JOURNAL OF APPROXIMATION THEORY **51**, 19–26 (1987)

Varying Norms and Constraints

CHARLES B. DUNHAM

Department of Computer Science,
University of Western Ontario, London, Ontario, Canada N6A 5B7

Communicated by E. W. Cheney

Received August 20, 1984; revised April 3, 1985

The effects of perturbing norm and constraints in linear approximation are considered. © 1987 Academic Press, Inc.

Let $\phi_1,\ldots,\phi_n$ be linearly independent and define the linear approximating function

$$L(A)=\sum_{i=1}^{n} a_i\phi_i.$$

Let $\|\ \|_k$ be a seminorm on the space C and let K_k be a nonempty closed subset of n-space. The kth problem of constrained approximation is, given $f\in C$, to find a parameter A^k to minimize $e_k(A)=\|f-L(A)\|_k$ subject to the constraint $A\in K_k$.

We generalize a perturbation result of Kripke [3] for approximation with no constraint (i.e., $K_k=n$-space). We use more elementary arguments.

THEOREM. *Let $\|\ \|_0$ be a norm on C and for each $g\in C$, let $\|g\|_k\to\|g\|_0$. Let $A^k\in K_k$ and $A^k\to A^0$ imply $A^0\in K_0$. Let every element of K_0 be a limit point of a sequence $\{A^k\}$, $A^k\in K_k$. Let A^k be best in the* kth *problem of constrained approximation. Then $\{A^k\}$ has an accumulation point and any accumulation point A^0 is best with respect to $\|\ \|_0$, K_0.*

Proof. Define

$$\|A\|_c=\max\{|a_i|: 1\leqslant i\leqslant n\}.$$

Suppose $\|A^k\|_c$ is unbounded. By taking a subsequence if necessary we can assume $\|A^k\|_c\geqslant k$. Define $B^k=A^k/\|A^k\|_c$; then $\|B^k\|_c=1$. By taking a subsequence if necessary, we can assume $B^k\to B^0$. Now

$$\begin{aligned}\|f-L(A^k)\|_k&\geqslant\|L(A^k)\|_k-\|f\|_k=\|A^k\|_c\,\|L(B^k)\|_k-\|f\|_k\\&\geqslant k\,\|L(B^k)\|_k-\|f\|_k.\end{aligned}\tag{1}$$

0021-9045/87 $3.00
Copyright © 1987 by Academic Press, Inc.
All rights of reproduction in any form reserved.

Now

$$\|L(B^k)\|_k \geqslant \|L(B^0)\|_k - \|L(B^k - B^0)\|_k, \tag{2}$$

$$\|L(B^k - B^0)\|_k \leqslant \|B^k - B^0\|_c \sum_{j=1}^{n} \|\phi_j\|_k. \tag{3}$$

For all k sufficiently large $\|\phi_j\|_k \leqslant 2\,\|\phi_j\|_0$; hence by (3),

$$\|L(B^k - B^0)\|_k \to 0;$$

hence by (2)

$$\liminf_{k \to \infty} \|L(B^k)\|_k \geqslant \|L(B^0)\|_0.$$

By (1) we have

$$\|f - L(A^k)\|_k \to \infty. \tag{4}$$

But for any coefficient D^0 and $\{D^k\} \to D^0$, $D^k \in K_k$,

$$\limsup_{k \to \infty} \|f - L(D^k)\|_k \leqslant \|f - L(D^0)\|_0. \tag{5}$$

To prove this,

$$\begin{aligned} \|f - L(D^k)\|_k &\leqslant \|f - L(D^0)\|_k + \|L(D^0 - D^k)\|_k \\ &\leqslant \|f - L(D^0)\|_k + \sum_{i=1}^{n} |d_i^0 - d_i^k|\, \|\phi_i\|_k. \end{aligned}$$

Now (5) contradicts (4), so $\|A^k\|_c$ is bounded. Hence A^k has an accumulation point A^0. By taking a subsequence if necessary we can assume $\{A^k\} \to A^0$. Our hypotheses on K_k and K_0 ensure that $A^0 \in K_0$. Finally, suppose A^0 is not best with respect to $\|\ \|_0$ and K_0. Then there is $B \in K_0$ and $\varepsilon > 0$ with

$$\|f - L(B)\|_0 < \|f - L(A^0)\|_0 - \varepsilon,$$

which implies for all k sufficiently large and $\{A^k\} \to A^0$, $\{B^k\} \to B$,

$$B^k \in K_k,$$

$$\|f - L(B^k)\|_k < \|f - L(A^k)\|_k - \varepsilon/2. \tag{6}$$

To derive (6),

$$\begin{aligned}
&\|f-L(B^k)\|_k-\|f-L(A^k)\|_k \\
&\qquad \leqslant \|f-L(B)\|_k+\|L(B-B^k)\|_k-\|f-L(A^0)\|_k+\|L(A^k-A^0)\|_k \\
&\qquad \leqslant \|f-L(B)\|_k-\|f-L(A^0)\|_k \\
&\qquad\quad +\sum_{i=1}^{n}|b_i-b_i^k|\;\|\phi_i\|_k+\sum_{i=1}^{n}|a_i^0-a_i^k|\;\|\phi_i\|_k.
\end{aligned}$$

The first two terms tend to $\|f-L(B)\|_0-\|f-L(A^0)\|_0$ and the last two terms tend to 0. The proof is finished as (6) contradicts optimality of A^k in the k-problem.

Remark. With a little more work we can prove

$$\|f-L(A^k)\|_k\to\|f-L(A^0)\|_0. \tag{*}$$

Suppose not; then in view of (5) it suffices to prove

$$\liminf_{k\to\infty}\|f-L(A^k)\|_k\geqslant\|f-L(A^0)\|_0 \tag{7}$$

Suppose not; then we can assume

$$\|f-L(A^k)\|_k<\|f-L(A^0)\|_0-\delta. \tag{8}$$

But

$$\begin{aligned}
\|f-L(A^k)\|_k &\geqslant \|f-L(A^0)\|_k-\|L(A^k-A^0)\|_k \\
&\geqslant \|f-L(A^0)\|_k-\sum_{i=1}^{n}|a_i^k-a_i^0|\;\|\phi_i\|_k.
\end{aligned}$$

This contradicts (8), proving (7); (*) may prove essential in computational work, as $\|\ \|_0$ may be difficult to compute, e.g., discretization.

Remark. If f has a unique best approximation $L(A^0)$ in the 0-problem, $\|L(A^0)-L(A^k)\|_0\to 0$.

It is expected that in most mathematical applications of the theorem either the norm will be fixed or the constraints will be fied, but both will be perturbed in computation.

To fully exploit the heorem we need results on nearby norms. Perhaps because Kripke's result is little known, such results have not appeared. Indeed, papers have appeared subsequently [5, 9, 10] with results that could be derived in one line from Kripke's result. Kripke cites discrete L_p norms, $1\leqslant p\leqslant\infty$, as limits of $L_{p(k)}$ norms. The same holds for $1\leqslant p<\infty$

and intervals providing functions are continuous. It is clear that if we let C be the continuous functions on an interval I containing intervals $I_0, I_1, \ldots, I_k, \ldots$ then the limit of L_p norms on intervals $\{I_k\} \to I_0$ is the L_p norm on interval I_0 (compare with the author's [8]). Let C be the continuous functions in an interval I and $w_0, w_1, \ldots, w_k, \ldots$ be positive continuous weight function on I; then the limit of w_k-weighted L_p norms on I is the w_0-weighted L_p norm on I (with care this result might be applicable to some w_0's with zeros). The above three observations might be combined. Let C be the continuous functions on interval I and $\{Q_k\}$ a sequences of quadrature rules such that $Q_k(g) \to \int g$ for $g \in C$: if we fix p in $[1, \infty)$ and choose $\|g\|_k = [Q_k(|g|^p)]^{1/p}$, then $\|g\|_k$ tends to the L_p norm of g on I. Kripke [3, p. 104] thought some such result held. For some sequences of quadrature rules it is only necessary that g be Riemann integrable. It is known that the limit of $L_{p(k)}$ norms for $g \in C[0, 1]$ is the L_∞ norm [12, Problem 4]. Extensions to Bacopoulos-type norms [11] and Moursund-type problems [13, 14] are straightforward.

The hypotheses of sentences 2 and 3 of the theorem are related to the hypotheses of [2]. It is shown in Appendix 2 that they are the most general possible. Levasseur has perturbation hypotheses in the appendix to his thesis [4]. They do not cover varying interpolatory constraints or varying restricted range, but ours do.

In some problems constraints do not change, that is, $K_0 = K_1 = \cdots = K_k = \cdots$ and the hypotheses on the constraints are automatically satisfied. These problems include unchanging interpolatory constraints, for example, Lagrange-type or Hermite-type, and co-positive, co-monotone, co-convex constraints [1]. With varying interpolatory constraints matters may be less satisfactory.

EXAMPLE 1. Approximate $f(x) = x^2$ by linear combinations of $\{1, x\}$. We require interpolation on $\{0, 1/k\}$: the only interpolant is $k^3 x$ and no limit exists as $k \to \infty$. But any multiple of x interpolates f at zero.

EXAMPLE 2. Approximate $f(x) = x^2$ by multiples of x. We require interpolation on $\{1/k\}$: apply the rest of the above example.

In Example 1 we have coalescing of nodes and in Example 2 a non-Haar approximant.

We now establish the hypotheses of the main theorem for the case in which C is the continuous functions on compact X, L satisfies the Haar uniqueness condition, and points of interpolation do not coalesce. Let $\{x_i^k\} \to x_i^0$, $i = 1, \ldots, j$, and

$$K_k = \{A\colon L(A)(x_i^k) = f(x_i^k),\ i = 1, \ldots, j\}$$

where j is a fixed number between 1 and n. Let $A^k \in K_k$ and $\{A^k\} \to A^0$. By uniform convergence of $L(A^k)$ to $L(A^0)$ we must have $L(A^0)(x_i^0) = f(x_i^0)$, $i = 1,..., j$, establishing the hypotheses of the second sentence of the main theorem. Let Y be a set of $n-j$ points distinct from $\{x_i^0,..., x_j^0\}$. Let $A^0 \in K_0$ be given and choose A^k to satisfy

$$L(A^k)(x_i^k) = f(x_i^k) \qquad i = 1,..., j$$

$$L(A^k)(y) = L(A^0, y) \qquad y \in Y.$$

Apply the theorem of the Appendix.

With a number of unchanging restricted range constraints [1, p. 62ff] on function or derivatives (a sample one is

$$\mu \leqslant L^{(r)}(A) \leqslant v),$$

$K_0 = \cdots = K_k = \cdots$ and the hypotheses on K_0, K_k are satisfied automatically.

Consider one changing restricted range constraint on function or derivative, namely,

$$\mu_k \leqslant L^{(r)}(A) \leqslant v_k.$$

Let $\mu_k \to \mu_0$ uniformly and $v_k \to v_0$ uniformly. Providing K_0 is nonempty, the hypothesis of the second sentence of the theorem is satisfied: suppose without loss of generality that $L^{(r)}(A^0)(x) < \mu_0(x) - \varepsilon$; then for all k sufficiently large

$$L^{(r)}(A^k)(x) < \mu_k(x) - \varepsilon/2, \tag{9}$$

a contradiction. It does not appear possible to verify the hypothesis of the third sentence of the theorem without further assumptions: we might assume $\mu_k \leqslant \mu_0$, $v_0 \leqslant v_k$, which makes it automatic. Combining several changing restricted range constraints on functions and derivatives is straightforward.

The case of a changing restricted range *on one side* can be handled without very restrictive assumptions. Let C be the continuous functions on X, a compact subset of the real line and

$$K_k = \{A\colon L^{(r)}(A) \geqslant \mu_k\}, \tag{10}$$

and $\mu_k \to \mu_0$ uniformly on X.

LEMMA. *Let there exist B such that $L^{(r)}(B) > 0$ on X. Then any $A \in K_0$ is a limit point of a sequence $\{A^k\}$, $A^k \in K_k$.*

Proof. For conciseness we suppress the superscript (r) on L. Suppose not then there is $\delta > 0$ such that for all $A^k \in K_k$

$$\|A - A^k\|_c > \delta \tag{11}$$

for all k sufficiently large. Now let λ_k be the smallest number $\lambda \geqslant 0$ such that

$$L(A + \lambda B) \geqslant \mu_k. \tag{12}$$

λ_k is well defined as $L(A + \lambda B)$ becomes indefinitely large as $\lambda \to \infty$ since $L(B) > 0$ and the infimum of λ's $\geqslant 0$ satisfying (12) must satisfy (12). If $\{\lambda_k\}$ had zero as an accumulation point, (11) would be contradicted so we will assume there is $\varepsilon > 0$ such that $\lambda_k \geqslant \varepsilon$ for all k sufficiently large. By taking a subsequence, we can assume $\{\lambda_k\} \to \lambda_0 \geqslant \varepsilon$. By choice of λ_k, $L(A + \lambda_k B)$ must graze μ_k; that is, there is x_k such that

$$L(A + \lambda_k B)(x_k) = \mu_k(x_k) \tag{13}$$

for if this did not happen, we could reduce λ_k and still satisfy (12). By compactness of X, the sequence $\{x_k\}$ has an accumulation point x_0; assume $\{x_k\} \to x_0$. As $L(A + \lambda_k B) \to L(A + \lambda_0 B)$ uniformly on X, $L(A + \lambda_0 B)(x_0) = \mu_0(x_0)$ by (13). But as $L(A) \geqslant \mu_0$ and $L(B) > 0$ we have a contradiction, proving the lemma.

The lemma establishes the hypothesis of the third sentence of the theorem. The hypothesis of the second sentence is handled by preceding arguments.

The general case of changing restraints on two sides, if it can be handled at all, will require more subtle analysis.

EXAMPLE. Let C be the continuous functions on $X = [0, 1]$. Choose

$$\begin{aligned} v_k(x) &= 0 \qquad & 0 \leqslant x \leqslant 1/k \\ &= x - 1/k \qquad & x > 1/k \\ \mu_k &= -v_k \end{aligned}$$

By Dini's theorem $v_k \to v_0 = x$ uniformly on X and μ is similar. Now let us consider approximation by first-degree polynomials; then $\{0\}$ is the only approximant in the restricted range of μ_k, v_k. But the approximants in the restricted range of μ_0, v_0 are $\{rx \colon |r| \leqslant 1\}$.

This example suggests that no general theory is possible if restraining functions touch.

Even if restraining functions do not touch, a general theory may not be possible without conditions on approximants. Take the above example, except $\mu_0 = \mu_k = -1$ and approximants are multiples of x. $\{0\}$ is the only nonnegative approximation restrained by μ_k, ν_k but $\{rx : 0 \leqslant r \leqslant 1\}$ is restrained by μ_0, ν_0. If we completely neglect the lower restraint, we have an example for which the $L^{(r)}(B)$ condition of the preceding lemma is necessary.

A condition on approximants which does guarantee that the hypotheses of the theorem are satisfied is the ASSUMPTION of the author's paper [2, 196].

APPENDIX 1: VARYING INTERPOLATORY CONSTRAINTS

THEOREM. *Let L have the Haar property. Let $L(A^k)$ take the values $y_1^k, \ldots, y_n^k$ at points $x_1^k, \ldots, x_n^k$ (all distinct), respectively. Let $y_j^k \to y_j^0$, $j = 1, \ldots, n$, and $x_j^k \to x_j^0$, $j = 1, \ldots, n$. Then $A^k \to A^0$.*

Proof. First we must show $\{A^k\}$ is bounded. Suppose not; then by taking a subsequence if necessary we can assume $\|A^k\|_c > k$. By standard arguments, e.g., that of Rice [6, 24–25], this leads to a contradiction. Let B be any accumulation point of $\{A^k\}$ not equal to A^0; then by taking a subsequence if necessary we can assume $\{A^k\} \to B$. By uniform convergence of $L(A^k)$ to $L(B)$ we must have $L(B)(x_j^0) = y_j^0$, $j = 1, \ldots, n$. But by the Haar assumption interpolation is unique.

Remark. Tornheim [7] proved this for X an interval and real values.

Remark. This could be generalized to Hermite-type interpolation by replacing the Haar condition.

APPENDIX 2: NECESSITY OF HYPOTHESES

If we do not have the hypotheses of the second and third sentences of the main theorem satisfied, the conclusions of the theorem do not hold. Suppose (to contradict the second sentence) $D^k \in K_k$ and $\{D^k\} \to D^0$ not in K_0. Approximate $f = L(D^0)$; then by (5) $\{A^k\}$ D^0 for A^k best in the k-problem. Suppose (to contradict the third sentence) $D^0 \in K_0$ is not a limit point of any sequence $\{D^k\}$, $D^k \in K_k$. Approximate $f = L(D^0)$; then A^k best in the k-problem implies that $\|A^k - D^0\|_c$ is bounded away from zero for k large.

ACKNOWLEDGMENTS

The author thanks the referee and Dr. C. Z. Zhu for helpful criticism. This work was assisted by a grant from NSERC, Canada.

REFERENCES

1. B. L. CHALMERS AND G. D. TAYLOR, Uniform approximation with constraints, *Jahresber. Deutsch. Math. Verein* **81** (1979), 49–86.
2. C. B. DUNHAM, Nearby linear Chebyshev approximation under constraints, *J. Approx. Theory* **42** (1984), 193–199.
3. B. R. KRIPKE, Best approximation with respect to nearby norms, *Numer. Math.* **6** (1964), 103–105.
4. K. M. LEVASSEUR, "Best Approximation with Respect to Two Objectives," dissertation, University of Rhode Island, Kingston, 1980.
5. J. FISCHER, The convergence of the best discrete linear L_p approximation as $p \to 1$, *J. Approx. Theory* **39** (1983), 374–385.
6. J. RICE, "The Approximation of Functions," Vol. 1, Addison–Wesley, Reading, Mass., 1964.
7. L. TORNHEIM, On n-parameter families of functions and associated convex functions, *Trans. Amer. Math. Soc.* **69** (1950), 459–469.
8. C. B. DUNHAM, Alternating approximation on shorter intervals, *Studia Sci. Math. Hungar.* **14** (1979), 409–413.
9. T. J. RIVLIN AND E. W. CHENEY, A comparison of uniform approximations on an interval and a finite subset thereof, *SIAM J. Numer. Anal.* **3** (1966), 311–320.
10. N. ABDELMALEK, Linear L_1 approximation for a discrete point set and L_1 solutions of overdetermined linear equations. *J. Assoc. Comput. Mach.* **18** (1971), 41–47.
11. A. BACOPOULOS, On approximation by sums of Chebyshev norms, *Numer. Math.* **16** (1970), 243–247.
12. W. RUDIN, "Real and Complex Analysis," 2nd ed., McGraw–Hill, New York, 1974.
13. C. B. DUNHAM, Linear Chebyshev approximation, *Aequationes Math.* **10** (1974), 40–45.
14. D. G. MOURSUND, Chebyshev approximation of a function and its derivatives, *Math. Comput.* **18** (1964), 382–389.