JOURNAL OF APPROXIMATION THEORY 75, 312–324 (1993)

Rate of Convergence of the Discrete Pólya-1 Algorithm*

A. G. EGGER

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

AND

G. D. TAYLOR

Department of Mathematics, Colorado State University, Fort Collins, Colorado 80523, U.S.A.

Communicated by Günther Nürnberger

Received November 15, 1991; accepted in revised form August 3, 1992

The rate of convergence of the discrete Pólya-1 algorithm is studied. Examples are given to show that the rates derived are sharp.

Let V be a finite dimensional subspace of $\mathbb{R}^n$ and fix $z \in \mathbb{R}^n \setminus V$. Given a norm, $\|\cdot\|$, on $\mathbb{R}^n$, $v^* \in V$ is a best approximation from V to z if

$$\|v^* - z\| = \min\{\|v - z\| : v \in V\}.$$

In this setting the existence of a best approximation is immediate. Of course, different norms may give rise to different best approximations. The dependence of best approximations on the norm in use has been studied in a variety of contexts. For example, [1] and [10] are general studies of the effects of perturbing the norm on best approximation problems.

The p-norms, given by

$$\|x\|_p = \left[\sum_i^n |x_i|^p\right]^{1/p}, 1 \leqslant p < \infty, \qquad \text{and} \qquad \|x\|_\infty = \max_i |x_i|$$

form a well-known parameterized family of norms on $\mathbb{R}^n$. Denote by l^p the space $\mathbb{R}^n$ with the p-norm. In the l^p family of Banach spaces, selecting a value of p corresponds to a choice of norm. Each such choice determines

* Supported in part by the ARO under Grant DAAL03-86-K-0175, the NSF under Grant ATM-8814541, and the ONR under Grant N000014-88-K-0214.

0021-9045/93 $5.00

a different best approximation problem. Discussions of the relative merits of specific values of p date from the 18th century [2].

For $1 < p < \infty$ the corresponding p-norm is strictly convex, so that there is a unique solution to the best approximation problem for this p. Denote this solution by x^p. For $p = 1$ and $p = \infty$ solutions to this best approximation problem need not be unique. The problem of the dependence on p of best approximations has been extensively studied. Of particular interest has been the behavior of the arc x^p as $p \to \infty$. The taking of such a limit is referred to as the Pólya algorithm and was first considered by Pólya in a related setting [9]. In the subspace setting it is known that the Pólya algorithm converges and that in general the rate of this convergence is $O(1/p)$ [3, 4].

The behavior of the arc x^p as $p \to 1$ has also been studied [6–8, 12]. Taking this limit is known as the Pólya-1 algorithm. The Pólya-1 algorithm converges in a very general setting, including the subspace problem under consideration here. Aside from an example and a conjecture [4] little is known about the rate at which the Pólya-1 algorithm converges. In the following, the rate of convergence is developed. In contrast to the Pólya algorithm rate, it is shown that the rate of convergence of x^p depends heavily upon the set L of l^1 best approximations. To see this consider the following examples:

EXAMPLE 1. In $\mathbb{R}^3$, let $z = (0, 0, 1)$ and let $V = \{(a, a, a): a \in \mathbb{R}\}$. Here the l^1 best approximation is unique and is the median $(0, 0, 0)$. To find x^p, there is no point in considering $a > 1$ or $a < 0$. So we minimize $2a^p + (1-a)^p$ over $[0, 1]$. Differentiating gives $2a^\delta - (1-a)^\delta = 0$, where $\delta = p - 1$. Thus $(1-a) = 2^{1/\delta} a$ or $2^{-1/\delta} = a/(1-a)$. For p near 1, $\frac{1}{2} \leqslant 1 - a \leqslant 1$, so $a = O(2^{-1/\delta})$. Thus $x^p \to x^1$ at an exponential rate.

When the set L is not a singleton, a slower rate of convergence may hold.

EXAMPLE 2. In $\mathbb{R}^4$, let $z = (2, 1, 0, 0)$ and $V = \{a(1, 1, 1, -1): a \in \mathbb{R}\}$. Here $L = \{a(1, 1, 1, -1): a \in [0, 1]\}$. Consider the strict best approximation $x^1 = a^1(1, 1, 1, -1)$, the limit of x^p as $p \to 1$. On L, a^1 minimizes $\psi(r) = (2-r)\ln(2-r) + (1-r)\ln(1-r) + 2r\ln(r)$. (See Theorem 1.) Now $\psi'(r) = 2\ln(r) - \ln\{(2-r)(1-r)\}$, yielding critical values 0, 1, and $\frac{2}{3}$. Since x^1 lies in the relative interior of L [6], $a^1 = \frac{2}{3}$. Write $x^p = a^p(1, 1, 1, -1)$. Since $x^p \to x^1$, we know that for small $p > 1$, $\frac{1}{2} < a^p < \frac{7}{9}$. Note that for values of r between 0 and 1, $\psi_p(r) = \|z - r(1, 1, 1, -1)\|_p^p = (2-r)^p + (1-r)^p + 2(r)^p$. Then $\psi_p'(r) = -p((2-r)^\delta + (1-r)^\delta - 2(r)^\delta)$, where $\delta = p - 1$. Then $\psi_p'(\frac{2}{3}) = -p((\frac{4}{3})^\delta + (\frac{1}{3})^\delta - 2(\frac{2}{3})^\delta) = -p3^{-\delta}(2^\delta - 1)^2 < 0$. This forces $a^p > \frac{2}{3}$ for small $p > 1$. Thus, for small p, $2 > p > 1$, $\frac{7}{9} > a^p > \frac{2}{3}$.

Therefore, for such small $p>1$, we may write $a^p = \frac{2}{3} + \varepsilon_p/3$, where $0 < \varepsilon_p < \frac{1}{3}$. Hence ε_p satisfies

$$((4-\varepsilon_p)/3)^\delta + ((1-\varepsilon_p)/3)^\delta - 2((2+\varepsilon_p)/3)^\delta$$
$$= 0 = (4-\varepsilon_p)^\delta + (1-\varepsilon_p)^\delta - 2(2+\varepsilon_p)^\delta.$$

Thus, $4^\delta - (4-\varepsilon_p)^\delta + 1 - (1-\varepsilon_p)^\delta + 2(2+\varepsilon_p)^\delta - 2(2^\delta) = (2^\delta - 1)^2$. Now apply the Mean Value Theorem individually to the expressions $(4-x)^\delta$, $(1-x)^\delta$, $(2+x)^\delta$, and 2^x all centered at $x=0$ to get constants c_i, $4-\varepsilon_p < c_1 < 4$, $1-\varepsilon_p < c_2 < 1$, $2 < c_3 < 2+\varepsilon_p$, and $0 < c_4 < \delta$ such that

$$\delta\varepsilon_p(c_1^{\delta-1} + c_2^{\delta-1} + c_3^{\delta-1}) = (\delta 2^{c_4} \ln 2)^2.$$

Thus, $\varepsilon_p = \delta 2^{2c_4}(\ln^2 2)(c_1^{\delta-1} + c_2^{\delta-1} + c_3^{\delta-1})^{-1}$ and there exist positive constants A and B such that $A \leqslant 2^{2c_4}(\ln^2 2)(c_1^{\delta-1} + c_2^{\delta-1} + c_3^{\delta-1})^{-1} \leqslant B$ for p in this range. Hence $4A\delta \leqslant \|x^p - x^1\|_1 \leqslant 4B\delta$ for small $p>1$. Thus, x^p converges linearly to x^1 as $p \to 1$.

We now show that this dichotomy in rates holds in general. As above, denote by L the set of all l^1 best approximations from V to z. For $r \in \mathbb{R}$, we know that $r\ln(r) \to 0$ as $r \to 0^+$. Hence we identify $(0\ln(0))$ with 0 and, for $x \in \mathbb{R}^n$, define the function $\psi(x)$ by

$$\psi(x) = \sum_{i=1}^{n} |x_i - z_i| \ln |x_i - z_i|.$$

The limiting behavior of the net $\{x^p : p > 1\}$ is described in the following theorem.

THEOREM 1. [6, 8, 12]. *Under the above hypotheses, there exists $v \in L$ such that $\lim_{p\to 1} x^p = v$. Furthermore, v is in the relative interior of L and is the unique minimizer of ψ on L.*

The element v is known as the *natural best approximation* or the *strict best approximation* of $z \in \mathbb{R}^n$ from V. Our interest is in the rate at which the best l^p approximations converge to the natural best approximation. Since this rate is unaffected by translation and scaling, we assume that $v=0$ and that $\|z\|_\infty < 1/2e^2$ holds. Define $\Omega = \{i : z_i = 0\}$. The following lemmas describe the zero structure of vectors near $v=0$.

LEMMA 1. *If $v \in L$ and $i \in \Omega$ then $v_i = 0$.*

LEMMA 2. *For some $\rho > 0$ and $\varepsilon > 0$, the set $W = \{x \in \mathbb{R}^n : \|x\| < \rho\}$ has the following property: For $x \in W$ and $i \notin \Omega$, $|x_i - z_i| > \varepsilon$ and* $\operatorname{sgn}(z_i - x_i) = \operatorname{sgn}(z_i)$.

Lemma 1 follows from the minimality of $\psi(0)$ on L. Suppose $v \in L$ and $v_i \neq 0$ for some $i \in \Omega$. Then for sufficiently small λ, we would have $\lambda v \in L$ and $\psi(\lambda v) < \psi(0)$. This contradicts Theorem 1. Lemma 2 is a simple consequence of the continuity of each coordinate as a function of x.

The smoothness and strict convexity of the p-norms, $1 < p < \infty$, yield well-known uniqueness and characterization results for the corresponding l^p best approximation problems. While the 1-norm is not smooth, it does possess one-sided directional derivatives [11]. For x and $y \in \mathbb{R}^n$, $\|y\|_1 = 1$, define

$$D_y(x) = \lim_{t \to 0^+} \frac{\|x + ty\|_1 - \|x\|_1}{t}.$$

$D_y(x)$ is well defined for each such x, y pair and has the explicit formulation

$$D_y(x) = \sum_{S^c} \operatorname{sgn}(x_i)\, y_i + \sum_{S} |y_i|,$$

where $S = S_x = \{i : x_i = 0\}$ and S^c denotes the complement of S in $\{1, ..., n\}$.

Consider $w \in W \cap L$. By Lemmas 1 and 2, $w_i = 0$ for $i \in \Omega$ and $\operatorname{sgn}(z_i - w_i) = \operatorname{sgn}(z_i)$ for $i \notin \Omega$. Hence

$$\begin{aligned} D_v(z - w) &= \sum_{i \notin \Omega} \operatorname{sgn}(z_i - w_i)\, v_i + \sum_{i \in \Omega} |v_i| \\ &= \sum_{i \notin \Omega} \operatorname{sgn}(z_i)\, v_i + \sum_{i \in \Omega} |v_i| = D_v(z). \end{aligned}$$

This gives:

LEMMA 3. *Let $w \in W \cap L$ and $v \in \mathbb{R}^n$. Then $D_v(z - w) = D_v(z)$.*

For $v \in L$, Lemma 1 requires that $\operatorname{supp}(v) \subseteq \Omega^c$. The following lemma provides a partial converse.

LEMMA 4. *Suppose that $v \in V$ and $\operatorname{supp}(v) \subseteq \Omega^c$. Then $\lambda v \in L$ for small $\lambda > 0$.*

Proof: There is no loss in assuming $\|v\|_1 = 1$. Hence,

$$\begin{aligned} D_v(z) &= \sum_{i \notin \Omega} \operatorname{sgn}(z_i)\, v_i + \sum_{i \in \Omega} |v_i| = \sum_{i \notin \Omega} \operatorname{sgn}(z_i)\, v_i \\ &= -\textstyle\sum_{i \notin \Omega} \operatorname{sgn}(z_i)(-v_i) = -D_{-v}(z). \end{aligned}$$

This implies that $D_v(z) = D_{-v}(z) = 0$. Indeed, if not, then one must be negative. Without loss of generality, assume $D_v(0) < 0$. Then the definition of $D_v(z)$ requires that $\|\beta v - z\|_1 < \|z\|_1$ for small $\beta > 0$. This contradicts the l^1 optimality of 0. Now, for $|\lambda|$ sufficiently small, $\lambda v \in W$ and therefore $\text{sgn}(z_i) = \text{sgn}(z_i - \lambda v_i)$, for all $i \notin \Omega$. Thus,

$$\begin{aligned}\|z-\lambda v\|_1 &= \sum_{i=1}^{n} |z_i - \lambda v_i| = \sum_{i \notin \Omega} |z_i - \lambda v_i| \\ &= \sum_{i \notin \Omega} \text{sgn}(z_i - \lambda v_i)(z_i - \lambda v_i) = \sum_{i \notin \Omega} \text{sgn}(z_i)(z_i - \lambda v_i) \\ &= \|z\|_1 - \lambda D_v(z) = \|z\|_1,\end{aligned}$$

implying that $\lambda v \in L$. ∎

Directional derivatives provide a bound on the approximation error, $\|x - z\|_1$, near L as follows:

LEMMA 5. *Let* $w \in L \cap W$, $v \in V$ *with* $\|v\|_1 = 1$, *and* $D = \min(D_v(z), D_{-v}(z))$. *Then* $D \geqslant 0$ *and*

$$\theta(\lambda) = \frac{\|z - w + \lambda v\|_1 - \|z - w\|_1}{|\lambda|} \geqslant D.$$

Proof. Note that if $D < 0$, one of the directional derivatives would be negative. As in Lemma 4, this would contradict the l^1 optimality of 0. Fix $\lambda > 0$ and let $0 < t \leqslant 1$ hold. Then,

$$\begin{aligned}\|z - w + t\lambda v\|_1 - \|z - w\|_1 &= \|t(z - w + \lambda v) + (1 - t)(z - w)\|_1 - \|z - w\|_1 \\ &\leqslant t\,\|z - w + \lambda v\|_1 + (1 - t)\|z - w\|_1 - \|z - w\|_1 \\ &= t(\|z - w + \lambda v\|_1 - \|z - w\|_1).\end{aligned}$$

Hence,

$$\begin{aligned}\|z - w + \lambda v\|_1 - \|z - w\|_1 &= \lambda \lim_{\lambda t \to 0^+} \frac{\|z - w + \lambda t v\|_1 - \|z - w\|_1}{\lambda t} \\ &= \lambda D_v(z - w) = \lambda D_v(z).\end{aligned}$$

Thus, for $\lambda > 0$ then $\theta(\lambda)/|\lambda| \geqslant D_v(z) \geqslant D$ holds. Likewise, for the case $\lambda < 0$ essentially the same argument shows that $\theta(\lambda)/|\lambda| \geqslant D_{-v}(z) \geqslant D$ holds. ∎

Note that Lemma 5 is a directional strong uniqueness result at w in the direction of $v \in V$ whenever $D > 0$ holds. That is,

$$\|z - w + \lambda v\|_1 \geqslant \|z - w\|_1 + D\|w - \lambda v - w\|_1$$

holds for all $|\lambda|$ sufficiently small. For the special case of $w=0$, this has the form

$$\|z+\lambda v\|_1 \geqslant \|z\|_1 + D\|\lambda v\|_1.$$

If $D=0$, then no such directional strong uniqueness result exists. In fact, if $D=0$ occurs, then for small $\lambda>0$ both λv and $-\lambda v$ are in L. Indeed, suppose that $D_v(z)=0$. Then for $\lambda>0$ sufficiently small, $-\lambda v \in W$ so that

$$\begin{aligned}\|z+\lambda v\|_1 &= \sum_{i\notin\Omega} \operatorname{sgn}(z_i)(z_i+\lambda v_i) + \lambda \sum_{i\in\Omega} |v_i| \\ &= \|z\|_1 + \lambda \sum_{i\notin\Omega} \operatorname{sgn}(z_i)\, v_i + \lambda \sum_{i\in\Omega} |v_i| \\ &= \|z\|_1 + \lambda D_v(z) = \|z\|_1.\end{aligned}$$

Hence, $-\lambda v \in L$ and since 0 is in the relative interior of L, it follows that λv must be in L for sufficiently small $\lambda>0$. Thus, the lack of a local directional strong uniqueness estimate in this case corresponds to approaching 0 through L locally. On the other hand, if v is perpendicular to $K=\operatorname{span}(L)$ then a directional strong uniqueness estimate at 0 in the direction of v will hold. Rephrased, this implies that the approximation error must grow no more slowly than some fixed linear rate for all directions in $K^{\perp}$. This is established in Lemma 6.

LEMMA 6. *For arbitrary w and v satisfying $w\in W\cap L$, $v\in V$, with $\|v\|_1=1$ and $v\perp K$, there exists $k_0>0$ such that*

$$\|z-w+\lambda v\|_1 \geqslant \|z-w\|_1 + k_0|\lambda| \qquad \text{for each} \quad \lambda\in\mathbb{R}.$$

Proof. For $\lambda\neq 0$ Lemma 5 implies that

$$\|z-w+\lambda v\|_1 - \|z-w\|_1 \geqslant |\lambda|\, D.$$

Now, we claim that there exists $k_0>0$ such that $D\geqslant k_0$ for all $v\in V$ with $\|v\|_1=1$ and $v\perp K$. Indeed, assume that $D=D_v(z)$ without loss of generality. If D is not uniformily bounded away from zero we may construct a convergent sequence v_n from V such that $\|v_n\|_1=1$, $v_n\perp K$, and $D_{v_n}(z)<1/n$. Suppose $\lim_{n\to\infty} v_n = v^*$, which can be realized by passing to subsequences if necessary. Then $v^*\in V$, $v^*\perp K$, and $\|v^*\|_1=1$. We claim that $D_{v^*}(z)=0$. To see this, let $x\in V$. Then

$$D_x(z) = \sum_{i\in\Omega} |x_i| + \sum_{i\in\Gamma_x} |x_i| - \sum_{i\in\Psi_x} |x_i| + \sum_{i\in\Delta_x} |x_i|,$$

where $\Gamma_x = \{i: x_i z_i > 0\}$, $\Psi_x = \{i: x_i z_i < 0\}$, and $\Delta_x = \{i: i \notin \Omega \text{ and } x_i = 0\}$. Of course the final term contributes 0 to the expression. For large n, $\Gamma_{v^*} = \Gamma_{v_n}$ and $\Psi_{v^*} = \Psi_{v_n}$ so that

$$D_{v_n}(z) = \sum_{i \in \Omega} |v_{ni}| + \sum_{i \in \Gamma_{v^*}} |v_{ni}| - \sum_{i \in \Psi_{v^*}} |v_{ni}| + \sum_{i \in \Delta_{v^*}} \pm |v_{ni}|$$

and

$$D_{v^*}(z) = \sum_{i \in \Omega} |v_i^*| + \sum_{i \in \Gamma_{v^*}} |v_i^*| - \sum_{i \in \Psi_{v^*}} |v_i^*| + \sum_{i \in \Delta_{v^*}} |v_i^*|.$$

This implies that $|D_{v_n}(z) - D_{v^*}(z)| \leqslant \|v_n - v^*\|_1$. Hence $D_{v^*}(z) = 0$. Then, as in the comments following Lemma 5, $\lambda v^* \in L$ for small $|\lambda|$. This contradicts the fact that $v^* \perp K$. The result now follows, since both v and $-v$ satisfy the hypotheses. ∎

As before, let x^p be the best l^p approximation from V to z. Let w^p and v^p be the projections of x^p onto K and $K^\perp$, respectively. Then $x^p = w^p + v^p$. We will require the following inequalities:

Lemma 7. *There exist constants k_0, k_2, and k_3 so that for small $p > 1$ and $\delta = p - 1$,*

$$0 \geqslant k_0 \|v^p\|_1 + \delta\{n \|v^p\|_1 \ln \|v^p\|_1 + k_2 \|w^p\|_1^2\} - k_3\{\|w^p\|_1 + \|v^p\|_1\} \delta^2.$$

Proof. On $W \cap L$, $\psi(v) = \sum_{i=1}^n |v_i - z_i| \ln |v_i - z_i|$ reduces to $\psi(v) = \sum_{i \in \Omega^c} |v_i - z_i| \ln |v_i - z_i|$, where $|v_i - z_i| > \varepsilon$ for each $i \in \Omega^c$. Since 0 is in the relative interior of the polyhedral set L there exists ζ, $\rho > \zeta > 0$, so that the set $Q = \{v: v \in K, \|v\|_1 < \zeta\} \subseteq L$, where ρ is from Lemma 2. Write $\Omega^c = \Omega^+ \cup \Omega^-$, where $z_i > 0$ on Ω^+ and $z_i < 0$ on Ω^-. For $v \in K$, with $\|v\|_1 = 1$, compute the derivatives of $\psi(tv)$ for $|t| < \zeta$.

$$\begin{aligned}
\frac{d}{dt}\psi(tv) &= \frac{d}{dt}\sum_{\Omega^c} |tv_i - z_i| \ln |tv_i - z_i| \\
&= \frac{d}{dt}\sum_{i \in \Omega^+} (z_i - tv_i) \ln (z_i - tv_i) + \frac{d}{dt}\sum_{i \in \Omega^-} (tv_i - z_i) \ln (tv_i - z_i) \\
&= \sum_{i \in \Omega^+} -v_i[\ln(z_i - tv_i) + 1] + \sum_{i \in \Omega^-} v_i[\ln(tv_i - z_i) + 1],
\end{aligned}$$

and

$$\frac{d^2}{dt^2}\psi(tv) = \sum_{i \in \Omega^+} \frac{v_i^2}{(z_i - tv_i)} + \sum_{i \in \Omega^-} \frac{v_i^2}{(tv_i - z_i)}.$$

Evaluating this expression at $t=0$ yields

$$\frac{d^2}{dt^2}\psi(tv)|_{t=0}=\sum_{i\in\Omega^c}\frac{v_i^2}{|z_i|}.$$

Hence there exists $k_1>0$ such that for $v\in K$ and $\|v\|_1$ sufficiently small, $\psi(v)\geqslant k_1\|v\|_1^2$. Since $x^p\to 0$, we know that $w^p\in Q\subseteq L$ for p near 1. Then, for such values of p, Lemma 6 implies that

$$\|x^p-z\|_1-\|z\|_1=\left\|w^p+\|v^p\|_1\left(\frac{v^p}{\|v^p\|_1}\right)-z\right\|_1-\|w^p-z\|_1\geqslant k_0\|v^p\|_1. \quad (1)$$

Similarly, for small $p>1$, there exists $k_2>0$ such that $\psi(x^p)-\psi(0)\geqslant n\|v^p\|_1\ln\|v^p\|_1+k_2\|w^p\|_1^2$. Indeed, by previous scaling $\|z\|_1\leqslant 1/(2e^2)$. Set $\alpha=\min\{|z_i|: i\in\Omega^c\}$. Then $\alpha>0$ and there exists $p_0>1$ such that $p_0>p>1$ implies that $w^p\in Q$ and $\max_{1\leqslant j\leqslant n}\{|x_j^p|, |v_j^p|, |w_j^p|\}<(\alpha/10)$. Thus for any index $i\notin\Omega$ we have that

$$0.9\alpha<\max_i\{|z_i-x_i^p|, |z_i-v_i^p|, |z_i-w_i^p|\}<1/e^2.$$

The desired inequality follows from the fact that both $g(x)=-x\ln x$ and $h(x)=x\ln^2 x$ are strictly increasing on $[0, 1/e^2]$. To see this, observe that for such p

$$\psi(x^p)=\sum_{\Omega}|v_i^p|\ln|v_i^p|+\sum_{\Omega^c}|x_i^p-z_i|\ln|x_i^p-z_i|.$$

This implies that

$$\begin{aligned}\psi(x^p)-\psi(0)=&\sum_{\Omega}|v_i^p|\ln|v_i^p|\\&+\sum_{\Omega^c}\{|x_i^p-z_i|\ln|x_i^p-z_i|-|w_i^p-z_i|\ln|w_i^p-z_i|\}\\&+\sum_{\Omega^c}\{|w_i^p-z_i|\ln|w_i^p-z_i|-|z_i|\ln|z_i|\}. \quad (2)\end{aligned}$$

In the second summation, the Mean Value Theorem implies that there exists c_i^p between x_i^p and w_i^p such that

$$\sum_{\Omega^c}\{|x_i^p-z_i|\ln|x_i^p-z_i|-|w_i^p-z_i|\ln|w_i^p-z_i|\}=\sum_{\Omega^c}(1+\ln|z_i-c_i^p|)\,v_i^p.$$

Since $\ln \|v^p\| \to -\infty$ there exists p_1, with $p_0 > p_1 > 1$, such that for $p_1 > p > 1$ and $i \in \Omega^c$ it follows that $1 + \ln |z_i - c_i^p| \geq \ln \|v^p\|_1$. Hence, for $p_1 > p > 1$, the first two summations in (2) yield

$$\sum_{\Omega} |v_i^p| \ln |v_i^p| + \sum_{\Omega^c} \{|x_i^p - z_i| \ln |x_i^p - z_i| - |w_i^p - z_i| \ln |w_i^p - z_i|\} \geq n \|v^p\|_1 \ln \|v^p\|_1. \tag{3}$$

Consider now the final summation in (2). Since $w^p \in L$ the function $\psi(tw^p)$ has a local minimum at $t = 0$ and

$$\psi(w^p) - \psi(0) = \sum_{\Omega^c} \{|w_i^p - z_i| \ln |w_i^p - z_i| - |z_i| \ln |z_i|\} = \frac{1}{2} \sum_{\Omega^c} \frac{(w_i^p)^2}{|z_i - \gamma_i w_i^p|},$$

for some set γ_i, $0 < \gamma_i < 1$. The final term above is just the remainder term of a first order Taylor expansion for $\psi(w^p)$ expanded about $t = 0$. Hence, using the equivalence of norms on $\mathbb{R}^n$ implies that the final summation satisfies

$$\psi(w^p) - \psi(0) \geq k_2 \|w^p\|_1^2 \tag{4}$$

for some $k_2 > 0$. Combining (3) and (4) yields

$$\psi(x^p) - \psi(0) \geq n \|v^p\|_1 \ln \|v^p\|_1 + k_2 \|w^p\|_1^2. \tag{5}$$

We now bound the difference $\|x^p - z\|_p^p - \|z\|_p^p$. By the p-norm optimality of x^p this difference must be negative. Now expand $\|z\|_p^p$ and $\|x^p - z\|_p^p$ into Taylor series about 1 to obtain

$$\|x^p - z\|_p^p = \|x^p - z\|_1 + \delta\psi(x^p) + \frac{\delta^2}{2} \ln^2 |x_i^p - z_i| + \delta^2 \sum_{r=2}^{\infty} \frac{\delta^{r-2}}{r!} \sum_{\Omega^c} |x_i^p - z_i| \ln^r |x_i^p - z_i| \tag{6}$$

and

$$\|z\|_p^p = \|z\|_1 + \delta\psi(0) + \frac{\delta^2}{2} \ln^2 |z_i| + \delta^2 \sum_{r=2}^{\infty} \frac{\delta^{r-2}}{r!} \sum_{\Omega^c} |z_i| \ln^r |z_i|, \tag{7}$$

where $\delta = p - 1$ and the convergence in each series in uniform. To subtract (7) from (6) consider first the difference in the $\|\cdot\|_1$ terms. By (1),

$\|x^p - z\|_1 - \|z\|_1 \geqslant k_0 \|v^p\|_1$. Similarly, (5) bounds the $\delta\psi$ terms. To bound the series terms invoke the Mean Value Theorem to get

$$\delta^2 \sum_{r=2}^{\infty} \left[\frac{\delta^{r-2}}{r!} \sum_{\Omega^c} |x_i^p - z_i| \ln^r |x_i^p - z_i| - |z_i| \ln^r |z_i| \right]$$
$$= \delta^2 \sum_{r=2}^{\infty} \left[\frac{\delta^{r-2}}{r!} \sum_{\Omega^c} (\ln^r |\theta_i x_i^p - z_i| + r \ln^{r-1} |\theta_i x_i^p - z_i|)\, x_i^p \right]$$

for some set θ_i, $0 < \theta_i < 1$. By our restrictions on p, $|\theta_i x_i^p - z_i| \geqslant 4\alpha/5$. Hence there exists $k_3 > 0$ such that the above difference is bounded above by $k_3 \|x^p\|_1 \delta^2$. Combining these terms yields

$$0 \geqslant \|x^p - z\|_p^p - \|z\|_p^p$$
$$\geqslant k_0 \|v^p\|_1 + \delta\{n \|v^p\|_1 \ln \|v^p\|_1 + k_2 \|w^p\|_1^2\} - k_3(\|x^p\|_1)\, \delta^2.$$

Since $\|x^p\|_1 \leqslant \|v^p\|_1 + \|w^p\|_1$ we have

$$0 \geqslant \|x^p - z\|_p^p - \|z\|_p^p$$
$$\geqslant k_0 \|v^p\|_1 + \delta\{n \|v^p\|_1 \ln \|v^p\|_1 + k_2 \|w^p\|_1^2\} - k_3\{\|w^p\|_1 + \|v^p\|_1\}\, \delta^2,$$

which is the desired result. ∎

We can now prove the main result of this paper.

THEOREM 2. *The net x^p converges to the natural best approximation at a rate no worse than $O(p-1)$.*

Proof. By Lemma 7, there exist positive constants k_0, k_2, k_3, and $p_1 > 1$ so that if $p_1 > p > 1$,

$$0 \geqslant k_0 \|v^p\|_1 + \delta\{n \|v^p\|_1 \ln \|v^p\|_1 + k_2 \|w^p\|_1^2\} - k_3\{\|w^p\|_1 + \|v^p\|_1\}\, \delta^2.$$

By replacing k_0 by some $k_4 > 0$, we may absorb the final term into the first and find $p_2 > 1$ so that for $p_2 \geqslant p > 1$

$$0 \geqslant k_4 \|v^p\|_1 + \delta\{n \|v^p\|_1 \ln \|v^p\|_1 + k_2 \|w^p\|_1^2\} - k_3 \|w^p\|_1\, \delta^2$$

and

$$\|v^p\|_1 \leqslant |\ln \|v^p\|_1|$$

hold with the second inequality following from the fact that $\|v^p\|_1 \to 0$ as $p \to 1^+$. Set $\beta = \exp(-k_4/(2n))$ and $\eta = (1+\beta)/2$ and note that $0 < \beta < \eta < 1$ holds. Thus, there exists p_3, $1 < p_3 < \min(p_2, 1 + e^{-1})$, such that

$k_4/(2k_3(p-1)^2) \leqslant (\eta/\beta)^{1/(p-1)}$ and $\eta \leqslant ((p-1)/2)^{p-1}$ hold for $1 < p \leqslant p_3$. Now for a given p, $p_3 > p > 1$, suppose that

$$|n\delta \, \|v^p\|_1 \ln \|v^p\|_1| > k_3 \delta^2 \, \|w^p\|_1 \tag{8}$$

holds. Then $0 \geqslant k_4 \|v^p\|_1 + \delta 2n \|v^p\|_1 \ln \|v^p\|_1 + \delta k_2 \|w^p\|_1^2$ and so $0 \geqslant k_4 \|v^p\|_1 + \delta 2n \|v^p\|_1 \ln \|v^p\|_1$. This implies that $\beta^{1/\delta} \geqslant \|v^p\|_1$. Note also that (8) implies that $\beta^{1/\delta} \geqslant (2k_3\delta^2 \|w^p\|_1/k_4)$ holds since $|x \ln x|$ is increasing on $(0, e^{-1})$. Thus, η satisfies $\eta^{1/\delta} \geqslant \|v^p\|_1$ and

$$\eta^{1/\delta} = (\eta/\beta)^{1/\delta} \beta^{1/\delta} \geqslant (\eta/\beta)^{1/\delta} (2k_3\delta^2 \|w^p\|_1/k_4) \geqslant \|w^p\|_1.$$

From this it follows that x^p, corresponding to this p, satisfies $\|x^p\|_1 \leqslant \|v^p\|_1 + \|w^p\|_1 \leqslant 2\eta^{1/\delta}$. Since x^x is decreasing from 1 on $(0, e^{-1})$ it follows by the restrictions placed on η and p_3 above that $\|x^p\|_1 \leqslant \delta$ also holds in this case.

On the other hand, if (8) does not hold for a given p, $1 < p < p_3$, then

$$|n\delta \, \|v^p\|_1 \ln \|v^p\|_1| \leqslant k_3 \delta^2 \, \|w^p\|_1$$

implies $0 \geqslant k_4 \|v^p\|_1 + \delta k_2 \|w^p\|_1^2 - 2k_3 \|w^p\|_1 \delta^2$ and hence $0 \geqslant k_2 \|w^p\|_1 - 2k_3\delta$. Thus, $\|w^p\|_1$ is $O(\delta)$. In this case we also have by our choice of p_2 that

$$\|v^p\|_1^2 \leqslant \|v^p\|_1 \, |\ln \|v^p\|_1| \leqslant k_3 \delta \, \|w^p\|_1/n$$

so that $\|v^p\|_1$ is $O(\delta)$ and $\|x^p\|_1$ is $O(\delta)$, $\delta = p-1$, as desired. ∎

Note that if (8) holds for all p near 1 then convergence of at least exponential rate holds. This must always be the case if $x^p \perp K$ for all p sufficiently close to 1. This yields the following theorem:

THEOREM 3. *If $x^p \perp K$ for all p sufficiently close to* 1 *then there exists* γ, $1 > \gamma > 0$, *such that x^p converges to the natural best approximation at a rate no worse than* $O(\gamma^{1/(p-1)})$.

For the special case in which L is a singleton, Theorem 3 yields the following:

COROLLARY 1. *If L is a singleton there exists* γ, $1 > \gamma > 0$, *such that x^p converges to the natural best approximation at a rate no worse than* $O(\gamma^{1/(p-1)})$.

The examples given earlier illustrate these rates and show the rates to be sharp. The following example shows that these results need not hold in general finite dimensional L^1 subspace approximation problems.

EXAMPLE 3. Consider the 1-dimensional problem of approximating $f(x)=1$ on $[0, 1]$ from the subspace of functions $V=\{ax: a \in \mathbb{R}\}$. For $p>1$ it is immediate that there exists a unique best approximation $x^p = a^p x$. That is,

$$\int_0^1 |a^p x - 1|^p \, dx = \min_{a \in \mathbb{R}} \int_0^1 |ax - 1|^p \, \mathrm{dx}.$$

Furthermore, it is easily seen that $a^p \geqslant 1$ for $p>1$. Thus, finding best approximations is equivalent to minimizing $H_p(r)$, $r \geqslant 1$, $p \geqslant 1$, where

$$\begin{aligned} H_p(r) &= \int_0^1 |rx-1|^p \, dx \\ &= \int_0^{1/r} (1-rx)^p \, dx + \int_{1/r}^1 (rx-1)^p \, dx = \frac{(1+(r-1)^{p+1})}{(p+1)r}. \end{aligned}$$

Now $H_p'(r) = (-1+(pr+1)(r-1)^p)/((p+1)\, r^2)$. Thus, for $p=1$, it is easily seen that the problem,

$$\min_{r \geqslant 1} \int_0^1 |rx-1| \, dx,$$

has a unique solution $a^1 = \sqrt{2}$. Since $a^p \to a^1$, we need only consider $1.4 \leqslant r \leqslant 1.5$ for small $p \geqslant 1$. For small $p \geqslant 1$, a^p is a solution to $(pr+1)(r-1)^p - 1 = 0$. Note that

$$\begin{aligned} (pr+1)(r-1)^p - 1 &= (r+1)(r-1)(r-1)^{p-1} - 1 + (p-1)\, r(r-1)^p \\ &= [(r+1)(r-1)](1-(2-r))^{p-1} \\ &\quad -1 + (p-1)\, r(r-1)^p. \end{aligned}$$

Applying the Mean Value Theorem to $(1-x)^{p-1}$ then yields

$$\begin{aligned} (pr+1)(r-1)^p - 1 &= (r+1)(r-1)[1-(p-1)(1-\zeta)^{p-2}(2-r)] - 1 \\ &\quad + (p-1)\, r(r-1)^p, \end{aligned}$$

where ζ is between 0 and $2-r$. For the values of r of interest, $0 \leqslant \zeta \leqslant 0.6$ since $1.4 \leqslant a^p \leqslant 1.5$ here. Thus, for small $p \geqslant 1$, $(1-\zeta)^{p-2} \in [1, 2.5]$, $(1-\zeta)^{p-2}(2-a^p) \in [0.5, 1.5]$, and $a^p(a^p-1)^p \in [(1.4)(0.4)^{3/2}, (1.5)(0.5)^{3/2}] \subseteq [0.3, 0.75]$. Now $H_p'(a^p)=0$ implies that

$$\begin{aligned} 0 &= (pa^p+1)(a^p-1)^p - 1 \\ &= (a^p+1)(a^p-1)[1-(p-1)(1-\zeta)^{p-2}(2-a^p)] \\ &\quad -1 + (p-1)\, a^p(a^p-1)^p. \end{aligned}$$

Hence

$$(a^p+1)(a^p-1)-1=(p-1)([(a^p)^2-1](1-\zeta)^{p-2}(2-a^p)-a^p(a^p-1)^p)$$
$$=(p-1)\,\omega^p.$$

Using the above estimates, ω^p, defined in the previous equation, can be seen to be bounded. That is, there exist positive constants C and D such that $C\leqslant\omega^p\leqslant D$. Now $(a^p)^2=2+(p-1)\,\omega^p$ and so that $a^p=(2+(p-1)\,\omega^p)^{1/2}$. Finally, expanding $(1+\alpha)^{1/2}$ we may write $a^p=\sqrt{2}+(p-1)\,\gamma^p$ where there exist positive constants J and K with $J<\gamma^p<K$. Thus, we have a linear rate of convergence even through L is a singleton.

It remains open whether this rate holds in general in $C[0, 1]$, or whether even slower convergence may occur. Also open is the question of the effect of constraints on the rate of convergence. The Pólya-1 algorithm is known to converge as long as the approximating set is convex. However, it is not know whether the imposition of constraints slows or accelerates convergence.

References

1. P. M. Anselone and J. Davis, Perturbations of best approximation problems, *Number Math.* **21** (1973), 63–69.
2. P. Bloomfield and W. L. Steiger, "Least Absolute Deviations," Birkhäuser, Boston, 1983.
3. J. Descloux, Approximations in L^p and Chebychev approximations, *J. SIAM* **11** (1963), 1017–1026.
4. A. G. Egger and R. Huotari, Rate of convergence of the discrete Pólya algorithm, *J. Approx. Theory* **60** (1990), 24–30.
5. A. Egger and G. D. Taylor, Dependence on p of the best L^p approximation operator, *J. Approx. Theory* **49** (1987), 274–282.
6. J. Fischer, The convergence of the best discrete linear L_p approximation as $p\to 1$, *J. Approx. Theory* **39** (1983), 374–385.
7. D. Jackson, Note on the median of a set of numbers, *Bull. Amer. Math. Soc.* **28** (1921), 160–164.
8. D. Landers and L. Rogge, Natural choice of L_1-approximants, *J. Approx. Theory* **33** (1981), 268–280.
9. G. Pólya, Sur un algorithme toujours convergent pour obtenir les polynômes de meilleure approximation de Tchebycheff pour une fonction continue quelconque, *C.R. Acad. Sci. Paris* **157** (1913), 840–843.
10. J. Peetre, Approximation of norms, *J. Approx. Theory* **3** (1970), 243–260.
11. A. Pinkus, "On L^1 Approximation," Cambridge Univ. Press, Cambridge, 1989.
12. V. Stover, "The strict Approximation and Continuous Selections for the Metric Projection," Ph.D. dissertation, UCSD, 1981.