Best Approximants in L¹ Space II*

Toshitada SHINTANI** and Tsuyoshi ANDO***

(Received July 24, 1990)

Abstract

This is a study of the best approximants in L^1 space by *real analysis*.

Consider the *real* Banach space L^1 on a probability space (Ω, A, P). Given a σ-subalgebra B of A, let $L^1(B)$ denote the subspace consisting of all functions measurable with respect to B. Given a function $f \in L^1$, a function $g \in L^1(B)$ is called a *best approximant* of f if g has the minimum distance to f :

$$\|f-g\| = \inf\{\|f-h\| \,;\, h \in L^1(B)\}.$$

1. The set of best approximants of f is not empty but contains the *maximum*, denoted by $U_B(f)$, the Freudenthal spectral representation of which is given explicitly by

$$U_B(f) = \int_{-\infty}^{\infty} \lambda d g_\lambda,$$

where g is the characteristic function of the set $\{1/2 < E_B(\chi_{\{f \leq \lambda\}})\}$, E_B being the *conditional expectation* with respect to B.

2. Let $\{B_n\}$ be a sequence of σ-subalgebras. If the sequence $\{E_{B_n}(h)\}$ converges to $E_B(h)$ *almost everywhere* for every $h \in L^1$, then

$$\limsup_{n \to \infty} U_{B_n}(f) \leq U_B(f).$$

1. Introduction and Preliminaries

This paper concerns the Banach space L^1 of real-valued integrable functions on a probability space (Ω, A, P). Given a σ-subalgebra B of the σ-algebra A, let $L^1(B)$ stand for the subspace of L^1, consisting of all those functions that are measurable with respect to B. In an unpublished paper [4] H. Kudo pointed out that each integrable function admits a *best approximant* in $L^1(B)$, that is, there is a function in $L^1(B)$ with minimum distance to the given function.

The purpose of this paper is to study the property and the structure of best approximants. The main results are summarized in the following. First of all, the Kudo's result mentioned above is incorporated in a more general theorem that every sequence in $L^1(B)$ that minimizes asymptotically the distance to a given function is weakly compact. Here a sequence is said to be *weakly compact* if every subsequence admits a limiting function in the weak topology of the Banach space L^1. The set of best approximants in $L^1(B)$ of an integrable function f is shown to contain the maximum and the minimum, denoted by $U_B(f)$ and $V_B(f)$ respectively, in the following sense : a function g in $L^1(B)$ is a best approximant of the function f if and only if $V_B(f) \leq g \leq U_B(f)$. The non-linear operators U_B and V_B possess several remarkable properties, which make it possible to give explicit spectral

* 1975年6月22日ドイツ連邦共和国オーバーウォルフハッハ高等数学研究所で一部講演

** 数学，一般教科助教授

*** 北海道大学応用電気研究所応用数学部門教授（同所長）

representation of $U_B(f)$ and $V_B(f)$ in terms of that of f. As a consequence, the distance of a function f to the subspace $L^1(B)$ is provided with a convenient expression. In the final section convergence of a sequence of σ-subalgebras are taken into consideration. The exact definitions of several convergence concepts are given at the end of this section. The following convergence theorem will be proved. If a sequence $\{B_n\}$ of σ-subalgebras converges strongly to a σ-subalgebra B_∞ and if each g_n is a best approximant in $L^1(B_n)$ of a given function f then the sequence $\{g_n\}$ is weakly compact, every weak limiting function of this sequence is a best approximant in $L^1(B_\infty)$ of the function f, and the sequences $\{\max(g_n, U_{B\infty}(f))\}$ and $\{\min(g_n, V_{B\infty}(f))\}$ converge strongly to $U_{B\infty}(f)$ and $V_{B\infty}(f)$ respectively. Furthermore under the almost everywhere convergence of the sequence $\{B_n\}$ to B_∞ it is shown that

$$V_{B_\infty}(f) \leq \liminf_{n\to\infty} g_n \leq \limsup_{n\to\infty} g_n \leq U_{B_\infty}(f)$$

In the rest of this section, the terminologies, the notations and the important tools used in this paper are presented. Functions are denoted by f, g, h, etc. As the study depends largely on order-theoretic consideration, it is convenient to use some notations in the theory of vector lattices :

$$f \vee g := \max(f,g), \qquad f \wedge g := \min(f,g)$$

and

$$f^+ := \max(f,0) \qquad f^- := \max(-f,0).$$

Remark the familiar formula :

$$f+g = f\vee g + f\wedge g,$$

in particular,

$$f = f^+ - f^- \text{ and } |f| = f^+ + f^-.$$

A constant function with value α is denoted by α itself. The *indicator* (i. e. the characteristic function) of a set D is denoted by χ_D. For convenience, the notation $\{f \leq g\}$ (resp. $\{f < g\}$) will stand for the set of those points $\omega \in \Omega$ such that $f(\omega) \leq g(\omega)$ (resp. $f(\omega) < g(\omega)$). The *signature function* of a function f is defined by

$$\mathrm{sgn}(f) := \chi_{\{0<f\}} - \chi_{\{f<0\}}.$$

The norm of a function is always the L^1-norm :

$$\|f\| := \int_\Omega |f| dP.$$

Remark that any two functions are identified if they coincide with each other almost everywhere. The convergence in the norm topology is refered to *strong convergence*. The distance from a function f to the subspace $L^1(B)$ is denoted by $d(f, B)$, i. e.

$$d(f,B) := \inf\{\|f-h\| ; h \in L^1(B)\}.$$

Then a function g in $L^1(B)$ is, by definition, a *best approximant* of the function f if

$$\|f-g\| = d(f,B).$$

The dual of the Banach space L^1 is canonically realized by the space L^∞ of essentially bounded measurable functions (see [5 ; § 4.2]), and the *weak topology* is always understood with respect to the pairing (L^1, L^∞). For instance, a sequence $\{g_n\}$ in L^1 converges weakly to g_∞ if and only if for every h in L^∞

$$\lim_{n\to\infty} \int h \cdot g_n dP = \int h \cdot g_\infty dP.$$

The most important fact concerning the weak topology is the following criterion on weak compactness : a sequence $\{g_n\}$ is weakly compact if and only if it is bounded in norm and *equi-continuous* in the sense that $\|\chi_{D_n} \cdot g_n\|$ converges to 0 whenever $P(D_n)$ tends to 0 as $n \to \infty$.

In the study of this paper a dominant role is played by conditional expectation operators. The *conditional expectation* $E_B(f)$ of an integrable function f with respect to a σ-subalgebra B is defined as the uniquely determined function in $L^1(B)$ such that

$$\int_D E_B(f) dP = \int_D f dP \qquad \text{for every } D \text{ in } B.$$

The conditional expectation operator E_B is a linear projection of the space L^1 onto $L^1(B)$. This operator is known to possess the following remarkable properties (see [5 ; § 4.3]) :
(*Semi-multiplicativity*) : $E_B(f \cdot g) = E_B(f) \cdot g$ if $f \in L^1$ and $g \in L^\infty \cap L^1(B)$ or if $f \in L^\infty$ and $g \in L^1(B)$,
(*Monotonousness*) : $E_B(g) \leq E_B(f)$ if $g \leq f$,
(*Symmetry*) : $\int E_B(f) \cdot g dP = \int f \cdot E_B(g) dP$ if $f \in L^1$ and $g \in L^\infty$.

The conditional expectation of the indicator $\chi_{\{f \leq g\}}$ will be denoted by $P_B(f \leq g)$. The notation $P_B(f < g)$ has the corresponding meaning.

It is natural to introdue convergence concepts in the set of σ-subalgebras with the aid of conditional expectation operators. A sequence $\{B_n\}$ of σ-subalgebras is said to converge *strongly* (resp. *almost everywhere*) to a σ-subalgebra B_∞ if for every integrable function f the sequence $\{E_{B_n}(f)\}$ converges strongly (resp. almost everywhere) to the function $E_{B_\infty}(f)$. Almost everywhere convergence implies strong one. The most important fact concerning convergence of σ-subalgebras is the Martingale Convergence theorem (see [5 ; § 4.5]) : each monotone sequence $\{B_n\}$ of σ-subalgebras converges almost everywhere to a σ-algebra B_∞, where B_∞ is the intersection of all B_n or the smallest σ-subalgebra containing all B_n according as the sequence is decreasing or increasing.

2. Best Approximants

In this section B is a fixed σ-subalgebra and f is a fixed integrable function. A sequence $\{g_n\}$ in $L^1(B)$ is called an *optimal sequence* of f if the norm $\|f - g_n\|$ converges to the distance $d(f, B)$ from f to $L^1(B)$. In contrast to the case of L^p-approximation $(1 < p < \infty)$ (see [1]) the weak convergence of an optimal sequence is not immediate, for the unit ball of the Banach space L^1 is not weakly compact.

Theorem 1. *Each optimal sequence is weakly compact and every weak limiting function of the sequence is a best approximant. In particular, the set of best approximants is non-empty.*

Proof. Let $\{g_n\}$ be an optimal sequence of the given function f. First of all, the optimality implies boundedness in norm :

$$\limsup_{n\to\infty} \|g_n\| \leq \|f\| + \limsup_{n\to\infty} \|f - g_n\| \leq \|f\| + d(f, B).$$

Since the subspace $L^1(B)$ is isometric to the Banach space $L^1(\Omega, B, P)$, the weak compactness of $\{g_n\}$ in $L^1(B)$ will follow from (and is equivalent to) its weak compactness in $L^1(\Omega, B, P)$, or as mentioned in § 1, from its equi-continuity. Thus it suffices to prove that $\lim_{n\to\infty} \|\chi_{D_n} \cdot g_n\| = 0$ whenever $\lim_{n\to\infty} P(D_n) = 0$ $(D_n \in B)$. This is proved as follows. Since

$$\|\chi_{D_n} \cdot g_n\| + \|f - \chi_{D_n^c} \cdot g_n\| \leq \|f - g_n\| + 2\|\chi_{D_n} \cdot f\|$$

and

$$\lim_{n\to\infty} \|\chi_{D_n}\cdot f\|=0,$$

the optimality implies

$$\lim_{n\to\infty} \sup\|\chi_{D_n}\cdot g_n\|+\inf_n \|f-\chi_{D_n^c}\cdot g_n\|\leq d(f,B).$$

On the other hand, since $\chi_{D_n^c}\cdot g_n$ together with g_n and χ_{D_n} belongs to $L^1(B)$, the definition of distance implies

$$\inf_n\|f-\chi_{D_n^c}\cdot g_n\|\geq d(f,B).$$

These together yield the convergence of $\|\chi_{D_n}\cdot g_n\|$ to 0, so that the equi-continiuty has been established. Finally let g be any weak limiting function of the sequence $\{g_n\}$. Then it is well-known in the theory of Banach space that

$$\|f-g\|\leq\lim_{n\to\infty} \sup\|f-g_n\|.$$

Now the optimality of the sequence $\{g_n\}$ yields

$$\|f-g\|=d(f,B),$$

that is, g is a best approximant. This completes the proof.

The next theorem gives a characterization for a function in $L^1(B)$ to be a best approximant. Recall that $P_B(f<g)$ denotes the conditional expectation of the indicator $\chi_{\{f<g\}}$.

Theorem 2. *A function* g *in* $L^1(B)$ *is a best approximant of the given function f if and only if*

$$P_B(f<g)\leq 1/2 \text{ and } P_B(g<f)\leq 1/2.$$

Proof. By definition g is a best approximant if and only if

$$\|f-g+h\|\geq\|f-g\| \text{ for every } h \text{ in } L^1(B).$$

Since for each fixed h the function

$$\phi(\varepsilon):=\|f-g+\varepsilon h\| \quad (0\leq\varepsilon<\infty)$$

is convex and has the non-decreasing right-derivative, the latter condition turns out to be equivalent to the following :

$$\lim_{\varepsilon\to 0+}\varepsilon^{-1}\{\|f-g+\varepsilon h\|-\|f-g\|\}\geq 0 \quad (h\in L^1(B)).$$

Since $\varepsilon^{-1}\{|f-g+\varepsilon h|-|f-g|\}$ is majorated by $|h|$

and converges, as $\varepsilon\longrightarrow 0+$, almost everywhere to

$$\mathrm{sgn}\,(f-g)\cdot h+\{1-|\mathrm{sgn}\,(f-g)|\}\cdot|h|,$$

an application of the Lebesgue dominated convergence theorem (cf. [5 ; § 2.3]) yields

$$\lim_{\varepsilon\to 0+}\varepsilon^{-1}\{\|f-g+\varepsilon h\|-\|f-g\|\}$$
$$=\int\mathrm{sgn}\,(f-g)\cdot hdP+\int\{1-|\mathrm{sgn}\,(f-g)|\}\cdot|h|dP.$$

Now it follows by the semi-multiplicativity of the conditional expectation operator that

$$\int\mathrm{sgn}\,(f-g)\cdot hdP=\int E_B[\mathrm{sgn}\,(f-g)]\cdot hdP$$
$$=\int\{P_B(g<f)-P_B(f<g)\}\cdot\{h^+-h^-\}dP$$

and

$$\int\{1-|\mathrm{sgn}\,(f-g)|\}\cdot|h|dP$$
$$=\int\{1-P_B(g<f)-P_B(f<g)\}\cdot\{h^+ + h^-\}dP.$$

Therefore the condition for g to be a best approximant of f is that for every $h\in L^1(B)$

$$\int\{1-2P_B(f<g)\}\cdot h^+dP+\int\{1-2P_B(g<f)\}\cdot h^-dP\geq 0,$$

which is equivalent to the condition :

$$P_B(f<g)\leq 1/2 \quad \text{and} \quad P_B(g<f)\leq 1/2.$$

This completes the proof.

Lemma 3. *For each integrable function f there is the maximum of those functions g in $L^1(B)$ such that $P_B(f<g)\leq 1/2$.*

Proof. Denote by M the family of all those functions g in $L^1(B)$ such that $P_B(f<g)\leq 1/2$. First, M is non-empty. In fact, fix a best approximant g_0, the existence of which is guaranteed by Theorem 1. Then g_0 belongs to M by Theorem 2. Second, M is stable under the lattice operation $\vee$. To see this, take g_1, $g_2\in M$ and let D denote the set $\{g_1>g_2\}$. Then D belongs to the σ-subalgebra B and $g_1\vee g_2=\chi_D\cdot g_1+\chi_{D^c}\cdot g_2$ so that Theorem 2 and the semi-multiplicativity of the conditional expectation operator yield

$$P_B(f<g_1\vee g_2)=\chi_D\cdot P_B(f<g_1)+\chi_{D^c}\cdot P_B(f<g_2)\leq 1/2,$$

hence $g_1\vee g_2$ belongs to M. Now since every upper-directed bounded sequence in L^1 is a Cauchy sequence, the proof of the theorem will be completed if the subfamily $\{g\in M: g_0\leq g\}$ is closed and bounded in norm. To this end, take $g\in M$ with $g_0<g$. Then Theorem 2 and the monotonousness of the conditional expectation operator imply

$$P_B(g<f)\leq P_B(g_0<f)\leq 1/2$$

while $P_B(f<g)\leq 1/2$ by the definition of M. Therefore g is a best approximant by Theorem 2. This consideration shows that the subfamily in question coincides with the family of all best approximants greater than or equal to g_0, hence is closed and bounded in norm (indeed, bounded in norm by $2\|f\|$).

The maximum in the assertion of Lemma 3 will be denoted by $U_B(f)$. Since $P_B(g<f)=P_B(-f<-g)$, it follows immdeiately that $-U_B(-f)$ is the minimum of all functions g in $L^1(B)$ such that $P_B(g<f)<1/2$. This minimum will be denoted by $V_B(f)$, i.e.

$$V_B(f)=-U_B(-f).$$

Theorem 4. *A function* g *in $L^1(B)$ is a best approximant if and only if*

$$V_B(g)\leq g\leq U_B(f).$$

Proof. In view of Theorem 2 and the definition of $U_B(f)$ and $V_B(f)$ every best approximant satisfies the inequalities in question. If conversely $V_B(f)\leq g\leq U_B(f)$ then

$$P_B(f<g)\leq P_B[f<U_B(f)]\leq 1/2$$

and

$$P_B(g<f)\leq P_B[V_B(f)<f]\leq 1/2.$$

Now the assertion follows again from Theorem 2.

In the manuscript [4] H. Kudo constructed a best approximant and showed that every function g in $L^1(B)$ such that $P_B(f<g)\leq 1/2$ and $P_B(g<f)\leq 1/2$ is a best approximant. Though he did not obtain so complete description as Theorem 2, the idea of Lemma 3 is due to him.

Usually an integrable function admits many best approximants ; in general, $U_B(f)$ does not coincide with $V_B(f)$. Theorem 1 indicates that every optimal sequence admits a subsequence which is weakly convergent to a function between $V_B(f)$ and $U_B(f)$. Here strong convergence is not expected, for the family $\{g\in L^1(B)\ ;\ V_B(f)\leq g\leq U_B(f)\}$ is not strongly compact.

The following simple lemma is useful to an improvement of Theorem 1 as well as in the next section.

Lemma 5. *If a function g in $L^1(B)$ is a best approximant and a set D belongs to the σ-subalgebra B, then*

$$\|\chi_D\cdot(f-g)\|=\inf\{\|\chi_D\cdot(f-h)\|\,;\,h\in L^1(B)\}.$$

Proof. For every $h\in L^1(B)$ the function $\chi_D\cdot h+\chi_{D^c}\cdot g$ belongs to $L^1(B)$ so that by the definition of a best approximant

$$\|f-(\chi_D\cdot h+\chi_{D^c}\cdot g\|\geq d(f,B)=\|f-g\|.$$

On the other hand, the additivity property of the L^1-norm yields

$$\|f-(\chi_D\cdot h+\chi_{D^c}\cdot g)\|=\|\chi_D\cdot(f-h)\|+\|\chi_{D^c}\cdot(f-g)\|$$

and

$$\|f-g\|=\|\chi_D\cdot(f-g)\|+\|\chi_{D^c}\cdot(f-g)\|.$$

These together prove the assertion.

Theorem 6. If $\{g_n\}$ is an optimal sequence, then the sequences $\{g_n\vee U_B(f)\}$ and $\{g_n\wedge V_B(f)\}$ converge strongly to $U_B(f)$ and $V_B(f)$ respectively.

Proof. For every n the set $D_n:=\{g_n>U_B(f)\}$ belongs to the σ-subalgebra B and

$$g_n\vee U_B(f)=\chi_{D_n}\cdot g_n+\chi_{D_n^c}\cdot U_B(f).$$

Then Lemma 5 implies

$$\begin{aligned}\|f-g_n\|&=\|\chi_{D_n}\cdot(f-g_n)\|+\|\chi_{D_n^c}\cdot(f-g_n)\|\\&\geq\|\chi_{D_n}\cdot(f-g_n)\|+\|\chi_{D_n^c}\cdot(f-U_B(f))\|\\&=\|f-g_n\vee U_B(f)\|\geq d(f,B),\end{aligned}$$

hence the sequence $\{g_n\vee U_B(f)\}$ as well as $\{g_n\}$ is optimal.

Now in view of Theorems 1 and 4 the sequence $\{g_n\vee U_B(f)\}$ is weakly compact and every weak limiting function is a best approximant, which is majorated by $U_B(f)$. This leads to the relation :

$$\begin{aligned}&\limsup_{n\to\infty}\|g_n\vee U_B(f)-U_B(f))\|\\&=\limsup_{n\to\infty}\int g_n\vee U_B(f)dP-\int U_B(f)dP\leq 0.\end{aligned}$$

The proof for the convergence of $\{g_n\wedge V_B(f)\}$ is similar.

3. Operators U_B and V_B

In this section B is a fixed σ-subalgebra. Recall that $U_B(f)$ (resp. $V_B(f)$) is the maximum (resp. minimum) of all best approximants of f. The mapping $f \longmapsto U_B(f)$ (or $V_B(f)$) is non-linear but it has many remarkable properties.

Theorem 7. *The non-linear operator U_B has the following properties :*

(a) $U_B(af+h)=aU_B(f)+h$ *for* $a\geq 0$ *and* $h\in L^1(B)$.

(b) $U_B(f_1)< U_B(f_2)$ *whenever* $f_1\leq f_2$.

(c) $U_B(\chi_D\cdot f)=\chi_D\cdot U_B(f)$ *for* $D\in B$.

(d) $U_B(f\wedge a)=U_B(f)\wedge a$ *and* $U_B(f\vee a)=U_B(f)\vee a$ *for real* a.

(e) $\|U_B(g)\|\leq 2|f|$.

The assertions (a)~(e) *are valid with* U_B *replaced by* V_B.

Proof. The assertion (a) results immediately from the definition of U_B. Recall that by Lemma 3 the function $U_B(f)$ is the maximum of those functions g in $L^1(B)$ such that $P_B(f<g)\leq 1/2$ where $P_B(f<g)$ is the conditional expectation, with respect to B, of the indicator $\chi_{\{f<g\}}$. Now the assertion (b) is a consequence of the monotonousness of the conditional expectation operator. To see (c), take any $D\in B$. Since the semi-multiplicativity of the conditional expectation operator yields

$$P_B[\chi_D\cdot f<\chi_D\cdot U_B(f)]=\chi_D\cdot P_B< U_B(f)],$$

the characterization of U_B mentioned above implies

$$\chi_D\cdot U_B(f)\leq U_B(\chi_D\cdot f)$$

Since this implies, for every $n>0$,

$$P_B[\chi_D\cdot f< n\chi_{D^c}\cdot U_B(\chi_D\cdot f)]\leq P_B[\chi_D\cdot f< U_B(\chi_D\cdot f)]\leq 1/2$$

and since $\chi_{D^c}\cdot U_B(\chi_D\cdot f)$ belongs to $L^1(B)$, Lemma 3 yields

$$n\chi_{D^c}\cdot U_B(\chi_D\cdot f)\leq U_B(\chi_D\cdot f) \qquad (n=1,2,\cdots),$$

which together with the part proved above implies

$$0=\chi_{D^c}\cdot\chi_D\cdot U_B(f)\leq\chi_{D^c}\cdot U_B(\chi_D\cdot f)\leq\frac{1}{n}U_B(\chi_D\cdot f),$$

hence

$$U_B(\chi_D\cdot f)=\chi_D U_B(\chi_D\cdot f).$$

The corresponding formula with D replaced by D^c is

$$U_B(\chi_{D^c}\cdot f)=\chi_{D^c}\cdot U_B(\chi_{D^c}\cdot f).$$

Then since by Theorem 4 and the semi-multiplicativity of the conditional expectation operator

$$\begin{aligned}&P_B[f< U_B(\chi_D\cdot f)+U_B(\chi_{D^c}\cdot f)]\\&=P_B[\chi_D\cdot f< U_B(\chi_D\cdot f)]+P_B[\chi_{D^c}\cdot f< U_B(\chi_{D^c}\cdot f)]\\&\leq(\chi_D+\chi_{D^c})/2=1/2,\end{aligned}$$

Lemma 3 yields

$$U_B(\chi_D\cdot f)+U_B(\chi_{D^c}\cdot f)\leq U_B(f)$$

hence

$$U_B(\chi_D\cdot f)=\chi_D\cdot U_B(\chi_D\cdot f)\leq \chi_D\cdot U_B(f).$$

This together with the reversed inaquality obtained earlier proves (c).

To prove (d), remark first of all that (a) and (b), combined with Lemma 3, imply

$$U_B(f\wedge a)\leq U_B(f)\wedge a \text{ and } U_B(f\vee a)\geq U_B(f)\vee a.$$

Since $U_B(f)\wedge a$ belongs to $L^1(B)$ and

$$P_B[f\wedge a< U_B(f)\wedge a]\leq P_B[f< U_B(f)]\leq 1/2,$$

the definition of U_B in Lemma 3 implies $U_B(f)\wedge a\leq U_B(f\wedge a)$. Now to complete the proof of (d), it remain to show that the set $D:=\{U_B(f\vee a)>U_B(f)\vee a\}$ has measure zero. Obiously D belongs to the σ-subalgebra B and

$$D\cap\{f< U_B(f\vee a)]\subseteq\{f\vee a< U_B(f\vee a)\}$$

so that the semi-multiplicativity and the monotonousness of the conditional expectation operator yield

$$\chi_D\cdot P_B[f< U_B(f\vee a)]\leq P_B[f\vee a< U_B(f\vee a)]\leq 1/2,$$

hence by Lemma 3 and (c)

$$\chi_D\cdot U_B(f\vee a)\leq U_B(\chi_D\cdot f)=\chi_D\cdot U_B(f)\leq \chi_D\cdot[U_B(f)\vee a],$$

which leads to $P(D)=0$ by the definition of D.

To see (e), it suffices to prove that

$$\int_D |U_B(f)|dP\leq 2\int_D E_B(|f|)dP \qquad (D\in B).$$

But since $U_B(f)$ is a best approximant, Lemma 5 yields

$$\|\chi_D\cdot[f-U_B(f)]\|\leq\|\chi_D\cdot f\|$$

so that the definition and the semi-multiplicativity of the conditional expectation operator imply

$$\|\chi_D\cdot U_B(f)\|\leq 2\|\chi_D\cdot f\|=2\|E_B(\chi_D\cdot f)\|=2\|\chi_D\cdot E_B(f)\|,$$

which is just the expected inequality.

The assertions (a)~(e) with U_B replaced by V_β follow immediately from the relation : $V_B(h)=-U_B(-h)$. This completes the proof.

These remarkable properties of the operator U_B make it possible to give a concrete representation of $U_B(f)$ in terms of the given function f. Such a representation will be most conveniently established with the aid of the *spectral theory* that is familiar in the theory of vector lattices.

A one-parameter family $\{g_\lambda;-\infty<\lambda<\infty\}$ of indicators in L^1 is called a *spectral family* if it is non-decreasing, i.e. $g_{\lambda_1}\leqq g_{\lambda_2}$ whenever $\lambda_1\leq\lambda_2$, and g_λ converges strongly to 0 or 1 according as $\lambda\longrightarrow-\infty$ or $\longrightarrow\infty$. A spectral family $\{g_\lambda\}$ introduces the notion of integral in a natural way : for $-\infty<a<\beta<\infty$ the *inegral* $\int_a^\beta\lambda dg_\lambda$ is defined as the strong limit of the Riemann – Stieltjes sum

$$\sum_{j=1}^n \xi_j(g_{\lambda_j}-g_{\lambda_{j-1}})$$

as $\max_j(\lambda_j-\lambda_{j-1})$ tends to 0, where

$$a=\lambda_0<\lambda_1<\cdots<\lambda_n=\beta \text{ and } \lambda_{j-1}\leq\xi_j\leq\lambda_j.$$

Just as in the case of the numerical Riemann – Stieltjes integral the following estimate is valid :

$$\left|\int_\alpha^\beta \lambda dg_\lambda - \sum_{j=1}^n \xi_j(g_{\lambda_j} - g_{\lambda_{j-1}})\right| \leq \max_{1\leq j\leq n}(\lambda_j - \lambda_{j-1}).$$

When $\int_\alpha^\beta \lambda dg_\lambda$ is strongly convergent as $\alpha \longrightarrow -\infty$ and $\beta \longrightarrow \infty$, the limit will be denoted by $\int_{-\infty}^\infty \lambda dg_\lambda$.

To each integrable function f there is assigned its *canonical* spectral family $\{f_\lambda\}$, where f_λ is the indicator of the set $\{f \leq \lambda\}$. This spectral family is *right-continuous* in the sense that f_μ converges strongly to f_λ as μ tends to λ from the right. The function f itself is recaptured as the integral :

$$f = \int_{-\infty}^\infty \lambda df_\lambda.$$

Conversely if f is given as the integral $\int_{-\infty}^\infty \lambda dg_\lambda$ with respect to a spectral family $\{g_\lambda\}$ then each f_λ in the canonical spectral family is determined as the strong limit of g_μ when μ tends to λ from the right. These are only a reformulation of the Lebesgue integration theory.

Lemma 8. *Let* $-\infty < \alpha < \beta < \infty$ *and* $\varepsilon > 0$. *For each integrable function* f *and for each partition of the interval* $[\alpha, \beta]$ *such that*

$$\alpha = \lambda_0 < \lambda_1 < \cdots < \lambda_m = \beta \text{ with } \max_{1\leq j\leq m}(\lambda_j - \lambda_{j-1}) \leq \varepsilon,$$

$$|U_B(f) - \sum_{j=1}^m (g_{\lambda_j} - g_{\lambda_{j-1}}) - \alpha g_\alpha - \beta(1 - g_\beta)|$$
$$\leq \varepsilon + 2E_B[(f-\alpha)^- + (f-\beta)^+]$$

and

$$\left||U_B(f) - f| - \sum_{j=1}^{m-1}(\lambda_{j+1} - \lambda_j)|f_{\lambda_j} - g_{\lambda_j}|\right|$$
$$\leq 2\varepsilon + 2(f-\alpha)^- + 2(f-\beta)^+ + 2E_B[(f-\alpha)^- + (f-\beta)^+],$$

where f_λ *and* g_λ *are the indicators of the set* $\{f \leq \lambda\}$ *and* $\{\frac{1}{2} < P_B(f \leq \lambda)\}$ *respectively.*

Proof. Since Theorem 7 implies

$$U_B(f) = [U_B(f) \vee \alpha] \wedge \beta - [U_B(f) - \alpha]^- + [U_B(f) - \beta]^+$$
$$= U_B[(f \vee \alpha) \wedge \beta] + U_B[-(f-\alpha)^-] + U_B[(f-\beta)^+]$$

it follows that

$$|U_B(f) - U_B[(f \vee \alpha) \wedge \beta]|$$
$$\leq 2E_B[(f-\alpha)^- + (f-\beta)^+].$$

Let f' $= \sum_{j=1}^m \lambda_j(f_{\lambda_j} - f_{\lambda_{j-1}}) + \alpha f_\alpha + \beta(1 - f_\beta)$

and g' $= \sum_{j=1}^m \lambda_j(g_{\lambda_j} - g_{\lambda_{j-1}}) + \alpha g_\alpha + \beta(1 - g_\beta)$.

Since obviously

$$|(f \vee \alpha) \wedge \beta - f'| \leq \max_{1\leq j\leq m}(\lambda_j - \lambda_{j-1}) \leq \varepsilon,$$

Theorem 7 yields

$$|U_B[(f \vee \alpha) \wedge \beta] - U_B(f')| \leq \varepsilon,$$

hence

$$|U_B(f) - U_B(f')| \leq \varepsilon + 2E_B[(f-\alpha)^- + (f-\beta)^+].$$

Then to prove the first assertion, it suffices to show $U_B(f') = g'$. To this end, let

$$C_j := \{f < \lambda_j\} \text{ and } D_j := \{\tfrac{1}{2} < P_B(f \leq \lambda_j)\} \quad j = 0, 1, \cdots, m.$$

Since obviously

$$f'=\sum_{j=1}^{m}\lambda_j\chi_{C_j\cap C^c_{j-1}}+\alpha\chi_{C_0}+\beta\chi_{C^c_m}$$

and

$$g'=\sum_{j=1}^{m}\lambda_j\chi_{D_j\cap D^c_{j-1}}+\alpha\chi_{D_0}+\beta\chi_{D^c_m}$$

it follows that

$$\{f'<g'\}=\bigcup_{j=1}^{m}[C_{j-1}\cap D_j\cap D^c_{j-1}]\cup[C_{m-1}\cap D^c_m]$$

and

$$\{g'<f'\}=\bigcup_{j=1}^{m}[D_{j-1}\cap C_j\cap C^c_{j-1}]\cup[D_{m-1}\cap C^c_m].$$

Since every D_j belongs to the σ-subalgebra B, the semi-multiplicativity of the conditional expectation operator yields, on the basis of definition of D's,

$$P_B(f'<g')=\sum_{j=1}^{m}P_B(f\leq\lambda_{j-1})\cdot\chi_{D_j\cap D^c{j-1}}+P_B(f\leq\lambda_{m-1})\cdot\chi_{D^c m}\leq 1/2,$$

which implies by Lemma 3 the inequality : $g'\leq U_B(f')$. On the other hand, the semi-multiplicativity of the conditional expectation operator again yields

$$\chi_{\{\lambda_j<U_B(f')\}}\cdot P_B(f'\leq\lambda_j)\leq P_B[f'<U_B(f')]\leq 1/2,$$

in other words

$$\{\lambda_j<U_B(f')\}\subseteq\{P_B(f'\leq\lambda_j)\leq 1/2\}.$$

Since by the construction of f'

$$\{f'<\lambda_j\}=C_j=\{f<\lambda_j\} \qquad j=0,1,\cdots,m-1$$

the definition of g' implies

$$\{\lambda_j<U_B(f')\}\subseteq D^c_j=\{\lambda_j<g'\} \quad j=0,1,\cdots,m-1.$$

This implies $U_B(f')\leq g'$ and consequently $U_B(f')=g'$. In fact, otherwise there is a λ such that the set $\{\lambda<U_B(f')\}$ is not contained in the set $\{\lambda<g'\}$. Since both $U_B(f')$ and g' are majorated by λ_m and minorated by λ_0, there is $0<j<m-1$ such that $\lambda_j\leq\lambda<\lambda_{j+1}$. Then since by the definition of g' the set $\{\lambda<g'\}$ coincides with the set $\{\lambda_j<g'\}$, the set $\{\lambda_j<U_B(f')\}$ is not contained in $\{\lambda_j<g'\}$, a contradiction.

To see the second assertion, remark that, when applied to the algebra A instead of B, the first assertion just proved yields

$$|f-f'|\leq\varepsilon+2(f-\alpha)^-+2(f-\beta)^+,$$

so that

$$\left||U_B(f)-f|-(g'-f')\right|$$
$$\leq 2\varepsilon+2(f-\alpha)^-+2(f-\beta)^++2E_B[(f-\alpha)^-+(f-\beta)^+].$$

Now it remains to prove that

$$|g'-f'|=\sum_{j=1}^{m-1}(\lambda_{j+1}-\lambda_j)|f_{\lambda_j}-g_{\lambda_j}|.$$

To this end, note that

$$g'-f'=\sum_{j=1}^{m-1}(\lambda_{j+1}-\lambda_j)(f_{\lambda_j}-g_{\lambda_j})$$

and

$$\operatorname{sgn}(g'-f')=\sum_{j=1}^{m}\chi_{C_{j-1}\cap D_j\cap D^c_{j-1}}+\chi_{C_{m-1}\cap D^c_m}$$
$$-\sum_{j=1}^{m}\chi_{D_{j-1}\cap C_j\cap C_{j-1}{}^c}-\chi_{D_{m-1}\cap C^c_m}.$$

Then it is readily seen that for $0\leq k\leq m-1$

$$(f_{\lambda_k}-g_{\lambda_k})\cdot\operatorname{sgn}(g'-f')$$
$$=\chi_{C_k\cap D^c_k}+\chi_{D_k\cap C_k{}^c}=|f_{\lambda_k}-g_{\lambda_k}|$$

and consequently

$$|g'-f'|=\sum_{k=1}^{m-1}(\lambda_{k+1}-\lambda_k)(f_{\lambda_k}-g_{\lambda_k})\cdot\operatorname{sgn}(g'-f')$$
$$=\sum_{k=1}^{m-1}(\lambda_{k+1}-\lambda_k)|f_{\lambda_k}-g_{\lambda_k}|.$$

This completes the proof.

Theorem 9. *For each integrable function f its maximum best approximant $U_B(f)$ and the minimum one $V_B(f)$ admit the following integral representations :*

$$U_B(f)=\int_{-\infty}^{\infty}\lambda dg_\lambda \qquad \text{and} \qquad V_B(f)=\int_{-\infty}^{\infty}\lambda dh_\lambda.$$

where g_λ and h_λ are the indicators of the sets $\{1/2<P_B(f\leq\lambda)\}$ and $\{1/2\leq P_B(f\leq\lambda)\}$ respectively. Further the distance from the function f to the subspace $L^1(B)$ is explicitly given by the following formula :

$$d(f,B)=\int_{-\infty}^{\infty}[1/2-\|1/2-P_B(f\leq\lambda)\|]d\lambda.$$

Froof. Given $\varepsilon>0$, take $\alpha<\infty<\beta$ such that

$$\|\chi_{\{f\leq\alpha\}}\cdot f\|+|\chi_{\{f>\beta\}}\cdot f\|\leq\varepsilon$$

and take a partition : $\alpha=\lambda_0<\lambda_1<\cdots<\lambda_m=\beta$ with $\max\limits_{1\leq j\leq m}(\lambda_j-\lambda_{j-1})\leq\varepsilon$. Then obviously

$$|\int_\alpha^\beta\lambda dg_\lambda-\sum_{j=1}^{m}\lambda_j(g_{\lambda_j}-g_{\lambda_{j-1}})|\leq\varepsilon$$

while by Lemma 8

$$|U_B(f)-\sum_{j=1}^{m}\lambda_j(g_{\lambda_j}-g_{\lambda_{j-1}})|$$
$$\leq\varepsilon+2E_B[(f-\alpha)^-+(f-\beta)^+]+|\alpha|g_\alpha+\beta(1-g_\beta).$$

Remark that by the definition of the conditional expectation

$$\|E_B[(f-\alpha)^-+(f-g)^+]\|=\|(f-\alpha)^-+(f-\beta)^+\|$$
$$\leq\|\chi_{[f\leq\alpha]}\cdot f\|+\|\chi_{[f>\beta]}\cdot f\|$$

and by the definition of g_λ

$$|\alpha|\cdot\|g_\alpha\|\leq2|\alpha|\cdot\|P_B(f\leq\alpha)\|\leq2\|\chi_{[f\leq\alpha]}\cdot f\|$$

and

$$\beta\|1-g_\beta\|\leq2\beta\|P_\beta(\beta<f)\|\leq2\|\chi_{[f>\beta]}\cdot f\|.$$

These inequalities yield

$$\|U_B(f)-\int_\alpha^\beta\lambda dg_\lambda\|\leq6\varepsilon$$

and the integral representation for $U_B(f)$ follows immediately.

The integral representation for $V_B(f)$ follows from the relation :

$$V_B(f)=-U_B(-f).$$

To get the expression for the distance $d(f,B)$, let $\{f_\lambda\}$ be the canonical spectral family of f, that is, for each λ, f_λ is the indicator of the set $\{f\leq\lambda\}$. Since $d(f,B)=\|U_B(f)-f\|$by definition, Lemma 8 implies, just as above,

$$\left|d(f,B)-\sum_{j=1}^{m-1}(\lambda_{j+1}-\lambda_j)\|f_{\lambda_j}-g_{\lambda_j}\|\right|\leq 6\varepsilon$$

hence

$$d(f,B)=\int_{-\infty}^{\infty}\|f_\lambda-g_\lambda\|d\lambda.$$

Now it remains to show that

$$\|f_\lambda-g_\lambda\|=1/2-\|1/2-P_B(f\leq\lambda)\|.$$

Remark that $P_B(f\leq\lambda)=E_B(f_\lambda)$ and

$$|1/2-P_B(f\leq\lambda)|=\{E_B(f_\lambda)-1/2\}g_\lambda+\{1/2-E_B(f_\lambda)\}\cdot(1-g_\lambda).$$

Since g_λ belongs to $L^\infty\cap L^1(B)$, the semi-multiplicativity of the conditional expectation operator yields

$$\|f_\lambda-g_\lambda\|=\int[f_\lambda\cdot(1-g_\lambda)+(1-f_\lambda)\cdot g_\lambda]dP$$
$$=\int[E_B(f_\lambda)\cdot(1-g_\lambda)+\{1-E_B(f_\lambda)\}\cdot g_\lambda]dP$$
$$=1/2-\int|1/2-E_B(f_\lambda)|dP=1/2-\|1/2-P_B(f\leq\lambda)\|.$$

This completes the proof.

The formula for the distance in Theorem 9 was found, in its primitive form, by Kudo [3, 4].

4. Convergence

In this section $B_1, B_2, \cdots, B_\infty$ are σ-subalgebras. Recall that the sequence $\{B_n\}$ converges strongly (resp. almost everywhere) to B_∞ if for every function f in L^1 the sequence $\{E_{B_n}(f)\}$ converges strongly (resp. almost everywhere) to $E_{B_\infty}(f)$. Kudo [3] proved that the strong convergence of $\{B_n\}$ to B_∞ results from the following weaker condition :

$$\log_{n\to\infty}\|E_{B_n}(f)\|=\|E_{B_\infty}(f)\| \qquad (f\in L^1).$$

On the other hand, Becker [2] pointed out that the strong convergence is a consequence of the following weaker condition : for every f in L^1 the sequence $\{E_{B_n}(f)\}$ converges weakly to $E_{B_\infty}(f)$.

Convergence problems of best approximants in the case of L^p spaces $(1<p<\infty)$ were discussed in a previous paper [1].

Theorem 10. *Suppose that a sequence $\{B_n\}$ of σ-subalgebras converges strongly to a σ-subalgebra B_∞. If each g_n is a best approximant in $L^1(B_n)$ of one and the same function f, then the sequence $\{g_n\}$ is weakly compact and the sequence $\{g_n\vee U_{B_\infty}(f)\}$ (resp. $\{g_n\wedge V_{B_\infty}(f)\}$) converges strongly to $U_{B_\infty}(f)$ (resp. $V_{B_\infty}(f)$). Every weak limiting function of the sequence $\{g_n\}$ is a best approximant in $L^1(B_\infty)$ of the function f.*

Proof. Since strong convergence implies equi-continuity and since

$$|g_n|\leq|g_n\vee U_{B_\infty}(f)|+|g_n\wedge V_{B_\infty}(f)|$$

the weak compactness of the sequence $\{g_n\}$ will follow from the strong convergence of $\{g_n\vee U_{B_\infty}(f)\}$ and $\{g_n\wedge V_{B_\infty}(f)\}$, as remarked in § 1. To prove the strong convergence of the sequence $\{g_n\vee U_{B_\infty}(f)\}$

to $U_{B_\infty}(f)$, take according to Lemma 8 ε, $\gamma>0$ and a partition $-\gamma=\lambda_0<\lambda_1<\cdots<\lambda_m=\gamma$ such that

$$|U_{B_n}(f)-\sum_{j=1}^{m-1}(\lambda_j-\lambda_{j+1})g_{\lambda_j}^{(n)}-\gamma|$$
$$\leq\varepsilon+2E_{B_n}[(f-\gamma)^+ +(f+\gamma)^-] \qquad (n=1,2,\cdots),$$

where $g_\lambda^{(n)}$ is the indicator of the set $\{1/2<P_B(f\leq\lambda)\}$. Since for every λ, by hypothesis, $P_{Bn}(f\leq\lambda)$ converges strongly to $P_{B_\infty}(f\leq\lambda)$ as $n\longrightarrow\infty$, by taking a subsequence if necessary, it can be assumed that for every $j=0, 1, \cdots, m$ the sequence $\{P_{B_n}(f\leq\lambda_j)\}$ converges almost everywhere to $P_{B_\infty}(f\leq\lambda_j)$ as $n\longrightarrow\infty$. Further it can be assumed that the sequence $\{E_{Bn}[(f-\gamma)^+ +(f+\gamma)^-]\}$ is dominated by some integrable function and converges almost everywhere to $E_{B\infty}[(f-\gamma)^+ +(f+\gamma)^-]$. Since $g_n\leq U_{B_n}(f)$ by Theorem 4, it follows that

$$g_n\vee U_B(f)-U_{B_\infty}(f)$$
$$=[g_n-U_{B_\infty}(f)]^+\leq[U_{B_n}(f)-U_{B_\infty}(f)]^+$$
$$\leq\sum_{j=1}^{m-1}(\lambda_{j+1}-\lambda_j)[g_{\lambda_j}^{(\infty)}-g_{\lambda_j}^{(n)}]+2\varepsilon$$
$$+2E_{B_n}[(f-\gamma)^+ +(f+\gamma)^+]+2E_{B_\infty}[(f-\gamma)^+ +(f+\gamma)^-].$$

Then

$$0\leq\limsup_{n+\infty}[g_n\vee U_{B_\infty}(f)-U_{B_\infty}(f)]$$
$$\leq 2\varepsilon+4E_{B_\infty}[(f-\gamma)^+ +(f+\gamma)^-]$$

as well as

$$\limsup_{n\to\infty}\|g_n\vee U_{B_\infty}(f)-U_{B_\infty}(f)\|$$
$$\leq 2\varepsilon+4\|(f-\gamma)^+\|+4\|(f+\gamma)^-\|.$$

Now the strong and almost everywhere convergence follows by letting $\gamma\longrightarrow\infty$ and $\varepsilon\longrightarrow 0$. Remind that strong convergence takes place for the sequence $\{g_n\vee U_{B_\infty}(f)\}$ itself while almost everywhere convergence can occur, in general, only for some subsequence unless the sequence $\{B_n\}$ itself converges almost everywhere to B_∞. The strong convergence of the sequence $\{g_n\wedge V_{B_\infty}(g)\}$ to $V_{B_\infty}(f)$ is proved similarly. Finally, to complete the proof of the assertion, suppose that the sequence $\{g_n\}$ converges weakly to some g_∞. Since for each function h in L^∞ the sequence $\{E_{B_n}(h)\}$ converges strongly, hence in measure, to $E_{B_\infty}(h)$, the equi-continiity of the sequence $\{g_n\}$ yields

$$\lim_{n\to\infty}\int E_{B_n}(h)\cdot g_n dP=\int E_{B_\infty}(h)\cdot g_\infty dP.$$

On the other hand, the symmetry of the conditional expectation operator implies

$$\lim_{n\to\infty}\int E_{B_n}(h)\cdot g_n dP=\lim_{n\to\infty}\int h\cdot E_{B_n}(g_n)dP=\int h\cdot g_\infty dP.$$

These together yield $g_\infty=E_{B_\infty}(g_\infty)$, so that g_∞ belongs to $L^1(B_\infty)$. The weak convergence of the sequence $\{g_n\}$ to g_∞ implies

$$d(f,B_\infty)\leq\|f-g_\infty\|\leq\liminf_{n\to\infty}\|f-g_n\|=\liminf_{n\to\infty}d(f,B_n).$$

Finally, for each function g in $L^1(B_\infty)$ the strong convergence of $\{E_{B_n}(g)\}$ to g implies

$$\|f-g\|=\lim_{n\to\infty}\|f-E_{B_n}(g)\|\geq\limsup_{n\to\infty}d(f,B_n).$$

These together yield that g_∞ is a best approximant in $L^1(B_\infty)$. The theorem has been completely proved.

Theorem 11. *If a sequence $\{B_n\}$ of σ-subalgebras converges almost everywhere to a σ-subalgebra B_∞, then for each integrable function f*

$$V_{B_\infty}(f)\leq\liminf_{n\to\infty}V_{B_n}(f)\leq\limsup_{n\to\infty}U_{B_n}(f)\leq U_{B_\infty}(f).$$

Froof. Since for every λ the sequences $\{P_{B_n}(f \leq \lambda)\}$ and $\{E_{B_n}[(f-\lambda)^+ + (f+\lambda)^-]\}$ converge almost everywhere to $P_{B_\infty}(f \leq \lambda)$ and $E_{B_\infty}[(f-\lambda)^+ + (f+\lambda)^-]\}$ respectively, the proof of Theorem 10 proceeds, not with a subsequence but with the sequence $\{B_n\}$ itself, and the assertion follows with g_n substituted by $U_{B_n}(f)$ or $V_{B_n}(f)$.

Corollary 12. *Suppose that $\{B^n\}$ is a monotone sequence of σ-sugalgebras, and let B_∞ be the intersection of all B_n's or the smallest σ-subalgebra containing all B_n's according as the sequence $\{B_n\}$ is decreasing or increasing. If each g_n is a best approximant in $L^1(B_n)$ of one and the same function f, then the sequence $\{g_n\}$ is weakly compact and*

$$V_{B_\infty}(f) \leq \lim_{n\to\infty} \inf g_n \leq \lim_{n\to\infty} \sup g_n \leq U_{B_\infty}(f).$$

Every weak limiting function of this sequence is a best approximant in $L^1(B_\infty)$ of the function f.

This follows immdeiately from Theorems 10 and 11, for the sequence $\{B_n\}$ converges almost everywhere to B_∞ by the Martingale theorem, mentioned in § 1.

References

1. Ando, T., Amemiya, I. : Almost everywhere convergence of prediction sequences in L_p, Z. Wahrscheinlichkeitstheorie verw. Geb. 4, 113−120 (1965).
2. Becker, R. : Ensembles compacts de tribus, Z. Wahrscheinlichkeitstheorie verw. Geb. 29, 229−234 (1974).
3. Kudo, H. : A note on the strong convergence of σ-algebras, Ann. of Probability 2, 76−83 (1974).
4. Kudo, H. : An appendix to [3] (unpublished).
5. Neveu, J. : Bases Mathematiques du Calcul des Probabilitēs, 2ieme ed., Paris, Masson et c^{ie}, 1970.

T. Shintani
Department of Mathematics
Tomakomai Technical College
Tomakomai/Hokkaido
Japan

T. Ando
Division of Applied Mathematics
Research Institute of Applied Electricity
Hokkaido University
Sapporo
Japan