JOURNAL OF APPROXIMATION THEORY 60, 24–30 (1990)

Rate of Convergence of the Discrete Pólya Algorithm

ALAN EGGER AND ROBERT HUOTARI

Department of Mathematics, Idaho State University, Pocatello, Idaho 83209, U.S.A.

Communicated by F. W. Cheney

Received October 1, 1987

In approximating an arbitrary point of $\mathbf{R}^n$ from a fixed subspace, it is known that the net of l^p-best approximations, $1 \leqslant p < \infty$, converges to the strict uniform best approximation. It is shown that this convergence occurs at a rate no worse than $1/p$. It is also shown by example that this rate may be achieved.

INTRODUCTION

For $\mathbf{x} = (x^1, ..., x^n)$ in $\mathbf{R}^n$, the l^p norms, $1 \leqslant p \leqslant \infty$, are defined by

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x^i|^p \right)^{1/p} \qquad \text{for} \quad 1 \leqslant p < \infty$$

and

$$\|\mathbf{x}\|_\infty = \max_{1 \leqslant i \leqslant n} |x^i|.$$

If K is a convex subset of $\mathbf{R}^n$ and $\mathbf{z}$ is in $\mathbf{R}^n \setminus K$, we say that $\mathbf{x}_p \in K$, $1 \leqslant p \leqslant \infty$, is a *Best Approximation* from K to $\mathbf{z}$ with respect to the norm $\|\cdot\|_p$ if

$$\|\mathbf{x}_p - \mathbf{z}\|_p = \inf_{x \in K} \|\mathbf{x} - \mathbf{z}\|_p.$$

For $1 < p < \infty$, it is known that $\mathbf{x}_p$ is unique. If, in addition, K is an affine subspace of $\mathbf{R}^n$, then the net $\{\mathbf{x}_p : 1 < p < \infty\}$ converges to a vector $\mathbf{s} \in K$ as $p \to \infty$ [1]. Here, $\mathbf{s}$ is a distinguished l^∞-Best Approximation characterized by the following. Let $X_\infty = \{\mathbf{x} \in K: \|\mathbf{x} - \mathbf{z}\|_\infty = \inf_{y \in K} \|\mathbf{y} - \mathbf{z}\|_\infty\}$. For each $\mathbf{x} \in X_\infty$, let $\tau(\mathbf{x})$ be the vector whose components are given by $|x^i - z^i|$, $i = 1, ..., n$, arranged in nonincreasing order. The *Strict Uniform Approximation* is the unique $\mathbf{s} \in X_\infty$ with $\tau(s)$ minimal in the lexicographic ordering on X_∞. Thus, an application of the *Pólya Algorithm* [9] (i.e., the calculation of $\lim_{p \to \infty} \mathbf{x}_p$) would enable us to compute the "best" of the l^∞-Best Approximations.

0021-9045/90 $3.00

The need to estimate $\lim_{p\to\infty} \mathbf{x}_p$ naturally leads to questions regarding the rate of convergence of $\mathbf{x}_p$ to $\mathbf{s}$. In [4] an extrapolation scheme is proposed for estimating $\mathbf{s}$ from $\mathbf{x}_p$ and it is shown that when the Uniform Best Approximation from an affine subspace is strongly unique, the rate of convergence of $\mathbf{x}_p$ to $\mathbf{s}$ is at worst $1/p$. Of course, if $\mathbf{s}$ is strongly unique, then $\mathbf{s}$ is unique, so a Linear Programming technique could be used to compute $\mathbf{s}$. In the absence of uniqueness, Linear Programming may fail to return the strict approximation. The purpose of the present note is to obtain a convergence estimate without assuming that there is a unique best uniform approximation. This rate estimate could then be used in extrapolatory schemes in general discrete approximation problems. To motivate our result and show that it is sharp, we begin by discussing the following example.

EXAMPLE. Suppose that $\mathbf{z}=0$ and $K=P$, the hyperplane defined by

$$P=\{(x^1, ..., x^n)\colon a_1x^1+\cdots+a_nx^n=1\},$$

where each a_i is positive. Let $\mathbf{x}_p$ and $\mathbf{s}$ be as defined above and let $\mathbf{r}_p=\mathbf{x}_p-\mathbf{s}$. Then P may be represented by $P=V+\mathbf{s}$, where V is a subspace of $\mathbf{R}^n$, $\mathbf{r}_p\in V$, and $\mathbf{r}_p\to 0$ as $p\to\infty$. We claim that, for $1<p<\infty$, $\mathbf{x}_p$ is in the first octant of $\mathbf{R}^n$. Indeed, if $x_p^j<0$, for some j, let $\mathbf{y}=(|x_p^1|, ..., |x_p^n|)$. Then $\|\mathbf{y}\|_p=\|\mathbf{x}_p\|_p$ but $\sum_{i=1}^n a_i y^i>1$, which implies that P intersects the open ball $\{\mathbf{x}\colon \|\mathbf{x}\|_p<\|\mathbf{y}\|_p\}$, a contradiction. When $\mathbf{x}$ is in P and in the first octant,

$$\|\mathbf{x}\|_p^p=\left(a_1^{-1}\left(1-\sum_{i=2}^n a_ix^i\right)\right)^p+(x^2)^p+\cdots+(x^n)^p. \tag{1}$$

For $2\leqslant i\leqslant n$, the partial derivative with respect to x^i of $\|\cdot\|_p$ vanishes at $\mathbf{x}_p$, i.e.,

$$0=-pa_i(a_1)^{-1}\left(1-\sum_{j=2}^n a_jx_p^j\right)^{p-1}+p(x_p^i)^{p-1}, \tag{2}$$

so

$$x_p^i=(a_i)^{1/(p-1)}(a_1)^{-p/(p-1)}\left(1-\sum_{j=2}^n a_jx_p^j\right). \tag{3}$$

By (2), for $2\leqslant j\leqslant n$,

$$x_p^i(a_i)^{-1/(p-1)}\,(a_1)^{p/(p-1)}=x_p^j(a_j)^{-1/(p-1)}\,(a_1)^{p/(p-1)},$$

so

$$x_p^j = (a_j/a_i)^{1/(p-1)}\, x_p^i.$$

Plugging these values of x_p^j, $2 \leqslant j \leqslant n$, into (3), we obtain

$$x_p^i = (a_i)^{1/(p-1)} \left(\sum_{j=1}^{n} (a_j)^{p/(p-1)} \right)^{-1}, \qquad 2 \leqslant i \leqslant n. \tag{4}$$

If we write (1) as

$$\|\mathbf{x}\|_p^p = (x^1)^p + \cdots + (x^{n-1})^p + \left(a_n^{-1} \left(1 - \sum_{i=1}^{n-1} a_i x^i \right) \right)^p$$

a similar calculation shows that (4) also holds for $i = 1$.

By (4), $\lim_{p \to \infty} \mathbf{x}_p = \mathbf{x}_\infty$, where

$$x_\infty^j = \left(\sum_{i=1}^{n} a_i \right)^{-1}, \qquad 1 \leqslant j \leqslant n. \tag{5}$$

By [1], $\mathbf{x}_\infty = \mathbf{s}$, the strict uniform Best Approximation to 0. We will now investigate the behavior of $p\,|r_p^i| = p\,|x_p^i - s^i|$ as $p \to \infty$. We assume without loss of generality that $i = 1$. By (4) and (5),

$$pr_p^1 = \frac{p[(a_1)^{1/(p-1)} \sum_{j=1}^n a_j - \sum_{j=1}^n (a_j)^{p/(p-1)}]}{[\sum_{j=1}^n (a_j)^{p/(p-1)}][\sum_{j=1}^n a_j]}.$$

Since $\sum (a_j)^{p/(p-1)} \to \sum a_j$, we need only consider the numerator, which we call $\varphi(p)$, of the last fraction. Dividing the top and bottom of $\varphi(p)$ by p and applying L'Hôpital's rule, we have $\lim_{p \to \infty} \varphi(p) = \lim_{p \to \infty} \psi(p) \cdot \lim_{p \to \infty} \eta(p)$, where $\psi(p) = p^2/(p-1)^2$ and

$$\eta(p) = (a_1)^{1/(p-1)} \log a_1 \sum_{j=1}^{n} a_j - \sum_{j=1}^{n} (a_j)^{p/(p-1)} \log a_j.$$

Clearly $\lim_{p \to \infty} \psi(p) = 1$ and

$$\lim_{p \to \infty} \eta(p) = \sum_{j=2}^{n} a_j (\log a_1 - \log a_j). \tag{6}$$

Thus $r_p^1 = O(1/p)$. If $r_p^k = o(1/p)$ for each k, $1 \leqslant k \leqslant n$, then the kth version of (6) vanishes for each k, which implies that $\sum_{j=1}^n a_j \log a_j = \log a_1 \sum_{j=1}^n a_j = \cdots = \log a_n \sum_{j=1}^n a_j$ so $a_1 = a_2 = \cdots = a_n$. Thus, in every case except this trivial case (where $r_p^k = 0$ for every k and every p), the convergence of $\mathbf{r}_p$ to 0 is not faster than that of $1/p$.

We now show that $r_p^i = O(1/p)$ in the more general context where $\mathbf{z}=0$ and $K=H$, where H is any affine subspace of $\mathbf{R}^n$. Suppose $\{p_k\}_{k=1}^{\infty}$ is an unbounded increasing sequence in $\mathbf{R}$. By [3], we may suppose without loss of generality that the error vector, $\mathbf{r}_{p_k} = \mathbf{x}_{p_k} - \mathbf{s}$, is nonzero for each k. Let $\mathbf{u}_k = \mathbf{r}_{p_k}/\|\mathbf{r}_{p_k}\|_\infty$ and $\lambda_k = \|\mathbf{r}_{p_k}\|_\infty$. Since $\|\mathbf{u}_k\|_\infty = 1$ for each $\mathbf{k}$, a subsequence of $\{\mathbf{u}_k\}$ must converge to some $\mathbf{u} \in V = H - \mathbf{s}$, with $\|\mathbf{u}\|_\infty = 1$. By relabeling, we may assume that $\mathbf{u}_k \to \mathbf{u}$. If we let $\mathbf{v}_k = \mathbf{u}_k - \mathbf{u}$, then $\mathbf{v}_k \to 0$ as $k \to \infty$ and $\mathbf{x}_{p_k} = \mathbf{s} + \lambda_k \mathbf{u}_k = \mathbf{s} + \lambda_k(\mathbf{u} + \mathbf{v}_k)$. In our discussion of the rate of convergence of $\mathbf{r}_{p_k}$ to 0, we will refer to several constants and subsets of $\{1, ..., n\}$. They are defined as

$$\begin{aligned}
&A = \{i\colon u^i \neq 0\}, && a = \max_{i \in A} |s^i|, && \\
&B = \{i \in A\colon |s^i| = a\}, && b = \max_{i \in B} |u^i|, && b_1 = \min_{i \in B} |u^i|, \\
&C = \{i \in B\colon u^i s^i > 0\}, && c = \min_{i \in C} |u^i|. &&
\end{aligned}$$

Each of A, B, and C is nonempty, $a, b > 0$ and $0 < b_1$, $c < \infty$. Indeed, $\|\mathbf{u}\|_\infty = 1$, so $A \neq \emptyset$. Suppose $a = 0$. For any $\alpha \in \mathbf{R}$,

$$\begin{aligned}
\|\mathbf{x}_{p_k} + \alpha\mathbf{u}\|_{p_k}^{p_k} &= \|\mathbf{s} + \lambda_k(\mathbf{u} + \mathbf{v}_k) + \alpha\mathbf{u}\|_{p_k}^{p_k} \\
&= \sum_{i=1}^{n} |s^i + \lambda_k(u^i + v_k^i) + \alpha u^i|^{p_k} \\
&= \sum_{s^i \neq 0} |s^i + \lambda_k v_k^i|^{p_k} + \sum_{s^i = u^i = 0} |\lambda_k v_k^i|^{p_k} \\
&\quad + \sum_{s^i = 0 \neq u^i} |\lambda_k(u^i + v_k^i) + \alpha u^i|^{p_k}.
\end{aligned}$$

Observe that only the last sum depends on α. Since $\|\mathbf{u}\|_\infty = 1$, the set $\{i\colon s^i = 0 \neq u^i\}$ is nonempty and, for large p_k, $\operatorname{sign}(\lambda_{p_k}(u^i + v_k^i)) = \operatorname{sign}(u^i)$. Thus, for α sufficiently small and negative, each term in the last sum is reduced, contradicting the minimality of $\|\mathbf{x}_{p_k}\|_{p_k}$. Thus $a > 0$. Now it is clear that $B \neq \emptyset$, $b > 0$, and $0 < b_1 < \infty$. If $C = \emptyset$, then $u^i s^i \leqslant 0$ whenever $|s^i| \geqslant a$, $1 \leqslant i \leqslant n$, and $u^j s^j < 0$ for some $j \in B$. If $i \notin A$, $|s^i + \delta u^i| = |s^i|$. For $i \in B$ and for sufficiently small $\delta > 0$, $|s^i + \delta u^i| \leqslant a$ and $|s^j + \delta u^j| < a$. Thus $\mathbf{s} + \delta\mathbf{u}$ is a uniform best approximation which is less than $\mathbf{s}$ in the lexicographic ordering, a contradiction. Thus $C \neq \emptyset$ and $0 < c < \infty$.

Because $\mathbf{x}_p$ is the Best Approximation with respect to $\|\cdot\|_p$, the derivative of $\|\mathbf{x}\|_{p_k}^{p_k}$ in the direction $\mathbf{u}$ must vanish at $\mathbf{x}_{p_k}$, i.e.,

$$\sum_{i=1}^{n} |x_{p_k}^i|^{p_k - 1} \operatorname{sign}(x_{p_k}^i) u^i = \sum_{i \in A} |x_{p_k}^i|^{p_k - 1} \operatorname{sign}(x_{p_k}^i) u^i = 0.$$

By the definition of C, $\operatorname{sign}(x^i_{p_k})u^i$ is equal to $|u^i|$ for $i \in C$ and $-|u^i|$ for $i \in B \setminus C$, when p_k is sufficiently large. Then the above equation can be written

$$\sum_{i \in A \setminus B} |x^i_{p_k}|^{p_k - 1} \operatorname{sign}(x^i_{p_k})u^i + \sum_{i \in B} \theta\, |x^i_{p_k}|^{p_k - 1}\, |u^i| = 0, \tag{7}$$

where $\theta = -1$ if $i \in B \setminus C$ and $\theta = 1$ if $i \in C$. We can now state our main result.

THEOREM. *In the above context, there exist M and p_0 such that $\|\mathbf{x}_p - \mathbf{s}\|_\infty \leqslant M/p$ for all $p > p_0$.*

Proof. Suppose the theorem is false. Then there is an unbounded increasing sequence $\{p_k\}$ such that $p_k \lambda_k = p_k \|\mathbf{r}_{p_k}\|_\infty \to \infty$. We will derive a contradiction by showing that, in this case, the left-hand side of (7) would eventually be positive.

If $i \in A \setminus B$, then there is a number $\omega \in [0, 1)$ such that $|x^i_{p_k}| \leqslant \omega a$, when p_k is sufficiently large. Thus

$$\left| \sum_{i \in A \setminus B} |x^i_{p_k}|^{p_k - 1} \operatorname{sign}(x^i_{p_k})u^i \right| \leqslant n(\omega a)^{p_k - 1}. \tag{8}$$

Choose $\rho \in (0, \frac{1}{2})$. Since $|v^i_k| \to 0$, $|v^i_k| < \min(\rho b_1, \rho c)$ for sufficiently large k. Then

$$\begin{aligned}
\sum_{i \in B \setminus C} |x^i_{p_k}|^{p_k - 1}\, |u^i| &\leqslant \sum_{i \in B \setminus C} b\, |a - \lambda_k(b_1 - |v^i_k|)|^{p_k - 1} \\
&\leqslant nb\, |a - \lambda_k b_1(1 - \rho)|^{p_k - 1} \\
&= nba^{p_k - 1} \left\{ \left| 1 - \frac{b_1(1-\rho)}{a} \lambda_k \right|^{\lambda_k^{-1}} \right\}^{\lambda_k(p_k - 1)}.
\end{aligned}$$

Thus, for any $\varepsilon_1 > 0$ with $\exp[a^{-1} b_1(\rho - 1)] + \varepsilon_1 < 1$ and for sufficiently large k,

$$\sum_{B \setminus C} |x^i_{p_k}|^{p_k - 1}\, |u^i| \leqslant nba^{p_k - 1} \left\{ \exp\left[\frac{b_1(\rho - 1)}{a}\right] + \varepsilon_1 \right\}^{\lambda_k(p_k - 1)}. \tag{9}$$

Similarly,

$$\begin{aligned}
\sum_{i \in C} |x^i_{p_k}|^{p_k - 1}\, |u^i| &\geqslant \sum_{i \in C} c\, |a + \lambda_k c(1 - \rho)|^{p_k - 1} \\
&\geqslant ca^{p_k - 1} \left| 1 + \frac{c(1-\rho)}{a} \lambda_k \right|^{\lambda_k^{-1} \lambda_k(p_k - 1)}
\end{aligned}$$

so, for any $\varepsilon_2 > 0$ with $\exp[a^{-1}c(1-\rho)] - \varepsilon_2 > 1$ and for sufficiently large k,

$$\sum_{i \in C} |x^i_{p_k}|^{p_k - 1} |u^i| \geqslant c a^{p_k - 1} \left\{ \exp\left[\frac{c(1-\rho)}{a}\right] - \varepsilon_2 \right\}^{\lambda_k(p_k - 1)}. \qquad (10)$$

Since $\lambda_k(p_k - 1) \to \infty$ and each of ω and ρ is less than 1, the quantity in (10) dominates those in (8) and (9), so (7) is eventually positive. This contradiction establishes the theorem.

Remarks. The problem we have addressed concerns the approximation of real-valued functions on a finite discrete domain. If the domain in question is a compact interval, the rate bound $(\log p)/p$ is known to hold and be optimal [6, 8].

Although the discrete Póya algorithm need not converge for a general convex approximating set [2], it remains open whether the convergence, when it occurs, must occur at rate no worse than $1/p$, or may occur arbitrarily slowly. A general context in which convergence occurs is described in [5].

Also of interest are the qualitative convergence properties of the net $\{\mathbf{x}_p\}$. Although it is known [3] that r^i_p need not in general be monotone, it is true that in the above example, r^i_p is ultimately monotone. This may be seen as follows. Differentiating (4) with respect to p, we see that $(d/dp)(x^1_p) = 0$ if and only if

$$\log a_1 \sum_{j=1}^{n} (a_j)^{p/(p-1)} - \sum_{j=1}^{n} (a_j)^{p/(p-1)} \log a_j = 0. \qquad (11)$$

As $p \to \infty$, $p/(p-1) \to 1$. Since a branch of each of the complex functions $f_j(z) = (a_j)^z$ is analytic in a neighborhood of $z = 1$, (11) holds for all p (i.e., x^1_p is constant), or (11) holds for at most finitely many p in any neighborhood of $p = \infty$ (i.e., x^1_p is eventually monotone). If this monotonicity property holds for arbitrary affine subspaces, it would have desirable consequences relating to the extrapolation of $\mathbf{s}$ from $\mathbf{x}_p$.

Dual to the question we have just addressed is that of the behavior of r^i_p as p decreases to one. If q is the dual index defined by $1/p + 1/q = 1$, then the natural conjecture is that $r^i_p = O(1/q) = O((p-1)/p)$ as $p \downarrow 1$. In the case where $K = P$, an even stronger statement can be made, viz., the convergence is exponential. To see that this is true, let $\beta = \max\{a_i : 1 \leqslant i \leqslant n\}$, $\gamma = \max\{a_i : a_i < \alpha\}$ and choose α so that $\gamma < \alpha < \beta$. Looking at (4), we see that, as $p \to 1$, $x^1_p \to 0$ if $a_1 \neq \beta$, and $x^1_p \to (ka_1)^{-1}$ if $a_1 = \beta$, where k is the number of times the value β is assumed in the list $a_1, \ldots, a_n$. By algebraic manipulation and L'Hôpital's rule, we see that, in either case, $(\beta/\alpha)^{p/(p-1)} r_p^1 \to 0$ as $p \to 1$ so

$$r_p^1 = O[(\alpha/\beta)^{p/(p-1)}].$$

The convergence, as $p \to 1$, of x_p is discussed in [7].

References

1. J. Descloux, Approximations in L^p and Chebychev approximations, *J. Soc. Ind. Appl. Math.* **11** (1963), 1017–1026.
2. A. Egger and R. Huotari, The Pólya algorithm on convex sets, *J. Approx. Theory* **56** (1989), 212–216.
3. A. Egger and G. Taylor, Dependence on p of the best L^p approximation operator, *J. Approx. Theory* **49** (1987), 274–282.
4. R. Fletcher, J. Grant, and M. Hebden, Linear minimax approximation as the limit of best L_p-approximation, *SIAM J. Numer. Anal.* **11** (1974), 123–136.
5. R. Huotari, D. Legg, and D. Townsend, The Pólya algorithm on cylindrical sets, *J. Approx. Theory* **53** (1988), 335–349.
6. A. Kroo, On the convergence of Pólya's algorithm, *J. Approx. Theory* **30** (1980), 139–148.
7. D. Landers and L. Rogge, Natural choice of L_1-approximants, *J. Approx. Theory* **33** (1981), 268–280.
8. J. Peetre, Approximation of norms, *J. Approx. Theory* **3** (1970), 243–260.
9. G. Pólya, Sur un algorithm touours convergent pour obtenir les polynômes de meilleure approximation de Tchebycheff pour une fonction continue quelconque, *C.R. Acad. Sci. Paris* **157** (1913), 840–843.