JOURNAL OF APPROXIMATION THEORY 11, 283–284 (1974)

# Letter to the Editor

## A Remark on an Algorithm for Best $L^p$ Approximation

*Communicated by Oved Shisha*

In [1] Karlovitz discusses an algorithm for computing best approximation in the $L^p$ norm when $p$ is an even integer $\geqslant 4$. This algorithm is also applicable under the weaker hypothesis, $p \geqslant 2$. In fact, it is a special case of the following descent algorithm for unconstrained optimization.

Let $G$ be a real-valued function on $R^n$ which satisfies the following conditions:

(a) $\inf_{x \in R^n} G(x) > -\infty$,
(b) $\lim_{x \to \infty} G(x) = \infty$,
(c) $G \in C^2(R^n)$ and the Hessian matrix $H(x) = (\partial^2 G(x)/\partial x_i \, \partial x_j)$ has the property that whenever $y^T H(x)\, y = 0$ for $x, y \in R^n$, then $y = 0$. Equivalently, given any $R > 0$, there is a $\lambda > 0$ such that

$$| y^T H(x)\, y | \geqslant \lambda y^T y, \qquad y \in R^n, \quad x^T x \leqslant R^2. \tag{1}$$

Clearly, any function satisfying (a), (b), and (c) takes on its minimum in $R^n$ at unique point $z$

$$G(z) = \min_{x \in R^n} G(x).$$

Let $x_0 \in R^n$, and let $\beta$ be a nonzero real number. We define two sequences $x_k$ and $y_k$ in $R^n$ as follows. Given $x_k$, define $y_k$ as the solution of the equation

$$\beta \nabla G(x_k) = H(x_k)\, y_k, \qquad k = 0, 1, 2, \ldots . \tag{2}$$

Then set $x_{k+1} = x_k + \lambda_k y_k$, where $\lambda_k$ is given by

$$G(x_k + \lambda_k y_k) = \min_{-\infty < \lambda < \infty} G(x_k + \lambda y_k). \tag{3}$$

THEOREM. *Let $G$: $R^n \to R^1$ satisfy* (a), (b), *and* (c). *Then the sequences defined above converge; $x_k$ converges to $z$ and $y_k$ converges to zero.*

*Proof.* Since

$$G(x_{k+1}) \leqslant G(x_k + \lambda y_k), \qquad \lambda \in R^1, \tag{4}$$

we have $G(x_{k+1}) \leqslant G(x_k) \leqslant \cdots \leqslant G(x_0)$. Thus, condition (b) implies that there exists an $R > 0$ such that $x_k^T x_k < R$. Define $M = \max_{x^T x \leqslant R^2} (\nabla G(x))^T \nabla G(x)$; then from Eq. (2), inequality (1) and the Cauchy–Schwarz inequality, we have

$$\lambda y_k^T y_k \leqslant \beta M^{1/2} (y_k^T y_k)^{1/2},$$
$$y_k^T y_k \leqslant \beta^2 M / \lambda^2.$$

Thus, both $x_k$ and $y_k$ have convergent subsequences. We suppose that they converge to $\bar{x}$ and $\bar{y}$, respectively. In view of (2) and (4), it follows that $\beta \nabla G(\bar{x}) = H(\bar{x})\, \bar{y}$ and $\bar{y}^T \nabla G(\bar{x}) = 0$. Hence, we obtain $\bar{y}^T H(\bar{x})\, \bar{y} = 0$, and so, by property (c), $\bar{y} = 0$. This implies $\nabla G(\bar{x}) = 0$. Thus $\bar{x} = z$.

APPLICATION. Best approximation in $L^p$ for $p \geqslant 2$.

The function to be minimized is

$$G(x_1, \ldots, x_n) = \int_T \left| f(t) - \sum_{i=1}^n x_i \phi_i(t) \right|^p dt$$
$$= \left\| f - \sum_{i=1}^n x_i \phi_i \right\|_{L^p}^p .$$

If we use the notation in [1],

$$\| g \|_{L^2, W}^2 = \int_T g^2(t)\, W(t)\, dt,$$

then, for $p \geqslant 2$,

$$y^T H(x)\, y^T = p(p-1) \left\| \sum_{i=1}^n y_i \phi_i \right\|_{L^2, W}^2$$

where $W = | f - \sum_{i=1}^n x_i \phi_i |^{p-2}$.

Let $\beta = p - 1$, $g_k = \sum_{i=1}^n (x_k)_i\, \phi_i$, and $h_k = g_k + \sum_{i=1}^n (y_k)_i\, \phi_i$. Then (2) is equivalent to the statement that $h_k$ is the best $L^2$ approximation to $f$ with respect to the weight $W_k = | f - g_k |^{p-2}$, while (4) is equivalent to choosing $\lambda_k$ so as to minimize $\| f - g_k - \lambda(h_k - g_k) \|_{L^p}$.

This is the algorithm described in [1].

## REFERENCES

1. L. A. KARLOVITZ, Construction of nearest points in the $L^p$, $p$ even, and $L^\infty$ norms. I, *J. Approximation Theory* 3 (1970), 123–127.

CHARLES MICCHELLI

*IBM T. J. Watson Research Center*
*Yorktown Heights, New York*

*Printed in Belgium*