Proceedings of the Edinburgh Mathematical Society (1984) 27, 31–45 ©

THE L_1-VERSION OF THE DILIBERTO–STRAUS ALGORITHM IN $C(T\times S)$

by W. A. LIGHT and S. M. HOLLAND

(Received 16th November, 1982)

Introduction

Let (S,Σ,μ) and (T,Θ,ν) be two measure spaces of finite measure where we assume S,T are compact Hausdorff spaces and μ,ν are regular Borel measures. We construct the product measure space $(T\times S,\Phi,\sigma)$ in the usual way. Let $G=[g_1,g_2,\ldots,g_p]$ and $H=[h_1,h_2,\ldots,h_m]$ be finite dimensional subspaces of $C(S)$ and $C(T)$ respectively where G and H are also Chebyshev with respect to the L_1-norm. Note that a subspace Y of a normed linear space X is Chebyshev if each $x\in X$ possesses exactly one best approximation $y\in Y$. For example, in $C(S)$ with the L_1-norm, the subspace of polynomials of degree at most n is a Chebyshev subspace. This is an old theorem of Jackson. Now set

$$U=C(T)\otimes G,\quad V=H\otimes C(S),\quad W=U+V.$$

It is easy to prove [7] that U and V are proximinal subspaces of $C(T\times S)$. That is, every $f\in C(T\times S)$ possesses at least one best approximation from U or from V. A metric selection $P_U\colon C(T\times S)\to U$ is a mapping which associates each $f\in C(T\times S)$ with one of its best approximations in U. The metric selection P_V is similarly defined.

We shall investigate the behaviour of the Diliberto–Straus algorithm [2]. This algorithm generates a sequence of functions $\{f_n\}$ and may be described by taking $f_0=f$ and setting

$$f_n=f_{n-1}-P_U(f_{n-1}-P_Vf_{n-1}),\qquad n=1,2,3,\ldots$$

This algorithm is essentially the "alternating method" of Von Neumann [11] although his discussion centred on the Hilbert space setting. This investigation extends work in [8] where G and H were both restricted to being 1-dimensional. Other papers directly related to this work are [4], [7] and [9]. We shall show that if one or other of G and H is one dimensional and if f satisfies certain hypotheses than the norms of the iterates converge to $\operatorname{dist}(f,W)=\inf_{w\in W}\|f-w\|$. In contrast to the Diliberto–Straus paper [2] (which dealt with the L_∞-case) it is impossible to obtain convergence of these L_1-norms to $\operatorname{dist}(f,W)$ for all $f\in C(T\times S)$ as is shown by an example from [9].

It will be convenient to assume throughout that S and T have measure 1. We shall use unadorned norm symbols $\|.\|$ to denote the L_1-norm on $T\times S$ while $\|.\|_S$ and $\|.\|_T$

will denote L_1-norms on S and T respectively. We shall also need to assume that the bases in G and H are such that they each form one half of a bi-orthonormal set of bases for G, G^* and H, H^*. Thus

$$g_1, \ldots, g_p \in G; \phi_1, \ldots, \phi_p \in G^*$$

$$1 = \|g_i\|_S = \|\phi_i\|_\infty \qquad \langle g_i, \phi_j \rangle = \delta_{ij} \qquad 1 \leqq i, j \leqq p$$

$$h_1, \ldots, h_m \in H; \psi_1, \ldots, \psi_m \in H^*$$

$$1 = \|h_i\|_T = \|\psi_i\|_\infty \qquad \langle h_i, \psi_j \rangle = \delta_{ij} \qquad 1 \leqq i, j \leqq m$$

The algorithm

Given $f \in C(T \times S)$ the sections f_t and f_s are defined by the equation $f_t(s) = f(t, s) = f_s(t)$. $f_t \in C(S)$ for all $t \in T$. Every section f_t possesses a unique best approximation $g \in G$ because G is Chebyshev. We claim that when these "sectional" approximations are pieced together, we obtain an element of U. Before we can establish this we need a result from [4], whose proof we provide on account of its brevity.

Lemma 1 *There exists a function g in C(S) such that for each u in U,*

(i) $|u(t, s)| \leqq g(s)\|u_t\|_S$
(ii) $\|u_s\|_T \leqq g(s)\|u\|$

for all $t \in T$, $s \in S$.

Proof. Set $d_j^{-1} = \inf_{C_i \in \mathbf{R}} \|\sum_{i \neq j} C_i g_i + g_j\|_S$. Since the g_i are linearly independent we have $d_j^{-1} > 0$ for $j = 1, 2, \ldots, n$. Let $u = \sum_1^p x_i g_i$ and let $T_j = \{t \in T: x_j(t) \neq 0\}$. Then for t in T_j we have

$$\left\|\sum_1^p x_i(t) g_i\right\|_S = |x_j(t)| \left\|\sum_{i=1}^p \frac{x_i(t)}{x_j(t)} g_i\right\|_S \geqq |x_j(t)| d_j^{-1}.$$

Thus for all $t \in T |x_j(t)| \leqq d_j \|u_t\|_S$. Now

$$|u(t, s)| \leqq \sum_1^p |x_i(t)|\,|g_i(s)| \leqq \|u_t\|_S \sum_1^p d_i |g_i(s)|.$$

Choosing $g = \sum_1^p d_i |g_i|$ gives (i), while for (ii) we need only observe

$$\|u_s\|_T = \int_T |u(t, s)| dv \leqq \int_T g(s) \|u_t\|_S \, dv = g(s)\|u\|.$$

Let A be the best approximation operator for G, i.e.

$$\|\varphi - A\phi\|_S \leqq \|\phi - g\|_S \quad \text{for all } g \in G$$

https://doi.org/10.1017/S0013091500022094 Published online by Cambridge University Press

where $\phi\in C(S)$. By a well-known result [5], A is a continuous operator from $C(S)$ onto G. Consider the map A_U on $C(T\times S)$ defined by

$$(A_U f)(t,s)=(Af_t)(s).$$

We claim that the range of A_U is indeed U.

Lemma 2. *Let $f\in C(T\times S)$. Then,*

$$A_U f(t,s)=\sum_{i=1}^{p} x_i(t)g_i(s)$$

where $x_i\in C(T)$.

Proof. From the construction of A_U it is clear that A_U has the correct form and we need only check the claim that $x_i\in C(T)$. Let $f\in C(T\times S)$. Since T and S are compact, f is uniformly continuous on $T\times S$ and so given $\delta_1>0$ there is a δ_2 such that

$$|f(s,t')-f(s,t)|\leqq\delta_1 \text{ for } |t'-t|\leqq\delta_2 \text{ and all } s\in S,$$

so that $\|f_{t'}-f_t\|_S\leqq\delta_1$ for $|t'-t|\leqq\delta_2$. Now by the continuity of A, given $\varepsilon>0$ there is a $\delta>0$ such that

$$\|Af_{t'}-Af_t\|_S\leqq\varepsilon \quad \text{for} \quad \|f_t-f_{t'}\|_S\leqq\delta$$

or

$$\|Af_{t'}-Af_t\|_S\leqq\varepsilon \quad \text{for} \quad |t'-t|\leqq\delta'$$

i.e.

$$\left\|\sum_{i=1}^{p}\{x_i(t')-x_i(t)\}g_i\right\|_S\leqq\varepsilon \quad \text{for} \quad |t'-t|\leqq\delta'.$$

Now from the proof of Lemma 1 we can see that if $x_i(t')\neq x_i(t)$ then

$$\left\|\sum_{1}^{p}\{x_i(t')-x_i(t)\}g_i\right\|_S=|x_j(t')-x_j(t)|\left\|\sum_{i=1}^{p}\frac{x_i(t')-x_i(t)}{x_j(t')-x_j(t)}g_i\right\|_S$$

$$\geqq|x_j(t')-x_j(t)|\,d_j^{-1}.$$

Hence

$$|x_j(t')-x_j(t)|\leqq d_j\varepsilon \qquad \text{for } |t'-t|\leqq\delta'$$

and so $x_j\in C(T)$. □

Clearly the map $A_V: C(T\times S)\to V$ can be similarly defined. We remark that both A_U and A_V are metric selections. For a fixed $f\in C(T\times S)$ we can define mappings $B_U: V\to U$

and $B_V: U \to V$ by

$$B_U v = A_U(f - v) \quad \text{and} \quad B_V u = A_V(f - u).$$

Lemma 3. *The mappings B_U and B_V are continuous.*

Proof. We shall only establish that B_U is continuous. Consider first the mapping $\phi: V \times T \to G$ given by

$$\phi(v, t) = A(f - v)_t.$$

ϕ is clearly continuous since A is continuous and the operation of taking t-sections of functions in $C(T \times S)$ is also continuous. Now fix $v_0 \in V$. Then given $\varepsilon > 0$ we choose δ_t to be the largest real number such that

$$\|A(f - v_0)_t - A(f - v)_t\|_S < \varepsilon \quad \text{whenever } \|(v_0)_t - v_t\|_S < \delta_t.$$

Now by Lemma 1 applied to V rather than U, there is a function $h \in C(T)$ such that

$$\|v_t\|_S \leqq h(t)\|v\| \qquad \text{for all } t \in T.$$

Hence we can modify δ_t to be the largest real number such that

$$\|A(f - v_0)_t - A(f - v)_t\|_S < \varepsilon \quad \text{whenever } \|v_0 - v\| < \delta_t,$$

i.e.,

$$\|\phi(v_0, t) - \phi(v, t)\|_S < \varepsilon \quad \text{whenever } \|v_0 - v\| < \delta_t.$$

Set $\delta = \inf_{t \in T} \delta_t$. We claim that $\delta > 0$. Suppose to the contrary we have $\delta = 0$. Set

$$\operatorname{Osc}(\phi, t, I) = \sup_{v_1, v_2 \in I} |\phi(v_1, t) - \phi(v_2, t)|.$$

Then $\operatorname{Osc}(\phi, t, B(v_0, \delta_t)) \geqq \varepsilon$ where $B(v_0, \delta_t) = \{v: \|v_0 - v\| \leqq \delta_t\}$. Now by our hypothesis there is a sequence $\{t_n\} \in T$ such that $\delta_{t_n} \downarrow 0$. By the compactness of T we may as well assume that $t_n \to t^* \in T$. Now $\operatorname{Osc}(\phi, t_n, B(v_0, \delta_{t_n})) \geqq \varepsilon$ violates the continuity of ϕ at (v_0, t^*). We have thus shown that given $\varepsilon > 0$ there is a $\delta > 0$ such that

$$\|A(f - v_0)_t - A(f - v)_t\|_S < \varepsilon \quad \text{whenever } \|v_0 - v\| < \delta.$$

Integrating over T gives the required result

$$\|A_u(f - v_0) - A_u(f - u)\| < \varepsilon \quad \text{whenever } \|v_0 - v\| < \delta. \quad \square$$

Note that the continuity of the metric selection in general is not being asserted here since A_U is being restricted to act on V. Nevertheless, Lemma 3 provides a class of functions and a L_1-metric selection which is continuous on that (infinite-dimensional) class.

We are now in a position to define the algorithm for a fixed $f\in C(T\times S)$ we set

$$\begin{aligned} f_1 &= f \\ f_2 &= f_1 - A_U f_1 \\ f_3 &= f_2 - A_V f_2 \\ &\vdots \\ f_{2n} &= f_{2n-1} - A_U f_{2n-1} \\ f_{2n+1} &= f_{2n} - A_V f_{2n} \\ &\vdots \end{aligned}$$

If we also put

$$G_n=\sum_{i=1}^{n} A_U f_{2i-1}, \qquad H_n=\sum_{i=1}^{n} A_V f_{2i}, \qquad H_0=G_0=0,$$

then the algorithm can be rewritten

$$\left.\begin{aligned} f_{2n-1} &= f-H_{n-1}-G_{n-1} \\ f_{2n} &= f-H_{n-1}-G_n \end{aligned}\right\} \quad n=1,2,\dots.$$

It is easy to establish that

$$\left.\begin{aligned} H_n &= A_V(f-G_n)=B_V(G_n) \\ G_n &= A_U(f-H_{n-1})=B_U(H_{n-1}) \end{aligned}\right\} \quad n=1,2,\dots.$$

Our main question can now be stated more carefully. Given an $f\in C(T\times S)$ we want to know when $\|f_n\|\to \operatorname{dist}(f,W)$. It will become clear shortly that $\{\|f_n\|\}$ is a decreasing sequence of real numbers bounded below by $\operatorname{dist}(f,W)$ so that we are really asking—when is this lower bound attained?

Preliminary results

We shall need several results which are not in the mainstream of our argument. For convenience we collect them in this section. The first is a characterisation theorem of James [**6**].

Theorem 4. *In order that* 0 *be a best L_1-approximation to an $f\in L_1(S)$ from some linear subspace K it is necessary and sufficient that $\int k\operatorname{sgn} f\le\int_{Z(f)}|k|$ for all $k\in K$. Here $Z(f)$ denotes the set of points where $f(s)=0$.*

Corollary 5. *If K is a one-dimensional subspace spanned by a non-negative function k*

https://doi.org/10.1017/S0013091500022094 Published online by Cambridge University Press

then 0 *is a best* L_1*-approximation to an* $f \in L_1(S)$ *if and only if*

$$\int_{P(f)} k \leqq \tfrac{1}{2}\int_S k \qquad \textit{and} \qquad \int_{N(f)} k \leqq \tfrac{1}{2}\int_S k.$$

Here $N(f)$ *and* $P(f)$ *denote the sets where* f *is negative or positive respectively.*

This result is an elementary special case of Theorem 4, and we omit the proof.

Lemma 6. *Suppose* K *is a one-dimensional subspace of* $L_1(S)$ *generated by a non-negative function* k_0. *Then there exists a metric selection* $\pi: L_1(S) \to K$ *with the following properties*

(i) $\pi(f + \lambda k_0) = \pi f + \lambda k_0$ *for all* $\lambda \in \mathbb{R}, f \in L_1(S)$.

(ii) $\|f - \pi f\|_S \leqq \|f\|_S$ *for all* $f \in L_1(S)$.

(iii) $\pi f_1 \geqq \pi f_2$ *whenever* $f_1 \geqq f_2, f_1, f_2 \in L_1(S)$.

(iv) $\|\pi f_1 - \pi f_2\|_{S,\infty} \leqq \|f_1 - f_2\|_{S,\infty}$ *when* $f_1, f_2 \in C(S)$.

Proof. We begin by observing that only (iv) needs the requirement that S has a Borel measure defined on it. Further (i) and (ii) are elementary. To establish (iii) we firstly recall that the set of best approximations to f_1 from K is convex and closed. If K is generated by k_0 then this set corresponds to some interval $[\alpha_1, \beta_1] \subset \mathbb{R}$ in the sense that ck_0 is a best approximation to f_1 if and only if $c \in [\alpha_1, \beta_1]$. Now suppose $f_2 \geqq f_1$ with corresponding interval $[\alpha_2, \beta_2]$. Suppose further that $\alpha_2 < \alpha_1$. Then since $\alpha_2 k_0$ is *not* a best approximation to f_1 we have by Corollary 5 either

$$\int_{P(f_1 - \alpha_2 k_0)} k_0 > \tfrac{1}{2}\int_S k_0 \quad \text{or} \quad \int_{N(f_1 - \alpha_2 k_0)} k_0 > \tfrac{1}{2}\int_S k_0.$$

Now if the former obtains then $\int_{P(f_2 - \alpha_2 k_0)} k_0 > \frac{1}{2}\int_S k_0$ since $f_2 \geqq f_1$. In the latter case $\int_{N(f_1 - \alpha_2 k_0)} k_0 > \frac{1}{2}\int_S k_0$. These two conditions violate the fact that $\alpha_2 k_0$ and $\alpha_1 k_0$ are best approximations to f_2 and f_1 respectively. A similar argument shows $\beta_1 \leqq \beta_2$. Now defining $\pi f_1 = \lambda k_0$ where λ is the midpoint of $[\alpha_1, \beta_1]$ gives the result. The proof of (iv) is as follows. Firstly, (iv) involves no selection since in this case πf is unique for $f \in C(S)$ [10; p. 235] so

$$-\|f_1 - f_2\|_{S,\infty} \leqq f_1 - f_2 \leqq \|f_1 - f_2\|_{S,\infty}$$

or

$$-\|f_1 - f_2\|_{S,\infty} + f_2 \leqq f_1 \leqq f_2 + \|f_1 - f_2\|_{S,\infty}.$$

Now by parts (iii) and (i)

$$\pi f_2 - \|f_1 - f_2\|_{S,\infty} \leqq \pi f_1 \leqq \pi f_2 + \|f_1 - f_2\|_{S,\infty}$$

or

$$\|\pi f_1 - \pi f_2\|_{S,\infty} \leqq \|f_1 - f_2\|_{S,\infty}. \quad \square$$

Note that we cannot relax the condition that K be generated by a non-negative function for if k_0 has both $P(k_0)$ and $N(k_0)$ of positive measure then we cannot have $\alpha k_0 \geqq \alpha_1 k_0$ unless $\alpha = \alpha_1$.

We now give three results which appeared in [7]. We provide the proofs of two for completeness, the third being elementary.

Lemma 7. *The subspace W is closed in $C(T\times S)$. Hence each element $w\in W$ has the representation $w=u+v$ with $u\in U$, $v\in V$ and $\|u\|+\|v\|\leqq \beta\|w\|$, β constant.*

Proof. Recalling our bi-orthonormal bases we define

$$(Pf)(t,s)=\sum_{i=1}^{p}\langle f_t,\phi_i\rangle g_i(s) \qquad f\in C(T\times S)$$

$$(Qf)(t,s)=\sum_{i=1}^{p}\langle f^s,\psi_i\rangle h_i(t) \qquad f\in C(T\times S).$$

These are bounded linear projections onto U and V respectively. It is easily verified that $PQ=QP$. By [3], p. 481] $P+Q-PQ$ is a projection of $C(T\times S)$ onto W. W is therefore closed and given $w\in W$ defining $u=Pw-PQw$ and $v=Qu$ completes the proof. □

Lemma 8. *If $v\in V$ then*

$$\sup_t \|v_t\|_S \leqq \|v\| \sum_{i=1}^{m} \|h_i\|_{T,\infty}.$$

Proof. Let $v=\sum_{i=1}^{m} y_i h_i$. By the bi-orthonormality property

$$y_i(s)=\int v(t,s)\psi_i(t)\,dv.$$

Therefore, since $\|\psi_i\|_\infty = 1$ we have

$$\|y_i\|_S=\int |y_i(s)|\,d\mu=\int \left|\int v(t,s)\psi_i(t)\,dv\right| d\mu$$

$$\leqq \iint |v(t,s)|\,d\mu\,dv=\|v\|.$$

This gives

$$\sup_t \|v_t\|_S=\sup_t \int \left|\sum_{i=1}^{m} y_i(s)h_i(t)\right| ds$$

$$\sup_t \|v_t\|_S \leqq \sup_t \sum |h_i(t)|\,\|y_i\|_S \leqq \|v\| \sum_{i=1}^{m} \|h_i\|_{T,\infty}.$$

Lemma 9. *If $f\in C(T\times S)$* then $\|f_t\|_S\leqq \|f\|_\infty$.

https://doi.org/10.1017/S0013091500022094 Published online by Cambridge University Press

We shall now show that the sequence $\{f_n\}$, generated by our algorithm has cluster points. The proof is surprisingly intricate when compared with similar results from [**8**]. We begin by observing that the mechanism by which $w=u+v$ is shown to have a representation $w=u'+v'$ where $\|u'\|+\|v'\|\leqq\beta\|w\|$ in Lemma 7 is in fact to take an appropriate $z\in U\cap V$ and set $u'=u-z$, $v'=v+z$. Now given our algorithm we set

$$f_{2n-1}=f-H_{n-1}-G_{n-1}=f-(H_{n-1}-z_n)-(G_{n-1}+z_n)=f-H''_{n-1}-G''_n$$

$$f_{2n}=f-H_{n-1}-G_n=f-(H_{n-1}-z_n)-(G_n+z_n)=f-H''_{n-1}-G'_n$$

$$f_{2n+1}=f-H_n-G_n=f-(H_n-z_n)-(G_n+z_n)=f-H'_n-G'_n$$

where z_n is chosen in $U\cap V$ so that

$$\|H'_n\|+\|G'_n\|\leqq\beta\|H'_n+G'_n\|\leqq\beta\|H_n+G_n\|.$$

Notice that

$$\begin{aligned}G'_n=G_n+z_n&=A_U(f-H_{n-1})+z_n\\&=A_U(f-(H_{n-1}-z_n))\\&=A_U(f-H''_{n-1}).\end{aligned}$$

Similarly, $H'_n=A_V(f-G'_n)$ and $H''_{n-1}=A_V(f-G''_{n-1})$.

Note further that A_Uf produces a best approximation to f out of U since

$$\|(A_Uf)_t-f_t\|_S=\|Af_t-f_t\|_S\leqq\left\|\sum_1^p\lambda_i(t)g_i-f\right\|_S\qquad\text{for } \lambda_i(t)\in\mathbb{R}.$$

Integrating over T gives

$$\|A_Uf-f\|\leqq\|f-u\|\qquad\text{for all } u\in U.$$

Similarly

$$\|A_Vf-f\|\leqq\|f-v\|\qquad\text{for all } v\in V.$$

Now from $f_{2n}=f_{2n-1}-A_Uf_{2n-1}$ we see that $\|f_{2n}\|\leqq\|f_{2n-1}\|$ and in a similar way $\|f_{2n+1}\|\leqq\|f_{2n}\|$ i.e. the sequence $\{\|f_n\|\}$ is decreasing. Hence

$$\|f-H''_{n-1}-G''_{n-1}\|\leqq\|f\|$$

or

$$\|H''_{n-1}+G''_n\|\leqq 2\|f\|$$

https://doi.org/10.1017/S0013091500022094 Published online by Cambridge University Press

and similarly

$$\|H''_{n-1}+G'_n\|\leqq 2\|f\|$$

$$\|H'_n+G'_n\|\leqq 2\|f\|.$$

Using the fact that $\|H'_n\|+\|G'_n\|\leqq\beta\|H_n+G_n\|$ we see that

$$\|H'_n\|\leqq 2\beta\|f\|,\ \|G'_n\|\leqq 2\beta\|f\|.$$

Now since $\|G'_n\|$ is bounded we have

$$\|H''_{n-1}\|\leqq\|H''_{n-1}+G'_n\|+\|G'_n\|\leqq(2\beta+2)\|f\|$$

and again

$$\|G''_{n-1}\|\leqq\|H''_{n-1}+G''_{n-1}\|+\|H''_{n-1}\|\leqq(2\beta+4)\|f\|.$$

Hence $\|H'_n\|$, $\|G'_n\|$, $\|H''_{n-1}\|$, and $\|G''_{n-1}\|$ are all bounded by $\gamma\|f\|$.

Lemma 10. *Let $f\in C(T\times S)$ and let G have metric selection satisfying condition (iv) of Lemma 6. Then the sequence $\{G'_n\}$ has cluster points in U with respect to the supremum norm on $C(T\times S)$.*

Proof. Consider $|G'_n(s_1,t_1)-G'_n(s,t)|$

$$\leqq|G'_n(s_1,t_1)-G'_n(s_1,t)|+|G'_n(s_1,t)-G'_n(s,t)|$$

$$\leqq\|G'_n(\cdot,t_1)-G'_n(\cdot,t)\|_{S,\infty}+\|G'_n(s_1,\cdot)-G'_n(s,\cdot)\|_{T,\infty}$$

$$=\|A(f-H''_{n-1})_{t_1}-A(f-H''_{n-1})_t\|_{S,\infty}+\|G'_n(s_1,\cdot)-G'_n(s,\cdot)\|_{T,\infty}$$

$$\leqq\|(f-H''_{n-1})_{t_1}-(f-H''_{n-1})_t\|_{S,\infty}+\|G'_n(s_1,\cdot)-G_n(s,\cdot)\|_{T,\infty}\text{ by Lemma 6}$$

$$\leqq\|f_{t_1}-f_t\|_{S,\infty}+\|(H''_{n-1})_{t_1}-(H''_{n-1})_t\|_{S,\infty}+\|(G'_n)_{s_1}-(G'_n)_s\|_{T,\infty}$$

$$\leqq\|f_{t_1}-f_t\|_{S,\infty}+\|(H''_{n-1})_{t_1}-(H''_{n-1})_t\|_{S,\infty}+\|(G'_n)_{s_1}-(G'_n)_s\|_{T,\infty}$$

$$=\|f_{t_1}-f_t\|_{S,\infty}+\left\|\sum_{i=1}^{m}y_i(\cdot)[h_i(t_1)-h_i(t)]\right\|_{S,\infty}+\left\|\sum_{i=1}^{p}x_i(\cdot)[g_i(s_1)-g_i(s)]\right\|_{T,\infty}$$

$$\leqq\|f_{t_1}-f_t\|_{S,\infty}+\max_i|h_i(t_1)-h_i(t)|\sum_{i=1}^{m}\|y_i\|_{S,\infty}+\max_i|g_i(s_1)-g_i(s)|\sum_{i=1}^{p}\|x_i\|_{T,\infty}.$$

In these last two expressions the y_i and x_i depend on n. We need to show that the sums $\sum_{i=1}^{m}\|y_i\|_{S,\infty}$ and $\sum_{i=1}^{p}\|x_i\|_{T,\infty}$ are uniformly bounded.

We already know $(G'_n)_t = A(f_t - (H''_{n-1})_t)$ so that

$$\begin{aligned}\|(G'_n)_t\|_S &\leqq \|(G'_n)_t - f_t + (H''_{n-1})_t\|_S + \|f_t - (H''_{n-1})_t\|_S \\ &\leqq 2\|f_t - (H''_{n-1})_t\|_S \\ &\leqq 2\|f_t\|_S + 2\|(H''_{n-1})_t\|_S \\ &\leqq 2\|f\|_\infty + 2\|(H''_{n-1})_t\|_S \quad \text{by Lemma 9} \\ &\leqq 2\|f\|_\infty + 2\|H''_{n-1}\| \sum_{i=1}^{m} \|h_i\|_{T,\infty} \quad \text{by Lemma 8} \\ &\leqq 2\|f\|_\infty + 2\gamma\|f\| \sum_{i=1}^{m} \|h_i\|_{T,\infty} \quad \text{by the remarks preceding this lemma} \\ &\leqq 2\|f\|_\infty \left(1 + \gamma \sum_{i=1}^{m} \|h_i\|_{T,\infty}\right).\end{aligned}$$

Now from the proof of Lemma 1 we obtain

$$|x_i(t)| \leqq d_j \|(G'_n)_t\|_S \leqq 2d_j \|f\|_\infty \left(1 + \gamma \sum_{i=1}^{m} \|h_i\|_{T,\infty}\right),$$

which gives the boundedness of $\|x_i\|_{T,\infty}$.

Similarly, $(H''_{n-1})_s$ is the best approximation from H to $f_s - (G''_{n-1})_s$ to that

$$\begin{aligned}\|(H''_{n-1})_s\|_T &\leqq \|(H''_{n-1})_s - f_s + (G''_{n-1})_s\|_T + \|f_s - (G''_{n-1})_s\|_T \\ &\leqq 2\|f_s - (G''_{n-1})_s\|_T \\ &\leqq 2\|f_s\|_T + 2\|(G''_{n-1})_s\|_T \\ &\leqq 2\|f\|_\infty + 2\|G''_{n-1}\| \sum_{i=1}^{p} \|g_i\|_{S,\infty} \quad \text{by Lemmas 8, 9} \\ &\leqq 2\|f\|_\infty + 2\gamma\|f\| \sum_{i=1}^{p} \|g_i\|_{S,\infty} \\ &\leqq 2\|f\|_\infty \left(1 + \gamma \sum_{i=1}^{p} \|g_i\|_{S,\infty}\right).\end{aligned}$$

Again from the proof of Lemma 1 applied to an element of V rather than U we obtain

$$|y_i(s)| \leqq e_j \|(H''_{n-1})_s\|_T \leqq 2e_j \|f\|_\infty \left(1 + \gamma \sum_{i=1}^{p} \|g_i\|_{S,\infty}\right)$$

where

$$e_j^{-1} = \inf_{c_i \in \mathbf{R}} \left\| \sum_{i \neq j} c_i h_i + h_j \right\|_T .$$

The first of these two arguments shows further that $\|G_n'\|_\infty$ is bounded for $n = 1, 2, \ldots,$ and we may now conclude that $\{G_n'\}$ is a bounded equicontinuous family of functions in U. Since U is closed $\{G_n'\}$ has cluster points in U with respect to the supremum norm.

Theorem 11. *Let $f \in C(T \times S)$ and let G be one-dimensional generated by a non-negative function g. Then the sequence $\{f_n\}$ generated by the L_1-version of the Diliberto–Straus algorithm has cluster points with respect to the sup-norm topology on $C(T \times S)$.*

Proof. By Lemma 10 the sequence $\{G_n'\}$ has cluster points. Now $H_n' = A_V(f - G_n') = B_V(G_n')$ and B_V is continuous by Lemma 3. Thus if $G_{n_k}' \to C \in U$, $H_{n_k}' \to B_V(C)$ and so $f_{n_k} \to f^* = f - C - B_V(C)$. Since W is closed $C - B_V(C)$ lies in W. □

Finally, we have the following result from [**8**]. An alternative proof is provided.

Lemma 12. *Let (H, Δ, β) be a finite measure space and let $\{f_n\}$ be a convergent sequence in $L_1(H, \Delta, \beta)$ with limit e. Suppose $\{F_n\}$ is a sequence of measurable sets in H such that*

(i) $\int_{F_n} |f_n| \, d\beta \to 0$ *as* $n \to \infty$

(ii) $\beta(Z(e)) = 0$.

Then $\beta(F_n) \to 0$.

Proof. Let $E_m = \{h: e(h) > 1/m\}$. Then the E_m form an increasing sequence of sets in H with limit H. Suppose that the desired conclusion is false. Then by passing to subsequences if necessary we may assume $\beta(F_n) \geqq \delta > 0$ for all n. Choose E_m so that $\beta(E_m) \geqq 1 - \delta/2$. Then $\beta(E_m \cap F_n) \geqq \delta/2$. Now take $\varepsilon > 0$. Since $f_n \to e$ we deduce that there is an n_0 for which

$$\int_{F_n} |f_n - e| \, d\beta \leqq \varepsilon \quad \text{when } n \geqq n_0$$

or

$$\int_{F_n} |e| \, d\beta \leqq \int_{F_n} |f_n| \, d\beta + \varepsilon.$$

Now

$$\frac{1}{m} \beta(E_m \cap F_n) \leqq \int_{F_n} |e| \, d\beta \leqq \int_{F_n} |f_n| \, d\beta + \varepsilon.$$

Choosing n_1 such that

$$\int_{F_n} |f_n|\, d\beta \leqq \varepsilon \quad \text{for } n \geqq n_1$$

we have $(1/m)\beta(E_m \cap F_n) \leqq 2\varepsilon$ for large enough n.

This contradicts $\beta(E_m \cap F_n) \geqq \delta/2$. □

Convergence of the algorithm

We shall throughout this section assume that G is one-dimensional generated by a non-negative function g, $f \in C(T \times S)$ and that the sets of points at which f agrees with any member of W have measure zero, so that if

$$X_w = \{(t,s):(f - w)(t,s) = 0\}$$

then $\sigma(X_w) = 0$ for all $w \in W$. This condition is analogous to the idea of smoothness in L_1 spaces. We are requiring that each element $f - w$, $w \in W$ should be smooth (almost everywhere different from 0), see [6] for details.

Lemma 13. *Let* $F_n = \{(t,s): \operatorname{sgn} f_{n+1}(t,s) = -\operatorname{sgn} f_n(t,s)\}$, *and let* $\{f_{n_k}\}$ *be a convergent subsequence of the algorithm. Then* $\sigma(F_{n_k}) \to 0$.

Proof. Recall that $\{\|f_n\|_1\}$ is a decreasing sequence bounded below and hence convergent. Given $\varepsilon > 0$ take N sufficiently large so that $\|f_{2n-1}\| - \|f_{2n}\| < \varepsilon$ for all $n \geqq N$. By our assumption on f the functions f_n agree with members of W only on sets of measure zero and so by the characterisation theorem in the form given by Theorem 4 we obtain

$$\int g \operatorname{sgn} f_{2n}\, d\mu = 0 \text{ for almost all } t \in T \text{ and all } g \in G.$$

Furthermore $f_{2n} - f_{2n-1} \in U$ and so $f_{2n}(t,\cdot) - f_{2n-1}(t,\cdot) \in G$ for almost all $t \in T$. Hence

$$\int [f_{2n}(t,\cdot) - f_{2n-1}(t,\cdot)] \operatorname{sgn} f_{2n}(t,\cdot)\, d\mu = 0 \text{ for almost all } t \in T,$$

and so

$$\iint (f_{2n} - f_{2n-1}) \operatorname{sgn} f_{2n}\, d\sigma = 0.$$

This gives

$$\|f_{2n}\| - \iint f_{2n-1} \cdot \operatorname{sgn} f_{2n-1}\, d\sigma + 2 \iint_{F_{2n-1}} f_{2n-1} \cdot \operatorname{sgn} f_{2n-1} = 0.$$

But $\|f_{2n-1}\| - \|f_{2n}\| < \varepsilon$ so that we may conclude

$$2 \iint_{F_{2n-1}} |f_{2n-1}|\, d\sigma < \varepsilon.$$

A similar argument applies for F_{2n} so that $\iint_{Fn}|f_n|\to 0$. Now suppose $f_{n_k}\to e$. Then again e agrees with members of W only on sets of measure zero and so Lemma 12 can be applied to give $\sigma(F_{n_k})\to 0$. □

We shall now show that for the convergent subsequence $\{f_{n_k}\}$ with limit e, we have

$$\iint w \operatorname{sgn} e\, d\sigma = 0 \quad \text{for all } w\in W.$$

It will be sufficient to show

$$\iint u \operatorname{sgn} e\, d\sigma = 0 \quad \text{for all } u\in U$$

and

$$\iint v \operatorname{sgn} e\, d\sigma = 0 \quad \text{for all } v\in V$$

or again sufficient to demonstrate that

$$\int_T h \operatorname{sgn} e\, dv = 0 \quad \text{for all } h\in H, \text{ almost all } s\in S$$

and

$$\int_S g \operatorname{sgn} e\, d\mu = 0 \text{ for almost all } t\in T$$

when the former limits will follow from an application of the Fubini Theorem. If n_k is odd then

$$\int_T h \operatorname{sgn} f_{n_k}\, dv = 0 \quad \text{for all } h\in H, \text{ almost all } s\in S$$

while if n_k is even then

$$\int_S g \operatorname{sgn} f_{n_k}\, d\mu = 0 \text{ for almost all } t\in T.$$

Suppose n_k is even, $n_k=2p$ say. Then $\int_T h \operatorname{sgn} f_{n_k+1}\, dv=0$ for almost all $s\in S$. Set $F_n(t)=\{s:(s,t)\in F_n\}$ and $F_n(s)=\{t:(s,t)\in F_n\}$. Then

$$\begin{aligned}\int_S\left|\int_T h \operatorname{sgn} f_{n_k}\, dv\right| d\mu &= \int_S\left|\int_T h \operatorname{sgn} f_{2p+1}\, dv + 2\int_{F_{2p}(s)} h \operatorname{sgn} f_{2p+1}\, dv\right| d\mu\\ &= 2\int_S\left|\int_{F_{2p}(s)} h \operatorname{sgn} f_{2p+1}\, dv\right| d\mu\\ &\leqq 2\int_S\int_{F_{2p}(s)} |h|\, dv\, d\mu\\ &\leqq 2\|h\|_\infty \sigma(F_{n_k}).\end{aligned}$$

Now take k_0 sufficiently large to ensure that for a given ε and fixed $h \in H$

$$\left.\begin{array}{ll}\text{(i)} & \sigma(F_{n_k}) < (2\|h\|_\infty)^{-1}(\varepsilon/2) \\ \text{(ii)} & \left|\int_T h(\operatorname{sgn} e - \operatorname{sgn} f_{n_k})\, dv\right| < \frac{\varepsilon}{2}\end{array}\right\} \text{ for all } k \geqq k_0, \text{ and almost all } t \in T$$

Then

$$\int_S \left| \int_T h \operatorname{sgn} e\, dv \right| d\mu \leqq \int \left\{ \left| \int_T h \operatorname{sgn} f_{n_k}\, dv \right| + \frac{\varepsilon}{2} \right\} d\mu$$

$$\leqq 2\|h\|_\infty \sigma(F_{n_k}) + \varepsilon/2 < \varepsilon.$$

Thus

$$\int_T h \operatorname{sgn} e\, dv = 0 \quad \text{for almost all } s \in S.$$

In a similar manner we show that

$$\int_T g \operatorname{sgn} e\, d\mu = 0 \quad \text{for almost all } t \in T \text{ and all } g \in G.$$

With this preamble we are able to state and prove our main result.

Theorem 14. *Let G be a one-dimensional subspace generated by a non-negative function. Let $f \in C(T \times S)$ with the sets*

$$X_w = \{(t,s) : (f - w)(t,s) = 0\}$$

having $\sigma(X_w) = 0$ for all $w \in W$. Then the sequence $\{f_n\}$ generated by the L_1-version of the Diliberto–Straus algorithm has the property $\|f_n\| \downarrow \operatorname{dist}(f, W)$.

Proof. We shall show that if $\{f_{n_k}\}$ is a convergent subsequence with limit e then $\|e\| = \operatorname{dist}(f, W)$. Then we already know $\{\{\|f_n\|\}$ is a decreasing sequence and so the theorem will be proved. From the preamble we already know $\iint w \operatorname{sgn} e\, d\sigma = 0$ for all $w \in W$. Now

$$\|e + w\| \geqq \iint (e + w) \operatorname{sgn} e\, d\sigma$$

$$= \iint e \operatorname{sgn} e\, d\sigma = \|e\|.$$

Thus by the fact that W is closed and

$$e = \lim f_{n_k} = \lim f - w_{n_k} = f - w^*$$

we have

$$\operatorname{dist}(f, W) = \inf_{w \in W} \|f - w\| = \inf_{w_1 \in W} \|e + w_1\| \geqq \|e\| = \|f - w^*\| = \lim_{k \to \infty} \|f_{n_k}\| = \lim_{n \to \infty} \|f_n\|. \quad \square$$

Remarks

Other theorems of this nature are easily obtained after one has observed that the assumption that G be one-dimensional generated by a positive function is only needed to ensure that Lemma 6(iv) holds. An alternative approach is to assume that G has the property given in Lemma 6(iv) i.e. its metric selection is Lipshitz with constant unity on $C(T \times S)$. Apart from the one-dimensional subspaces generated by positive functions this phenomenon would seem to be somewhat rare.

REFERENCES

1. F. Deutsch, The alternating method of von Neumann, *Multivariate Approximation Theory*, eds. W. Schempp and K. Zeller (Birchhäuser Verlag, Basel, 1979).

2. 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.

3. N. Dunford and J. T. Schwartz, *Linear Operators, Part 1* (Interscience, New York, 1959).

4. S. M. Holland, W. A. Light and L. J. Sulley, On proximinality in $L_1(T \times S)$, *Proc. Amer. Math. Soc.*, to appear.

5. R. B. Holmes, *A Course on Optimization and Best Approximation* (Springer-Verlag, 1972).

6. R. C. James, Orthogonality and linear functionals in normed linear spaces, *Trans. Amer. Math. Soc.* **61** (1947), 265–292.

7. W. A. Light and E. W. Cheney, Some best approximation theorems in Tensor Product Spaces, *Math. Proc. Camb. Phil. Soc.* **8** (1981), 385–390.

8. W. A. Light, The Diliberto–Straus algorithm in $L_1(X \times Y)$, *J. Approx. Th.*, to appear.

9. W. A. Light, J. H. McCabe, G. H. Phillips and E. W. Cheney, The approximation of bivariate functions by sums of univariate ones using the L_1-metric, *Proc. Edinburgh Math. Soc.* **25** (1982), 173–181.

10. I. Singer, *Best Approximation in Normed Linear Spaces by Elements of Linear Subspaces* (Springer-Verlag, New York, Heidelberg, Berlin, 1970).

11. J. Von Neumann, *Functional Operators Part II*, The geometry of orthogonal spaces (Annals of Math. Studies, no. 22, Princeton University Press, 1950).

Mathematics Department
Texas A. & M. University
College Station
Texas