JOURNAL OF APPROXIMATION THEORY **51**, 54–67 (1987)

On Nonuniqueness in Nonlinear L_2-Approximation

IMMO DIENER*

Institut für Numerische und Angewandte Mathematik, University of Göttingen, D-3400 Göttingen, West Germany

Communicated by E. W. Cheney

Received March 7, 1985; revised October 10, 1985

It is shown that, under weak assumptions, nonlinear L_2-approximation problems generally have unbounded numbers of local best approximations. This includes the rational and the exponential families of approximating functions. In addition, for a certain class of approximating families, we construct functions with three global best approximations. The results apply, for instance, to exponential and rational approximating families with one nonlinear parameter. Finally, we extend results of Spiess and Braess on the finiteness of the number of local best approximations by rational functions.

1. GENERAL NONLINEAR L_2-APPROXIMATION

In this section we show that general nonlinear L_2-approximation problems have unbounded numbers of local best approximations. This extends results of Wolfe [7] for the special case of ordinary rational functions. We consider the Hilbert space $H := L_2[-1, 1]$. Let S be an open subset of $\mathbf{R}^M$ and A a twice Fréchet-differentiable map from S to H. Thus, elements of H are to be approximated by elements of $A(S) = \{A(x) \mid x \in S\}$. The first and second Fréchet-derivative of a transformation g at a point x will be denoted by $g'(x, \cdot)$ and $g''(x, \cdot, \cdot)$ respectively. For a function f out of H the functional $N_f(x)$ from S to $\mathbf{R}$ is defined by $N_f(x) := \|A(x) - f\|^2$. Then this functional is twice differentiable with respect to x. The span of vectors $x_1, \ldots, x_n$ is denoted by $\langle x_1, \ldots, x_n \rangle$.

In general a local minimum of $N_f(\cdot)$ at x_0 does not imply that $A(x_0)$ is a local best approximation to f with respect to $A(S)$. For this, some further conditions must be met, especially $A(x_0)$ must be normal. This condition is even sufficient in most cases. The interested reader may consult Wolfe's paper [7] for further details. He defines normality as follows:

* This paper summarizes the author's thesis [4].

0021-9045/87 $3.00

DEFINITION. A point $A(x) \in A(S)$ is called normal if A^{-1} exists on a neighborhood of $A(x)$, is continuous at $A(x)$, and $A'(x, \cdot)$ is one-to-one.

Now we want to show how for a given mapping A with certain properties one can get a result of the following general structure:

1.1. The number of isolated local best approximations to a function $f \in H$ cannot be bounded independently of f: For any natural number q there are distinct parameters $P_\nu \in S$, $\nu = 1,..., q$ and a function $f \in H$ such that $A(P_1),..., A(P_\nu)$ are isolated, local best approximations to f with respect to the approximating family $A(S)$.

We need a lemma of Wolfe [7],

1.2. LEMMA *Let M_i, $i = 1, 2,...$ be a sequence of finite dimensional subspaces of a Hilbert space H such that*

$$M_i \cap \left(\sum_{j=1}^{i-1} M_j \right) = \langle 0 \rangle \tag{1.3}$$

for all $i = 2, 3,...$. *Let $r_i \in M_i$ be given for $i = 1, 2,...$ where $r_i \neq 0$. Then for each $n \in N$,*

$$L_n := \bigcap_{i=1}^{n} (r_i + M_i^\perp)$$

is nonvoid.

We use this to show

1.4. LEMMA. *Let $P_\nu \in S$, $\nu = 1,..., q$ be parameters, satisfying $A(P_\nu) \neq 0$ for all $\nu = 1,..., q$. Let M_ν be linear finite dimensional subspaces of $L_2[-1, 1]$. Suppose that $A'(P_\nu, \cdot)$ is injective and*

$$A(P_\nu),\ A'(P_\nu, h),\ A''(P_\nu, h, h) \in M_\nu \tag{1.5}$$

for $\nu = 1,..., q$, and all $h \in \mathbf{R}^M$. Then, if the M_ν satisfy (1.3), *there is a function $f \in L_2[-1, 1]$ such that the functional $N_f(\cdot)$ has isolated local minima at the points $P_1,..., P_q$.*

Proof. Under the above hypotheses $L_q := \bigcap_{i=1}^{q} (A(P_i) + M_i^\perp)$ is not empty. Let f be a function in L_q. Then for any $h \in \mathbf{R}^M$ and any $\nu = 1,..., q$ we obtain

$$\tfrac{1}{2} N_f'(P_\nu, h) = (A(P_\nu) - f, A'(P_\nu, h)) = 0,$$

since f has a representation $f = A(P_\nu) - g_\nu$, with $(g_\nu, A'(P_\nu, h)) = 0$. Furthermore, for any $h \in \mathbf{R}^M$ and any $\nu = 1, \ldots, q$, we have

$$\tfrac{1}{2} N_f''(P_\nu, h, h) = \|A'(P_\nu, h)\|^2 + (A(P_\nu) - f, A''(P_\nu, h, h)).$$

The last term vanishes since f can be written in the form $f = A(P_\nu) + \hat{g}_\nu$, with a function $\hat{g}_\nu$ that satisfies $(\hat{g}_\nu, A''(P_\nu, h, h)) = 0$, and the first term is positive definite since $A'(P_\nu)$ is injective.

To apply this lemma one has to find suitable subspaces M_ν. One obvious way is to consider the subspaces generated by $A(P_\nu)$, $A'(P_\nu, h)$, and $A''(P_\nu, h, h)$. So we define for $P_\nu := (x_1^\nu, \ldots, x_M^\nu) \in S$ the spaces M_ν by

$$M_\nu := \left\langle A(P_\nu), \frac{\partial A}{\partial x_\mu}(P_\nu), \frac{\partial^2 A}{\partial x_\mu \, \partial x_{\boldsymbol{\mu}}}(P_\nu) \,\middle|\, \mu, \boldsymbol{\mu} = 1, \ldots, M \right\rangle.$$

These M_ν obviously satisfy (1.5) and we have

$$d_\nu := \dim M_\nu \leqslant 1 + M + \binom{M}{2} + M = \binom{M+2}{2}.$$

Let $\{\omega_j^\nu\}_{j=1}^{d_\nu}$ be a basis for M_ν. Then (1.3) will be satisfied if the functions

$$\omega_j^\nu; \qquad \nu = 1, \ldots, q; \qquad j = 1, \ldots, d_\nu \tag{1.6}$$

are linearly independent over $\mathbf{R}$.

So in general, to **prove 1.1** for a given A it is sufficient to find for any q certain parameters $P_\nu \in S$, $\nu = 1, \ldots, q$, such that $A(P_\nu)$ is normal and the functions (1.6) are linearly independent over $\mathbf{R}$. This can be done in many cases. We exhibit some examples:

1.7. Example. Set $S := \mathbf{R} \times \mathbf{R}^+ \times (0, 2\pi)$, $A(x_1, x_2, x_3) := x_1 \sin(x_2 t + x_3)$. One obtains that $A'(x)$ is injective iff $x_1 \neq 0$. For such parameters we have $d_\nu = 5$ and the functions

$$\omega_1^\nu = \sin(x_2^\nu t + x_3^\nu), \qquad \omega_2^\nu = \cos(x_2^\nu t + x_3^\nu),$$
$$\omega_3^\nu = t\omega_1^\nu, \qquad \omega_4^\nu = t\omega_2^\nu, \qquad \omega_5^\nu = t^2 \omega_1^\nu$$

constitute a basis for M_ν. If $x_1^\nu \neq 0$ for $\nu = 1, \ldots, q$ and $x_2^\nu \neq x_2^{\boldsymbol{\nu}}$ for $\nu \neq \boldsymbol{\nu}$, these functions are linearly independent over $\mathbf{R}$.

1.8. Example. Set $S := \mathbf{R}$ and, $A(x_1) = e^{x_1 t} + x_1 t$. $A'(x_1)$ is injective for all $x_1 \in S$ and we have $d_\nu = 3$. A basis is given by

$$\omega_1^\nu = te^{x_1^\nu t} + t, \qquad w_2^\nu = t^2 e^{x_1^\nu t}, \qquad \omega_3^\nu = A(x_1^\nu).$$

Thus, the functions (1.6) are linearly independent, iff all x_1^ν, $\nu = 1,..., q$ are distinct.

1.9. EXAMPLE. Set $S := \mathbf{R}^2$ and $A(x_1, x_2) := e^{x_1 t} + x_2 t$. Then $A'(x)$ is injective, iff $x_1 \neq 0$. We obtain $d_\nu = 4$ and the functions

$$\omega_1^\nu = e^{x_1^\nu t}, \quad \omega_2^\nu = te^{x_1^\nu t}, \quad \omega_3^\nu = t^2 e^{x_1^\nu t}, \quad \omega_4^\nu = t$$

constitute a basis for the M_ν. But we have $M_\nu \cap M_{\mathbf{v}} \supset \langle t \rangle$. So (1.3) is not satisfied and the above construction is not applicable. However, the difficulty in this case (and in similar ones) can be overcome by using the same trick Braess used in [1] to prove his generalization of Wolfe's Theorem 6 in [7]. This shall not be further amplified here.

1.10. *Remark.* Lemma 1.4 equally applies to the discrete case, i.e., if we take $\mathbf{R}^N$ as the Hilbert space H. In this case q can of cause not be arbitrarily large, because (1.3) can only hold if q is smaller than the dimension of H.

1.11. EXAMPLE. As an example we report without proof a result one obtains by applying 1.4 to discrete approximation in the family $\mathscr{R}_m^n$ of polynomial rational functions with the Euclidean norm:

Let D be a discrete set of N nodes and for $0 \leqslant n < m$ suppose $N^* \leqslant (N + m - n - 1)/3m$. For $\nu = 1,..., N^*$ let $r_\nu = p_\nu/q_\nu$ be rational functions in $\mathscr{R}_m^n(D)$. Suppose the r_ν are normal and all q_ν are of degree m. Suppose further, that q_ν and $q_{\mathbf{v}}$ have no common factors unless $\nu = \mathbf{v}$. Then there exists a function $f \in C(D)$ such that the r_ν for $\nu = 1,..., N^*$ are isolated local best approximations to f with respect to the Euclidean norm.

1.12. *Remark.* Note that the function f in (1.11) may not have a best approximation in $\mathscr{R}_m^n(D)$.

1.13. *Remark.* If P_ν, $\nu = 1,..., q$ are points from S such that the functions (1.6) are linearly independent, one can repeat the argument in the Hilbert space

$$H := \bigoplus_{\nu=1}^{q} M_\nu \subset L_2[-1, 1].$$

Then f can be choosen from H. It then has the form

$$f = \sum_{1 \leqslant \nu \leqslant q} \alpha_\nu A(P_\nu) + \sum_{\substack{1 \leqslant \mu \leqslant M \\ 1 \leqslant \nu \leqslant q}} \beta_{\nu\mu} \frac{\partial A}{\partial x_\mu}(P_\nu) + \sum_{\substack{1 \leqslant \mu, \boldsymbol{\mu} \leqslant M \\ 1 \leqslant \nu \leqslant q}} \gamma_{\nu\mu\boldsymbol{\mu}} \frac{\partial^2 A}{\partial x_\mu \partial x_{\boldsymbol{\mu}}}(P_\nu)$$

with certain real coefficients α_ν, $\beta_{\nu\mu}$, $\gamma_{\nu\mu\boldsymbol{\mu}}$.

2. The Rational and Exponential Case

It would be desirable to formulate 1.1 more precisely, but in general this turns out to be rather technical and tedious. So we confine ourselves to a special case which includes general rational and exponential approximation.

Notation. Suppose X and $Y = \{\psi_1, ..., \psi_m\}$, respectively, are n- and m-sets from $C[-1, 1]$ and 0 is not in X. We say Y is linearly independent over X, iff

$$\sum_{i=1}^{m} \alpha_i \phi_i \psi_i \equiv 0 \quad \text{on} \quad [-1, 1], \qquad \text{and} \qquad \phi_i \in X,\ \alpha_i \in \mathbf{R}$$

implies $\alpha_i = 0$ for $i = 1, ..., m$.

2.1. Theorem. *Let I be a compact real interval and let s be a fixed real number. Let the kernel $\hat{\gamma}$ be defined by*

$$\hat{\gamma}(x) := \begin{cases} (1 + (1-s)\,x)^{1/(1-s)} & \text{for} \quad s \neq 1; \\ e^x, & \text{for} \quad s = 1. \end{cases}$$

Let ψ_0 and φ_0 denote the constant Function 1 on I, and let $\{\psi_\nu\}_{\nu=1}^{m}$, and $\{\varphi_\mu\}_{\mu=1}^{n}$ be two systems of linearly independent functions from $C(I)$. Define two affine–linear functions l_1 and l_2 from $\mathbf{R}^n$ and $\mathbf{R}^m$, respectively, to $C[I]$ by

$$l_1(x) = \sum_{\mu=1}^{n} x_\mu \varphi_\mu + c_1 \varphi_0$$

$$l_2(y) = \sum_{\nu=1}^{m} y_\nu \psi_\nu + c_2 \psi_0.$$

Using $B_s := \{y \in \mathbf{R}^m \mid 1 + (1-s)\, l_2(y) > 0\}$, we consider the open and convex subset $S := \mathbf{R}^{nN} \times B_s^N$ of $\mathbf{R}^{N(n+m)}$ for a fixed natural number N. For $y \in B_s$, set $\gamma(y) := \hat{\gamma}(l_2(y))$ and for each parameter $P := (x^{1^T}, ..., x^{N^T}, y^{1^T}, ..., y^{N^T}) \in S$ define the approximating function by

$$A(P) := \sum_{k=1}^{N} l_1(x^k)\, \gamma(y^k).$$

Let the linear space L be defined by

$$L := \langle \varphi_\mu \psi_\nu \psi_{\mathbf{v}} \mid \mu = 0, ..., n;\ \nu, \mathbf{v} = 0, ..., m \rangle.$$

Suppose, that for $\omega = 1, ..., q \in \mathbf{N}$, certain points $P_\omega := (x_\omega^1, ..., x_\omega^N, y_\omega^1, ..., y_\omega^N) \in S$ are given such that for $\omega = 1, ..., q$ the functions $A'(P_\omega)$ are *injective,* $A(P_\omega) \neq 0$, *and the functions* $\gamma^{2s-1}(y_\omega^k)$ *for* $\omega = 1,\dots, q$ *and* $k = 1,\dots, N$ *are linearly independent over L.*

Then there exists a function f from

$$\bigoplus_{\omega,k} L \cdot \gamma^{2s-1}(y_\omega^k) \subset L_2(I),$$

such that the functional $N_f(\cdot)$ *has isolated local minima at the points* $P_1,\dots, P_q$.

Before we prove this theorem, we shall consider some examples.

2.2. EXAMPLE (Approximation with Rational Functions). We show how to get Wolfe's theorem [7, Theorem 6] from 2.1. We get $s = 2$, c_1, $c_2 = 0$, $N = 1$, $n = n_{\text{Wolfe}} + 1$, $m = m_{\text{Wolfe}}$, $\varphi_\mu = t^{\mu-1}$, $\psi_\nu = -t^\nu$. We get $B_s = \{y \in \mathbf{R}^m \mid \sum_{\nu=1}^m y_\nu t^\nu + 1 > 0\}$ and for $P = (x_1,\dots, x_n, y_1,\dots, y_m)$ we obtain

$$A(P) = \frac{x_1 + x_2 t + \cdots + x_n t^{n-1}}{1 + y_1 t + \cdots + y_m t^m} \in \mathscr{R}_m^{n-1}$$

For L we get

$$L = \langle t^{\mu-1} t^\nu t^{\mathbf{v}} \mid \mu = 1,\dots, n; \nu, \mathbf{v} = 0,\dots, m\rangle = \mathscr{P}_{2m+n-1}$$

and we have $\gamma^{2s-1}(y) = (1 + y_1 t + \cdots + y_m t^m)^{-3}$. The prepositions of 2.1 are surely satisfied if $A(P_\omega)$ is normal for $\omega = 1,\dots, q$ (then $A'(P_\omega)$ is injective), the denominators are relatively prime and have full degree m, and if L contains no polynomial of degree $\geqslant 3m$. This means that $2m + n - 1 < 3m$. These are exactly the premises that Wolfe needed for his theorem. Moreover, we can say something about the form of f: If the Q_i are the given denominator polynomials, then f can be chosen from $\bigoplus_{i=1}^q (\mathscr{P}_{2m+n-1})/Q_i^3$.

2.3. EXAMPLE (Exponential Approximation). We set $s = n = m = 1$, $\varphi_1 \equiv 1$, $\psi = \mathrm{id}_I$, $c_1 = c_2 = 0$ to get $S = \mathbf{R}^N \times \mathbf{R}^N$, $L = \mathscr{P}_2$ and

$$A(P) = \sum_{k=1}^N x^k e^{y^k t}.$$

If all $x_\mu \neq 0$ and all y_ν are distinct, then $A'(P)$ is injective. The functions $\gamma^{2s-1}(y_\omega^k) = e^{y_\omega^k t}$ are linearly independent over $L = \mathscr{P}_2$, if the parameters y_ω^k are distinct for $k = 1,\dots, N$ and $\omega = 1,\dots, q$.

2.4. EXAMPLE. Set $I = [a, b]$, $0 < a < b < \infty$, $s = m = n = N = 1$, $\varphi_1 \equiv 1$, $\psi_1 = \ln t$, $S = \mathbf{R} \times \mathbf{R}$, $L = \langle 1, \ln t, (\ln t)^2\rangle$ and

$$A(x, y) = xt^y$$

$A'(x, y)$ is injective, iff x does not vanish. The functions $\gamma^{2s-1}(y^k_\omega) = t^{y_\omega}$ are linearly independent over L, iff all y_ω are pairwise distinct.

2.5. *Proof of* 2.1. We consider the linear subspaces

$$M_\omega := \bigoplus_{k=1}^{N} L \cdot \gamma^{2s-1}(y^k_\omega)$$

of the Hilbert space $H := \bigoplus_{\omega=1}^{q} M_\omega$. It follows from the linear independence of the functions $\gamma^{2s-1}(y^k_\omega)$ over L that these sums are direct and the M_ω satisfy condition (1.3). In order to apply 1.4, we need to show that

$$A(P_\omega),\ A'(P_\omega, h),\ A''(P_\omega, h, h) \in M_\omega$$

for all $h \in \mathbf{R}^{N(n+m)}$. We have

$$M_\omega = \langle \varphi_\mu, \psi_\nu, \psi_{\mathbf{v}} \gamma^{2s-1}(y^k_\omega) \mid 1 \leqslant k \leqslant N, 0 \leqslant \mu \leqslant n, 0 \leqslant \nu, \mathbf{v} \leqslant m \rangle.$$

Then $0 \neq A(P_\omega) \in M_\omega$. Corresponding to $P := (x^1, ..., x^N, y^1, ..., y^N)$ we also partition h according to $h := (u^1, ..., u^N, v^1, ..., v^N)$ and define

$$\hat{l}_1(x) := l_1(x) - c_1 \varphi_0$$
$$\hat{l}_2(y) := l_2(y) - c_2 \psi_0.$$

Then

$$\begin{aligned}
A(P+h) - A(P) &= \sum_{k=1}^{N} l_1(x^k + u^k)\, \gamma(y^k + v^k) - \sum_{k=1}^{N} l_1(x^k)\, \gamma(y^k) \\
&= \sum_{k=1}^{N} \left(l_1(x^k) + \hat{l}_1(u_k) \right) \hat{\gamma}\left(l_2(y^k) + \hat{l}_2(v^k) \right) - \sum_{k=1}^{N} l_1(x^k)\, \gamma(y^k) \\
&= \sum_{k=1}^{N} \left(l_1(x^k) + \hat{l}_1(u_k) \right) \Big(\gamma(y^k) + \hat{\gamma}'(l_2(y^k))\, \hat{l}_2(v^k) \\
&\quad + \frac{1}{2} \hat{\gamma}''(l_2(y^k))\, \hat{l}_2(v^k)^2 + O(\|v^k\|^3) \Big) - \sum_{k=1}^{N} l_1(x^k)\, \gamma(y^k) \\
&= \sum_{k=1}^{N} \left(l_1(x^k)\, \hat{\gamma}'(l_2(y^k))\, \hat{l}_2(v^k) + \hat{l}_1(u^k)\, \gamma(y^k) \right) \\
&\quad + \sum_{k=1}^{N} \left(\hat{\gamma}'(l_2(y^k))\, \hat{l}_2(v^k)\, \hat{l}_1(u^k) + \frac{1}{2} l_1(x^k)\, \hat{\gamma}''(l_2(y^k))\, \hat{l}_2(v^k)^2 \right) \\
&\quad + O(\|h\|^2).
\end{aligned}$$

Because of the identities

$$\frac{d}{dx}\hat{\gamma}(x)=\hat{\gamma}^s(x),\ \frac{d^2}{dx^2}\hat{\gamma}(x)=s\gamma^{2s-1}(x)$$

we obtain

$$A'(P,h)=\sum_{k=1}^{N}\left(l_1(x^k)\,\gamma(y^k)\,\hat{l}_2(v^k)+\gamma(y^k)\,\hat{l}_1(u^k)\right)$$

$$A''(P,h,h)=\sum_{k=1}^{N}\left(\gamma^s(y^k)\,\hat{l}_1(u^k)\,\hat{l}_2(v^k)+\frac{s}{2}\gamma^{2s-1}(y^k)\,\hat{l}_2(v^k)^2\right).$$

From this the theorem follows immediately.

3. Nonuniqueness of Global Best Approximations

Several authors have published examples for functions with more than one global best L_2-approximation (g.b.a.). In the approximation with ordinary rational functions on a compact real interval, one uses a symmetry argument and the fact, that best approximations are normal [2]. We shall give a more general argument that works for nonsymmetric cases too, and gives three best approximations in symmetric cases.

Consider the L_2-approximation in $\mathscr{R}^n_m[-1,1]$. Local best approximations to functions $f\notin\mathscr{R}^{n-1}_{m-1}$ are always normal [3] and the subset of normal functions in $\mathscr{R}^n_m$ is not connected for $n\geqslant 0$.

3.1. Theorem. *Let C_0 and C_1 be two connected components of $\mathscr{R}^n_m\setminus\mathscr{R}^{n-1}_{m-1}$. Suppose f_0 and f_1 are functions in $L_2[-1,1]$ such that f_0 has a* g.b.a. *in C_0 and f_1 has a* g.b.a. *in C_1. Let $H\colon[0,1]\to L_2[-1,1]\setminus\mathscr{R}^{n-1}_{m-1}[-1,1]$ be a continuous curve from f_0 to f_1. Then there exists $\lambda^*\in[0,1]$, such that $H(\lambda^*)$ has two global best approximations in $\mathscr{R}^n_m$.*

Proof. Let C_2 be the set of normal functions from $\mathscr{R}^n_m$ that are neither in C_0 nor in C_1. Let

$$E(\lambda):=\inf_{r\in\mathscr{R}^n_m}\|H(\lambda)-r\|$$

be the distance from $H(\lambda)$ to $\mathscr{R}^n_m$ and

$$E_i(\lambda):=\inf_{r\in C_i}\|H(\lambda)-r\|,\qquad i=0,1,2$$

be the distance from $H(\lambda)$ to C_i. The functions E, E_i are continuous on $[0, 1]$ and for all $\lambda \in [0, 1]$ we have

$$\min\{E_0(\lambda), E_1(\lambda), E_2(\lambda)\} = E(\lambda).$$

For $\lambda = 0$ the minimum is attained at E_0 and for $\lambda = 1$ at E_1. Thus there exists $\lambda^* \in [0, 1]$ with $E_j(\lambda^*) = E_k(\lambda^*) = E(\lambda^*)$ and $j \neq k$. Since best approximations are normal, the corresponding best approximations must lie in C_j and C_k.

Using this result we can eliminate the restriction to odd r in [2, Theorem 4.1]:

3.2. THEOREM. *Let $n \geqslant 0$, $m \geqslant 1$. To posess only one local best approximation in $\mathscr{R}^n_m[-1, 1]$ is no generic property in $L_2[-1, 1]$.*

For the rest of this chapter we use $I := [-1, 1]$, $S := (-1, 1)$. Let $u: S \times I \to \mathbf{R}^+$ be a continnuous kernel function, positive on $S \times I$, such that zero is the only function orthogonal to every function in $u(S, \cdot)$. Let ϕ_i, $f: I \to \mathbf{R}^N$, $i = 1, 2, \ldots, N$, be linearly independent functions on I such that $\phi_1(t) f(t) \not\equiv 0$ on I. We define the approximating family by a mapping $A: S \times \mathbf{R}^N \to C[I]$ where

$$A(b, x) := u(b, \cdot) \sum_{i=1}^{N} x_i \phi_i.$$

For simplicity (although we do not need that much) we require that $b \mapsto A(b, \cdot)$ is injective and that any continuous function f on I has a best L_2-approximation in $A(S \times \mathbf{R}^N)$.

3.3. THEOREM. *Under the assumptions stated above, define for any $\lambda \in \mathbf{R}^N$ a continuous function f_λ by*

$$f_\lambda := \frac{1}{1 + \|\lambda\|} \left(f - \sum_{i=1}^{N} \lambda_i \phi_i \right).$$

Then

(i) *There exists a vector $\tilde{\lambda} \in \mathbf{R}^N$ such that $f_{\tilde{\lambda}}$ has two global best approximations in $A(S \times \mathbf{R}^N)$.*

(ii) *If $u(b, t) \equiv u(-b, -t)$ and the functions f, ϕ_i are even, then there exists a vector $\tilde{\lambda} \in \mathbf{R}^N$ such that $f_{\tilde{\lambda}}$ has three global best approximations in $A(S \times \mathbf{R}^N)$.*

Proof. 1. Zero is not a best approximation to any $f \not\equiv 0$ because otherwise $\omega(t) := \|t u(b, \cdot)\, \phi_1 - f\|^2$ would, for all $b \in S$, have a minimum at

$t=0$. But then $\phi_1 f$ would be orthogonal to all $u(b,\cdot)$, $b \in S$, contradicting the assumption that $\phi_1 f$ does not vanish identically and that zero is the only function orthogonal to all $u(b,\cdot)$, $b \in S$.

2. For $b \in S$ let $\lambda^*(b)$ be the solution of the linear system $(u(b,\cdot)\,\phi_j, f_{\lambda^*(b)})=0$, $j=1,\ldots,N$. Then λ^* exists, is uniquely determined and depends continuously on b, because the matrix of this system is a Gramian matrix of the linearly independent functions ϕ_i with respect to the positive weight function $u(b,\cdot)$. For any $b \in S$ the minimum of $\|A(b,x)-f_{\lambda^*(b)}\|^2$ where x ranges over $\mathbf{R}^N$ is attained for $x=0$.

3. Suppose that for any $\lambda \in \mathbf{R}^N$ there exists exactly one $b^*(\lambda) \in S$, such that $A(b^*(\lambda), x^*)$ is the g.b.a. to f_λ. Then b^* depends continuously on λ. But then $b^*(\mathbf{R}^N)$ is contained in a compact subinterval of S. To see this, assume that there exists a sequence $\{\lambda^\nu\}$ of vectors in $\mathbf{R}^N$ such that $b^*(\lambda^\nu)$ converges to $+1$ or -1 as ν goes to infinity. By choosing an appropriate subsequence if necessary, we may suppose that $\lambda^\nu/(1+\|\lambda^\nu\|)$ converges to a vector $\hat{\lambda} \in \mathbf{R}^N$. If $\|\lambda^\nu\|$ converges to $q \in \mathbf{R} \cup \{\infty\}$, then f_{λ^ν} converges to the continuous function

$$f_\infty := \frac{f}{1+q} + \sum_{i=1}^{N} \hat{\lambda}_i \phi_i$$

which does not vanish identically on I. Thus f_∞ has a unique g.b.a. $A(b^\infty, x^\infty)$ with $b^\infty \in S$ and $x^\infty \neq 0$.

4. Thus, if the g.b.a. to any f_λ is unique, the mapping $b^* \circ \lambda^*$ is continuous and maps S into a compact subset of S. It follows that $b^* \circ \lambda^*$ has a fixed point $\hat{b} \in S$. By construction zero would be a g.b.a. to $f_{\lambda^*(\hat{b})}$. This function, however, does not vanish identically, and this is a contradiction to part 1 of this proof. Then b^* does not depend continuously on λ and so there exists a $\tilde{\lambda} \in \mathbf{R}^N$ such that $f_{\tilde{\lambda}}$ has two g.b.a. in $A(S \times \mathbf{R}^N)$.

5. To prove (ii) we note that under the symmetry assumptions made, $A(b,x)$ is a g.b.a. to f_λ if and only if $A(-b,x)$ is also a g.b.a. to f_λ. Now we repeat the arguments 1 to 4 but restrict b to the interval $[0,1)$. It follows that there is a $\tilde{\lambda} \in \mathbf{R}^N$, such that $f_{\tilde{\lambda}}$ has two g.b.a. in $A([0,1) \times \mathbf{R}^N)$. One of them might have $b^*(\tilde{\lambda})=0$, but the other one can be paired with a g.b.a. in $A((-1,0) \times \mathbf{R}^N)$.

Obviously we can take for $u(b,t)$ the kernels e^{bt} and $(1-bt)^{-1}$ to obtain results for exponential or rational approximation and thus answering questions of Braess [1, 2].

3.4. Example. The following example has already been considered by Lamprecht [5]; he did however not give a rigorous proof. We consider L_2-approximation in $\mathscr{R}_1^0[-1,1]$ and choose $f(t) \equiv t^2$, $\phi_1(t) \equiv 1$. The theorem implies that there exists $\tilde{\lambda} \in \mathbf{R}$, such that $f_{\tilde{\lambda}} = t^2 - \tilde{\lambda}$ has three g.b.a. in $\mathscr{R}_1^0[-1, 1]$. A closer analysis shows that $\tilde{\lambda} \in (\frac{1}{3}, 1)$.

4. The Number of Critical Points in Rational L_2-Approximation

It is still an open problem whether for some $f \in L_2[-1, 1]$ the function $\|f - \cdot\|^2$ has infinitely many (local) minima in $\mathscr{R}_m^n$. Finiteness was reported for the case $m = 1$ in [7], but this information was based on a linguistic translation error of a result by Spiess [6]. We parametrize $\mathscr{R}_1^0[-1, 1]$ by $a \cdot u(b, \cdot)$ where $u(b, t) := \sqrt{(1 - b^2)/2}\,(1 + bt)^{-1}$ is the normalized kernel. Spiess proved that for $f \in L_2[-1, 1]$ the critical points of $N_f(a, b)$: $\|a \cdot u(b, \cdot) - f(\cdot)\|^2$ cannot accumulate at points (a^*, b^*) with $b^* \in (-1, 1)$.

Later Braess [1] showed that the number of critical points in $\mathscr{R}_1^0$ is finite if f is continuous on $[-1, 1]$ and $f(1) \cdot f(-1)$ is nonzero. In this section a much stronger result is obtained.

With the above parametrisation (a, b) is critical iff

$$a - (u(b), f) = 0 \qquad \text{and} \qquad a \cdot (u'(b), f) = 0.$$

We can restrict the computations to the case $a \neq 0$ and thus have to prove that $(u'(b), f)$ has only finitely many zeros for $b \in (-1, 1)$. We obtain

$$\frac{\partial u}{\partial b}(b, t) = \frac{-1}{\sqrt{2}\sqrt{1 - b^2}} \frac{b + t}{(1 + bt)^2}.$$

Thus the possible values for b are the zeros of F: $(-1, 1) \to \mathbf{R}$, defined by

$$F(b) := \int_{-1}^{1} f(t) \frac{t + b}{(1 + bt)^2}\, dt$$

4.1. Lemma. *Let f be a continuous function on $[-1, 1]$. Suppose $f(-1) \neq 0$ or there exists a constant $\delta \in (0, 1)$ such that f is monotonous in the interval $[-1, -1 + \delta]$. Then F: $(-1, 1) \to \mathbf{R}$, defined by*

$$F(b) := \int_{-1}^{1} f(t) \frac{t + b}{(1 + bt)^2}\, dt$$

has only finitely many zeros in the interval $(0, 1)$.

Proof. We may suppose $f(-1) = 0$ because the other case is covered by the result of Braess in [1] cited above. F is analytic on $(-1, 1)$. Thus, if the zeros of F had an accumulation point in $(-1, 1)$, F would be identically zero. This cannot happen for $f \in L_2[-1, 1]$ as has been shown by Spiess in [6]. Suppose the zeros accumulate at 1. Then, for each $n = 0, 1,\ldots$ a sequence b_i^n, $i = 0, 1,\ldots$ exists such that $b_i^n \to 1$ for $i \to \infty$ and $F^{(n)}(b_i^n) = 0$ for all i, $n \geqslant 0$. Let $K(b, t)$ denote the kernel

$$K(b, t) := \frac{t+b}{(1+bt)^2}.$$

For $n \geqslant 0$ induction gives

$$K_n(b, t) := \frac{\partial^n}{\partial b^n} K(b, t) = (-1)^n \frac{n!t^{n-1}((n+1)t^2 + bt - n)}{(1+bt)^{n+2}}$$

We now split $F^{(n)}(b)$ into two terms

$$F^{(n)}(b) = \left(\int_{-1}^{-1+\delta} + \int_{-1+\delta}^{1}\right)(f(t)\, K_n(b, t)\, dt)$$

and estimate roughly:

$$\begin{aligned}
\left\| \int_{-1+\delta}^{1} f(t)\, K_n(b, t)\, dt \right\|_{\infty, [0,1]} &\leqslant \|f\|_\infty \cdot \left\| \int_{-1+\delta}^{1} K_n(b, t)\, dt \right\|_{\infty, [0,1]} \\
&\leqslant \|f\|_\infty \cdot 2 \cdot \sup_{b \in (0,1)} \sup_{t \in [-1+\delta, 1]} |K_n(b, t)| \\
&\leqslant 2\|f\|_\infty |K_n(1, -1+\delta)| \\
&\leqslant 2\frac{(n+1)!}{\delta^{n+2}} \|f\|_\infty
\end{aligned}$$

Thus we have

$$F^{(n)}(b) \geqslant -2\|f\|_\infty \frac{(n+1)!}{\delta^{n+2}} + \int_{-1}^{-1+\delta} f(t)\, K_n(b, t)\, dt.$$

Now, $K_n(b, -1) = -(n!/(1-b)^{n+1})$ is less than zero and $K_n(b, \cdot)$ has exactly one zero in $(-1, 0)$, namely

$$z_n(b) := -\frac{1}{2n+2}(b + \sqrt{(2n+1)^2 - (1-b)}).$$

We find $z_n(b) \to -1$ for $b \to 1$ or $n \to \infty$. We may suppose, that f increases monotonously on $[-1, -1, +\delta]$. Thus for $z_n(b) \in (-1, -1+\delta)$ we get

$$\int_{-1}^{-1+\delta} f(t)\, K_n(b, t)\, dt \geqslant f(z_n(b)) \int_{-1}^{-1+\delta} K_n(b, t)\, dt.$$

The integral can asymptotically be estimated by

$$\int_{-1}^{-1+\delta} K_n(b, t)\, dt = \frac{\partial^n}{\partial b^n} \left\{ \ln(1-b+\delta b) - \ln(1-b) - \frac{\delta b(1+b)}{1-b+\delta b} \right\}$$

$$= O((1-b)^{-n})$$

for fixed $n \geqslant 1$ and $b \to 1$. So we obtain

$$F^{(n)}(b) \geqslant -2\|f\|_\infty \frac{(n+1)!}{\delta^{n+2}} + f(z_n(b))\, O((1-b)^{-n})$$

for $b \to 1$. Since $F^{(n)}(b_i^n) = 0$ we must have

$$f(z_n(b_i^n))\, O((1-b_i^n)^{-n}) < 2\|f\|_\infty \frac{(n+1)!}{\delta^{n+2}}$$

for $i \to \infty$. This implies $f(z_n(b)) \leqslant O((1-b)^{-n})$ for $b \to 1$. Because of $z_n(b) > -1 + (1-b)/(2n+2)$ this means we have $f(t) = O((1+t)^n)$ for $t \to -1$. So $f(t)$ vanishes faster than any polynomial for $t \to -1$. But then

$$c_n := \lim_{b \to 1} \int_{-1}^{-1+\delta} f(t)\, K_n(b, t)\, dt$$

exists for all n and this limit must be smaller than $2\|f\|_\infty (n+1)!/\delta^{n+2}$. We now show that under these circumstances F can be analytically continued beyond $b = 1$ and we again obtain a contradiction to the result of Spiess. We expand F into a series around $b = 1$ and show that this series has a positive radius of convergence:

$$F(b) := \sum_{n=0}^{\infty} \frac{(b-1)^n}{n!} \int_{-1}^{+1} K_n(1, t)\, f(t)\, dt.$$

The coefficients can be estimated by

$$\left| \int_{-1}^{+1} K_n(1, t)\, f(t)\, dt \right| = \left| \int_{-1}^{-1+\delta} + \int_{-1+\delta}^{+1} \right| \leqslant 4\|f\|_\infty \frac{(n+1)!}{\delta^{n+2}}.$$

Thus the series has radius of convergence at least δ.

Likewise F has only finitely many zeros in $(-1, 0)$, provided f is monotonous at the right end of $[-1, 1]$. Thus we obtain

4.2. THEOREM. *Let f be a continuous function on $[-1, 1]$. Suppose that for $x = +1$ and $x = -1$ either $f(x) \neq 0$ or there exists a neighbourhood $U(x)$ of x such that f is monotonous in $U(x) \cap [-1, 1]$. Then the function $\|f - \cdot\|^2$ has only finitely many critical points in $\mathscr{R}_1^0$.*

References

1. D. Braess, On Rational L_2-Approximation, *J. Approx. Theory* **18** (1976), 136–151.
2. D. Braess, "Über das Anzahlproblem bei der rationalen L_2-Approximation, *in*: Numerische Methoden der Approximationstheorie, Vol. 3" (L. Collatz, G. Meinardus, and H. Werner, Eds.), Birkhäuser Verlag, Basel and Stuttgart 109–121, 1979.
3. E. W. Cheney and A. A. Goldstein, Mean-square approximation by generalized rational functions, *Math. Z.* **95** (1967), 232–241.
4. I. Diener, "Zur Mehrdeutigkeit bei der nichtlinearen L_2-Approximation," dissertation, Göttingen, 1983.
5. G. Lamprecht, Mehrdeutigkeit bei der Approximation in der L_p-Norm mit Hilfe rationaler Funktionen, *Computing* **5** (1970), 349–355.
6. J. Spiess, "Eindeutigkeitssätze bei der nichtlinearen Approximation in strikt-konvexen Räumen," dissertation, Hamburg, 1969.
7. J. M. Wolfe, On the Unicity of Nonlinear Approximation in Smooth Spaces, *J. Approx. Theory* **12** (1974), 165–181.