & & 0 & & & \\ & & & 0 & & \\ & & & & I & \\ & & & & & I \end{bmatrix}, \quad P_{V_2^\perp} = I - P_{V_2} = \begin{bmatrix} S^2 & CS & & & & \\ CS & C^2 & & & & \\ & & 0 & & & \\ & & & I & & \\ & & & & 0 & \\ & & & & & I \end{bmatrix}$$

and finally, for the *product* $P_{V_1} P_{V_2}$ and the *commutator* $[P_{V_1}, P_{V_2}] := P_{V_1} P_{V_2} - P_{V_2} P_{V_1}$ we have:

$$P_{V_1} P_{V_2} = \begin{bmatrix} C^2 & CS & & & & \\ 0 & 0 & & & & \\ & & I & & & \\ & & & 0 & & \\ & & & & 0 & \\ & & & & & 0 \end{bmatrix}, \quad [P_{V_1}, P_{V_2}] = \begin{bmatrix} 0 & CS & & & & \\ -CS & 0 & & & & \\ & & 0 & & & \\ & & & 0 & & \\ & & & & 0 & \\ & & & & & 0 \end{bmatrix}.$$

Note that the C/S decomposition has a geometric interpretation. Let γ_i and σ_i denote the i-th (diagonal) elements of C and S, respectively. It follows that $\gamma_i^2 + \sigma_i^2 = 1$. As the γ_i and σ_i are positive and less than one, there are $\alpha_i \in (0, \pi/2)$ such that $\gamma_i = \cos(\alpha_i)$ and $\sigma_i = sin(\alpha_i)$. One can rearrange the leading two by two block matrices of P_{V_1} and P_{V_2} in the C/S decomposition into a block diagonal matrix with two by two blocks of the form

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \gamma_i^2 & \gamma_i \sigma_i \\ \gamma_i \sigma_i & \sigma_i^2 \end{bmatrix} = \begin{bmatrix} \gamma_i \\ \sigma_i \end{bmatrix} \begin{bmatrix} \gamma_i & \sigma_i \end{bmatrix}, \tag{4}$$

respectively. These are just the projections onto the first coordinate axis and onto the line with direction vector $\gamma_i e_1 + \sigma_i e_2$. The space V_2, which is the range of P_{V_2}, thus consists of the direct sum of one dimensional spaces which are at an angle of α_i to their corresponding subspaces of V_1, and the spaces $V_1 \cap V_2$ and $V_1^\perp \cap V_2$.

From the C/S decomposition and the characterisation of the angle in terms of the projections (3) one gets

$$c(V_1, V_2) = \| P_{V_1} P_{V_2} P_{(V_1 \cap V_2)^\perp} \| = \left\| \begin{bmatrix} C^2 & CS \\ 0 & 0 \end{bmatrix} \right\| = \|C\| = \max_i \gamma_i.$$

Thus the angle between two spaces is exactly the minimum of the angles between any two corresponding (according to equation 4) one-dimensional spaces in the C/S decomposition of their projections P_{V_i}.

The commutator $[P_{V_1}, P_{V_2}]$ sets all components outside of the first two blocks in the C/S decomposition to zero, rotates the components corresponding to the first blocks each by $\pi/2$ and then dilates the components with $\gamma_i\sigma_i$. The (spectral) norm of the commutator is

$$\|[P_{V_1}, P_{V_2}]\| = \left\| \begin{bmatrix} 0 & CS \\ -CS & 0 \end{bmatrix} \right\| = \|CS\| = \max_i \gamma_i\sigma_i.$$

It follows directly from $0 < \sigma_i < 1$ that

$$\|[P_{V_1}, P_{V_2}]\| < c(V_1, V_2).$$

Conversely, as $\gamma_i = c(V_1, V_2)$ for some i one gets the bound

$$c(V_1, V_2)\sqrt{1 - c(V_1, V_2)^2} \leq \|[P_{V_1}, P_{V_2}]\|.$$

If $c(V_1, V_2) \leq 1/\sqrt{2}$ the left hand side is monotone and in this case one gets a bound for the cosine of the angle in terms of the commutator as

$$c(V_1, V_2) \leq \frac{1 - \sqrt{1 - 4\|[P_{V_1}, P_{V_2}]\|^2}}{2}.$$

In summary, for the case of commuting projections, the C and S do not occur in the C/S decomposition, the commutator is zero and the cosine of the angle between the two spaces is $c(V_1, V_2) = 0$. If the angle between the two spaces is larger than $\arccos(1/\sqrt{2}) = \pi/4$ then one gets a lower bound for the angle (or and upper bound for the cosine) in terms of the norm of the commutator.

When the projections commute, the product of the projections is equal to the projection onto the intersection. More generally, the C/S decomposition provides a bound for how well the product approximates the projection onto the intersection by

$$\|P_{V_1}P_{V_2} - P_{V_1 \cap V_2}\| = \left\| \begin{bmatrix} C^2 & CS \\ 0 & 0 \end{bmatrix} \right\| = \|C\| = c(V_1, V_2). \tag{5}$$

Of course, this also follows directly from $c(V_1, V_2) = \|P_{V_1}P_{V_2}P_{(V_1 \cap V_2)^\perp}\|$ as $P_{V_1}P_{V_2}P_{(V_1 \cap V_2)^\perp} = P_{V_1}P_{V_2}(I - P_{V_1 \cap V_2}) = P_{V_1}P_{V_2} - P_{V_1 \cap V_2}$.

Corresponding to the C/S decomposition we now introduce for any vector $u \in H$ a partitioning by

$$u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \end{pmatrix}$$

and observe

$$\|u\|^2 = \|u_1\|^2 + \|u_2\|^2 + \|u_3\|^2 + \|u_4\|^2 + \|u_5\|^2 + \|u_6\|^2.$$

The norms of the projections P_{V_i} are

$$\|P_{V_1}u\|^2 = \|u_1\|^2 + \|u_3\|^2 + \|u_4\|^2, \quad \|P_{V_2}u\|^2 = \|Cu_1 + Su_2\|^2 + \|u_3\|^2 + \|u_5\|^2,$$

and the scalar product between the projections is

$$(P_{V_1}u, P_{V_2}u) = (u_1, C^2u_1 + CSu_2) + \|u_3\|^2 = (Cu_1, Cu_1 + Su_2) + \|u_3\|^2.$$

It follows that

$$\begin{aligned}|(P_{V_1}u, P_{V_2}u)| &\leq \|Cu_1\|\|Cu_1 + Su_2\| + \|u_3\|^2 \\ &\leq c(V_1, V_2)\sqrt{\|P_{V_1}u\|^2 - \|u_3\|^2 - \|u_4\|^2}\sqrt{\|P_{V_2}u\|^2 - \|u_3\|^2 - \|u_5\|^2} + \|u_3\|^2.\end{aligned}$$

2.2 "Two space" combination approximations

We now apply the observations from the previous subsection to get bounds on the errors of several approximations. Using the C/S decomposition one gets

$$\|P_{V_1}+P_{V_2}-P_{V_1}P_{V_2}-P_{V_1+V_2}\| = \left\|\begin{bmatrix} 0 & 0 \\ CS & S^2 - I \end{bmatrix}\right\| = \left\|\begin{bmatrix} 0 & 0 \\ CS & -C^2 \end{bmatrix}\right\| = c(V_1, V_2).$$

This is the error of the simplest multiplicative combination approximation.

If we apply the triangular inequality and the equation (5) from the previous subsection we get $\|P_{V_1} + P_{V_2} - P_{V_1\cap V_2} - P_{V_1+V_2}\| \leq 2c(V_1, V_2)$ as a bound for the ordinary combination approximation. However, by direct application of the C/S decomposition one gets

$$\|P_{V_1} + P_{V_2} - P_{V_1\cap V_2} - P_{V_1+V_2}\| = \left\|\begin{bmatrix} C^2 & CS \\ CS & -C^2 \end{bmatrix}\right\| = c(V_1, V_2).$$

Thus the (worst case) error of the combination approximation is the cosine of the angle between the two spaces, giving good results if the two spaces are fairly close to orthogonal. Interestingly, the multiplicative approximation does not appear to have any advantage over the additive version, in contrast to many practical cases where the multiplicative approximation is observed to yield better approximations. However, this is due to the fact that the worst case error was considered. Nevertheless, the results above do indicate that the additive combination approximation is competitive, in particular, it is better than the simple additive (Jacobi) approximation $P_{V_1} + P_{V_2}$ for which the worst case error is $\|P_{V_1} + P_{V_2} - P_{V_1+V_2}\| = 1$ even in the case of commuting projections, unless the two spaces V_1 and V_2 are orthogonal.

The combination approximation

$$T^c := P_{V_1} + P_{V_2} - P_{V_1\cap V_2}$$

is exact (i.e. $T^c = P_{V_1+V_2}$ in this case) if the spaces $V_1 \cap (V_1 \cap V_2)^\perp$ and $V_2 \cap (V_1 \cap V_2)^\perp$ are orthogonal. It turns out, that this approximation is also the best approximation in terms of the operator norm and one has

Proposition 1. *For any additive combination approximation* $T^a = c_1 P_{V_1} + c_2 P_{V_2} + c_{12} P_{V_1 \cap V_2}$ *one has*

$$c(V_1, V_2) \leq \|T^a - P_{V_1+V_2}\|.$$

Moreover, if $c_1 = c_2 = 1$ *and* $c_{12} = -1$ *one has*

$$\|T^c - P_{V_1+V_2}\| = c(V_1, V_2).$$

Proof. Observe that the error on the intersection $V_1 \cap V_2$ is $|c_1 + c_2 + c_{12} - 1|$ and, as $T^a - P_{V_1+V_2} = P_{V_1 \cap V_2}(T^a - P_{V_1+V_2}) + P^{\perp}_{V_1 \cap V_2}(T^a - P_{V_1+V_2})$ one has

$$\|T^a - P_{V_1+V_2}\| = \max\{|c_1 + c_2 + c_{12} - 1|, \|(c_1 P_{V_1} + c_2 P_{V_2} - P_{V_1+V_2}) P_{(V_1 \cap V_2)^{\perp}}\|\}.$$

It follows that the error is not increased if one replaces c_{12} with $1 - c_1 - c_2$. Thus while there might be optimal methods for which $c_{12} \neq 1 - c_1 - c_2$ by replacing the c_{12} one gets another optimal method. Consider in the following only methods for which $c_{12} = 1 - c_1 - c_2$.

The C/S decomposition provides orthogonal decompositions $V_1 = U_1 \oplus U_2$ and $V_2 = U_1 \oplus U_3$ where $U_2 = V_1 \cap V_2^{\perp}$ and $U_3 = V_2 \cap V_1^{\perp}$. The errors on U_2 and U_3 are (if these spaces are not null spaces) $|c_1 - 1|$ and $|c_2 - 1|$, respectively. Denote the error on U_1 by $g(c_1, c_2)$ and it follows that

$$g(c_1, c_2) \leq \|T^a - P_{V_1+V_2}\| \leq \max\{g(c_1, c_2), |c_1 - 1|, |c_2 - 1|\}.$$

Note that in the case of $c_1 = c_2 = 1$ one has $\|T^c - P_{V_1+V_2}\| = g(1, 1) = c(V_1, V_2)$.

In terms of the notation of the C/S decomposition one has

$$g(c_1, c_2) = \left\| \begin{bmatrix} c_1 I + c_2 C^2 - I & c_2 CS \\ c_2 CS & c_2 S^2 - I \end{bmatrix} \right\|.$$

By applying a reverse odd-even permutation (4) to the matrix inside this norm we get a block diagonal matrix with blocks

$$\begin{bmatrix} c_1 + c_2 \gamma_i^2 - 1 & c_2 \gamma_i \sigma_i \\ c_2 \gamma_i \sigma_i & c_2 \sigma_i^2 - 1 \end{bmatrix}.$$

With $\xi_1 = (c_1 + c_2)/2 - 1$ and $\xi_2 = (c_1 - c_2)/2$ the norms of the blocks can be seen to be

$$|\xi_1| + \sqrt{\gamma_i^2 (\xi_1 + 1)^2 + \sigma_i^2 \xi_2^2}$$

and it follows that

$$g(c_1, c_2) = |\xi_1| + \max_i \sqrt{\gamma_i^2 (\xi_1 + 1)^2 + \sigma_i^2 \xi_2^2}$$

which is a monotonically increasing function of $|\xi_2|$ and one gets

$$g(c_1, c_2) \geq |\xi_1| + c(V_1, V_2)|\xi_1 + 1|$$

and equality holds if $c_1 = c_2$. The minimal value of the right hand side is obtained for $\xi_1 = 0$ and it follows that

$$c(V_1, V_2) \leq \|T^a - P_{V_1+V_2}\|.$$

Note that the case $\xi_1 = 0$ corresponds to the classical combination technique for which it had been demonstrated earlier that the error is equal to $c(V_1, V_2)$. It follows thus that the "classical combination approximation" is optimal in the worst case sense. □

This suggests that maybe the worst case error analysis is not realistic. An alternative, corresponding to an average case scenario, is provided by the Frobenius norm, which we recall is defined for a matrix A as

$$\|A\|_F = \sqrt{\mathrm{tr}(AA^T)} = \sqrt{\sum_{i,j} a_{ij}^2}, \tag{6}$$

where $\mathrm{tr}(A)$ represents the trace, or sum of the diagonal elements, of A. From the C/S decomposition we get

$$\|P_{V_1} + P_{V_2} - P_{V_1 \cap V_2} - P_{V_1+V_2}\|_F^2 = 2\sum_{i=1}^{n} \gamma_i^2 \tag{7}$$

if n is the size of the blocks C and S. For the multiplicative approximation one gets half the error as

$$\|P_{V_1} + P_{V_2} - P_{V_1} P_{V_2} - P_{V_1+V_2}\|_F^2 = \sum_{i=1}^{n} \gamma_i^2.$$

For a general additive combination approximation it holds

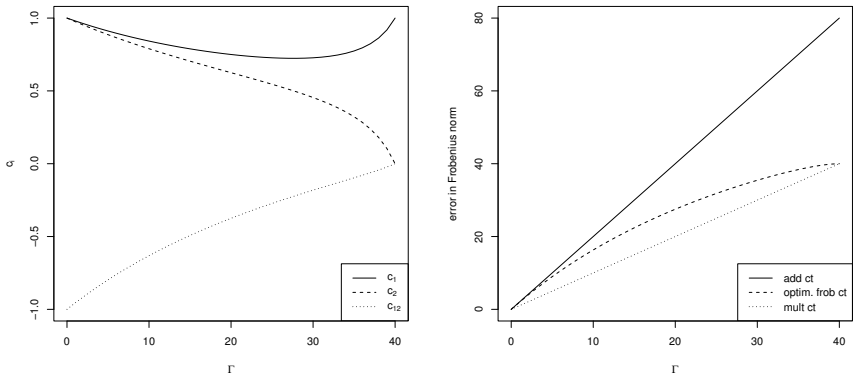
$$\|T^a - P_{V_1+V_2}\|_F^2 = 2c_1c_2\Gamma(V_1, V_2) + (c_1-1)^2 n_1 + (c_2-1)^2 n_2 + (c_1+c_2+c_{12}-1)^2 n_{12},$$

where n_i is the dimension of $V_i \cap (V_1 \cap V_2)^\perp$, $i = 1, 2$, n_{12} is the dimension of the intersection $V_1 \cap V_2$ and

$$\Gamma(V_1, V_2) = \sum_{i=1}^{n} \gamma_i^2 = \mathrm{tr}(P_{V_1} P_{V_2}) - n_{12}.$$

Note that n_i is *not* the size of the u_i in the C/S decomposition lemma, but rather the difference of the dimension of V_i and the dimension of $(V_1 \cap V_2)$. It follows that the best approximation with respect to the Frobenius norm must satisfy $c_{12} = 1 - c_1 - c_2$. To determine c_1 and c_2 one now needs to minimise a quadratic function

$$J(c_1, c_2) = n_1(c_1 - 1)^2 + n_2(c_2 - 1)^2 + 2\Gamma c_1 c_2$$



(a) coefficients c_1, c_2, and c_{12}

(b) error in the Frobenius norm

Figure 1: Examples for the dependence of the coefficients c_1, c_2, and c_{12} and the errors on Γ for $n_1 = 50$ and $n_2 = 40$.

where $\Gamma = \Gamma(V_1, V_2)$. The normal equations for this minimisation problem are

$$\begin{bmatrix} n_1 & \Gamma \\ \Gamma & n_2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} n_1 \\ n_2 \end{bmatrix}.$$

The solution is therefore

$$c_1 = n_2 \frac{n_1 - \Gamma}{n_1 n_2 - \Gamma^2}, \quad c_2 = n_1 \frac{n_2 - \Gamma}{n_1 n_2 - \Gamma^2},$$

and the minimum of the quadratic function J, i.e. the error of the approximation, is

$$\|T^{(\mathrm{a})} - P_{V_1+V_2}\|_F^2 = \min J(c_1, c_2) = \Gamma \frac{2n_1 n_2 - (n_1 + n_2)\Gamma}{n_1 n_2 - \Gamma^2}.$$

Since $2n_1 n_2 - (n_1 + n_2)\Gamma \leq 2(n_1 n_2 - \Gamma^2)$ this error is always smaller or equal to the error of the ordinary combination technique (7), it is equal only in the limit of $n = \Gamma$ for the special case $n_1 = n_2 = n$. Asymptotically with $\Gamma \to 0$ this gives the same error as the ordinary combination approximation as the spaces V_1 and V_2 get more and more orthogonal until we get the coefficients $c_1 = c_2 = 1$ and recover the ordinary combination technique for the case $\Gamma = 0$. In Figure 1(a) we show the dependence of the coefficients c_i on Γ, going from the ordinary combination technique for $\Gamma = 0$ to the case of $\Gamma = n$. Consider for this case a $u \in V_2 \cap V_1^{\perp}$. As there is no component in V_1 the value of c_1 will not have any influence and only combination coefficient c_2 is affecting the approximation.

However, since for a Γ close enough to $n < n_2$ one gets $c_2 < 0.5$ the combination technique results in a dilation by a factor < 0.5 so that the error is over 50% in any vector space norm. For the special case $n_1 \geq n_2 = n$ it follows from $\Gamma \to n$ that V_2 is more and more included in V_1 and for $\Gamma = n$ we recover the case $V_2 \subset V_1$.

Figure 1(b) shows the dependence of the error for the ordinary combination technique, the one with 'optimal' coefficients in the Frobenius norm, and the multiplicative method in relation to Γ. We see for $\Gamma \to N$ an increasing error for the methods and observe that the error of the 'optimal' (in the average) method converges to the error of the multiplicative method.

This shows that although we compute coefficients c_1, c_2, c_{12} for which the average case error is minimised it still can be large, especially for functions $u \in V_2 \cap V_1^{\perp}$, $n_1 > n_2$.

So far we considered the worst case scenario using the operator norm and an average case scenario using the Frobenius norm. In the following, consider an arbitrary, but fixed $u \in V$. One finds that

Proposition 2. *Let* $U_1 := V_1 \cap (V_1 \cap V_2)^{\perp}$, $U_2 := V_2 \cap (V_1 \cap V_2)^{\perp}$, $\gamma = \angle(P_{U_1}u, P_{U_2}u)$, $\alpha_1 = \angle(P_{U_1+U_2}u, P_{U_1}u)$, $\alpha_2 = \angle(P_{U_1+U_2}u, P_{U_2}u)$ *and*

$$e^2(c_1, c_2, c_{12}) = \|P_{V_1+V_2}u - c_1 P_{V_1}u - c_2 P_{V_2}u - c_{12} P_{V_1 \cap V_2}u\|^2,$$

$e_o^2 = \inf_{c_1,c_2,c_{12}} e^2(c_1, c_2, c_{12})$, *and* $e_c^2 = e^2(1, 1, -1)$. *Then*

$$e^2(c_1, c_2, c_{12}) = \left(x^T A x - 2b^T x + 1\right) \|P_{U_1+U_2}u\|^2 + (c_1 + c_2 + c_{12} - 1)^2 \|P_{V_1 \cap V_2}u\|^2$$

where $x = (c_1 \cos\alpha_1, c_2 \cos\alpha_2)^T$, $b = (\cos\alpha_1, \cos\alpha_2)^T$ *and*

$$A = \begin{bmatrix} 1 & \cos\gamma \\ \cos\gamma & 1 \end{bmatrix}.$$

Consequently, one has

$$e_c^2 = (1 - (\cos^2\alpha_1 - 2\cos\alpha_1 \cos\alpha_2 \cos\gamma + \cos^2\alpha_2))\|P_{U_1+U_2}u\|^2$$

and

$$e_c^2 = e_o^2 + \cos^2\gamma(\|P_{U_1+U_2}u\|^2 - e_o^2). \tag{8}$$

Furthermore, if $\gamma = \alpha_1 + \alpha_2$ *the* $P_{U_1+U_2}u$, $P_{U_1}u$ *and* $P_{U_2}u$ *are collinear and in this case one has* $e_o^2 = 0$ *and* $e_c^2 = \cos^2\gamma$. *The minimum is achieved for* $c_1 + c_2 = 1$ *in the case of* $\gamma = 0$ *and for*

$$c_1 = \frac{1 - \cos\gamma\cos\alpha_2/\cos\alpha_1}{\sin^2\gamma}, \quad c_2 = \frac{1 - \cos\gamma\cos\alpha_1/\cos\alpha_2}{\sin^2\gamma}$$

when $\gamma > 0$.

Proof. As the U_i are orthogonal to $V_1 \cap V_2$ one has

$$\begin{aligned} e^2(c_1, c_2, c_{12}) =& \|P_{U_1+U_2}u - c_1 P_{U_1}u - c_2 P_{U_2}u - c_{12} P_{U_1 \cap U_2}u\|^2 \\ &+ (c_1 + c_2 + c_{12} - 1)^2 \|P_{V_1 \cap V_2}u\|^2. \end{aligned} \tag{9}$$

As $\|P_{U_i}u\| = \cos\alpha_i \|P_{U_1+U_2}u\|$ and $(P_{U_1}u, P_{U_2}u) = \cos\gamma \|P_{U_1}u\| \|P_{U_2}u\|$ one gets

$$\|P_{U_1+U_2}u - c_1 P_{U_1}u - c_2 P_{U_2}u - c_{12} P_{U_1 \cap U_2}u\|^2 = \left(x^T A x - 2b^T x + 1\right) \|P_{U_1+U_2}u\|^2.$$

By inserting this relation and $c_1 = c_2 = -c_{12} = 1$ into (9) one directly gets the formula for e_c^2.

For the least error we consider the tripod defined by $P_{U_1}u$, $P_{U_2}u$ and $P_{U_1+U_2}u$. The triangle inequality on the sphere gives $\gamma \leq \alpha_1 + \alpha_2$ and the three corners of the tripod are in the same plane if $\gamma = \alpha_1 + \alpha_2$. Consider now the latter case for $\gamma = 0$, therefore we also have $\alpha_1 = \alpha_2 = 0$ and so $x = (c_1, c_2)^T$, $b = (1, 1)^T$ and the error (9) reduces here to

$$e^2(c_1, c_2, c_{12}) = (c_1 + c_2 - 1)^2 \|P_{U_1+U_2}u\|^2 + (c_1 + c_2 + c_{12} - 1)^2 \|P_{V_1 \cap V_2}u\|^2.$$

It follows that in this case $e_c^2 = \|P_{U_1+U_2}u\|^2$ and $e_o^2 = 0$ and the minimal error is achieved when $c_1 + c_2 = 1$.

If $0 < \gamma$ then the matrix A is nonsingular and the minimum of $x^T A x - 2b^T x + 1$ is $1 - b^T A^{-1} b$ which is achieved for $x = A^{-1}b$. Substituting the values for A and b one then gets

$$e_o^2 = \left(1 - \frac{\cos^2\alpha_1 - 2\cos\alpha_1 \cos\alpha_2 \cos\gamma + \cos^2\alpha_2}{\sin^2\gamma}\right) \|P_{U_1+U_2}u\|^2$$

and it follows directly that $e_c^2 = \cos^2\gamma \|P_{U_1+U_2}u\|^2 + e_o^2 \sin^2\gamma$ from which one gets the claimed relation between e_c and e_o. Since $x = A^{-1}b$ the minimum is achieved for

$$c_1 = \frac{1 - \cos\gamma\cos\alpha_2/\cos\alpha_1}{\sin^2\gamma}, \quad c_2 = \frac{1 - \cos\gamma\cos\alpha_1/\cos\alpha_2}{\sin^2\gamma}.$$

□

The optimal choice of the combination coefficients can thus provide substantial improvements over the traditional choice for small angles γ. Note that one gets a zero error whenever the three projections $P_{U_1}u$, $P_{U_2}u$ and $P_{U_1+U_2}u$ are collinear and the commutation property, while sufficient to guarantee this, is not necessary. Observe as well that the difference between the two methods is small when $\gamma \approx \pi/2$.

2.3 Numerical Experiments

We now consider an application from machine learning where we look at the problem of reconstructing a function from some sample evaluations.

Starting from a data set $(\underline{x}, y_i)_{i=1}^M, \underline{x} \in [0,1]^d, y_i \in \mathbb{R}$ we assume that a function $\hat{f}$ describes the relationship between $\underline{x}$ and y, i.e. $\hat{f}(\underline{x}) \approx y$. The goal is now to reconstruct this function out of a function space V based on the given training data to allow predictions of $\hat{f}(\underline{z})$ on new data points $\underline{z}$. To achieve a well-posed problem we employ Tikhonov-regularisation which results in the following variational problem

$$R(f) \underset{f \in V}{\longrightarrow} \min !$$

with

$$R(f) = \frac{1}{M} \sum_{i=1}^{M} (f(\underline{x}_i) - y_i)^2 + \lambda \|\nabla f\|^2. \tag{10}$$

The first term enforces closeness of the function f to the data, the second term results in a certain smoothness of f, and the regularisation parameter λ balances these two terms. The solution f of this variational problem can be viewed as the projection of $\hat{f}$ into the space V, i.e. $f = P_V \hat{f}$.

We approximate this variational problem through the use of grids $\Omega_{\underline{l}}$ with mesh size $h_i := 2^{-l_i}$ in dimensions i and employ piecewise linear so called hat functions

$$\phi_{\underline{l},\underline{j}}(\underline{x}) \ := \ \prod_{t=1}^{d} \phi_{l_t,j_t}(x_t)$$

on each grid $\Omega_{\underline{l}}$, where the one-dimensional basis functions $\phi_{l,j}(x)$ are defined as the so-called hat functions

$$\phi_{l,j}(x) = \begin{cases} 1 - |\frac{x}{h_l} - j|, & x \in [(j-1)h_l, (j+1)h_l] \\ 0, & \text{otherwise.} \end{cases}$$

The angle between two spaces is now computed with (3)

$$c(V_1, V_2) = \|P_{V_1} P_{V_2} P_{(V_1 \cap V_2)^\perp}\| = \|P_{V_1} P_{V_2} - P_{V_1 \cap V_2}\| = \sup_g \frac{\|P_{V_1} P_{V_2} g - P_{V_1 \cap V_2} g\|}{\|P_{V_2} g\|}$$

We apply a Monte-Carlo-approach in the following way: for fixed data positions $\underline{x}_i$ we take random function values to describe a function g on these and compute the expression $\gamma_g := \frac{\|P_{V_1} P_{V_2} g - P_{V_1 \cap V_2} g\|}{\|P_{V_2} g\|}$. If we repeat this a large number of times the maximum over all γ_g gives an estimation for the cosine c of the angle between the spaces V_i.

For simplicity we only consider examples in two dimensions. We give the cosines of the angle between the two spaces (grids) $\Omega_{i,0}$ and $\Omega_{0,i}$ and the corresponding angle, furthermore also the mean over all γ_g with the corresponding angle to give an average type result.

We use four data points in four different configurations, a) on the corners of the hypercube $[0,1]^2$, b) on the corners of the hypercube $[0.25, 0.75]^2$, c) four random (but fixed) points in $[0,1]^2$, and d) the points $[0.2*i, 0.2*i], i = 1, \ldots, 4$. The measured results are given in Tables 1 to 4, respectively.

level	λ	$c(\Omega_{i,0}, \Omega_{0,i})$	acos(c)	mean(γ_g)	acos(mean)
1	1.0	$2.5305 \cdot 10^{-13}$	90	$6.1478 \cdot 10^{-15}$	90
1	0.01	$1.4213 \cdot 10^{-10}$	90	$1.3251 \cdot 10^{-14}$	90
1	0.0001	$7.8201 \cdot 10^{-10}$	90	$5.7113 \cdot 10^{-11}$	90
2	1.0	$3.9943 \cdot 10^{-3}$	89.7	$1.1038 \cdot 10^{-3}$	89.9
2	0.01	$3.8734 \cdot 10^{-5}$	89.998	$1.1930 \cdot 10^{-5}$	89.9993
2	0.0001	$4.0470 \cdot 10^{-8}$	90	$1.2498 \cdot 10^{-8}$	90
3	1.0	$5.9229 \cdot 10^{-3}$	89.661	$1.6467 \cdot 10^{-3}$	89.9057
3	0.01	$5.6652 \cdot 10^{-5}$	89.997	$1.7488 \cdot 10^{-5}$	89.999
3	0.0001	$5.9168 \cdot 10^{-8}$	90	$1.8310 \cdot 10^{-8}$	90
4	1.0	$6.4531 \cdot 10^{-3}$	89.630	$1.7857 \cdot 10^{-3}$	89.8977
4	0.01	$6.1508 \cdot 10^{-5}$	89.997	$1.8889 \cdot 10^{-5}$	89.9989
4	0.0001	$6.4229 \cdot 10^{-8}$	90	$1.9787 \cdot 10^{-8}$	90
5	1.0	$6.6729 \cdot 10^{-3}$	89.618	$1.8456 \cdot 10^{-3}$	89.8943
5	0.01	$6.3201 \cdot 10^{-5}$	89.996	$1.9483 \cdot 10^{-5}$	89.9989
5	0.0001	$6.5993 \cdot 10^{-8}$	90	$2.0399 \cdot 10^{-8}$	90

Table 1: Four data points on the corners of the hypercube $[0, 1]^2$.

In case a) the data points are grid points of the involved grids, so for $\lambda = 0$ the problem reduces to an interpolation problem. We see that the spaces $\Omega_{i,0}, \Omega_{0,i}$ are all close to orthogonal and with $\lambda \to 0$ the deviation gets smaller. Here only the regularisation operator ∇ causes the (small) non-orthogonality.

The situation is different for the points $[0.25, 0.75]^2$ in case b), the (still small) deviation gets larger with $\lambda \to 0$. Now the data points are the main cause of non-orthogonality, the regularisation with ∇ reduces this effect.

The random position of the points in case c) result in a significantly different situation. For small λ we get a deviation of almost 45° for level 4, while for $\lambda = 1$ the regularisation operator dominates and allows only a small non-orthogonality. Note that for the average case we still have a deviation of 20° starting with level

level	λ	$c(\Omega_{i,0}, \Omega_{0,i})$	acos(c)	mean(γ_g)	acos(mean)
1	1.0	$4.2393 \cdot 10^{-13}$	90	$1.4388 \cdot 10^{-14}$	90
1	0.0001	$4.9508 \cdot 10^{-10}$	90	$1.8927 \cdot 10^{-11}$	90
2	1.0	$6.5475 \cdot 10^{-3}$	89.625	$2.2535 \cdot 10^{-3}$	89.871
2	0.0001	$1.6473 \cdot 10^{-2}$	89.056	$5.8373 \cdot 10^{-3}$	89.666
3	1.0	$8.1239 \cdot 10^{-3}$	89.535	$2.7523 \cdot 10^{-3}$	89.842
3	0.0001	$1.7102 \cdot 10^{-2}$	89.020	$6.2032 \cdot 10^{-3}$	89.645
4	1.0	$8.3019 \cdot 10^{-3}$	89.524	$2.8296 \cdot 10^{-3}$	89.838
4	0.0001	$1.7227 \cdot 10^{-2}$	89.013	$6.2177 \cdot 10^{-3}$	89.644

Table 2: Four data points on the corners of the hypercube $[0.25, 0.75]^2$.

level	λ	$c(\Omega_{i,0}, \Omega_{0,i})$	acos(c)	mean(γ_g)	acos(mean)
1	1.0	$9.8611 \cdot 10^{-3}$	89.435	$2.0645 \cdot 10^{-3}$	89.882
1	0.0001	$1.5160 \cdot 10^{-1}$	81.280	$3.4440 \cdot 10^{-2}$	88.026
2	1.0	$1.2087 \cdot 10^{-2}$	89.308	$3.0671 \cdot 10^{-3}$	89.824
2	0.0001	$6.0629 \cdot 10^{-1}$	52.678	$1.6507 \cdot 10^{-1}$	80.499
3	1.0	$1.5288 \cdot 10^{-2}$	89.124	$3.8938 \cdot 10^{-3}$	89.777
3	0.0001	$6.7697 \cdot 10^{-1}$	47.392	$1.8681 \cdot 10^{-1}$	79.233
4	1.0	$1.5478 \cdot 10^{-2}$	89.113	$4.1309 \cdot 10^{-3}$	89.763
4	0.0001	$6.9539 \cdot 10^{-1}$	45.942	$1.8967 \cdot 10^{-1}$	79.067

Table 3: Four random but fixed data points in $[0, 1]^2$.

level	λ	$c(\Omega_{i,0}, \Omega_{0,i})$	acos(c)	mean(γ_g)	acos(mean)
1	1.0	$9.3628 \cdot 10^{-4}$	89.946	$2.4777 \cdot 10^{-4}$	89.986
1	0.0001	$6.5799 \cdot 10^{-1}$	48.854	$2.0203 \cdot 10^{-1}$	78.345
2	1.0	$4.6057 \cdot 10^{-3}$	89.736	$1.3093 \cdot 10^{-3}$	89.925
2	0.0001	$6.6614 \cdot 10^{-1}$	48.230	$2.0902 \cdot 10^{-1}$	77.935
3	1.0	$5.8815 \cdot 10^{-3}$	89.663	$1.8220 \cdot 10^{-3}$	89.896
3	0.0001	$6.8919 \cdot 10^{-1}$	46.434	$2.1725 \cdot 10^{-1}$	77.453
4	1.0	$7.9936 \cdot 10^{-3}$	89.542	$2.3429 \cdot 10^{-3}$	89.866
4	0.0001	$6.9138 \cdot 10^{-1}$	46.260	$2.1773 \cdot 10^{-1}$	77.424

Table 4: Four points $[0.2 * i, 0.2 * i], i = 1, \ldots, 4$.

2. The situation is similar for the points $[0.2 * i, 0.2 * i], i = 1, \ldots, 4$, but note that here already for level 1 we observe a deviation of more than 40° for small λ.

Now let us consider one particular additive function $u = e^{-x^2} + e^{-y^2}$, which we want to reconstruct based on 5000 random data samples in the domain $[0, 1]^2$. We use the combination technique and optimized combination technique for the grids $\Omega_{i,0}, \Omega_{0,i}, \Omega_{0,0}$. For $\lambda = 10^{-4}$ and $\lambda = 10^{-6}$ we show in Figure 2 the value of the functional (10), in Table 5 the corresponding numbers for the residuals and the angle $\gamma = \angle(P_{U_1} u, P_{U_2} u)$ are given. We see that both methods diverge for higher levels of the employed grids, nevertheless as expected the optimized combination technique is always better than the normal one.

da muss $cos(\gamma)$ statt γ stehen)

We also show in Figure 2 the results for an optimized combination technique which involves all intermediate grids, i.e. $\Omega_{j,0}, \Omega_{0,j}$ for $1 \leq j < i$, as well. The generalisation of the computation of the coefficients is straightforward and is described in section 3.4. Here we do not observe rising values of the functional for higher levels but a saturation, i.e. higher refinement levels do not substantially change the value of the functional. This effect for grid based approaches for function reconstruction was already observed in [10] and is due to the fact that

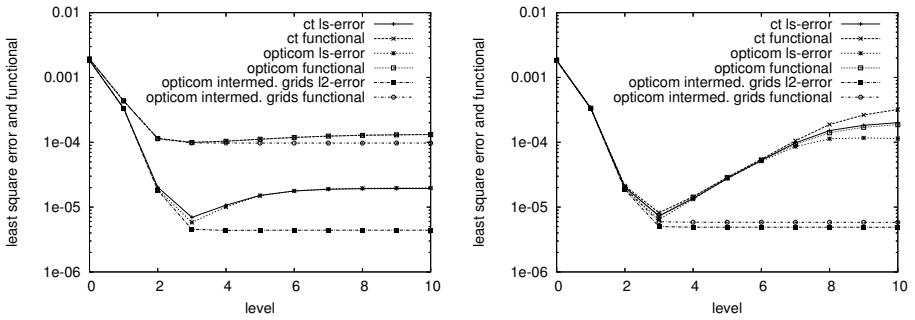


Figure 2: Value of the functional (10) and the least squares error on the data, i.e. $\frac{1}{M}\sum_{i=1}^{M}(f(\underline{x}_i)-y_i)^2$, for the reconstruction of $e^{-x^2}+e^{-y^2}$ for the combination technique and the optimised combination technique for the grids $\Omega_{i,0}, \Omega_{0,i}, \Omega_{0,0}$ and the optimised combination technique for the grids $\Omega_{j,0}, \Omega_{0,j}, 0 \leq j \leq i$ with $\lambda = 10^{-4}$ (left) and 10^{-6} (right).

level	γ	e_c^2	e_o^2
1	-0.012924	$3.353704 \cdot 10^{-4}$	$3.351200 \cdot 10^{-4}$
2	-0.025850	$2.124744 \cdot 10^{-5}$	$2.003528 \cdot 10^{-5}$
3	-0.021397	$8.209228 \cdot 10^{-6}$	$7.372946 \cdot 10^{-6}$
4	-0.012931	$1.451818 \cdot 10^{-5}$	$1.421387 \cdot 10^{-5}$
5	0.003840	$2.873697 \cdot 10^{-5}$	$2.871036 \cdot 10^{-5}$
6	0.032299	$5.479755 \cdot 10^{-5}$	$5.293952 \cdot 10^{-5}$
7	0.086570	$1.058926 \cdot 10^{-4}$	$9.284347 \cdot 10^{-5}$
8	0.168148	$1.882191 \cdot 10^{-4}$	$1.403320 \cdot 10^{-4}$
9	0.237710	$2.646455 \cdot 10^{-4}$	$1.706549 \cdot 10^{-4}$
10	0.285065	$3.209026 \cdot 10^{-4}$	$1.870678 \cdot 10^{-4}$

Table 5: Residual for the normal combination technique e_c^2 and the optimized combination technique, as well as the angle $\gamma = \angle(P_{U_1}u, P_{U_2}u)$.

after a certain discretisation level the error term on the data points cannot be further reduced by grids with a finer discretisation.

3 Multiple spaces

In this section we look at the combination of more than two spaces. In the first subsection, the most general case of spaces which form an intersection structure will be considered. It will be seen that if the set of spaces form a distributive lattice, then additive measures satisfy the inclusion/exclusion principle which is equivalent to a general combination formula and a first characterisation of the (classical) combination coefficients by a linear system of equations is provided. If the projections onto the subspaces commute they are an additive (operator valued) function of the hierarchical basis and thus the inclusion/exclusion principle holds. In the second subsection, a recursion for the combination coefficients is derived which is based on the Hasse diagram of the lattice of spaces.

For noncommuting operators the mapping which maps the sets of basis vectors of the subspaces onto the projection into that subspace is not additive. The the inclusion/exclusion principle cannot be applied here. However, even in this case the combination technique frequently provides a good approximation, this is in particular known for the finite element solution of the Poisson problem. On the other hand, the approximation can be poor for machine learning applications. Thus two alternatives for the combination coefficients are considered here. First, we present a method which determines an average approximation using the Frobenius norm of the error of the combination of projections and second, we consider the "opticom" method in which the "best possible" linear combination of projections onto the component spaces is selected. This results in a nonlinear approximation as both the projections onto the component spaces and the combination coefficients depend on the data. In a final subsection the error of a general combination technique is discussed and compared with the "opticom" method.

3.1 The combination formula in the context of partially ordered sets

Consider any finite collection of closed linear subspaces $V_1, \ldots, V_m$ of a Hilbert space H. This collection is a *partially ordered set* with respect to the subspace relation $\subset$. The collection forms an *intersection structure* or *closure system* if for every two V_i, V_j there is a V_k such that $V_k = V_i \cap V_j$. If, as usual, one introduces the sum of two spaces V_i and V_j to be

$$V_i + V_j = \{v_i + v_j \mid v_i \in V_i, v_j \in V_j\}$$

one can define the lattice of subspaces generated by the V_i which contains all V_i but also any sums and intersections of sums. In addition, include (if necessary) the space $V_\infty = \{0\}$ as the "zero" element of the lattice and $V = \sum_{j=1}^{m} V_j$ as the

"one". This lattice is *distributive* if for any three spaces U, V, W in this lattice one has the *distributive law*

$$U \cap (V + W) = U \cap V + U \cap W.$$

Note that not every lattice of spaces defined by $\cap$ and $+$ is distributive. Consider, in particular, the case of the one dimensional spaces U, V, W generated by the basis vectors $(1, 1), (0, 1)$ and $(1, 0)$, respectively. Note that in this case one has $U \cap (V + W) = U$ but $U \cap V + U \cap W = \{0\}$.

The configurations considered in the following are based on tensor products and will all lead to distributive lattices of subspaces. It is known that every distributive lattice is order isomorphic to a lattice of sets, see [6, Theorem 10.21]. The importance of this isomorphy is that for sets the *inclusion/exclusion principle* holds, see, e.g. [1.3.3,p.179][24]. In general this principle takes the form: For any intersection structure of sets $A_1, \ldots, A_m$ and any (additive) measure μ on these sets one has

$$\mu(A_k) = \sum_{i=1}^{m} c_i \mu(A_{\varphi(i,k)}),$$

where for $\varphi(i, k)$ holds $A_{\varphi(i,k)} = A_i \cap A_k$. As the mapping of the lattice of spaces V_i to the lattice of sets A_i is order preserving one has

$$V_i \cap V_k = V_{\varphi(i,k)}.$$

From the following discussion it can be seen that the c_i are the *combination coefficients* and thus this provides a linear system of equations relating the combination coefficients (which do not depend on the actual measure nor the specific sets A_k but just on the intersection structure). As the dimensions form a measure on the sets A_k one observes in particular the relation

$$\dim(V_k) = \sum_{i=1}^{m} c_i \dim(V_k \cap V_i).$$

Adding $V = \sum_{i=1}^{m} V_m$ to the collection of subspaces one still has an intersection structure and it follows that

$$\dim(V) = \sum_{i=1}^{m} c_i \dim(V_i).$$

Note that the only properties required for this to hold is that the subspaces form an intersection structure and the lattice generated by these subspaces is distributive. As $\dim(V_i) = \|P_{V_i}\|_F^2$ for the Frobenius norm $\|\cdot\|_F$ of the orthogonal projections P_{V_i} it follows directly that

$$\|P_{V_k}\|_F^2 = \sum_{i=1}^{m} c_i \|P_{V_{\varphi(i,k)}}\|_F^2$$

and

$$\|P_V\|_F^2 = \sum_{i=1}^{m} c_i \|P_{V_i}\|_F^2.$$

The simplest case of an intersection structure is totally ordered or a *chain*, i.e. for any $i, j = 1, \ldots, m$ one has either $V_i \subset V_j$ or $V_j \subset V_i$. We assume now that the V_i are numbered such that $V_i \subset V_{i+1}$. We introduce difference spaces as

$$W_i = V_{i+1} \ominus V_i := V_{i+1} \cap V_i^{\perp}$$

and observe the orthogonal decomposition

$$V_k = V_1 \oplus W_1 \oplus \cdots \oplus W_{k-1}, \quad k = 2, \ldots, m.$$

Let us introduce a basis $e_1, \ldots, e_N$ of V_m such that $e_1, \ldots, e_{N_k}$ form a basis for V_k for any $k = 1, \ldots, m$ or, equivalently, such that $e_{N_{k-1}+1}, \ldots, e_{N_k}$ forms a basis of W_k for $k = 1, \ldots, m-1$. Such a basis is called a *hierarchical basis* [27]. In this case the mapping of V_i onto the set of generating basis vectors e_i provides an isomorphy between the lattice V_i and the set of sets A_i.

The most general case considered here is the setting where the spaces V_i are tensor products of spaces which themselves form chains. More specifically, let $H = H^1 \otimes \cdots \otimes H^d$ be a tensor product Hilbert space and for every i let $V_1^i, \ldots, V_{m_i}^i$ be a chain of subspaces of H^i. Now let the collection of subspaces $V_1, \ldots, V_m$ be such that

$$V_k = V_{i_{1,k}}^1 \otimes \cdots \otimes V_{i_{d,k}}^d.$$

Assume that each V_j^i occurs as a tensor product factor in at least one of the spaces V_k. For each $i = 1, \ldots, d$ let $e_1^i, \ldots, e_{m_i}^i$ be a hierarchical basis of the $V_1^i, \ldots, V_{m_i}^i$. Then for each V_k one can form a basis which consists of tensor products

$$e_{i_1,\ldots,i_d} = e_{i_1}^1 \otimes \cdots \otimes e_{i_d}^d$$

and so each V_k is uniquely defined by the set A_k of basis vectors $e_{i_1,\ldots,i_d}$. From this it can be shown that the lattice generated by the sets V_i is distributive. The mapping which associates V_k with A_k provides an alternative isomorphism between the lattice of subspaces and sets. Furthermore any additive functions on the basis vectors can be used for the combination formulas above.

In the special situation where the basis vectors $e_{i_1,\ldots,i_d}$ are orthogonal any squared norm $\|P_{V_i} u\|^2$ is additive and one has for this case the formuli

$$\|P_{V_i} u\|^2 = \sum_{i=1}^{n} c_i \|P_{V_i \cap V_k} u\|^2, \quad i = 1, \ldots, m$$

as well as

$$\|P_V u\|^2 = \sum_{i=1}^{n} c_i \|P_{V_i} u\|^2.$$

The above relations for c_i form a linear system of equations which the c_i need to satisfy and they can be used to compute the c_i. As the sets or spaces form an intersection structure the m by m matrix occurring only depends on m parameters and thus is a structured matrix which is generated by the measures of the intersections on an intersection structure. It follows that the c_i satisfy the equations

$$\begin{bmatrix} n_1 & n_{\varphi(1,2)} & \cdots & n_{\varphi(1,m)} \\ n_{\varphi(1,2)} & n_2 & \cdots & n_{\varphi(2,m)} \\ \vdots & \vdots & \ddots & \vdots \\ n_{\varphi(1,m)} & n_{\varphi(2,m)} & \cdots & n_m \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix} = \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_m \end{bmatrix}. \tag{11}$$

Finally, we note that the classical combination formula can be viewed as an application of the inclusion/exclusion principle applied to the operator valued mapping defined on a lattice of linear spaces which maps the space V_i onto the orthogonal projection $P_{V_i} : H \to V_i$. This mapping is additive if the operators P_{V_i} commute.

3.2 Construction of the classical combination coefficients

Consider the case of pairwise commuting projectors, i.e, the case where

$$P_{V_i} P_{V_j} = P_{V_j} P_{V_i} = P_{V_{\varphi(i,j)}}, \quad \text{for all } i, j = 1, \ldots, m.$$

We assume in the following that the numbering of the spaces V_i is consistent with the partial order, i.e. that if $V_i \subset V_j$ then $i \geq j$. As the spaces form a lattice it follows that

$$\prod_{j=1}^{m} P_{V_j} = P_{\bigcap_{i=1}^{m} V_i} = P_{V_m}$$

as $\bigcap_{i=1}^{m} V_i = V_m$.

The partial order on the spaces V_i defines a partial order on the projections P_{V_i} and we say that $P_{V_i} < P_{V_j}$ if and only if $V_i \subset V_j$. As usual, the Hasse diagram of the partial order is the directed graph where the vertices are the P_{V_i} and there is an edge between P_{V_i} and P_{V_j} if $P_{V_i} < P_{V_j}$ and there is no P_{V_k} such that $P_{V_i} < P_{V_k} < P_{V_j}$. Note that the Hasse diagram has no cycles. We now introduce the *level* $\lambda(i)$ of a projection P_{V_i} in the Hasse diagram by setting $\lambda(m) = 0$ and $\lambda(i) = s$ if the shortest chain $P_{V_{i_0=m}} < P_{V_{i_1}} < \ldots < P_{V_{i_{s-1}}} < P_{V_{j=i_s}}$ is such that there are no P_{V_k} with $P_{V_{i_t}} < P_k < P_{V_{i_{t+1}}}$. In other words, the level is the distance in the Hasse diagram from the joint space V_m to the space V_i. It follows that if $V_i \subset V_j$ one has $\lambda(i) \leq \lambda(j)$.

Now introduce the level spaces V^k by

$$V^k = \sum_{\lambda(i) \leq k} V_i.$$

It follows that $V^k = \sum_{\lambda(i)=k} V_i$, and, in particular, one has $V^{\lambda(1)} = \sum_{i=1}^{m} V_i$ and $V^0 = V_m$. The orthogonal projections onto the V^k shall be denoted by P^k.

One can now show (see [19]) that

$$P^k = \sum_{\lambda(j)=k} P_{V_j}(I - P^{k-1}) + P^{k-1} \tag{12}$$

as all the $P_{V_j}(I-P^{k-1})$ and P^{k-1} are pairwise orthogonal. In the same reference it is shown that a combination formula exists, i.e. that for any k one has

$$P^k = \sum_{\lambda(i)\leq l} c_i^k P_{V_i} \tag{13}$$

for some c_i^k. Inserting this into (12) gives

$$P^k = \sum_{\lambda(i)=k} P_{V_i} + \sum_{\lambda(i)\leq k-1} c_i^{k-1} P_{V_i} - \sum_{\lambda(i)=k} P_{V_i} \sum_{\lambda(j)\leq k-1} c_j^{k-1} P_{V_j}$$

from which one gets

$$P^k = \sum_{\lambda(i)=k} P_{V_i} + \sum_{\lambda(i)\leq k-1} c_i^{k-1} P_{V_i} - \sum_{\lambda(i)=k} \sum_{\lambda(j)\leq k-1} c_j^{k-1} P_{V_{\varphi(j,k)}}.$$

Comparing this with (13) gives a recursion for the combination coefficients. In particular, one has $c_1^0 = 1$,

$$c_i^l = 1 \quad \text{for } \lambda(i) = l$$

and

$$c_i^l = c_i^{l-1} - \sum_{\lambda(k)\leq l-1} c_k^{l-1} \cdot \#\{j|\varphi(k,j) = i, \lambda(j) = l\} \quad \text{else.}$$

These recursions for a generalised combination technique can be found in [19]. The generalised combination technique has been mentioned in [14] and proven in [22, 23] with a different approach. With these recursions one can determine the combination coefficients for fairly general situations where the projections commute. The coefficients are uniquely determined by the Hasse diagram and do not depend on the dimensions of the particular spaces.

3.3 Combination coefficients for the best approximation relative to the Frobenius norm

While the combination coefficients in the previous section are all which is needed for the case of commuting projections one needs a different approach for non-commuting projections. Fundamentally, one would like to approximate an orthogonal projection P onto the space $V = V_1 + \cdots + V_m$ by a linear combination of the orthogonal projections onto the component spaces V_i. Consider here the approximation which minimises the error of the operator approximation in the

Frobenius norm. This corresponds to a best average case approximation and the combination coefficients c_i are obtained by minimising

$$J(c_1, \ldots, c_m) = \|P - \sum_{i=1}^{m} c_i P_{V_i}\|_F^2.$$

Inserting the definition $\|M\|_F^2 = \mathrm{tr}(M^T M)$ one gets

$$J(c_1, \ldots, c_m) = \sum_{i,j=1}^{m} c_i c_j \Gamma_{ij} - 2\sum_{i=1}^{m} c_i n_i + n,$$

where n is the dimension of V, $n_i = \mathrm{tr}(P_{V_i})$ are the dimensions of the spaces V_i, and $\Gamma_{ij} = \mathrm{tr}(P_{V_i} P_{V_j})$, in particular $\Gamma_{ii} = n_i$. The normal equations for this optimisation problem are then

$$\begin{bmatrix} n_1 & \Gamma_{12} & \cdots & \Gamma_{1m} \\ \Gamma_{21} & n_2 & \cdots & \Gamma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{m1} & \Gamma_{m2} & \cdots & n_m \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix} = \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_m \end{bmatrix} \tag{14}$$

One now gets the following bounds (using the Schwarz inequality for the matrix scalar product $\mathrm{tr}(A^T B)$):

$$\dim(V_i \cap V_j) = \mathrm{tr}(P_{V_i \cap V_j}) \leq \Gamma_{ij} \leq \min(n_i, n_j) \leq \sqrt{n_i n_j}.$$

These bounds provide estimates for the coefficients c_i, which, in general, would be hard to compute in practice as the Γ_{ij} are typically unknown. For the commuting case, the Γ_{ij} are all either 0 or $\mathrm{tr}(P_{\varphi(i,j)}) = n_{\varphi(i,j)}$ and the classical combination coefficients have to satisfy the normal equations. While this approach does require the determination of the Γ_{ij} – which can be a substantial computational problem – it is linear and the same coefficients can be used for any data. However, like for the classical combination coefficients, this method may lead to very poor approximations as well as has been demonstrated for the case of two spaces.

3.4 The opticom choice

Instead of using a method to provide a best approximant on average, we now attempt to find the best approximant for the given data. More specifically, the functional

$$J(c_1, \ldots, c_m) = \|Pf - \sum_{i=1}^{m} c_i P_{V_i} f\|^2$$

is minimised here. By simple expansion one gets

$$J(c_1, \ldots, c_m) = \sum_{i,j=1}^{m} c_i c_j (P_{V_i} f, P_{V_j} f) - 2\sum_{i=1}^{m} c_i \|P_{V_i} f\|^2 + \|Pf\|^2.$$

While this functional depends on the quantity Pf to be approximated, the location of the minimum of J does not. The best combination coefficients then satisfy

$$\begin{bmatrix} \|P_{V_1} f\|^2 & (P_{V_1} f, P_{V_2} f) & \cdots & (P_{V_1} f, P_m f) \\ (P_{V_2} f, P_{V_1} f) & \|P_{V_2} f\|^2 & \cdots & (P_{V_2} f, P_m f) \\ \vdots & \vdots & \ddots & \vdots \\ (P_m f, P_{V_1} f) & (P_m f, P_{V_2} f) & \cdots & \|P_m f\|^2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix} = \begin{bmatrix} \|P_{V_1} f\|^2 \\ \|P_{V_2} f\|^2 \\ \vdots \\ \|P_m f\|^2 \end{bmatrix}.$$

Thus this matrix has the typical structure of a normal equation matrix. The determination of the best combination coefficients by solving such a system actually creates little overhead. Nevertheless, in general a large increase in computational complexity is due to the need for the determination of the scalar products $(P_{V_i} f, P_{V_j} f)$ and the norms $\|P_{V_i} f\|^2$. The computation of the scalar products in particular is often difficult as it requires embedding the problem in a space which contains both V_i and V_j. For details in the case of an application in machine learning see [13].

3.4.1 Numerical Experiments

We again consider the machine learning application from section 2.3. Instead of employing only two grids and their intersection we use all grids which normally arise for level n of the sparse grid combination technique [12, 18], i.e. all grids $\Omega_{\underline{l}}$ with

$$|\underline{l}|_1 := l_1 + ... + l_d = n - q, \quad q = 0, .., d-1, \quad l_t \geq 0. \tag{15}$$

Note that the formula for the original combination technique is

$$f_n^c(\underline{x}) := \sum_{q=0}^{d-1} (-1)^q \binom{d-1}{q} \sum_{|\underline{l}|_1 = n-q} f_{\underline{l}}(\underline{x}).$$

In Figure 3 we give results using both the original combination technique and the optimal one for the two-dimensional data already considered in section 2.3. We show both the residual (10) and the least squares error. Again the ordinary combination technique diverges after level 3, whereas the residual for the optimal always declines, although only small amounts after level 6 or so. Note that for $\lambda = 10^{-6}$ and level 10 the residual for the optimised combination technique using the grids after (15) is $9.27 \cdot 10^{-7}$ as opposed to $5.83 \cdot 10^{-6}$ which we observed in section 2.3 for the optimised combination technique involving the grids $\Omega_{j,0}, \Omega_{0,j}, 0 \leq j \leq n$. We also observe that now the least squares error part of the functional (10) is still decreasing for higher levels, which is not the case for the grids considered in section 2.3.

3.5 Error bounds for the combination technique

In this subsection the errors of combination techniques are analysed in terms of geometric quantities, in particular, in terms of angles between spaces and

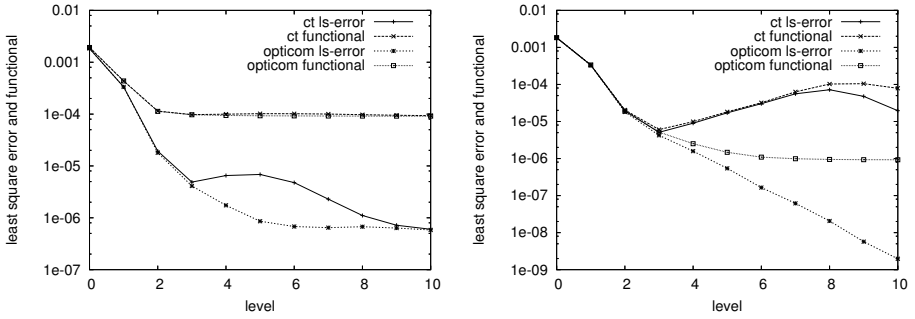


Figure 3: Value of the functional (10) and the least squares error on the data, i.e. $\frac{1}{M}\sum_{i=1}^{M}(f(\underline{x}_i)-y_i)^2$, for the reconstruction of $e^{-x^2}+e^{-y^2}$ for the combination technique and the optimised combination technique with $\lambda = 10^{-4}$ (left) and 10^{-6} (right) and level $n = 0,\ldots,10$.

between vectors. In a first part a bound for the worst case error of the classical combination technique is obtained based on results in [8]. This bound uses an error bound for the multiplicative combination technique. In a second part, error bounds for a general combination technique are obtained. In particular the best possible error which is achieved by the opticom method is determined and a formula for the difference between the optimal error and an error of a general combination technique is provided.

Let $U_1,\ldots,U_m$ be closed subspaces of the Hilbert space H and P_{U_i} be the orthogonal projections onto U_i. Furthermore, let $U := \bigcap_{i=1}^{m} U_i$ and set

$$E(U_1,\ldots,U_m) = \|P_{U_m}\cdots P_{U_1} - P_U\|.$$

Recall the definition of the cosine $c(V_1,V_2)$ of the angle between two spaces V_1 and V_2 from section 2.1. It follows from theorem 2.7 in [8] that

$$E(U_1,\ldots,U_m) \le \alpha(U_1,\ldots,U_m)$$

where

$$\alpha(U_1,\ldots,U_m) = \sqrt{1-\prod_{i=1}^{m-1}\left(1-c(U_i,\bigcap_{j=i+1}^{m}U_j)\right)}.$$

This provides a characterisation of E in terms of the geometric quantities $c(V_1,V_2)$. In the special case where $U_i\cap(U_i\cap U_j)^\perp$ is orthogonal to $U_j\cap(U_i\cap U_j)^\perp$ for every pair U_i,U_j one can see that $\alpha(U_1,\ldots,U_m)=0$ and $\alpha(U_1^\perp,\ldots,U_m^\perp)=0$.

Let, as before, $V_1,\ldots,V_n$ be closed subspaces of H and let $V=\sum_{i=1}^{n}V_i$. In deriving a bound for the worst case error of the (classical) combination technique one uses the *multiplicative combination approximation* which is defined as:

$$P_V \approx T^{\mathrm{mc}} = \sum_{k=1}^{n}\sum_{1\le i_1<\cdots i_k\le n}(-1)^k P_{V_{i_k}}\cdots P_{V_{i_1}}.$$

The multiplicative approximation satisfies

$$I - T^{\mathrm{mc}} = (I - P_{V_n}) \cdots (I - P_{V_1})$$

which allows the application of the above bound for the product of projections.

Proposition 3. *Let T^{mc} be the multiplicative combination approximation for $V = V_1 + \cdots + V_n$. Then one has*

$$\|T^{\mathrm{mc}} - P_V\| \leq \alpha(V_1^{\perp}, \ldots, V_n^{\perp}).$$

Proof. Let $U_i = V_i^{\perp}$ and $U = \bigcap_{i=1}^{n} U_i$. Then $P_V = I - P_U$ and one has

$$P_V - T^{\mathrm{mc}} = P_{U_n} \cdots P_{U_1} - P_U$$

and by the theorem of Deutsch and Hundal [8] one gets

$$\|T^{\mathrm{mc}} - P_V\| \leq \alpha(U_1, \ldots, U_n)$$

which is the claimed bound. □

The *combination approximation* can be written as

$$P_V \approx T^c = \sum_{k=1}^{n} \sum_{1 \leq i_1 < \cdots i_k \leq n} (-1)^k P_{V_{i_1} \cap \cdots \cap V_{i_k}}.$$

An application of the previous bound for the multiplicative combination technique yields:

Proposition 4. *Let T^c be the combination approximation for $V = V_1 + \cdots + V_n$. Then one has*

$$\|T^{\mathrm{c}} - P_V\| \leq \alpha(V_1^{\perp}, \ldots, V_n^{\perp}) + \sum_{k=2}^{n} \sum_{1 \leq i_1 < \cdots i_k \leq n} \alpha(V_{i_1}, \ldots, V_{i_k}).$$

Proof. First one observes that the combination approximation is obtained by approximating the products of projections onto the spaces V_i in the terms of T^{mc} by the projections onto the intersections of the spaces. Then one invokes the triangular inequality and the bound from the theorem 2.7 in [8] to get the desired result. Note that in the last bound one only needs to sum from $k = 2$. □

This concludes our discussion of worst case error bounds for the classical combination technique.

Consider now an arbitrary combination approximation defined by

$$T^a u = \sum_{i=1}^{n} c_i P_{V_i} u$$

of the projection of an arbitrary $u \in V$ into the space $V_1 + \cdots + V_m$, where $c = (c_1, \ldots, c_n)$ are the combination coefficients. The squared Euclidean norm of the error of this approximation is

$$e(c)^2 = \|P_{\sum V_i} u - \sum_{i=1}^{n} c_i P_{V_i} u\|^2.$$

By simple expansion one has

$$e(c)^2 = \sum_{i,j} c_i c_j (P_{V_i} u, P_{V_j} u) - 2 \sum_i c_i \|P_{V_i} u\|^2 + \|P_{\sum V_i} u\|^2.$$

This is a quadratic function of the coefficients and the smallest value of this function is obtained by standard means.

Let $\gamma_{i,j}$ be the angle between the vectors $P_{V_i} u$ and $P_{V_j} u$ and let α_i be the angle between $P_{V_i} u$ and $P_{\sum_j V_j} u$. It follows that

$$(P_{V_i} u, P_{V_j} u) = \cos \gamma_{ij} \|P_{V_i} u\| \|P_{V_j} u\|$$

and

$$\|P_{V_i} u\| = \cos \alpha_i \|P_{\sum V_j} u\|.$$

In terms of these geometric quantities the error is

$$e(c)^2 = \left(\sum_{i,j} c_i c_j \cos \gamma_{ij} \cos \alpha_i \cos \alpha_j - 2 \sum_i c_i \cos^2 \alpha_i + 1 \right) \|P_{\sum V_i} u\|^2.$$

A further simplification of this expression is obtained with the introduction of the vector x with components $x_i = c_i \cos \alpha_i$, the vector b with components $b_i = \cos \alpha_i$ and the matrix A with elements $a_{ij} = \cos \gamma_{ij}$ for $i \neq j$ and $a_{ii} = 1$. With this we get

$$e(c)^2 = \left(x^T A x - 2 b^T x + 1 \right) \|P_{\sum_i V_i} u\|^2. \tag{16}$$

Using some standard linear algebra it follows

Proposition 5. *Let A, x and b be defined as above and let A be invertible. The minimum of the squared norm of the combination error is achieved for $x = A^{-1} b$ and equals*

$$\min_c e(c)^2 = \left(1 - b^T A^{-1} b \right) \|P_{\sum V_i} u\|^2.$$

Furthermore, one has

$$e(c)^2 - \min_{c'} e(c')^2 = (Ax - b)^T A^{-1} (Ax - b) \|P_{\sum_i V_i} u\|^2.$$

Of course the case $x = A^{-1} b$ is nothing else than the opticom method discussed previously.

The errors can be reformulated in terms of the combination coefficients instead of x. If M is the matrix with elements $m_{ij} = \cos\alpha_i \cos\alpha_j \cos\gamma_{ij}$ if $i \neq j$ and $m_{ii} = \cos^2\alpha_i$ one has for the optimal combination coefficients c_{opt}:

$$e(c_{\text{opt}})^2 = \left(1 - c_{\text{opt}}^T M c_{\text{opt}}\right) \|P_{\sum V_i} u\|^2$$

and

$$e(c)^2 - e(c_{\text{opt}})^2 = (c - c_{\text{opt}})^T M (c - c_{\text{opt}}) \|P_{\sum V_i} u\|^2.$$

We end this section with a discussion of the qualitative types of precision which can be achieved with combination approximations. One can now distinguish four sets $S_1 \subset S_2 \subset S_3 \subset S_4$ of vectors defined by the performance of the combination technique with coefficients c_i. We say $u \in S_1$ if the linear combination of projections onto the partial spaces is equal to the projection onto the sum of the spaces, i.e. $T_u^c = P_{\sum V_i}$. If S_1 is the full space the corresponding combination technique is exact. If the projection operators onto the partial spaces V_i commute, the set S_1 is the full space for the classical combination technique, which equals the opticom method in this case.

If we assume that the spaces V_i are one-dimensional, then for the opticom method the set S_1 contains all vectors u which are not orthogonal to any of the V_i [21]. We will show an example later where u is orthogonal to one V_i and for which the opticom method is not exact.

Consider now any method for which the sum of the combination coefficients $c^T e = 1$. In this case one has $V_m \subset S_1$ where $V_m = V_1 \cap \ldots \cap V_m$ is the smallest subspace. One can see that this holds for the classical combination technique. As an exact approximation is always optimal this condition has to hold for the opticom coefficients and from this one gets a condition which links the α_i and the γ_{ij}:

$$e^T D^{-1} A^{-1} D e = 1,$$

where e is a vector with all components one and D is a diagonal matrix with entries $\cos\alpha_i$ such that $b = De$, and $x = Dc$.

The second set S_2 in the above collection is the set of elements for which the combination technique produces a best possible combination approximation. This is by design always achieved by the opticom method and so S_2 is the full space for the opticom method. However, other methods may also produce a best possible approximation for some vectors u. The following example illustrates this case. Assume that u is such that all the projections $P_{V_i} u$ are pairwise orthogonal and so $A = I$. From the above one has $x = b$ and so all $c_i = 1$. This can occur, e.g. when all the V_i are pairwise orthogonal but in other cases as well. If one has $V_i \cap V_j = 0$ for $i \neq j$ the combination coefficients are all one for the classical combination technique. While in general for non orthogonal V_i one may not get optimal results the combination approximation is optimal if the $P_{V_i} u$ are pairwise orthogonal. However, this does not necessarily mean that in this case the combination approximation is also exact. A simple example of this is where V_1 is spanned by $(1, 0)$ and where V_2 is spanned by $(1, 1)$ and where $u = (1, -1)$. As u is orthogonal on V_2, one has $A = I$ and

the optimal combination approximation is the sum of the projections onto the subspaces which is $(1, 0)$. As u is orthogonal to the second space this is the best combination approximation but the error is not zero, although $u \in V_1 + V_2$. One can easily construct similar examples for more spaces which shows that there are optimal reconstructions, in the sense of Proposition 5, which are obtained from the classical combination technique but which are not exact.

The set S_3 additionally contains vectors for which the combination technique provides an approximation which is close to optimal. A particular case where this might occur is when there is a basis which is close to orthogonal and spans the spaces V_i. In this case, there is a matrix A_0 and a vector b_0 such that $A_0^{-1}b_0$ results in the classical combination technique and where $A = A_0 + \epsilon A_1$ and $b = b_0 + \epsilon b_1$. The difference with respect to the opticom is in this case

$$e(c)^2 - e(c_{\text{opt}})^2 = (A_0^{-1}b_0 - A^{-1}b)^T A (A_0^{-1}b_0 - A^{-1}b).$$

It follows that

$$e(c)^2 - e(c_{\text{opt}})^2 = (AA_0^{-1}b_0 - b)^T A^{-1} (AA_0^{-1}b_0 - b)$$

and this is

$$e(c)^2 - e(c_{\text{opt}})^2 = \epsilon^2 (A_1 A_0^{-1}b_0 - b_1)^T A^{-1} (A_1 A_0^{-1}b_0 - b_1)^T.$$

Thus if the angles between the spaces are all $\pi/2 + O(\epsilon)$ the classical combination technique will provide an approximation of the order of $O(\epsilon)$ to the optimal one. As long as this error is acceptable the set S_3 can be chosen to be the full space.

The last set S_4 furthermore includes cases where the combination technique "fails" or, more concisely, is far from optimal. Consider the example of two spaces where the first space V_1 is spanned by $(1, 0)^T$ and the second space V_2 is spanned by $(a, 1)^T$. For any u which is not orthogonal to any V_i the opticom method is now exact. The classical combination approximation has $c_1 = c_2 = 1$. An elementary calculation shows that the norm of the error is $\sqrt{u_1^2 + u_2^2}$ asymptotically for $a \to \infty$. in a and so the set S_4 would contain all u with a large u_2. It follows that the classical combination technique can produce results which are far from optimal and this was confirmed by the experiments discussed earlier and in the previous section.

4 Conclusions

The use of finite element approaches for problems with functions of more than about four variables has been observed as computationally infeasible due to the curse of dimensionality. The sparse grid approximation has introduced a way to overcome this curse and has allowed the solution of problems with up to around ten independent variables. One cost incurred is that for many applications the matrices involved become fairly dense and have a less simple structure than matrices for standard finite element methods. This is a consequence of the now

necessary use of a hierarchical basis instead of the local basis functions used otherwise.

The combination technique has made a substantial impact as it did allow to approximate the sparse grid solution by a linear combination of the solutions on regular subgrids which generate the sparse grid. Thus for the solution only problems on regular grids need to be treated and the sparse grid approximation is obtained by a simple linear combination. For many problems, in particular the solution of partial differential equations, this approach was demonstrated to be highly effective. The error occurred by this combination approximation has been seen to be of the same order as the original sparse grid approximation error. However, in the PhD thesis of one of the authors it was found that some data mining applications show substantially larger errors of the combination technique. It is thought that this is due to correlations of the predictor variables.

In this work we study an approach which modifies the original combination technique so that the combination coefficients are chosen adaptively. An "optimal combination technique" is obtained which we suggest to call the "opticom" method. It is shown that this technique can substantially improve approximations for sparse grid fitting problems. However, one can see that (in some rare cases) the combination approach itself, even with flexible coefficients, has its limitations. One of the authors has thus suggested in the past to combine this approach with iteration to get a method which generalises Krylov space iterations. In the future we plan to study this method further, and, in particular, to develop a convergence theory. In the earlier work it has been suggested that this iterative method converges almost as fast as multiplicative Schwartz methods while maintaining the parallelism of additive methods. While there is a close connection between sparse grid and multigrid algorithms which has been explored by other authors, the combination technique thus provides a connection to domain decomposition techniques.

Traditional sparse grid approximations push the limit of treatable problems from up to four dimensional problems to over ten dimensional ones. Nevertheless the question remains how to deal with problems of hundreds or even thousands of variables which occur in machine learning and other fields like, e.g. biology. The suggested opticom method fits nicely with dimension adaptive approaches which are thought to be able to address this curse. However, there are severe computational issues to be resolved for such approaches, this is a field of active investigation.

In the work presented in this paper some of the foundations have been laid for the development of the theory and implementation of such very high-dimensional schemes which share some commonalities with ANOVA decompositions and additive models used in statistics. We hope that the numerical analysis of the combination methods will also benefit the work in statistics in these areas. At the heart of both techniques is the tensor product structure of many of these problems. The methods can be viewed as an effective compression scheme for tensor product spaces. The effectiveness of such a scheme does rely on smoothness properties which are higher than for ordinary approximation techniques. The consequences of these requirements have still to be fully understood. Another

aspect are the concentration phenomena which occur for many highdimensional measures. One of the authors has been studying how this might affect approximation and we believe that concentration will be an important component in highdimensional approximation theory and thus in the theory for approximation by sparse grid combination techniques.

References

[1] Jerzy K. Baksalary and Oskar Maria Baksalary. Idempotency of linear combinations of two idempotent matrices. *Linear Algebra and its Applications*, 321:3–7, December 2000.

[2] Oskar Maria Baksalary. Idempotency of linear combinations of three idempotent matrices, two of which are disjoint. *Linear Algebra and its Applications*, 388:67–78, September 2004.

[3] Åke Björck. *Numerical methods for least squares problems*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996.

[4] H.-J. Bungartz, M. Griebel, D. Röschke, and C. Zenger. Pointwise convergence of the combination technique for the Laplace equation. *East-West J. Numer. Math.*, 2:21–45, 1994.

[5] Hans-Joachim Bungartz and Michael Griebel. Sparse grids. *Acta Numerica*, 13:147–269, 2004.

[6] B. A. Davey and H. A. Priestley. *Introduction to lattices and order*. Cambridge University Press, New York, second edition, 2002.

[7] Frank Deutsch. Rate of convergence of the method of alternating projections. In *Parametric optimization and approximation (Oberwolfach, 1983)*, volume 72 of *Internat. Schriftenreihe Numer. Math.*, pages 96–107. Birkhäuser, Basel, 1985.

[8] Frank Deutsch and Hein Hundal. The rate of convergence for the method of alternating projections. II. *J. Math. Anal. Appl.*, 205(2):381–405, 1997.

[9] Kurt Friedrichs. On certain inequalities and characteristic value problems for analytic functions and for functions of two variables. *Trans. Amer. Math. Soc.*, 41(3):321–364, 1937.

[10] J. Garcke. *Maschinelles Lernen durch Funktionsrekonstruktion mit verallgemeinerten dünnen Gittern*. Doktorarbeit, Institut für Numerische Simulation, Universität Bonn, 2004.

[11] J. Garcke and M. Griebel. Data mining with sparse grids using simplicial basis functions. In F. Provost and R. Srikant, editors, *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, USA*, pages 87–96, 2001.

[12] J. Garcke, M. Griebel, and M. Thess. Data mining with sparse grids. *Computing*, 67(3):225–253, 2001.

[13] Jochen Garcke. Regression with the optimised combination technique. In W. Cohen and A. Moore, editors, *Proceedings of the 23rd International Conference on Machine Learning*, pages 321–328, 2006.

[14] T. Gerstner and M. Griebel. Numerical Integration using Sparse Grids. *Numer. Algorithms*, 18:209–232, 1998.

[15] T. Gerstner and M. Griebel. Dimension–Adaptive Tensor–Product Quadrature. *Computing*, 71(1):65–87, 2003.

[16] Gene H. Golub and Charles F. Van Loan. *Matrix computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, third edition, 1996.

[17] M. Griebel. A domain decomposition method using sparse grids. In A. Quarteroni, editor, *Contemporary Mathematics, Vol. 157, DDM6*, pages 255–261. American Mathematical Society, 1994.

[18] M. Griebel, M. Schneider, and C. Zenger. A combination technique for the solution of sparse grid problems. In P. de Groen and R. Beauwens, editors, *Iterative Methods in Linear Algebra*, pages 263–281. IMACS, Elsevier, North Holland, 1992.

[19] M. Hegland. Adaptive sparse grids. In K. Burrage and Roger B. Sidje, editors, *Proc. of 10th Computational Techniques and Applications Conference CTAC-2001*, volume 44 of *ANZIAM J.*, pages C335–C353, April 2003.

[20] M. Hegland, O. M. Nielsen, and Z. Shen. Multidimensional smoothing using hyperbolic interpolatory wavelets. *Electronic Transactions on Numerical Analysis*, 17:168–180, 2004.

[21] Markus Hegland. Additive sparse grid fitting. In *Proceedings of the Fifth International Conference on Curves and Surfaces, Saint-Malo, France 2002*, pages 209–218. Nashboro Press, 2003.

[22] S. Knapek. *Approximation und Kompression mit Tensorprodukt-Multiskalenräumen*. Doktorarbeit, Universität Bonn, April 2000.

[23] S. Knapek. Hyperbolic cross approximation of integral operators with smooth kernel. SFB Preprint 665, SFB 256, University of Bonn, 2000. `http://wissrech.ins.uni-bonn.de/research/pub/knapek/fourier.ps.gz`.

[24] Donald E. Knuth. *The art of computer programming. Vol. 1: Fundamental algorithms*. Second printing. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont, 1969.

[25] S. A. Smolyak. Quadrature and interpolation formulas for tensor products of certain classes of functions. *Dokl. Akad. Nauk SSSR*, 148:1042–1043, 1963. Russian, Engl. Transl.: Soviet Math. Dokl. 4:240–243, 1963.

[26] G. Wahba. *Spline models for observational data*, volume 59 of *Series in Applied Mathematics*. SIAM, Philadelphia, 1990.

[27] H. Yserentant. On the multi-level splitting of finite element spaces. *Numerische Mathematik*, 49:379–412, 1986.

[28] Christoph Zenger. Sparse grids. In *Parallel algorithms for partial differential equations (Kiel, 1990)*, volume 31 of *Notes Numer. Fluid Mech.*, pages 241–251. Vieweg, Braunschweig, 1991.