JOURNAL OF APPROXIMATION THEORY 38, 1–8 (1983)

The Diliberto–Straus Algorithm in $L_1(X \times Y)$

WILLIAM A. LIGHT

Mathematics Department, Cartmel College,
University of Lancaster, Lancaster, England

Communicated by Carl de Boor

Received January 14, 1981; revised May 5, 1982

1. INTRODUCTION

In 1951 Diliberto and Straus published a paper [3] in which they developed an algorithm for generating the closest point in $M = C(X) + C(Y)$ to any $f \in C(X \times Y)$, where $X = Y = [0, 1]$. In fact, they were only able to show that the iterates f_n in their algorithm had the properties

(i) $\|f_n\|_\infty \to \operatorname{dist}(f, M)$;

(ii) $\{f_n\}$ contains cluster points.

Later Aumann [1] showed that the iterates do in fact converge to a function $f - m$, where $m \in M$ and m is a closest point to f from M.

In this paper we consider the same problem in the space $L_1(X \times Y)$ with $M = L_1(X) + L_1(Y)$. Several results are already known about this setting—see [2] for details. In particular, if the natural generalisation of the algorithm to $L_1(X \times Y)$ is used, then there exist functions $f \in L_1(X \times Y)$ for which $\|f_n\|_1 \nrightarrow \operatorname{dist}(f, M)$. We shall investigate the conditions under which the convergence of $\|f_n\|_1$ to $\operatorname{dist}(f, M)$ holds.

2. THE ALGORITHM

Let (X, Σ, μ) and (Y, θ, ν) be two measure spaces of finite measure. We assume that X and Y are compact Hausdorff spaces and that μ and ν are regular Borel measures. It is convenient, and involves no sacrifice of generality, to suppose $\mu(X) = \nu(Y) = 1$. Let $(Z, \Phi, \sigma) = (X, \Sigma, \mu) \times (Y, \Theta, \nu)$. By identifying an element $g \in L_1(X)$ with the (equivalence class of the) function $\bar{g}(x, y) = g(x)$, we embed $L_1(X)$ in $L_1(Z)$. In the same way $L_1(Y)$ is embedded isometrically in $L_1(Z)$, and we henceforth do not distinguish between g and $\bar{g}$.

In the space $L_1(X)$ we define an operator A which produces best approx-

0021-9045/83 $3.00

imation by constants. Since such approximations are not unique, we let $I(g)$ denote the interval of all best constant approximations to g, i.e.,

$$r \in I(g) \quad \text{iff} \quad \|g-r\|_1 \leqslant \|g-c\|_1 \quad \text{for all} \quad c \in \mathbb{R}.$$

Then Ag is defined as the midpoint of $I(g)$. If $f \in L_1(X \times Y)$, then by the Fubini theorem, $f(x,y) \in L_1(X)$ for almost all $y \in Y$. We define $A_x f$ to be the function of y which results upon applying A to $f(\cdot, y)$. We define A_y similarly. It is not immediately clear where the ranges of A_x and A_y lie, but a result from [2] shows that $A_x: L_1(X \times Y) \to L_1(Y)$ and $A_y: L_1(X \times Y) \to L_1(X)$. It is then easy to see that, for example, A_x satisfies $\|f - A_x f\|_1 \leqslant \|f-h\|_1$ for all $h \in L_1(Y)$.

The generalisation of the Diliberto–Straus algorithm is now: given $f \in L_1(X \times Y)$ we set $f_1 = f$,

$$\begin{aligned} f_2 &= f_1 - A_y f_1, \\ f_3 &= f_2 - A_x f_2, \\ \vdots &= \vdots \quad\quad \vdots \\ f_{2n} &= f_{2n-1} - A_y f_{2n-1}, \\ f_{2n+1} &= f_{2n} - A_x f_{2n}. \end{aligned}$$

It is sometimes convenient to rephrase the algorithm by setting

$$G_n = \sum_{p=1}^{n} A_y f_{2p-1}, \qquad H_n = \sum_{p=1}^{n} A_x f_{2p}, \qquad n = 1, 2, \ldots$$

and $H_0 = 0$, when

$$\begin{aligned} f_{2n} &= f - G_n - H_{n-1}, \\ f_{2n+1} &= f - G_n - H_n, \qquad n = 1, 2, \ldots. \end{aligned}$$

It is now easy to see that since $A_y f_{2n} = 0$, we have $A_y(f - G_n - H_{n-1}) = 0$ or $A_y(f - H_{n-1}) - G_n = 0$, since G_n lies in the range of A_y. This gives $G_n = A_y(f - H_{n-1})$. By a similar argument we obtain $H_n = A_x(f - G_n)$.

The above reasoning has already used one of two useful results which were established in [2]. The first is the elementary observation that, for example, $A_x(f+h) = A_x f + h$ for all $h \in L_1(Y)$. Secondly, if f_1, f_2 lie in $C(X \times Y)$ and $f_1 \leqslant f_2$, then $A_x f_1 \leqslant A_x f_2$. These are consequences of corresponding results about the operator A. We have $A(g+c) = Ag + c$, where $g \in C(X)$, and c is any constant function in $C(X)$, and for $g_1, g_2 \in C(X)$ with $g_1 \leqslant g_2$, $Ag_1 \leqslant Ag_2$.

3. The Existence of Cluster Points

We shall for the moment content ourselves with cluster points. We need

LEMMA 3.1. *If* $f \in C(X \times Y)$, *then*

(i) $|(A_x f)(y_1) - (A_x f)(y_2)| \leqslant \sup_{x \in X} |f(x, y_1) - f(x, y_2)|$ *for all* $y_1, y_2 \in Y$;

(ii) $|(A_y f)(x_1) - (A_y f)(x_2)| \leqslant \sup_{y \in Y} |f(x_1, y) - f(x_2, y)|$ *for all* $x_1, x_2 \in X$.

Proof. These results have already appeared in [2]. For completeness we provide the proof of (i) here. We begin by observing that

$$\min_{x \in X} f(x, y) \leqslant (A_x f)(y) \leqslant \max_{x \in X} f(x, y).$$

The upper and lower bounds here are finite by the compactness of $X \times Y$ and so $A_x f$ is certainly in $L_\infty(X \times Y)$. Now we observe that since $f \in C(X \times Y)$

$$\begin{aligned} -\sup_{x \in X} |f(x, y_1) - f(x, y_2)| &\leqslant f(x, y_1) - f(x, y_2) \\ &\leqslant \sup_{x \in X} |f(x, y_1) - f(x, y_2)| \end{aligned}$$

for all $y_1, y_2 \in Y$. We may rewrite the above inequality as

$$\begin{aligned} -\sup_{x \in X} |f(x, y_1) - f(x, y_2)| + f(x, y_2) &\leqslant f(x, y_1) \\ &\leqslant \sup_{x \in X} |f(x, y_1) - f(x, y_2)| + f(x, y_2) \end{aligned}$$

Now using the properties of A and A_x mentioned at the end of Section 2, we obtain

$$\begin{aligned} &-\sup_{x \in X} |f(x, y_1) - f(x, y_2)| + (A_x f)(y_2) \\ &\qquad \leqslant (A_x f)(y_1) \leqslant \sup_{x \in X} |f(x, y_1) - f(x, y_2)| + (A_x f)(y_2), \end{aligned}$$

or

$$|(A_x f)(y_1) - (A_x f)(y_2)| \leqslant \sup_{x \in X} |f(x, y_1) - f(x, y_2)|.$$

THEOREM 3.2. *If* $f \in C(X \times Y)$, *then its associated sequence of iterates* $\{f_n\}$ *generated by the* L_1*-version of the Diliberto–Straus algorithm lies in* $C(X \times Y)$ *and has cluster points in the* $C(X \times Y)$*-topology.*

Proof. Using Lemma 3.1 we have

$$\begin{aligned}|H_n(y_1)-H_n(y_2)| &= |A_x(f-G_n)(y_1)-A_x(f-G_n)(y_2)| \\ &\leqslant \sup_{x\in X} |(f-G_n)(x,y_1)-(f-G_n)(x,y_2)| \\ &= \sup_{x\in X} |f(x,y_1)-f(x,y_2)|.\end{aligned}$$

A similar result holds for each G_n and it follows that $\{f_n\}$ is an equicontinuous sequence of functions. This sequence has cluster points by the Ascoli theorem.

4. Convergence of Norms

We now prove a result about the convergence of the norms of the elements in the L_1-version of the Diliberto–Straus algorithm to the distance from the function f to the subspace M. This is achieved by showing that the cluster points of the algorithm are themselves best approximations. Such a result will actually prove that to each continuous function there is a best approximation in M (a cluster point of its Diliberto–Straus sequence) which is again continuous. This is not really a new result, however, since it can easily be extracted from a result in [2].

We need the following elementary Lemma before we can proceed to our main result:

LEMMA 4.1. *Suppose (H,β) is a finite measure space, with $\{F_n\}$ a sequence of measurable sets in H such that $\int_{F_n}|f|\to 0$ for some function $f\in L_1(H)$ satisfying $\beta(N(f))=0$. Then $\beta(F_n)\to 0$. Here $N(f)=\{h\in H: f(h)=0\}$.*

Proof. Suppose the desired conclusion is false. Then, by passing to a subsequence if necessary, $\beta(F_n)>\delta$, where $\delta>0$. Again by passing to a second subsequence if need be, we can ensure that $\int_{F_n}|f|\leqslant 1/2^n$. Now let C be the set of all $x\in H$ such that x belongs to infinitely many of the F_n, i.e.,

$$C=\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty}F_n.$$

Now $\beta(\bigcup_{n=m}^{\infty}F_n)\geqslant\delta$ and so $\beta(C)\geqslant\delta$. Also

$$\int_C|f|\leqslant\int_{\bigcup_{n=m}^{\infty}F_n}|f|\leqslant\sum_{n=m}^{\infty}\frac{1}{2^n}=\frac{1}{2^{m-1}}.$$

Thus we conclude that $\int_C |f| = 0$ and hence that $f = 0$ a.e. on C. This contradicts $\beta(N(f)) = 0$. ∎

THEOREM 4.1. *Let $f \in C(X \times Y)$ and let the set of points at which f agrees with any member of M have measure zero. Then the iterates $\{f_n\}$ in the L_1-version of the Diliberto–Straus algorithm satisfy $\|f_n\|_1 \to \operatorname{dist}(f, M)$.*

Proof. We begin by showing that if $F_n = \{(x, y): \operatorname{sgn} f_{n+1}(x, y) = -\operatorname{sgn} f_n(x, y)\}$, then $\sigma(F_n) \to 0$. To see this, we first recall that $\{\|f_n\|_1\}$ forms a decreasing sequence bounded below and hence is convergent. Given $\varepsilon > 0$, take N sufficiently large so that $\|f_{2n-1}\|_1 - \|f_{2n}\|_1 < \varepsilon$ for all $n \geqslant N$. Then $f_{2n} = f_{2n-1} - A_y f_{2n-1}$.

Next we observe that if the set of points at which f agrees with any member of M has measure zero, the sequence $\{f_n\}$ inherits this property. Using this and the characterisation theorem for best approximation by constants (see [4]) we obtain

$$\int_Y \operatorname{sgn} f_{2n}(x, y)\, dv = 0 \qquad \text{for almost all } x \in X.$$

Now $f_{2n-1} - f_{2n} \in L_1(X)$ so that

$$\iint_{X \times Y} (f_{2n-1} - f_{2n}) \operatorname{sgn} f_{2n} = 0$$

and consequently

$$\iint f_{2n-1} \operatorname{sgn} f_{2n} = \iint |f_{2n}| = \|f_{2n}\|_1 \geqslant \|f_{2n-1}\|_1 - \varepsilon \qquad \text{for} \quad n \geqslant N.$$

Thus

$$\iint f_{2n-1} \operatorname{sgn} f_{2n} \geqslant \iint f_{2n-1} \operatorname{sgn} f_{2n-1} - \varepsilon \quad \text{or} \quad \iint (\operatorname{sgn} f_{2n-1} - \operatorname{sgn} f_{2n}) f_{2n-1} \leqslant \varepsilon.$$

Again by our assumption that f_n agrees with any member of M only on sets of measure zero, we have $\operatorname{sgn} f_{2n-1} = \operatorname{sgn} f_{2n}$ for almost all $(x, y) \in X \times Y \backslash F_{2n-1}$ and so

$$\int_{F_{2n-1}} (\operatorname{sgn} f_{2n-1} - \operatorname{sgn} f_{2n}) f_{2n-1} \leqslant \varepsilon$$

or

$$2 \iint_{F_{2n-1}} |f_{2n-1}| \leqslant \varepsilon.$$

A similar argument is valid for F_{2n} and hence we may conclude $\iint_{F_n} |f_n| \to 0$.

Now by Theorem 3.2, $\{f_n\}$ contains a uniformly convergent subsequence $\{f_{n_k}\}$ with limit e, say. Again e can only agree with elements of M on sets of measure zero and $\iint_{F_{n_k}} |e| \to 0$. Thus Lemma 4.1 may be applied to give $\sigma(F_{n_k}) \to 0$.

Now since $f_{n_k} \to e$ in the $C(Z)$-topology it is obvious that $\iint \operatorname{sgn} f_{n_k} \cdot e \, d\sigma \to \|e\|_1$. Furthermore, we assert that $\iint \operatorname{sgn} f_{n_k} \cdot m \, d\sigma \to 0$ for all $m \in M$. Given that this assertion is true, then our theorem will follow from the inequalities (true for any $m \in M$)

$$\|e + m\|_1 \geqslant \iint (e + m) \operatorname{sgn} f_n \, d\sigma$$

and so

$$\|e + m\|_1 \geqslant \lim_{n_k \to \infty} \iint \operatorname{sgn} f_{n_k}(e + m) \, d\sigma$$
$$= \|e\|_1.$$

It remains to prove our assertion that $\iint \operatorname{sgn} f_{n_k} \cdot m \, d\sigma \to 0$ for all $m \in M$. It will be sufficient to show that

$$\iint \operatorname{sgn} f_{n_k} \cdot u \, d\sigma \to 0 \qquad \text{for all } u \in L_1(X),$$

$$\iint \operatorname{sgn} f_{n_k} \cdot v \, d\sigma \to 0 \qquad \text{for all } v \in L_1(Y).$$

By the Fubini theorem, these assertions are equivalent to

$$\int_Y \operatorname{sgn} f_{n_k} \, dv \to 0 \qquad \text{for almost all } x \in X,$$

$$\int_X \operatorname{sgn} f_{n_k} \, d\mu \to 0 \qquad \text{for almost all } y \in Y.$$

There are now two cases depending on whether n_k is even or odd. We suppose that $n_k = 2p$, the case $n_k = 2p - 1$ being similar. Then $\int_Y \operatorname{sgn} f_{2p}(x, y) \, dv = 0$ for almost all $x \in X$, and by the Fubini theorem, if we set $\int_X \operatorname{sgn} f_{2p}(x, y) \, d\mu = r_{2p}(y)$ and $F_{2p-1}(y) = \{x\colon (x, y) \in F_{2p-1}\}$, then we obtain

$$\int_Y |r_{2p}(y)| \, dv = \int_Y \left| \int_X \operatorname{sgn} f_{2p}(x, y) \, d\mu \right| dv$$
$$= \int_Y \left| \int_X \operatorname{sgn} f_{2p-1}(x, y) \, d\mu + 2 \int_{F_{2p-1}(y)} \operatorname{sgn} f_{2p}(x, y) \, d\mu \right| dv$$

$$
\begin{aligned}
&= 2\int_Y \left| \int_{F_{2p-1}(y)} \operatorname{sgn} f_{2p}(x,y)\, d\mu \right| dv \\
&\leqslant 2\int_Y \int_{F_{2p-1}(y)} |\operatorname{sgn} f_{2p}(x,y)|\, d\mu\, dv \\
&= 2\sigma(F_{2p-1}).
\end{aligned}
$$

Now setting $r_{n_k}(y) = \int_X \operatorname{sgn} f_{n_k}(x,y)\, d\mu$ and $s_{n_k}(x) = \int_Y \operatorname{sgn} f_{n_k}(x,y)\, dv$, we have

$$\int_Y |r_{n_k}(y)|\, dv \leqslant 2\sigma(F_{n_{k-1}}) \quad \text{and} \quad \int_X |s_{n_k}(x)|\, d\mu \leqslant 2\sigma(F_{n_{k-1}}).$$

Of course, for each n_k one of these inequalities is trivial. We can now conclude

$$\int_Y |r_{n_k}(y)|\, dv \to 0 \qquad \text{as} \quad k \to \infty$$

and

$$\int_X |s_{n_k}(x)|\, d\mu \to 0 \qquad \text{as} \quad k \to \infty.$$

Again recalling the convergence $f_{n_k} \to e$, we have

$$\int_Y \left| \int_X \operatorname{sgn} e(x,y)\, d\mu \right| dv = 0, \qquad \int_X \left| \int_Y \operatorname{sgn} e(x,y)\, dv \right| d\mu = 0,$$

and this is sufficient to allow us to conclude

$$\int_X \operatorname{sgn} f_{n_k}(x,y)\, d\mu \to 0 \qquad \text{for almost all } y \in Y$$

$$\int_Y \operatorname{sgn} f_{n_k}(x,y)\, dv \to 0 \qquad \text{for almost all } x \in X.$$

5. Remarks

It is interesting to note that our approach follows that of Diliberto and Straus insofar as we exploit the same formula $\operatorname{dist}(f, M) = \sup_{f^* \in S(L_1^*) \cap M^\perp} |f^*(f)|$. However, Diliberto and Straus needed to develop first a subset of $M^\perp$ which had rather simple properties. They then showed that it was sufficient to take the sup over this "nice" subset. Here we have used the formula $\text{dist}(f, M) = \lim_{n\to\infty} \sup |f_n^*(f)|$, where $f_n^* \in S(L_1^*)$ and f_n^* converges weakly to a member of $M^\perp$.

The condition that the function differs from all members of M on every set of positive measure cannot in general be omitted. In [2] an example of a function f not satisfying this condition and for which $\|f\|_1 > \text{dist}(f, M)$ was constructed. The function had the further property that $A_x f = A_y f = 0$ so that the algorithm is stationary. In this case $\|f_n\|_1 \not\to \text{dist}(f, M)$.

This paper leaves open the problem of convergence of the algorithm. As in the $C(X \times Y)$ case, this seems to be a more difficult question than convergence of the norms.

Acknowledgments

The author wishes to acknowledge the many helpful remarks made by the referee.

References

1. G. Aumann, Uber approximative Nomographie II, *Bayer Akad. Wiss. Math. Natur. Kl. Sitzungsber.* (1959), 103–109.
2. E. W. Cheney, W. A. Light, J. H. McCabe, and G. M. Phillips, The approximation of bivariate functions by sums of univariate ones using the L_1-metric, *Proc. Edinburgh Math. Soc. (5)*, to appear.
3. S. P. Diliberto and E. G. Straus, On the approximation of a function of several variables by the sum of functions of fewer variables, *Pacific J. Math.* **1** (1951), 195–210.
4. R. C. James, Orthogonality and linear functionals in normed linear spaces, *Trans. Amer. Math. Soc.* **61** (1947), 265–292.