JOURNAL OF THE
CHUNGCHEONG MATHEMATICAL SOCIETY
Volume **24**, No. 2, June 2011

AN EXTENDED THEOREM FOR GRADIENTS AND SUBGRADIENTS

Hyang Joo Rhee*

Abstract. In this paper, we introduce certain concepts which we will provide us with a perspective and insight into the problem of calculating best approximations. The material of this paper will be mainly, but not only, used in developing algorithms for the one-sided and two-sided sided approximation problem.

1. Introduction

We first fix some notation. For a set B, Σ a σ-field of subsets of B, and ν a positive measure defined on Σ, i.e., $\nu(E) \geq 0$ for all $E \in \Sigma$. By $L^p(B,\nu), 1 \leq p \leq \infty$, we denote the set of all real-valued ν-measurable functions f defined on B for which $|f|^p$ is ν-integrable over B. We consider two functions of $L^p(B,\nu)$ as equivalent if they are equal ν a.e.. Under this convention $L^p(B,\nu)$ with norm

$$\begin{aligned} ||f||_p &= (\int_B |f(x)|^p \mathrm{d}\nu(x))^{\frac{1}{p}} \\ &= (\int_B |f|^p d\nu)^{\frac{1}{p}} \end{aligned}$$

is a normed linear space and in fact a Bancch space.

Let $L^\infty(B,\nu)$ is defined analogously with norm

$$||f||_\infty = \operatorname{ess\,sup}_{x\in B}|f(x)|$$

where the ess sup is the infimum of all real constants c for which $|f(x)| \leq c$, ν a.e.. Then $L^\infty(B,\nu)$ is also a Banach space. We are interested in the case $p = 1$. If $p = 1$, the dual space is not always given by this

Received April 12, 2011; Accepted June 01, 2011.
2010 Mathematics Subject Classification: Primary 41A28, 41A65.
Key words and phrases: gradient, subgradients.
This work was completed with the support by a fund of Duksung Women's University 2010.

equality. However we shall assume that ν is σ-finite, in which case we necessarily have

$$L^1(B,\nu)^* = L^\infty(B,\nu).$$

Let S be a n-dimensional subspace of $L^1(B,\nu)$ and we choose and fix a basis $s_1, s_2, \cdots, s_n$ for S. For each $f_j \in L^1(B,\nu), 1 \leq j \leq \ell$, set $F = \{f_1, f_2, \cdots, f_\ell\}$ and defined by

$$M(a) = \max_{1\leq j\leq \ell} ||f_j - \sum_{i=1}^{n} a_i s_i||_1$$

where $a = (a_1, a_2, \cdots, a_n) \in \mathbb{R}^n$.

Before proving the main Theorem, we need to pursue some technical facts, that M is continuous, convex and $\lim_{||a||\to\infty} M(a) = \infty$, where $||\cdot||$ is any norm on $\mathbb{R}^n$. By A. M. Pinkus[4], we know that $H(a) = ||f - \sum_{i=1}^n a_i s_i||$ is continuous, almost everywhere differentiable, convex and $\lim_{||a||\to\infty} H(a) = \infty$. From general considerations it follows that M is continuous, almost everywhere differentiable, convex and $\lim_{||a||\to\infty} M(a) = \infty$. Thus the unconstrained problem of determining a best approximation to F from S is equivalent to that of finding the minimum of a given convex function M. The study of this problem leads us to the important concepts of gradients and subgradients. The subgradients of M at a are defined as follows:

Definition 1.1. Let M be as above and $a \in \mathbb{R}^n$. A vector $g \in \mathbb{R}^n$ is said to be a subgradient to M at a if

$$M(b) \geq M(a) + (g, b-a)$$

for all $b \in \mathbb{R}^n$ where $(\cdot,\cdot)$ is the usual inner product of vectors in $\mathbb{R}^n$. We let $G(a)$ denote the set of subgradients to M at a.

Each elements of $G(a)$ corresponds to a supporting hyperplane to M at a. Since M is convex, $G(a)$ is non-empty. Furthermore, the set $G(a)$ is bounded, closed and convex for each $a \in \mathbb{R}^n$.

Definition 1.2. Let M be as above and $a \in \mathbb{R}^n$. If $G(a)$ is a singleton then this singleton is called the gradient to M at a.

Thus a gradient to M exists at a if and only if there is a unique supporting hyperplane to M at a.

Let us now deduce the usual simple criterion for determing when a^* is a minimum point of M. Such a minimum point exists.

LEMMA 1.3. *Let M be as above and $a^* \in \mathbb{R}^n$. Then a^* is a minimum point of M if and only if $0 \in G(a^*)$.*

Proof. If a^* is a minumum of M, then $M(b) - M(a^*) \geq 0$ for all $b \in \mathbb{R}^n$. Thus $M(b) - M(a^*) \geq (0, b - a^*)$ for all $b \in \mathbb{R}^n$. So $0 \in G(a^*)$.

Conversely, if $0 \in G(a^*)$, then $M(b) - M(a^*) \geq (0, b - a^*) = 0$ for all $b \in \mathbb{R}^n$. Thus $M(b) \geq M(a^*)$ for all $b \in \mathbb{R}^n$. □

2. Gradients and subgradients

Since $G(a)$ is a compact convex set, it is uniquely determined by its extreme points. These extreme points are related to one-sided directional derivatives as follows. Let W be an arbitrary subset of $\mathbb{R}^n$. A ray W is the union of the origin and the various rays(half-lines of the form $\{\lambda y | \lambda \geq 0\}$).

PROPOSITION 2.1. *Let M be as above and $a \in \mathbb{R}^n$. For each $d \in \mathbb{R}^n$*

$$\lim_{t \to 0^+} \frac{M(a + td) - M(a)}{t} = M'_d(a)$$

exists. Furthermore,

$$M'_d(a) = \max\{(g, d) : g \in G(a)\}.$$

Proof. Set

$$r(t) = \frac{M(a + td) - M(a)}{t}.$$

We verify that the above limit exists by proving that $r(t)$ is non-decreasing and bounded below on $(0, \infty)$.

Let $0 < s < t < \infty$. Since $\frac{s}{t} \in (0, 1)$ and M is convex,

$$M(a + sd) = M(\frac{s}{t}(a + td) + (1 - \frac{s}{t})a) \leq \frac{s}{t} M(a + td) + (1 - \frac{s}{t}) M(a).$$

Thus

$$t[M(a + sd) - M(a)] \leq s[M(a + td) - M(a)],$$

that is, $r(s) \leq r(t)$. Now, take any $g \in G(a)$. By definition,

$$M(a + td) - M(a) \geq (g, td) = t(g, d).$$

Thus $r(t) \geq (g, d)$ for all $t \in (0, \infty)$. Since $r(t)$ is bounded below, the desired limit exists.

The above also confirms that $M'_d(a) \geq (g,d)$ for every $g \in G(a)$. Since $G(a)$ is compact, we therefore have

$$M'_d(a) \geq \max\{(g,d) : g \in G(a)\}.$$

It remain to prove that equality holds.

Let $W_1, W_2 \subset \mathbb{R}^{n+1}$ be defined by

$$W_1 = \{(b,y) : b \in \mathbb{R}^n,\ y \geq M(b)\}$$
$$W_2 = \{(a+td,\ M(a)+tM'_d(a)) : t \geq 0\}.$$

Since M is convex, W_1 is convex and by definition, W_2 is ray. Since $r(t) \geq M'_d(a)$ for all $t > 0$, W_2 contains no point in the interior of W_1. However $(a, M(a)) \in W_1 \bigcap W_2$. Therefore, there exists a $\tilde{g} = (g, g_{n+1}) \in \mathbb{R}^{n+1} \backslash \{0\}$ such that

$$(g, b-a) + g_{n+1}(y - M(a)) \geq 0 \geq (g, td) + g_{n+1} t M'_d(a)$$

for all $b \in \mathbb{R}^n, y \geq M(b)$, and $t \geq 0$. If $b = a$, then $g_{n+1}(y - M(a)) \geq 0$ for all $y \geq M(a)$ which implies that $g_{n+1} \geq 0$. If $g_{n+1} = 0$, then $(g,\ b-a) \geq 0$ for all $b \in \mathbb{R}^n$, and so $g = 0$. It is a contradiction. Thus $g_{n+1} > 0$, set $g^* = -g/g_{n+1}$. Then

$$M(b) - M(a) \geq (g^*,\ b-a)$$

for all $b \in \mathbb{R}^n$, and

$$(g^*,\ d) \geq M'_d(a).$$

From the inequality, $g^* \in G(a)$. The second inequality implies that

$$max\{(g,\ d) : g \in G(a)\} = M'_d(a).$$

□

Based on the above, we call d a descent direction if $M'_d(a) < 0$. We can now discern the germ of an idea behind the construction of algorithms for this problem.

Theorem 2.2. *[4] Let S be a subspace of $L^1(B,\nu)$ and $f \in L^1(B,\nu) \backslash \overline{S}$. Then g^* is a best $L^1(B,\nu)$ approximation to f from S if and only if*

$$|\int_B sgn(f-g^*) g d\nu| \leq \int_{Z(f-g^*)} |g| d\nu$$

for all $g \in S$, where $Z(f-g^*) = \{x | f(x) = g^*(x)\}$.

The general plan is to find a good descent direction, and to employ this information in an efficient manner. Let us assume that we can find $G(a)$ or at least a descent direction. We end this paper by identifying $G(a)$.

THEOREM 2.3. *Let $a \in \mathbb{R}^n$. Then $G(a)$ is the set of all vector $g = (g_1, \cdots, g_n)$, where*

$$g_j = \int_{Z(F-\sum_{i=1}^n a_i s_i)} h s_j d\nu - \int_B sgn(f_{j_a} - \sum_{i=1}^n a_i s_i) s_j d\nu \quad j = 1, \cdots, n$$

and h is any $L^\infty(B, \nu)$ function satisfying $|h| \leq 1$ νa.e. on $Z(F - \sum_{i=1}^n a_i s_i)$ with

$$Z(F - \sum_{i=1}^n a_i s_i) = \bigcap_{j=1}^{\ell} Z(f_j - \sum_{i=1}^n a_i s_i)$$

and j_a is a subindex of f satisfying $M(a) = ||f_{j_a} - \sum_{i=1}^n a_i s_i||_1$.

Proof. Let $g = (g_1, \cdots, g_n)$ be as in the statement. For every $b \in \mathbb{R}^n$, it follows from Theorem 2.0.5. that

$$\begin{aligned}
(g,\, b-a) &= \sum_{j=1}^n g_j (b_j - a_j) \\
&= \int_{Z(F-\sum_{i=1}^n a_i s_i)} h(\sum_{j=1}^n b_j s_j - \sum_{j=1}^n a_j s_j) d\nu \\
&\qquad - \int_B sgn(f_{j_a} - \sum_{i=1}^n a_i s_i)(\sum_{j=1}^n b_j s_j - \sum_{j=1}^n a_j s_j) d\nu \\
&= \int_{Z(F-\sum_{i=1}^n a_i s_i)} h(\sum_{j=1}^n b_j s_j - f_{j_a}) d\nu \\
&\qquad + \int_B sgn(f_{j_a} - \sum_{i=1}^n a_i s_i)(f_{j_a} - \sum_{j=1}^n b_j s_j) d\nu \\
&\qquad - \int_B sgn(f_{j_a} - \sum_{i=1}^n a_i s_i)(f_{j_a} - \sum_{j=1}^n a_j s_j) d\nu \\
&\leq ||f_{j_a} - \sum_{j=1}^n b_j s_j||_1 - ||f_{j_a} - \sum_{j=1}^n a_j s_j||_1 \\
&\leq M(b) - M(a).
\end{aligned}$$

So each g is a subgradient to M at a. Let $\tilde{G}$ denote the set of all such g. Then $\tilde{G} \subseteq G(a)$, and $\tilde{G}$ is both convex and compact. If $\tilde{G} \neq G(a)$, there exists a $g^* \in G(a)$ and a $d \in \mathbb{R}^n$ for which

$$(g,\, d) < (g^*,\, d)$$

for all $g \in \tilde{G}$. Thus

$$max\{(g,\, d) : g \in \tilde{G}\} < max\{(g,\, d) : g \in G(a)\} = M_d'(a).$$

Let $g \in \tilde{G}$ be as in the statement of the proposition with $h = sgn(\sum_{j=1}^n d_j s_j)$ on $Z(F - \sum_{i=1}^n a_i s_i)$. Then

$$(g,\, d) = \int_{Z(F-\sum_{i=1}^n a_i s_i)} |\sum_{j=1}^n d_j s_j| d\nu - \int_B sgn(f_{j_a} - \sum_{j=1}^n a_j s_j) \sum_{j=1}^n d_j s_j d\nu.$$

Let us extend $h = sgn(\sum_{j=1}^n d_j s_j)$ on $Z(f_{j_a} - \sum_{i=1}^n a_i s_i)$, then by the theorem 2.0.5.,

$$\begin{aligned}(g,\, d) &= \int_{Z(f_{j_a}-\sum_{i=1}^n a_i s_i)} |\sum_{j=1}^n d_j s_j| d\nu - \int_B sgn(f_{j_a} - \sum_{j=1}^n a_j s_j) \sum_{j=1}^n d_j s_j d\nu \\ &= \lim_{t\to 0^+} r(t) \\ &= M_d'(a).\end{aligned}$$

This contradicts the above strict inequality and therefore $\tilde{G} = G(a)$. □

Note that d is a descent direction if and only if

$$\int_B sgn(f_{j_a} - \sum_{i=1}^n a_i s_i) \sum_{j=1}^n d_j s_j d\nu > \int_{Z(F-\sum_{i=1}^n a_i s_i)} |\sum_{j=1}^n d_j s_j| \nu.$$

This is yet another explanation of Theorem 2.0.5. Also note that M has a gradient at a if and only if $\nu(Z(F - \sum_{i=1}^n a_i s_i)) = 0$. It is then given by $g = (g_1, \cdots, g_n)$ where

$$g_j = -\int_B sgn(f_{j_a} - \sum_{i=1}^n a_i s_i) s_j d\nu,$$

$j = 1, \cdots, n$.

References

[1] L. Elsner, I. Koltracht and M. Neumann, *Convergence of sequential and asynchronous nonlinear paracontractions*, Numerische mathematik **62** (1992), 305–319.

[2] S. H. Park and H. J. Rhee, *Characterization of best simultaneous approximation for a compact set*, Bull K. Math. Soc. **33** (1996), 435–444.

[3] S. H. Park and H. J. Rhee, *One-sided best simultaneous L_1-approximation*, J. K. Math. Soc. **33** (1996), 155–167.

[4] A. Pinkus, *On L_1-approximation*, Cambridge Tracts in Mathematics **93**, Cambridge University Press, Cambridge-New York, 1989.

[5] H. J. Rhee, *An algorithm of the one-sided best simultaneous approximation*, K. Ann. Math. **24** (2007), 69–79.

[6] I. Singer, *Best Approximation in Normed Linear Spaces by Elements of Linear Subspaces*, Springer-Verlag, New York, 1970.

[7] A. S. Holland, B. N. Sahney and J. Tzimbalario, *On best simultaneous approximation*, Letter to the Editer, J. Approx. Theory **17** (1976), 187 - 188.

[8] H. J. Rhee , *Best simultaneous spproximations and one-sided best simultaneous approximations*, Doctoral Dissertation, Sogang University, Seoul, 1994.

*

College of Natural Sciences
Duksung Women's University
Seoul 132-714, Republic of Korea
E-mail: rhj@duksung.ac.kr