The Annals of Statistics
1987, Vol. 15, No. 2, 880–882

ON A CONJECTURE OF HUBER CONCERNING THE CONVERGENCE OF PROJECTION PURSUIT REGRESSION[1]

BY LEE K. JONES

University of Lowell

We generalize the projection pursuit procedure of Friedman and Stuetzle (abstract version) and prove strong convergence. This answers a question of Huber.

1. Introduction and preliminaries. Let (X, Y) be such that X is R^d valued, Y is R valued, and X is distributed according to the probability measure P in R^d. Assume the response surface $f(x) = E(Y|X = x)$ is in $L_2(P)$. Then the projection pursuit regression problem is to approximate f by a sum of ridge functions:

$$f(x) \sim \sum_1^m g_j(a_j^t x), \quad \text{where } a_j \in R^d,\ a_j^t a_j = 1.$$

The method of Friedman and Stuetzle [1] is as follows: Having determined functions $g_1, g_2, \ldots, g_{m-1}$ and unit vectors $a_1, a_2, \ldots, a_{m-1}$, choose a unit vector a_m and a function g_m to minimize

$$E\left[r_m(x) - g_m(a_m^t x)\right]^2, \quad \text{where } r_m = f - \sum_1^{m-1} g_j(a_j^t x).$$

As is shown in [2] the solution at stage m is given by

$$g_m(z) = E(r_m(X)|a_m^t X = z), \tag{1}$$

with a_m a minimizing direction [or as is demonstrated easily a maximizing direction for $E(g_m)^2$]. Huber establishes weak $L_2(P)$ convergence of the procedure [$r_m \to 0$ weakly in $L_2(P)$]. Also in the Comments of [2] Donoho and Johnstone announce a proof of strong convergence for P uniform on the unit ball or multivariate Gaussian. Huber mentions that mild smoothness assumptions are necessary to ensure the existence of a minimizing direction. To avoid this complication and also generalize the procedure we shall allow any direction at stage m to be chosen as long as

$$E\left(g_m(a_m^t X)\right)^2 > \rho \sup_{b^t b = 1} E\left(g_m(b^t X)\right)^2, \qquad \rho \text{ fixed}, 0 < \rho < 1. \tag{2}$$

In Section 2 we prove strong convergence for general P for the above class of procedures.

Received September 1985; revised July 1986.

[1] Research performed with the facilities of the Naval Research Laboratory and supported, in part, by ONR Contract N00014-84-C2245.

AMS 1980 *subject classification*. Primary 62H99

Key words and phrases. Projection pursuit, ridge function, strong convergence.

2. Proof of convergence. Following [2], given $a^tX = z$ the best choice of $g_m(z)$ would be the conditional expectation of r_m given $a^tX = z$. By taking expectations over z using the appropriate marginal of P, we get

$$E(r_m - g_m)^2 = E(r_m)^2 - E(g_m)^2$$

or by induction

$$E(r_m)^2 = E(f)^2 - \sum_1^{m-1} E(g_j)^2. \tag{3}$$

Now $\Sigma_1^\infty \|g_j\|^2 < \infty$, where $\|\cdot\|$ is norm for $L_2(P)$. If r_m converges to $r(x)$ in $L_2(P)$, then $r(x)$ must be 0, for if not there is some a of length one such that $E(g(a^tX))^2 = \delta > 0$ [this follows after first determining a such that $E(r(x)\exp(\omega i a^tX)) \neq 0$] with g the conditional expectation of $r(x)$ given a^tX. It follows by standard measure theory that for arbitrarily large m, $E(g_m)^2 > \rho(\delta/2)$, contradicting $\|g_j\| \to 0$. It suffices to show r_m is a Cauchy sequence in $L_2(P)$. First we give two simple lemmas. (Note that Lemma 2 will also follow from Kronecker's lemma; see [3, page 239].)

LEMMA 1. $|E(g_m r_n)| \le \rho^{-1/2}\|g_m\|\,\|g_n\|$.

PROOF. $|E(g_m r_n)| = |E_z(g_m E(r_n | a_m^t X = z))|$, where E_z denotes the expectation over the marginal defined by a_m. By the Schwarz inequality and the "near" optimality of g_n,

$$|E(g_m r_n)| \le \|g_m\|\left(E_z\left(E\left(r_n | a_m^t X = z\right)\right)^2\right)^{1/2}$$
$$\le \|g_m\|\left(\frac{1}{\rho}\|g_n\|^2\right)^{1/2}. \qquad \square$$

LEMMA 2. *Suppose* $S_1, S_2, \ldots$ *is a nonnegative sequence of reals such that* $\Sigma_1^\infty S_i^2 < \infty$. *Then* $\liminf_{N\to\infty} S_N \Sigma_1^N S_j = 0$.

PROOF. For any $\varepsilon > 0$ choose N such that $\Sigma_N^\infty S_j^2 < \varepsilon/2$. Since $S_i \to 0$, we can, by choosing $\bar{i}$ large enough, ensure that $S_{\bar{i}} \Sigma_1^N S_j < \varepsilon/2$. By letting S_i be the minimum term for $j = N+1, \ldots, \bar{i}$, we have

$$S_i \sum_1^i S_j = S_i \sum_1^N S_j + S_i \sum_{N+1}^{i}{}' S_j \le \varepsilon/2 + \sum_{N+1}^{i} S_j^2 < \varepsilon. \qquad \square$$

Now to finish the proof: If r_m is not Cauchy then $\exists\ \Delta > 0$ such that $\|r_M - r_{M+N}\| > \Delta$ for arbitrarily large M (and an associated N). Since $\|r_m\| \downarrow 1$ by (3), we may assume w.l.o.g. that $\|r_m\| \downarrow 1$ (by multiplying f appropriately). Hence for small γ we get $\|r_M\|^2 < 1 + \gamma$ and $\|r_M - r_{M+N}\| > \Delta$ for some M and N. Now choose K larger than $M + N$ such that $\|g_K\|\Sigma_1^K \|g_i\| < \gamma$. Note either $\|r_K - r_M\| \ge \Delta/2$ or $\|r_K - r_{M+N}\| \ge \Delta/2$. Let us assume the former; otherwise the argument is identical. We have

$$\|r_K - r_M\|^2 = \|r_K - (r_K + g_M + g_{M+1} + \cdots + g_{K-1})\|^2$$
$$\leq \|r_K\|^2 + \|r_M\|^2 - 2\|r_K\|^2 + 2\rho^{-1/2}\|g_K\| \sum_M^{K-1} \|g_l\| < (2 + 2\rho^{-1/2})\gamma,$$

which is impossible if $\gamma < \Delta^2/(8 + 8\rho^{-1/2})$ was chosen. This completes the proof. □

REFERENCES

[1] FRIEDMAN, J. H. and STUETZLE, W. (1981). Projection pursuit regression. *J. Amer. Statist. Assoc.* **76** 817–823.

[2] HUBER, P. J. (1985). Projection pursuit (with discussion). *Ann. Statist.* **13** 435–525.

[3] FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**, 2nd ed. Wiley, New York.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF LOWELL
1 UNIVERSITY AVENUE
LOWELL, MASSACHUSETTS 01854