Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Error estimates for two-dimensional Cross Approximation

by

Jan Schneider

Preprint no.: 5 2009

Error estimates for Cross Approximation in 2D

Jan Schneider
Max Planck Institute for Mathematics
in the Sciences, Leipzig

1 Abstract

In this paper we deal with the approximation of a given function f on $[0,1]^2$ by special bilinear forms $\sum_{i=1}^k g_i \otimes h_i$ via the so-called cross approximation. In particular we are interested in estimating the error function $f - \sum_{i=1}^k g_i \otimes h_i$ of the corresponding algorithm in the maximum norm. Because of the large amount of papers available that deal with similar matrix algorithms in applied situations without giving satisfactory error estimates, we concentrate on the theoretical issues of the problem in the language of functions. We connect it with related results from other areas of Analysis in an historical survey and give a lot of references. Our main result is the connection of the error of our algorithm with the error of best approximation by arbitrary bilinear forms.

2 Introduction and preliminaries

We are basically concerned with the following question:
Given a function $f : [0,1]^2 \to \mathbb{R}$, how well can we approximate it by something like

$$f \sim \sum_{i=1}^{k} g_i \otimes h_i, \tag{1}$$

i.e., by a finite sum of tensor products of one-dimensional functions (here we write $(g \otimes h)(x,y) = g(x)h(y)$)? The right-hand side of (1) is also called a bilinear form and the first famous result in this direction is due to Schmidt ([23]), who gave a complete answer in the case $f \in L_2$. A standard reference for questions in this area is [11], a nice survey can be found in [10].
In this paper we consider a very special choice of functions g, h in (1), namely the restriction of f itself to certain lines, as will be described in the sequel.

2.1 The Construction

Now we describe the algorithm CA2D and fix the notation. We are given an arbitrary function f on the unit square $[0,1]^2$. In the first step we choose the point $(x_1, y_1) \in [0,1]^2$ with $f(x_1, y_1) \neq 0$ and define the auxiliary function

$$f_1(x,y) = \frac{f(x_1,y)f(x,y_1)}{f(x_1,y_1)}.$$

Then it is easy to see that

$$f_1(x,y) = f(x,y) \quad \text{for all } (x,y) \in C_1 = \{(x,y) \in [0,1]^2 : x = x_1 \vee y = y_1\}.$$

Hence, for the remainder we have

$$R_1 = f - f_1 = 0 \qquad \text{on } C_1.$$

Now we want to approximate the remainder function R_1 by the same idea. Therefore, we choose $(x_2, y_2) \in [0,1]^2$ with
$(f - f_1)(x_2, y_2) \neq 0$ and define

$$f_2(x,y) = \frac{(f-f_1)(x,y_2)(f-f_1)(x_2,y)}{(f-f_1)(x_2,y_2)}.$$

Then we verify

$$f_2 = f - f_1 = R_1 \quad \text{on } C_2 = \{(x,y) \in [0,1]^2 : x = x_2 \vee y = y_2\}$$

and $f_2 = 0$ on C_1, hence

$$f_1 + f_2 = f \quad \text{on } G_2 = C_1 \cup C_2 \quad \text{and so} \quad R_2 = f - f_1 - f_2 = 0 \quad \text{on } G_2.$$

We go on with this scheme and define for $j \in \mathbb{N}$ the iterative expression

$$f_j(x,y) = \frac{\left(f - \sum_{i=1}^{j-1} f_i\right)(x_j, y)\left(f - \sum_{i=1}^{j-1} f_i\right)(x, y_j)}{\left(f - \sum_{i=1}^{j-1} f_i\right)(x_j, y_j)},$$

where the pivot points (x_j, y_j) are always chosen such that

$$\left(f - \sum_{i=1}^{j-1} f_i\right)(x_j, y_j) \neq 0.$$

For a detailed discussion about how to choose those points we refer to the next subsection. The function given by

$$F_k(x,y) = \sum_{j=1}^{k} f_j(x,y)$$

is the resulting k-th interpolation function of f via this algorithm, the two-dimensional cross approximation (CA2D). We observe that F_k has the property of the right hand-side of (1) as a sum of products of one-dimensional functions. By repeating the same arguments as before, one can prove

$$F_k(x,y) = f(x,y) \quad \text{for all } (x,y) \in G_k = \bigcup_{j=1}^{k} C_j,$$

where $C_j = \{(x,y) \in [0,1]^2 : x = x_j \vee y = y_j\}$, hence,

$$R_k(x,y) := f(x,y) - F_k(x,y) = 0 \quad \text{for all } (x,y) \in G_k \tag{2}$$

for $k \in \mathbb{N}$. So one can think of R_k as a function that lives inside of small rectangles and vanishes on their edges. By the above construction we have

$$R_k(x,y) = R_{k-1}(x,y) - \frac{R_{k-1}(x,y_k)R_{k-1}(x_k,y)}{R_{k-1}(x_k,y_k)}, \tag{3}$$

which also shows, how to recursively implement CA2D.

2.2 Questions

Our main goal is to estimate the maximum norm of the CA2D error function R_k by the error of best approximation by bilinear forms. This will be done in section 5 for functions either belonging to the space $C([0,1]^2)$ of continuous functions or to the space $L_\infty([0,1]^2)$ of bounded functions. There are some questions related to this, for example: What are the influences of smoothness and structural properties of f? We discuss those issues in section 3.
The point we want to treat now is the choice of the pivot elements (x_j, y_j). At the first glance it seems reasonable to choose the remaining maximum of the error in $[0,1]^2$ (full pivoting), i.e.

$$\left(f - \sum_{i=1}^{j-1} f_i\right)(x_j,y_j) = \max_{(x,y)\in[0,1]^2} \left(f - \sum_{i=1}^{j-1} f_i\right)(x,y) \neq 0,$$

to minimize the error after the next step. But as soon as it comes to the implementation, one of course intends to avoid full pivoting. Therefore, another alternative was considered (partial pivoting), see [7], Chapter 4, where the x-coordinates of the crosses are chosen randomly such that $R_k(x_{k+1}, y)$ does not vanish for all $y \in [0,1]$ and the y-coordinates as the maxima on the line. When we implemented CA2D for testing we used an even more restrictive algorithm (special pivoting), where the x_k are determined by the following procedure: $x_1 = 1/2, x_2 = 1/4, x_3 = 3/4$ and going on from left to right by always dividing the remaining intervals in the middle. The y-coordinates are again the maxima on the line. (One has to be careful here with symmetric functions !). We will see in sections 4 and 5 that there is also another pivot strategy of interest called maximal-volume concept, see [13], we will discuss it at the end of 4.1. However, the results of the following section are independent of such a strategy.

3 Basic results

We start with an observation already mentioned as formula (2) in 2.1. Using the notation introduced there we will formulate it as Proposition and refer to it later on as interpolation property.

Proposition 3.1 *For any function $f : [0,1]^2 \to \mathbb{R}$, we have*

$$R_k(x,y) = 0 \quad \textit{for all } (x,y) \in G_k.$$

The next result takes an a priori knowledge about structural properties of the underlying function into account. We say that a function f has separation rank k, if one can represent it as

$$f(x,y) = \sum_{i=1}^{k} g_i(x)h_i(y)$$

and there is no such representation with reduced summing order. We call the following the rank property.

Proposition 3.2 *Let f have separation rank k. Then CA2D reproduces f after k steps exactly, that means*

$$R_k = f - F_k = 0 \quad \textit{on } [0,1]^2.$$

A matrix version of this result can be found in [7](Chapter 4).
Proof We will prove that $R_{k'} = f - F_{k'}$ has separation rank $k - k'$ for $k' \in \{0, 1, \ldots, k\}$ by induction. For $k' = 0$ there is nothing to prove, so let for $k' < k$ the function $R_{k'}$ have separation rank $k - k'$. We define

$$V = \operatorname{span}\{R_{k'}(\cdot, y) : y \in [0,1]\}$$

and

$$V' = \text{span}\{R_{k'+1}(\cdot, y) : y \in [0,1]\}.$$

We know that $\dim V = k - k'$ and want to show $\dim V' = k - k' - 1$. For each $\bar{y} \in [0,1]$ we write with the notation of section 2.1 (formula (3))

$$R_{k'+1}(x, \bar{y}) = R_{k'}(x, \bar{y}) - \frac{R_{k'}(x, y_{k'+1}) R_{k'}(x_{k'+1}, \bar{y})}{R_{k'}(x_{k'+1}, y_{k'+1})}$$

and see, that both terms on the right hand-side belong to V, hence $V' \subset V$. Furthermore, we know $R_{k'}(\cdot, y_{k'+1}) \in V$ but because $R_{k'}(x_{k'+1}, y_{k'+1}) \neq 0$ and $R_{k'+1}(x_{k'+1}, y) = 0$ for all $y \in [0,1]$, there is no representation of $R_{k'}(\cdot, y_{k'+1})$ as a linear combination of functions $R_{k'+1}(\cdot, y)$, hence $R_{k'}(\cdot, y_{k'+1}) \notin V'$. It follows $\dim V' < \dim V$ and because those dimensions can differ at most by one, we get $\dim V' = \dim V - 1 = k - k' - 1$. Now we know that for all $\bar{y} \in [0,1]$ we have a representation

$$R_{k'+1}(x, \bar{y}) = \sum_{i=1}^{k-k'-1} \alpha_{i,\bar{y}} \varphi_i(x)$$

with coefficients $\alpha_{i,\bar{y}}$ and functions $\varphi_i(x)$. If we now identify $\psi_i(y) = \alpha_{i,y}$, we conclude that $R_{k'+1}$ has separation rank $k - (k'+1)$.

$\square$

This result tells us that the algorithm is exact, if f has already a tensor product structure as in (1), even if f is not smooth at all. Besides that one would expect that CA2D converges faster if f shares some nice smoothness properties. To get an explicit estimate for a more general class of functions determined by the smoothness, we follow the basic idea appearing for polynomial interpolation on an interval. For that we define

$$\omega_k(x) = \prod_{i=1}^{k} (x - x_i)$$

and denote by $f_x^{(k)}$ the k-th partial derivative of f with respect to x.

Proposition 3.3 *Let $f \in C^k([0,1]^2)$. Then the error of CA2D can be estimated by*

$$|R_k(x, y)| \leq \frac{|\omega_k(x)|}{k!} 2^k \sup_{x \in [0,1]} |f_x^{(k)}(x, y)|. \tag{4}$$

Proof We fix $(\bar{x}, \bar{y}) \in [0,1]^2$ and define

$$F(x) = R_k(x, \bar{y}) - K \omega_k(x).$$

Now we determine K such that $F(\bar{x}) = 0$. Then F has at least $n+1$ zeros in $[0,1]$, hence $F^{(k)}(\xi) = 0$ for $\xi \in [0,1]$. We find

$$K = \frac{R_k^{(k)}(\xi, \bar{y})}{k!},$$

where the derivative is with respect to the first variable, and because of $|R_k^{(k)}(x,y)| \leq 2^k |f_x^{(k)}(x,y)|$ we can estimate

$$|R_k(\bar{x},\bar{y})| = |K\omega_k(\bar{x})| \leq \frac{|\omega_k(\bar{x})|}{k!} 2^k |f_x^{(k)}(\xi,\bar{y})|,$$

which finishes the proof.

□

Let us discuss this result. First one observes, that it is basically one-dimensional, where the behavior in the other direction is not taken into account. So one can do the same argumentation with respect to y and the corresponding assertion would also be true. That means the error R_k is bounded by the expression (4) with respect to x or y, therefore,

$$|R_k(x,y)| \leq \frac{2^k}{k!} \min\left(|\omega_k^x(x)| \sup_{x\in[0,1]} |f_x^{(k)}(x,y)|, |\omega_k^y(y)| \sup_{y\in[0,1]} |f_y^{(k)}(x,y)| \right).$$

But that of course doesn't change the quality of the estimate. We tested the algorithm with our special pivoting for the function $f(x,y) = \exp(-xy)$, where $|f_x^{(k)}(x,y)| \leq 1$. After 15 steps the error of CA2D was $5.2479 \cdot 10^{-15}$, where estimate (4) gives $2.8422 \cdot 10^{-14}$. This seems reasonable, but the situation changes dramatically if the partial derivatives of f are not uniformly bounded. We tested also $f(x,y) = \sin(10xy)$, where the corresponding factor in (4) grows like 10^k, here the algorithm gave after 15 steps an error of $2.8084 \cdot 10^{-11}$ but our estimate gives 28.422. That means, derivatives of f itself can not be a suitable factor in the error estimate, but of course smoothness should influence it somehow. So observing that R_k does not change after "nice" transformations $\Phi : [0,1]^2 \to [0,1]^2$, not the derivatives of f itself, but the "smallest possible" derivatives after a suitable transformation Φ are of interest. So we conclude

$$\|R_k\|_\infty \leq \frac{2^k}{k!} \inf_\Phi \left(\min\left(\left\| \omega_k^x \frac{\partial^k}{\partial x^k}(f\circ\Phi) \right\|_\infty, \left\| \omega_k^y \frac{\partial^k}{\partial y^k}(f\circ\Phi) \right\|_\infty \right) \right).$$

Unfortunately we can not see a way of simplifying this expression and we can also not test it.
When we realized that direct error estimates would need new techniques, we started to search intensively through the literature for similar ideas. The most important results of this process are presented in the next section.

4 Historical survey on related questions

This survey is not meant to be complete or even ordered in time. It simply shows, how the literature influenced this work. We start with the most recent interest in approximation schemes by low rank matrices.

4.1 Asymptotically smooth functions

In [5] Bebendorf and later in [6] Bebendorf and Rjasanov considered the approximation of matrices $A = (a_{ij})_{i,j=1}^n$ generated by a function f if one assigns $a_{ij} = f(x_i, y_j)$ on a sufficiently fine grid $(x_i, y_j)_{i,j=1}^n$ in the corresponding domain. Such a function was assumed to be asymptotically smooth. We will repeat the definition now in a form that fit to our purposes.

Definition 4.1 *A function $f : [a, b] \times [c, d] \to \mathbb{R}$ is called asymptotically smooth if there are constants $C_1, C_2 > 0$ and $s \leq 0$ such that for all $\alpha, \beta \in \mathbb{N}_0$*

$$|\partial_x^\alpha \partial_y^\beta f(x, y)| \leq C_1 l! C_2^l |x - y|^{s-l}, \quad l = \alpha + \beta.$$

Compare also [15] (Definition 4.2.5). In addition to that they assumed

$$|c - d| \leq \eta \operatorname{dist}([a, b], [c, d])$$

for some $0 < \eta < C_2^{-1}$. In other words the function was considered off the diagonal $y = x$, which is quite different from our original question in $[0, 1]^2$. An important class of examples is given by the fundamental solutions of elliptic equations. The kernels $\log |x - y|$ with $s = 0$ or $|x - y|^{-a}$ with $s = a$ are prominent representatives. As mentioned in [15] asymptotically smooth functions are also called Calderon-Zygmund kernels, see also [12] and [16]. Unfortunately, the focus of the authors in those papers lies on the operators generated by such functions, but not on the functions itself. It would be desirable to clarify what kind of function spaces, maybe in the sense of microlocal analysis by Moritoh/Yamada [21] and Kempka [18] or even in the sense of varying smoothness [24], would be the right scale for these kernels. But that is not done within this work.
Using a result about high-dimensional Lagrange interpolation Bebendorf proved in [5] (Theorem 4) the following estimate for the error of CA2D off the diagonal with partial pivoting

$$|R_k(x, y)| \leq C_k \operatorname{dist}^s([a, b], [c, d]) \eta^k,$$

where $C_k = C_1 C_2^k (1 + 2^k) C_3$. This seems satisfactory since the factor η^k suggests an exponential decay. However, a closer look at the number C_k together with the condition $\eta < C_2^{-1}$ destroys this hope. But it was still a big improvement in terms of explicit error estimates for CA2D in comparison with earlier results concerning for example so-called Pseudoskeleton Approximations by Goreinov, Tyrtyshnikov and Zamarashkin, see [14]. Also in [5] Bebendorf mentions the maximum-volume concept to control the error of CA2D. This concept proposes to choose the pivots (x_i, y_i) such that the absolute value of the determinant $\det (f(x_i, y_i))_{i,j=1}^k$ is maximal. That is of course practically not acceptable but because of the nice formula

$$\prod_{i=0}^{k} R_i(x_{i+1}, y_{i+1}) = \det (f(x_i, y_i))_{i,j=1}^{k+1}, \tag{5}$$

for all $k \in \mathbb{N}$ (which you can also find in [5], Lemma 2), we can see, that partial pivoting is the best strategy with respect to maximal determinants if we want to keep all previous pivots fixed. Much work concerning asymptotically smooth functions and the maximal volume concept in connection with the numerical application was done for example by Tyrtyshnikov, see [28], where also some more references are given.
It is one aim of this paper to extend the class of functions for which CA2D error estimates are available. In the next subsections we examine older examples in literature that are already very close to our purposes.

4.2 Totally positive functions

Already more than thirty years ago, Micchelli and Pinkus wrote an very interesting paper [20] concerning the approximation problem (1) in mixed p, q-norms. The main assumption on their functions was total positivity. Here we repeat the definition.

Definition 4.2 *A real valued kernel $K(x, y)$ continuous on $[0, 1]^2$ is called totally positive if all its Fredholm minors*

$$K\begin{pmatrix} s_1, \dots, s_m \\ t_1, \dots, t_m \end{pmatrix} = \det\left(K(s_i, t_i)\right)_{i,j=1}^m = \begin{vmatrix} K(s_1, t_1) & \cdots & K(s_1, t_m) \\ \vdots & & \vdots \\ K(s_m, t_1) & \cdots & K(s_m, t_m) \end{vmatrix}$$

are nonnegative for $0 \leq s_1 < \cdots < s_m \leq 1, 0 \leq t_1 < \cdots < t_m \leq 1$ and all $m \geq 1$.

For further details about total positivity see [17], where also many examples are given. Micchelli and Pinkus were concerned with finding the best approximation by bilinear forms, i.e.

$$E_{p,q}^n(K) = \inf\left\{\left|K - \sum_{i=1}^n u_i \otimes v_i\right|_{p,q}\right\},$$

where the infimum is taken over all $u_1, \dots, u_n \in L^p[0, 1]$ and $v_1, \dots, v_n \in L^q[0, 1]$ and

$$|K|_{p,q} = \left(\int_0^1 \left(\int_0^1 |K(x, y)^q dy\right)^{p/q} dx\right)^{1/p}, \tag{6}$$

for $1 \leq p, q \leq \infty$. That is exactly our problem (1) in these mixed norms restricted to those special functions. Before we state their results we need some preparation. By the notation in [20] let

$$E(x, y) = \frac{K\binom{x, \tau_1, \dots, \tau_n}{y, \xi_1, \dots, \xi_n}}{K\binom{\tau_1, \dots, \tau_n}{\xi_1, \dots, \xi_n}}.$$

By using Laplace extension twice we see

$$E(x,y) = K(x,y) - \sum_{i,j=1}^{n} c_{ij} K(x,\xi_i) K(\tau_j, y),$$

where

$$c_{ij} = (-1)^{i+j} \frac{K\binom{\tau_1,\dots,\tau_{j-1},\tau_{j+1},\dots,\tau_n}{\xi_1,\dots,\xi_{j-1},\xi_{j+1},\dots,\xi_n}}{K\binom{\tau_1,\dots,\tau_n}{\xi_1,\dots,\xi_n}}.$$

Remark 4.3 *It is easy to see that $E(x,y)$ is nothing else than our error function $R(x,y)$ for CA2D after n steps, compare with* (5) *and the construction given in 2.1. This observation will play the central role in section 5.*

The question remains, how the points $(\xi_i, \tau_i) \in [0,1]^2$ were chosen. We clarify that by stating the first result given in [20].
For $0 = s_0 < s_1 < \cdots < s_n < s_{n+1} = 1$ let

$$h_s(x) = (-1)^i, \quad s_i \le x < s_{i+1}$$

be the corresponding step function according to the n-partition $s = (s_0, \dots s_{n+1})$ of $[0,1]$. The set of those partitions may be denoted by Λ_n. Given a nondegenerate totally positive kernel K there exists a n-partition ξ of $[0,1]$ such that for any other $t \in \Lambda_n$

$$\|Kh_\xi\|_1 := \int_0^1 \left| \int_0^1 K(x,y) h_\xi(y) dy \right| dx \le \int_0^1 \left| \int_0^1 K(x,y) h_t(y) dy \right| dx =: \|Kh_t\|_1.$$

Moreover, Kh_ξ has exactly n distinct zeros in $(0,1)$ at $(\tau_1, \dots, \tau_n) \in \Lambda_n$ and

$$\operatorname{sgn} Kh_\xi = h_\tau, \quad \operatorname{sgn} K^T h_\tau = h_\xi.$$

This is a very helpful result in finding good estimates for $E^n_{p,q}(K)$, but one should be aware that the choice of pivots here is not constructive. The additional assumption on K of being nondegenerate just means that each of the sets of functions $\{K(s_1,y), \dots, K(s_m,y)\}$ and $\{K(x,t_1), \dots, K(x,t_m)\}$ are linearly independent for all choices $s,t \in \Lambda_m$ and all $m \ge 1$.
Now let's state the interesting results obtained in [20] which generalize a former work on n-widths [19].
The first one concerns the case $p = q = 1$ and says that for a nondegenerate totally positive kernel K

$$\begin{aligned} E^n_{1,1}(K) &= \int_0^1 \int_0^1 |E(x,y)| dx dy = \|Kh_\xi\|_1 \\ &= \int_0^1 \int_0^1 \left| K(x,y) - \sum_{i=1}^{n} K(x,\xi_i) \sum_{j=1}^{n} c_{ij} K(\tau_j, y) \right| dx dy \end{aligned}$$

holds. Here (τ_i, ξ_i) are defined as in the result before. In other words, this choice of tensor product approximation as in CA2D is optimal in this norm. The proof of that is really nice and uses the Hobby-Rice Theorem.
After that Micchelli and Pinkus generalized this to all values $p \in [1, \infty]$ and related the question to n-widths of certain subspaces of L^p. Here we briefly recall their notation. The Kolmogorov n-width is defined by

$$d_n(U, X) = \inf_{X_n} \sup_{x \in U} \inf_{y \in X_n} \|x - y\|,$$

where U is a subset of the normed linear space X and X_n any n-dimensional subspace of X. The subspaces of interest here are

$$\mathcal{K}_p = \{Kh : \|h\|_p \leq 1\}.$$

Now, the result states that for a nondegenerate totally positive kernel K

$$\begin{aligned} E^n_{p,1}(K) &= \left(\int_0^1 \left(\int_0^1 |E(x,y)| dy \right)^p dx \right)^{1/p} = d_n(\mathcal{K}_\infty, L^p[0,1]) \\ &= \left(\int_0^1 \left(\int_0^1 \left| K(x,y) - \sum_{i=1}^n K(x, \xi_i) \sum_{j=1}^n c_{ij} K(\tau_j, y) \right| dy \right)^p dx \right)^{1/p} \end{aligned}$$

holds. Here for the choice of points (τ_i, ξ_i) an analogue result as above was used, so we have optimality of the construction as in CA2D again, although we have an existence assertion only for the pivot points.
In [22], chapter V, some further work is done in this direction, but the question of error estimates of our specific construction slipped out of interest. All those results are contained, generalized and considered under a more complex framework in [8]. To conclude this section we remark that these results connect the error of CA2D with the error of best approximation by bilinear forms, as in (1), for certain mixed L_p-norms. If one has asymptotic estimates for the error of best approximation by bilinear forms in such spaces, now one can make direct use of it, see section 6. In the next subsection we find a hint how to connect the best approximation error and the one of CA2D also for other norms than mixed L_p, namely for the sup-norm in the first place.

4.3 Exact annihilators

At the beginning of the eighties M.-B. A. Babaev, see [1] and [2], introduced the concept of an exact annihilator (EA) of a set of functions and used it to solve several problems in approximation theory. In particular, he gave two-sided estimates for the best approximation by bilinear forms, as in (1), using the operator norm of such an EA corresponding to the underlying spaces. We start with repeating his notion of an EA of a set $G \subset X(T)$, where $T = [0,1]^2$ and either $X = C(T)$, the space of continuous functions, or $X = L_{p,q}(T)$, the space of integrable functions normed as in (6). We keep his notation as far as possible.

Definition 4.4 *Let $M \in \mathbb{N}$ and $\Theta = T^M$. An exact annihilator (EA) of the set G is a continuous operator $\nabla : X(T) \longrightarrow X(\Theta)$, such that*

$$f \in G \quad \textit{if, and only if,} \quad (\nabla f)(\theta) = 0 \quad \forall \theta \in \Theta.$$

Because our main goal in this paper is to get information about the error of CA2D in the maximum norm we concentrate now on the cases $X = C$ and $X = L_\infty$ separately.

4.3.1 The case $C([0,1]^2)$

All what follows in this part can be found in [1]. For $\theta = (x_1, \dots, x_M, y_1, \dots, y_M) \in \Theta$ we define the operator $\overset{M}{\nabla}_*$ by

$$(\overset{M}{\nabla}_* f)(\theta) = \begin{cases} \frac{(\overset{M}{\nabla} f)(\theta)}{\|\overset{M-1}{\nabla} f\|_{C(T^{M-1})}}, & (\overset{M}{\nabla} f)(\theta) \neq 0, \\ 0, & (\overset{M}{\nabla} f)(\theta) = 0. \end{cases},$$

where $(\overset{M}{\nabla} f)(\theta) = \det\left(f(x_i, y_i)\right)_{i,j=1}^M$.

Theorem 4.5 *For each $M \geq 2$ the operator $\overset{M}{\nabla}_*$ is an EA of the following set of bilinear forms*

$$G = G_C^{M-1}(T) = \left\{ g = \sum_{i=1}^{M-1} \varphi_i(x)\psi_i(y), \quad \varphi_i, \psi_i \in C([0,1]) \right\}.$$

Here we already see that some close connection between this operator and the error of CA2D seems possible. If we interpret $\theta = (x_1, \dots, x_M, y_1, \dots, y_M) \in \Theta$ as the set of pivot coordinates and keep Remark 4.3 in mind, obviously the Theorem above reminds us of the rank property of CA2D in Proposition 3.2. Now we state the main result of [1] from which we can establish the connection of the errors of CA2D and best approximation, i.e,

$$E(f, G)_{C(T)} = \inf_{g \in G} \|f - g\|_{C(T)}.$$

Theorem 4.6 *For any function $f \in C(T)$ we have*

$$1/M^2 \| \overset{M}{\nabla}_* f\|_{C(T^M)} \leq E(f, G)_{C(T)} \leq \| \overset{M}{\nabla}_* f\|_{C(T^M)}.$$

This result is one of the keys to our main results in section 5. Because CA2D does not require continuity we assume only boundedness in the next part.

4.3.2 The case $L_\infty([0,1]^2)$

In [3] Babaev used the concept of an exact annihilator to attack the problem of estimating the best approximation by bilinear forms in mixed L_p spaces including L_∞. We state his results in this part in full generality even though we are most interested in the L_∞ versions, which we will use afterwards to find a concrete error estimate for CA2D.

For $\theta = (x_1, \dots, x_M, y_1, \dots, y_M) \in \Theta$ we define the operator $\overset{M}{\nabla}_+$ by

$$(\overset{M}{\nabla}_+ f)(\theta) = \begin{cases} \frac{(\overset{M}{\nabla} f)(\theta)}{\|\overset{M-1}{\nabla} f\|_{L_{p,q}(T^{M-1})}}, & (\overset{M}{\nabla} f)(\theta) \neq 0, \\ 0, & (\overset{M}{\nabla} f)(\theta) = 0. \end{cases},$$

where $(\overset{M}{\nabla} f)(\theta) = \det\left(f(x_i, y_i)\right)_{i,j=1}^M$ has the same meaning as in the previous case.

Theorem 4.7 *For each $M \geq 2$ the operator $\overset{M}{\nabla}_+$ is an EA of the following set of bilinear forms*

$$B = B_{p,q}^{M-1}(T) = \left\{ g = \sum_{i=1}^{M-1} \varphi_i(x)\psi_i(y), \quad \varphi_i \in L_p([0,1]), \psi_i \in L_q([0,1]) \right\}.$$

Now we state the part of the main result of [3] that we can use for the connection of the errors of CA2D and best approximation, i.e,

$$E(f,B)_{L_{p,q}(T)} = \inf_{g \in B} \|f - g\|_{L_{p,q}(T)}.$$

Theorem 4.8 *For any function $f \in L_{p,q}(T)$ with $0 < p, q \leq \infty$ we have*

$$A_{M,p,q}(f) b_{M,p,q} \| \overset{M}{\nabla}_+ f\|_{L_{p,q}(T^M)} \leq E(f,B)_{L_{p,q}(T)},$$

where

$$A_{M,p,q}(f) = \frac{\| \overset{M-1}{\nabla} f\|_{L_{p,q}(T^{M-1})}}{\|f\|_{L_{p,q}(T)}^{M-1}}$$

and

$$b_{M,p,q} = \left(\frac{2^{Mp^*} - 1}{2^{p^*} - 1} M! \right)^{-1/p^*}, \quad p^* = \min(1, p, q).$$

We don't believe that these constants can not be improved, especially if one imposes more properties of the function f, but we did not yet succeed in proving it.

5 Main results

Now we are in the position to combine everything we learned from the literature and state our main results. Let the points $(x_1, y_1), \dots, (x_k, y_k)$ according to CA2D be chosen, such that

$$\left|(\overset{k}{\nabla} f)(\theta)\right| = \left|\det\left(f(x_i, y_i)\right)_{i,j=1}^{k}\right| = \left|f\begin{pmatrix} x_1, \dots, x_k \\ y_1, \dots, y_k \end{pmatrix}\right|$$

is maximal with respect to $\theta = (x_1, \dots, x_k, y_1, \dots, y_k) \in T^k$. This is the maximal-volume concept already discussed at the end of 4.1.

Theorem 5.1 *Let $R_k(x, y)$ be the remainder function of CA2D after k steps with the above choice of pivots, then we have*

$$|R_k(x, y)| \leq (k+1)^2 E(f, G)_{C(T)}.$$

Proof By Remark 4.3 we have the following identity

$$|R_k(x, y)| = \left| \frac{f\binom{x, x_1, \dots, x_k}{y, y_1, \dots, y_k}}{f\binom{x_1, \dots, x_k}{y_1, \dots, y_k}} \right|.$$

Because of the special choice of pivots we can for $\widehat{\theta} = (x, x_1, \dots, x_m, y, y_1, \dots, y_m)$ write

$$|R_k(x, y)| = \left| \frac{(\overset{k+1}{\nabla} f)(\widehat{\theta})}{\|\overset{k}{\nabla} f\|_{C(T^k)}} \right|.$$

Using Definition 4.4 and Theorem 4.6 we can conclude

$$|R_k(x, y)| = \left|(\overset{k+1}{\nabla}_* f)(\widehat{\theta})\right| \leq \|\overset{k+1}{\nabla}_* f\|_{C(T^{k+1})} \leq (k+1)^2 E(f, G)_{C(T)}.$$

□

The proof of Theorem 5.1 allows an immediate generalization in terms of the pivot strategy. Let now the points $(x_1, y_1), \dots, (x_k, y_k)$ be chosen, such that

$$\|\overset{k}{\nabla} f\|_{C(T^k)} \leq \tau \left| f\begin{pmatrix} x_1, \dots, x_m \\ y_1, \dots, y_m \end{pmatrix} \right|$$

for a real number $\tau \geq 1$.

Corollary 5.2 *With the above notation we have*

$$|R_k(x, y)| \leq \tau (k+1)^2 E(f, G)_{C(T)}.$$

Finally, we established an estimate of the error of CA2D for continuous functions from above by the error of best approximation by arbitrary bilinear forms. If there would be an explicit estimate of $E(f,G)_{C(T)}$ for special functions f (say smooth) available, one could immediately plug it in here to obtain a concrete estimate for CA2D. For the next case, we will follow this idea in section 6. Now we state the analogue of Theorem 5.1 for the L_∞-norm. Let the points $(x_1,y_1),\dots,(x_k,y_k)$ according to CA2D be chosen, such that

$$\left|(\overset{k}{\nabla} f)(\theta)\right| = \| \overset{k}{\nabla} f\|_{L_\infty(T^k)}.$$

Then we can state:

Theorem 5.3 *Let $R_k(x,y)$ be the remainder function of CA2D after k steps with the above choice of pivots, then we have*

$$|R_k(x,y)| \le (2^{k+1}-1)(k+1)! \frac{\|f\|^k_{L_\infty(T)}}{\| \overset{k}{\nabla} f\|_{L_\infty(T^k)}} E(f,B)_{L_\infty(T)}.$$

The idea of the proof is exactly the same as for Theorem 5.1, one can follow it line by line. Also in analogy to the case of $C([0,1]^2)$ we can formulate an easy modification. Let now the points $(x_1,y_1),\dots,(x_k,y_k)$ be chosen, such that

$$\| \overset{k}{\nabla} f\|_{L_\infty(T^k)} \le \tau \left|(\overset{k}{\nabla} f)(\theta)\right|$$

for a real number $\tau \ge 1$.

Corollary 5.4 *With the above notation we have*

$$|R_k(x,y)| \le \tau(2^{k+1}-1)(k+1)! \frac{\|f\|^k_{L_\infty(T)}}{\| \overset{k}{\nabla} f\|_{L_\infty(T^k)}} E(f,B)_{L_\infty(T)}.$$

Because of the bad looking constants this result lost some beauty. Nevertheless, we use it in the next section to show how explicit error estimates for CA2D can be produced.

6 Best approximation

In this section we complement the results obtained in the previous section by two-sided error estimates for the best approximation by bilinear forms available in the literature. We concentrate here on the papers published by Babaev (see

[4]) and Temlyakov (see for example [25]-[27]). They were concerned with the following question: What is the exact asymptotic behavior of the quantities

$$\tau_M(f)_{p_1,p_2} = \inf_{u_i,v_i;i=1,\dots,M} \left\| f(x,y) - \sum_{i=1}^{M} u_i(x)v_i(y) \right\|_{p_1,p_2}$$

and

$$\tau_M(F)_{p_1,p_2} = \sup_{f\in F} \tau_M(f)_{p_1,p_2}$$

for various choices of function classes F ? Here we keep the notation for best approximation used by these authors.
Babaev concentrated on the unit ball of the classical Sobolev class $W_q^r(T)$, where Temlyakov treated periodic functions f defined on the 2d-dimensional torus π_{2d} belonging to a Sobolev class with bounded mixed derivatives. We will formulate some of their results in a common notation.
A typical result in Temlyakovs papers, obtained in [25] for $p_1 = p_2 = p$, looks like

$$\tau_M(W^{\mathbf{r}}_{q,\alpha})_p \sim \begin{cases} M^{-2r+1/q-1/p}, & 1 \le q \le p \le 2, r > 1/q - 1/p, \\ M^{-2r}, & 2 \le q, p \le \infty, r > 1/2, \\ M^{-2r+1/q-1/2}, & 1 \le q < 2 < p \le \infty, r > 1/q. \end{cases},$$

for $\mathbf{r} = (r_1, r_2) = r$.
For our purposes the results of Babaev ([4]) fit better to our needs. He found

$$\tau_M(W_q^r)_p \sim M^{-r}$$

for all $2 \le q \le p \le \infty$ and $r > 2/q - 1/p$.
Now we combine this result with Corollary 5.4 to establish a quantitative error estimate of CA2D in the L_∞-norm.

Theorem 6.1 *With the assumptions of Corollary 5.4 we have for all f belonging to the unit ball of $W_\infty^r(T)$*

$$|R_k(x,y)| \le c\tau(2^{k+1}-1)(k+1)! \frac{(k+1)^{-r}}{\| \overset{k}{\nabla} f\|_{L_\infty(T^k)}}.$$

This estimate surely suffers again from the k-dependence of the constants. But one could argue the following way : If we assume f to be very smooth, we can reach a very large r in the estimate. Since we know by experiments that CA2D converges very fast for nice functions, we only need to consider small values of k and the terms blowing up with k would not destroy the nice flavor of the estimate. But of course it is desirable to improve the constants, which we postpone to further work.

Acknowledgements. The author thanks Prof. W. Hackbusch (Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany) as well as Prof. E. Novak and Prof. W. Sickel (both Friedrich Schiller University, Jena, Germany) for inspiring discussions.

References

[1] Babaev, M.-B. A.: Best Approximation by Bilinear Forms. Mat. Zametki, Vol. 46, No. 2, 21-33, 1989.

[2] Babaev, M.-B. A.: Exact Annihilators and their Applications in Approximation Theory. Trans. Acad. Sci. Azerb. Ser. Phys.-Tech. Math. Sci. 20 (2000), no. 1, Math. Mech., 17–24, 233.

[3] Babayev, M.-B. A.: On estimation of the best approximation by bilinear forms in space with mixed norm. Trans. Acad. Sci. Azerb. Ser. Phys.-Tech. Math. Sci. 24 (2004), no. 4, Math. Mech., 23–32.

[4] Babaev, M.-B. A.: Approximation of the Sobolev classes W_q^r of functions of several variables by bilinear forms in L_p for $2 \leq q \leq p \leq \infty$. (Russian) Mat. Zametki 62 (1997), no. 1, 18–34; translation in Math. Notes 62 (1997), no. 1-2, 15–29 (1998).

[5] Bebendorf, M.: Approximation of boundary element matrices. Numer. Math. (2000)86: 565-589.

[6] Bebendorf, M.; Rjasanow, S.: Adaptive Low-Rank Approximation of Collocation Matrices. Computing 70, 1-24 (2003).

[7] Boerm, S.; Grasedyck, L.; Hackbusch, W.: Hierarchical Matrices. Max Planck Institute for Mathematics in the Sciences, Lecture note 21/2003.

[8] Brown, A.L.: Finite Rank Approximations to Integral Operators which satisfy certain Total Positivity conditions. Jour. Approx. Theo. **34**, 42-90 (1982).

[9] Carvajal, O.A.; Chapman, F.W.; Geddes, K.O.: Hybrid symbolic-numeric integration in multiple dimensions via tensor-product series. ISSAC'05, 84–91 (electronic), ACM, New York, 2005.

[10] Cheney, E.W.: The Best Approximation of Multivariate Functions by Combinations of Univariate Ones. Approximation theory IV (College Station, Tex., 1983), 1-26, Academic Press, New York, 1983.

[11] Cheney, E.W.; Light, W.A.: Approximation Theory in Tensor Product Spaces. Lecture Notes in Mathematics, 1169. Springer-Verlag, 1985.

[12] Coifman, R.; Meyer, Y.: Wavelets and Operators II, Calderon-Zygmund and multilinear operators. Cambridge Studies in Advanced Mathematics, 48. Cambridge University Press, Cambridge, 1997.

[13] Goreinov, S.A.; Tyrtyshnikov, E.E.: The Maximal-Volume Concept in Approximation by Low-Rank Matrices. Contemp. Math. Vol. **280** 2001.

[14] Goreinov, S.A.; Tyrtyshnikov, E.E.; Zamarashkin, N.L.:A Theory of Pseudoskeleton Approximations. Lin. Algeb. Appl. **261** (1997), 1-21.

[15] Hackbusch, W.: Hierarchische Matrizen - Algorithmen und Analysis. Lecture Notes of Max Planck Institute for Mathematics in the Sciences, Leipzig 2007.

[16] Jaffard, S. and Meyer, Y.: Wavelet Methods for Pointwise Regularity and Local Oscillations of Functions. Mem. Amer. Math. Soc. **123** (1996), No. 587.

[17] Karlin, S.: Total Positivity, Vol 1. Stanford Univ. Press, Stanford, Calif., 1968.

[18] Kempka, H.:Atomic, molecular and wavelet decomposition of generalized 2-microlocal Besov spaces. to appear in Jour. Function Spaces Appl. (2008).

[19] Micchelli, C.A.; Pinkus, A.: On n-Widths in L^∞. Trans. Amer. Math. Soc. **234**, 139-174 (1977).

[20] Micchelli, C.A.; Pinkus, A.: Some Problems in the Approximation of functions of Two Variables and n-Widths of Integral Operators. Jour. Approx. Theo. **24**, 51-77 (1978).

[21] Moritoh, S.; Yamada, T.: Two-microlocal Besov spaces and wavelets. Rev. Mat. Iberoamericana **20** (2004), 277-283.

[22] Pinkus, A.: N-Widths in Approximation Theory. Ergebnisse der Mathematik und ihrer Grenzgebiete, 3.Folge, Band 7, Springer 1985.

[23] Schmidt, E.: Zur Theorie der linearen und nicht linearen Integralgleichungen, I., Math. Ann. **63** (1907), 433-476.

[24] Schneider, J.: Function spaces of varying smoothness I. Math. Nachr. 280, No.16, 1801-1826 (2007)

[25] Temlyakov, V.N.: On Best Bilinear Approximations of Periodic Functions of Several Variables. Soviet Math. Dokl. Vol.33 (1986), No.1.

[26] Temlyakov, V.N.: Approximation of Periodic Functions of Several Variables by Bilinear Forms. Math. USSR Izvestiya Vol.28 (1987), No.1.

[27] Temlyakov, V.N.: Approximation of Periodic Functions of Several Variables by Combinations of Functions Depending on Fewer Variables. Proc. Stek. Inst. 1987, Issue 4.

[28] Tyrtyshnikov, E.E.: Tensor approximations of matrices generated by asymptotically smooth functions. Sbor. Math. (2003) **194**, No.6, 147-160.