Available at WWW.MATHEMATICSWEB.ORG
POWERED BY SCIENCE DIRECT
Journal of Approximation Theory 124 (2003) 89–95
JOURNAL OF Approximation Theory
http://www.elsevier.com/locate/jat

Convergence of the weak dual greedy algorithm in L_p-spaces ☆

M. Ganichev and N.J. Kalton*

Department of Mathematics, University of Missouri, 201 Mathematical Sciences Building, Columbia, MO 65211, USA

Received 15 December 2002; accepted in revised form 26 June 2003

Communicated by Amos Ron

Abstract

We prove that the weak dual greedy algorithm converges in any subspace of a quotient of L_p when $1<p<\infty$.

A subset D of a (real) Banach space X is called a *dictionary* if

(i) D is *normalized* i.e. if $g \in D$ implies $||g|| = 1$.
(ii) D is *symmetric* i.e. $D = -D$.
(iii) D is *fundamental* i.e. $[D] = X$.

Given $x \in X$ we are interested in algorithms which generate a sequence of approximations by n-term linear combinations of members of the dictionary. Many examples of such algorithms have been introduced and studied in approximation theory. We refer to the paper of Temlyakov [11] for a survey of possible algorithms. A desirable feature of a given algorithm is that the sequence of approximations always converge to x (i.e. the algorithm converges). Surprisingly, relatively few general convergence theorems are known for most of the basic algorithms available. In this paper we consider the so-called weak dual greedy algorithm (WDGA).

☆ The authors were partially supported by NSF Grants DMS-9870027 and DMS-0244515.
*Corresponding author. Fax: +573-882-1869.
E-mail addresses: ganichev@math.missouri.edu (M. Ganichev), nigel@math.missouri.edu (N.J. Kalton).

0021-9045/$ - see front matter
doi:10.1016/S0021-9045(03)00133-3

The weak dual greedy algorithm is a natural generalization to Banach spaces of the so-called pure greedy algorithm (PGA) and its modification the weak greedy algorithm (WGA) for Hilbert spaces. The (PGA) was introduced and first studied by Huber [3]; its convergence was shown by Jones [4]. For the fact that the (WGA) converges in a Hilbert space see [8]; more general results are given in [9] and [7]. Very little is known about the convergence of the (WDGA) for an arbitrary dictionary in a Banach space; see [2]. In [11] it is conjectured that the (WDGA) converges whenever X is a uniformly smooth Banach space with power-type modulus of smoothness. Our main theorem in this paper is that for any subspace of a quotient of L_p when $1<p<\infty$ the (WDGA) converges for any dictionary, thus proving a special case of the conjecture in [11]. As noted by one of the referees the convergence of the (WDGA) in L_p for $1<p<\infty$ was previously unknown even for the dictionary consisting of the Haar basis.

For any $x\in X$ we define the *descent rate* associated to the dictionary D by

$$\rho_D(x) = \sup_{t>0}\sup_{g\in D}\frac{||x|| - ||x-tg||}{t} = \sup_{g\in D}\lim_{t\to 0+}\frac{||x|| - ||x-tg||}{t}. \tag{1}$$

By the Hahn–Banach theorem

$$\rho_D(x) = \sup_{\substack{||x^*||=1 \\ x^*(x)=||x||}}\sup_{g\in D} x^*(g). \tag{2}$$

We will usually deal with Banach spaces with a Gateaux differentiable norm, i.e. such that for each $x\in X\backslash\{0\}$ there is a unique $x^*\in X$ with $x^*(x)=||x||$ and $||x^*||=1$. We denote this functional by F_x. The map $x\to F_x$ is norm to weak*-continuous on $X\backslash\{0\}$; see [1, p. 7]. We set $F_0=0$ for notational convenience. Thus in this case we have

$$\rho_D(x) = \sup_{g\in D} F_x(g). \tag{3}$$

Suppose X has a Gateaux differentiable norm. Let us describe the *weak dual greedy algorithm* (*WDGA*) with parameter $0<c<1$. Suppose $x\in X$. We construct a sequence $(g_n)_{n=1}^{\infty}$ with $g_n\in D$ and a sequence $(t_n)_{n=1}^{\infty}$ of reals with $t_n\geqslant 0$. Let $x_0=x$ and construct $(x_n)_{n=0}^{\infty}, (g_n)_{n=1}^{\infty}, (t_n)_{n=1}^{\infty}$ inductively as follows. For each $n\geqslant 1$ pick $g_n\in D$ so that

$$F_{x_{n-1}}(g_n)\geqslant c\rho_D(x_{n-1}). \tag{4}$$

Pick $t_n\geqslant 0$ so that

$$||x_{n-1}-t_ng_n|| = \min_{t\geqslant 0}||x_{n-1}-tg_n||. \tag{5}$$

Finally set

$$x_n = x_{n-1}-t_ng_n. \tag{6}$$

Thus the n-term approximation to x is given by $\sum_{k=1}^{n} t_k g_k$ and the error is given by x_n. The (WDGA) is said to converge at x if $\lim_{n\to\infty} x_n = 0$ and hence $x = \sum_{n=1}^{\infty} t_n g_n$. The (WDGA) (with parameter c) is said to converge if it converges for every $x \in X$.

Let us remark that Temlyakov [11] considers this algorithm for a sequence of parameters $(c_n)_{n=1}^{\infty}$ with $c_n > 0$ replacing c. Thus in place of (4) one has

$$F_{x_{n-1}}(g_n) \geqslant c_n \rho_D(x_{n-1}). \tag{7}$$

A necessary and sufficient condition in Hilbert spaces for convergence of the (WDGA) with a sequence $(c_n)_{n=1}^{\infty}$ of parameters is given in [10].

Lemma 1. *Let X be a Banach space with a Gateaux differentiable norm and let D be a dictionary in X. Suppose $x = x_0 \in X$ and $0 < c < 1$. Suppose further that $(x_n)_{n=0}^{\infty}$, $(g_n)_{n=1}^{\infty}$ and $(t_n)_{n=1}^{\infty}$ are sequences with $g_n \in D$, $t_n > 0$ which satisfy* (4) *and* (6) *but not necessarily* (5). *Suppose that*

$$\frac{\|x_{n-1}\| - \|x_n\|}{t_n} \geqslant c\rho_D(x_{n-1}) \qquad n \geqslant 1. \tag{8}$$

Then if $\sum_{n=1}^{\infty} t_n = \infty$ we have $\lim_{n\to\infty} x_n = 0$ and

$$x = \sum_{n=1}^{\infty} t_n g_n.$$

Proof. Let $s_n = t_1 + \cdots + t_n$. Then we note that

$$\sum_{n=2}^{\infty} \log \frac{s_n - t_n}{s_n} = -\infty$$

and so

$$\sum_{n=1}^{\infty} \frac{t_n}{s_n} = \infty.$$

Now since $\|x_n\|$ is monotone decreasing the series $\sum_{n=1}^{\infty}(\|x_{n-1}\| - \|x_n\|)$ is convergent. We deduce the existence of a sequence (n_k) such that

$$\lim_{k\to\infty} \frac{s_{n_k+1}(\|x_{n_k}\| - \|x_{n_k+1}\|)}{t_{n_k+1}} = 0.$$

Let $\varepsilon_k = s_{n_k}\rho_D(x_{n_k})$. By (8), since $s_{n_k} < s_{n_k+1}$, we have $\lim_{k\to\infty} \varepsilon_k = 0$. Note in particular

$$\lim_{k\to\infty} \rho_D(x_{n_k}) = 0. \tag{9}$$

Now if $0 \leqslant l \leqslant n_k - 1$

$$\left| F_{x_{n_k}}\left(\sum_{j=l+1}^{n_k} t_j g_j \right) \right| \leqslant \sum_{j=1}^{n_k} t_j \rho_D(x_{n_k}) \leqslant \varepsilon_k.$$

Hence

$$|F_{x_{n_k}}(x_l) - ||x_{n_k}||| \leqslant \varepsilon_k, \qquad 0 \leqslant l \leqslant n_k.$$

Let x^* be any weak*-cluster point of the sequence $(F_{x_{n_k}})_{k=1}^{\infty}$. Then if $L = \lim_{n\to\infty} ||x_n||$ we have

$$x^*(x_l) = L \qquad 0 \leqslant l < \infty.$$

If $L \neq 0$ we will obtain a contradiction. In this case $x^* \neq 0$ and so $\sup_{g\in D} x^*(g) = \theta > 0$. But then

$$\limsup_{k\to\infty} \sup_{g\in D} F_{x_{n_k}}(g) \geqslant \theta.$$

This gives a contradiction to (9). □

The key to the proof of the main theorem is the following simple inequality. If $a \in \mathbb{R}$ we write $\mathrm{sgn}(a) = a/|a|$ when $a \neq 0$ and $\mathrm{sgn}0 = 0$.

Lemma 2. *Suppose $1 < p < \infty$. There is $C_p > 0$ such that for any real numbers a and b*

$$b|a+b|^{p-1}sgn(a+b) - b|a|^{p-1}sgn(a) \leqslant C_p(|a+b|^p - pb|a|^{p-1}sgn(a) - |a|^p).$$

Proof. By homogeneity it is enough to consider the case $a = 1$. Note that $|1+b|^p - pb - 1 \geqslant 0$ with equality only at $b = 0$. Let

$$\varphi(b) = \frac{b(|1+b|^{p-1}\mathrm{sgn}(1+b) - 1)}{|1+b|^p - p\ b - 1}, \qquad b \neq 0.$$

Then

$$\lim_{b\to 0} \varphi(b) = \frac{2}{p},$$

$$\lim_{b\to\infty} \varphi(b) = 1,$$

$$\lim_{b\to-\infty} \varphi(b) = 1.$$

Since the function $\varphi(b)$ is continuous these estimates imply an upper bound $\varphi(b) \leqslant C_p$ for all $b \neq 0$ and the lemma follows. □

Let us say that a Banach space X with a Gateaux differentiable norm has *property* Γ if there is a constant $0 < \gamma \leqslant 1$ such that if $x, y \in X$ and $F_x(y) = 0$ then

$$||x+y|| \geqslant ||x|| + \gamma F_{x+y}(y). \tag{10}$$

As pointed out by one of the referees, this condition has been considered previously in the context of greedy algorithms by Livshits [6, Theorem 1.2] although his formulation is somewhat different.

We recall that if X is a Banach space and E is a closed subspace then the quotient space X/E is a Banach space under the norm

$$||x+E|| = \inf_{e \in E} ||x+e||.$$

If X is reflexive (or more generally if E is reflexive) then the infimum is attained, i.e.

$$||x+E|| = \min_{e \in E} ||x+e||.$$

In the case $p \geqslant 2$ the following proposition was essentially proved in [6], Corollary 1.3.

Proposition 3. *If* $1<p<\infty$, *every quotient of a subspace of* L_p *has property* Γ.

Proof. We first show that $L_p(0,1)$ has property Γ. Suppose $x, y \in L_p(0,1)$ and $F_x(y)=0$. Then by Lemma 2,

$$\begin{aligned} & y(s)|x(s)+y(s)|^{p-1}\operatorname{sgn}(x(s)+y(s)) \\ & \quad \leqslant C_p(|x(s)+y(s)|^p - |x(s)|^p) + (1-pC_p)y(s)|x(s)|^{p-1}\operatorname{sgn}(x(s)). \end{aligned}$$

We have

$$\int y(s)|x(s)|^{p-1}\operatorname{sgn}(x(s))\, ds = 0$$

and so by integration we have

$$\int_0^1 y(s)|x(s)+y(s)|^{p-1}\operatorname{sgn}(x(s)+y(s))ds \leqslant C_p(||x+y||^p - ||x||^p).$$

Thus (noting that $||x+y|| \geqslant ||x||$ since $F_x(y)=0$),

$$\begin{aligned} ||x+y||^{p-1}F_{x+y}(y) &= \int_0^1 |x(s)+y(s)|^{p-1}\operatorname{sgn}(x(s)+y(s))y(s)\, ds \\ &\leqslant C_p(||x+y||^p - ||x||^p) \\ &\leqslant pC_p||x+y||^{p-1}(||x+y|| - ||x||) \end{aligned}$$

and Γ follows with $\gamma = (pC_p)^{-1}$.

It is clear property Γ passes to subspaces, so we prove it also passes to quotients at least for reflexive spaces. Suppose X has property Γ and is reflexive. Let Y be a quotient, i.e. $Y = X/E$ for some subspace E of X. Let $Q: X \to Y$ be the quotient map $x \to x+E$. If $x, y \in Y$ with $F_x(y)=0$, we may pick $u, w \in X$ so that $Qu = x$, $Qw = x+y$ and $||u|| = ||x||$, $||w|| = ||x+y||$. Note Y also has a Gateaux differentiable norm and furthermore we have $F_u = F_x \circ Q$ and $F_w = F_{x+y} \circ Q$. Hence $F_u(w-u) = F_x(y) = 0$ and so

$$||x+y|| = ||w|| \geqslant ||u|| + \gamma F_w(w-u) = ||y|| + \gamma F_{x+y}(y)$$

and the proposition follows. $\square$

Theorem 4. *Suppose X is a quotient of a subspace of L_p for some $1<p<\infty$. Then the (WDGA) converges for any dictionary and any parameter.*

Proof. We note that if X is a quotient of a subspace of L_p for $1<p<\infty$ has property Γ by Proposition 3; also the norm on X is Fréchet differentiable since X is uniformly smooth (this follows quickly from the duality properties of uniform smoothness and uniform convexity, see [5, p. 61]).

We therefore show that if X has property Γ and a Fréchet differentiable norm then the (WDGA) with parameter $0<c<1$ converges for every dictionary D and every $x\in X$. We shall use the fact that if the norm is Fréchet differentiable then the map $x\to F_x$ is norm-continuous on $X\backslash\{0\}$ (see e.g. [1, p. 7, Proposition 1.8]).

Let $x = x_0$ and suppose $(x_n)_{n=0}^{\infty}$, $(g_n)_{n=1}^{\infty}$ and $(t_n)_{n=1}^{\infty}$ are selected according to (4), (5) and (6). If $t_n = 0$ for any n then $\rho_D(x_{n-1}) = 0$ and so since D is fundamental, $x_k = 0$ for $k\geqslant n-1$. Thus we consider the case $t_n>0$ for all n. We note that $F_{x_n}(g_n) = 0$ (by (5)) and so

$$||x_{n-1}||\geqslant||x_n|| + \gamma t_n F_{x_{n-1}}(g_n)$$

and hence,

$$c\gamma\rho_D(x_{n-1})\leqslant\frac{||x_{n-1}|| - ||x_n||}{t_n}.$$

By Lemma 1 (with c replaced by $c\gamma$) we have $\lim_{n\to\infty} x_n = 0$ if $\sum_{n=1}^{\infty} t_n = \infty$. To complete the proof we consider the case $t_n>0$ for all n but $\sum_{n=1}^{\infty} t_n<\infty$ and $\lim_{n\to\infty} x_n = x_\infty\neq 0$. By the Fréchet differentiability of the norm we have $\lim_{n\to\infty}||F_{x_n} - F_{x_\infty}|| = 0$. Thus $\lim_{n\to\infty}||F_{x_n} - F_{x_{n-1}}|| = 0$. But observe that $F_{x_n}(g_n) = 0$ by (5) and so $\lim_{n\to\infty} F_{x_{n-1}}(g_n) = 0$. This implies by (4) that $\lim_{n\to\infty}\rho_D(x_{n-1}) = 0$ and so $F_{x_\infty}(g) = 0$ for every $g\in D$, which of course contradicts the fact that D is fundamental. □

It is possible to glean a little more from this argument. Let us introduce another algorithm which we call the *modified dual greedy algorithm* (*MDGA*) with parameter $0<c<1$ as follows. Given $x\in X$ let (x_n), (g_n) and (t_n) be chosen according to (4) and (6) but with (5) replaced by

$$F_{x_{n-1}-t_n g_n}(g_n) = cF_{x_{n-1}}(g_n). \tag{11}$$

Thus in the (MDGA) we do not choose t_n to minimize the error but in general we make a smaller choice of t_n, selecting a point at which the rate of decrease of $||x_{n-1} - tg_n||$ has fallen to a fixed fraction of its initial rate of descent.

Theorem 5. *Suppose that X is a Banach space with Gateaux differentiable norm and D is a dictionary in X. Then the (MDGA) converges (for any parameter $0<c<1$) provided either the norm is Fréchet differentiable or D is relatively norm compact.*

Proof. The argument is similar to the preceding theorem. Suppose (x_n), (g_n) and (t_n) are selected according to (4), (11) and (6). As before the case $\sum_{n=1}^{\infty} t_n = \infty$ is resolved by Lemma 1. In fact,

$$\|x_{n-1}\| - \|x_n\| \geqslant t_n F_{x_n}(g_n) = c t_n F_{x_{n-1}}(g_n) \geqslant c^2 t_n \rho_D(x_{n-1}).$$

Thus we can apply Lemma 1 (replacing c by c^2) to deduce that $\lim_{n\to\infty} x_n = 0$.

We can therefore suppose $t_n > 0$ for all n but $\sum_{n=1}^{\infty} t_n < \infty$. We again suppose $\lim_{n\to\infty} x_n = x_\infty \neq 0$.

Now in either case of the theorem we have

$$\lim_{n\to\infty} \sup_{g\in D} |F_{x_n}(g) - F_{x_\infty}(g)| = 0. \tag{12}$$

If the norm is Fréchet differentiable this follows since the map $x \to F_x$ is norm continuous on $X\backslash\{0\}$. If D is relatively norm compact it follows since F_{x_n} converges to F_{x_∞} weak* by Gateaux differentiability of the norm. This means that

$$\lim_{n\to\infty} \sup_{g\in D} |F_{x_n}(g) - F_{x_{n-1}}(g)| = 0$$

and hence

$$(1-c) \lim_{n\to\infty} F_{x_{n-1}}(g_n) = 0.$$

Thus $\lim_{n\to\infty} \rho_D(x_n) = 0$ which implies by (12) that $\rho_D(x_\infty) = 0$ which is a contradiction. □

References

[1] R. Deville, G. Godefroy, V. Zizler, Smoothness and Renormings in Banach Spaces, Pitman Monographs and Surveys in Pure and Applied Mathematics, Vol. 64, Longman Scientific and Technical, Harlow, 1993.

[2] S.J. Dilworth, D. Kutzarova, V.N. Temlyakov, Convergence of some greedy algorithms in Banach spaces, J. Fourier Anal. Appl. 8 (2002) 489–505.

[3] P.J. Huber, Projection pursuit, Ann. Statist. 13 (1985) 435–525.

[4] L.K. Jones, On a conjecture of Huber concerning the convergence of projection pursuit regression, Ann. Statist. 15 (1987) 880–882.

[5] J. Lindenstrauss, L. Tzafriri, Classical Banach Spaces, Vol. II, Function Spaces, Springer, Berlin, 1979.

[6] E.D. Livshits, On convergence of greedy algorithms in Banach spaces, Mat. Zametki 73 (2003) 371–389.

[7] E.D. Livshits, V.N. Temlyakov, On the convergence of a weak greedy algorithm, (Russian) Tr. Mat. Inst. Steklova 232 (2001) 236–247; translation in Proc. Steklov Inst. Math. 232(1) (2001) 229–239.

[8] L. Rejto, G.G. Walter, Remarks on projection pursuit regression and density estimation, Stochastic Anal. Appl. 10 (1992) 213–222.

[9] V.N. Temlyakov, Weak greedy algorithms, Adv. Comput. Math. 12 (2000) 213–227.

[10] V.N. Temlyakov, A criterion for convergence of weak greedy algorithms, Adv. Comput. Math. 17 (2002) 269–280.

[11] V.N. Temlyakov, Nonlinear methods of approximation, Found. Comput. Math. 3 (2003) 33–107.