On the Convergence of an Algorithm for Discrete $L_p$ Approximation

Jerry M. Wolfe

Department of Mathematics, University of Oregon, Eugene, Oregon, USA

Summary. The convergence properties of an algorithm for discrete $L_p$ approximation ($1 \leq p < 2$) that has been considered by several authors are studied. In particular, it is shown that for $1 < p < 2$ the method converges (with a suitably close starting value) to the best approximation at a geometric rate with asymptotic convergence constant $2 - p$. A similar result holds for $p = 1$ if the best approximation is unique. However, in this case the convergence constant depends on the function to be approximated.

Subject Classifications: AMS(MOS) 65D15; CR: 5.13.

The problem of constructing best linear $L_p$ approximations has been considered by many authors. For $p = 2$, the problem may be solved, of course, by the solution of a single system of linear equations, while for $2 < p < \infty$, a damped version of Newton's method will converge to the solution (see [1] or [2]). The case $p = \infty$ can be solved effectively by several methods (see [3] for example) and linear programming techniques have proved to be efficient for the case $p = 1$ [4]. The case $1 < p < 2$ has been troublesome, however. For one thing, methods requiring second derivatives will not even be defined for certain values of the parameters. Of course, the problem is a smooth convex programming problem and is thus susceptible to general convex programming procedures such as the method of steepest descent [5]. While such methods are always convergent for a convex problem, they often exhibit slow convergence. It seems reasonable that taking into account the special structure of the problem one could devise a robust and reasonably fast method.

Recently, Merle and Spath [6] have done a detailed empirical study of an algorithm for discrete $L_p$ approximation for $1 \leq p < 2$. They concluded that the algorithm was quite satisfactory and seemed to converge in all cases. Also in [7] Ekblom discussed the use of a damped Newton's method for these values of $p$. In [9] Watson gave an explanation for the observed behavior of the algorithm of Merle and Spath by comparing it with Newton's method. The purpose of this paper is to analyze in more detail the convergence properties of this algorithm.
The principal result is that the method converges for a suitably close initial guess and that the rate of convergence is typically geometric for $1 < p < 2$ with an asymptotic convergence constant of $2 - p$. Superlinear convergence occurs in certain cases. For $p = 1$ the convergence is also geometric (if the best approximation is unique) but the asymptotic convergence constant depends on the function approximated. It is also proved that the iterates defined by the algorithm are bounded.

Preliminaries

Let $X = \{x_1, \ldots, x_N\}$ and $Y = \{y_1, \ldots, y_N\}$ be given sets of real numbers and let \{\(\Phi_1(x), \ldots, \Phi_n(x)\)\} be a Haar family of dimension $n$ on some interval containing $X$. For each $A = (a_1, \ldots, a_n) \in \mathbb{R}^n$ let $L(A, x) = \sum_{i=1}^{n} a_i \Phi_i(x)$ and let $Z_A = \{j \mid L(A, x_j) = y_j\}$. Then given $p$ with $1 \leq p < 2$ we wish to find an $A^* \in \mathbb{R}^n$ that minimizes the function $\psi(A) = \sum_{j=1}^{N} |L(A, x_j) - y_j|^p$. We shall denote the difference $L(A, x_j) - y_j$ by $E(A, x_j)$, $j = 1, \ldots, N$. Also, to avoid degeneracy, we shall assume that for each $A \in \mathbb{R}^n$, the error $E(A, x_j)$ is not zero for at least $n$ values of $j$.

Remark 1. It will be evident that the results of this paper could still be obtained under weaker assumptions than the Haar property. The Haar assumption has been made to simplify the statements and proofs of several of the results that follow. Also, a thorough study of the use of $L_p$ approximation in statistics may be found in [10].

Algorithm

Algorithm. Let $A_0$ be chosen as the minimizer of the expression $\sum_{j=1}^{N} E^2(A, x_j)$ as $A$ ranges over $\mathbb{R}^n$. Then for $v = 0, 1, \ldots$ proceed as follows:

(a) If $E(A_v, x_j) \geq 0$ for all $j$, choose $A_{v+1}$ to minimize $\sum_{j=1}^{N} w_{v,j} (L(A, x_j) - y_j)^2$ where $w_{v,j} = |E(A_v, x_j)|^{p-2}$ for $j = 1, \ldots, N$.

(b) If $Z_v = \{j \mid E(A_v, x_j) = 0\}$, define $A_{v+1}$ as the minimizer of the expression $\sum_{j \notin Z_v} w_{v,j} (L(A, x_j) - y_j)^2$ defined on the set of $A$'s in $\mathbb{R}^n$ that satisfy $E(A, x_j) = 0$ for all $j \notin Z_v$. Here again $w_{v,j} = |E(A_v, x_j)|^{p-2}$ for $j \notin Z_v$.

We should note here that in its original form, only part (a) of the algorithm was considered with the algorithm being undefined if $Z_v = \emptyset$ for some $v$. To make the method defined for all $A \in \mathbb{R}^n$, Ekblom ([7]) proposed solving a slightly perturbed problem while Merle and Spath defined $w_{v,j}$ to be $1/\delta$ for a fixed $\delta > 0$ whenever $E(A_v, x_j)$ became too small. In practice, the existence of roundoff error requires that some procedure be implemented to preserve stability when the error $E(A_v, x_j)$ becomes small.