The Annals of Statistics
1980, Vol. 8, No. 4, 789–801

CHARACTERIZING THE CONSISTENT DIRECTIONS OF LEAST SQUARES ESTIMATES[1]

BY CHIEN-FU WU

University of Wisconsin, Madison

Given a sequence of $p \times 1$ vectors $\mathbf{v} = \{v_i\}_{i=1}^{\infty}$ such that $M_n = \sum_{i=1}^{n} v_i v_i'$ is positive definite for some n, the linear space $\{u : u' M_n^{-1} u \to 0 \text{ as } n \to \infty\}$ is characterized in terms of the limiting properties of $\mathbf{v}$. This characterization result is applied to give a necessary and sufficient condition for the asymptotic consistency of any best linear unbiased estimator in terms of the limiting properties of the design sequence. For the polynomial regression model, it can be further related to the geometry of the polynomial system.

1. Introduction. A linear model is given by

$$y = \sum_{i=0}^{p-1} \theta_i f_i(x) + \varepsilon = \theta' f(x) + \varepsilon \tag{1.1}$$

where θ and $f(x)$ are $p \times 1$ vectors, x is from a set $\mathcal{X}$ on which linear model (1.1) is valid, the random error ε has mean 0, variance σ^2 and errors corresponding to different observations are uncorrelated. For one dimensional polynomial regression, $f(x)' = (1, x, x^2, \cdots, x^{p-1})$; for multiple linear regression, $f(x) = x$, x is a $p \times 1$ vector. If $y_1, y_2, \cdots, y_n$ are observed at $x_1, x_2, \cdots, x_n$ and $M_n = \sum_{i=1}^{n} f(x_i) f(x_i)'$ is a nonsingular $p \times p$ matrix, the least squares estimate of θ is given by

$$\hat{\theta} = M_n^{-1} X_n' y \tag{1.2}$$

where $X_n' = [f(x_1), \cdots, f(x_n)]$ and $y' = (y_1, \cdots, y_n)$. An important optimality property of $\hat{\theta}$ is provided by the Gauss-Markov Theorem: $b'\hat{\theta} = \sum_{i=0}^{p-1} b_i \hat{\theta}_i$ for any $b' = (b_0, \cdots, b_{p-1})$ has the smallest variance among all the linear unbiased estimators of $b'\theta$. As for the asymptotic properties of $\hat{\theta}$, it is surprising that some of the basic results have been proved only recently. Lai, Robbins and Wei (1978) proved that $\hat{\theta} \to \theta$ a.s. when $M_n^{-1} \to 0$ and $\{\varepsilon_i\}_{i=1}^{\infty}$ are i.i.d. with mean zero and variance σ^2. The proof of asymptotic normality of $b'\hat{\theta}$ is relatively easier. Huber (1973) gave one such result which allows the dimension p of θ to go to infinity as n goes to infinity. But if M_n^{-1} does not necessarily go to zero, what typically happens is that $E(b'\hat{\theta} - b'\theta)^2 = b' M_n^{-1} b \to 0$ for certain "directions" b only. Such a direction b is called a *consistent direction for the least squares estimate* $\hat{\theta}$. The main purpose of this paper is to characterize the space of consistent directions for $\hat{\theta}$ and to investigate the implications of this characterization. The consistency question of the nonlinear least squares estimator is treated in Jennrich (1969) and Wu (1979).

Received December 1977; revised April 1979.

[1]Research supported by U. S. Army under contract no. DAAG29-75-C0024.

AMS 1970 *subject classifications.* Primary 62J05: secondary 62F20.

Key words and phrases. Asymptotic consistency, best linear unbiased estimator, consistent direction, consistency region, polynomial regression.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics.* www.jstor.org

If the random errors $\{\varepsilon_i\}_{i=1}^{\infty}$ are assumed to be uncorrelated with mean zero and variance σ^2, $\mathrm{Var}(b'\hat{\theta}) = E(b'\hat{\theta} - b'\theta)^2 = b'M_n^{-1}b \to 0$ implies $b'\hat{\theta} \to b'\theta$ in probability; on the other hand, if $b'M_n^{-1}b$ does not converge to zero, it must converge to a positive number c (since $b'M_n^{-1}b$ is nonincreasing) which implies that $b'\hat{\theta} \to b'\theta$ in probability is not true. Therefore, under model (1.1), $b'M_n^{-1}b \to 0$ and the weak consistency of $b'\hat{\theta}$ are equivalent. If we further assume that $\{\varepsilon_i\}_1^{\infty}$ are i.i.d. with mean 0 and variance σ^2, from the result of Lai et al. (1978), $b'M_n^{-1}b \to 0$ implies $b'\hat{\theta} \to b'\theta$ a.s. Therefore, $b'M_n^{-1}b \to 0$ and the strong consistency of $b'\hat{\theta}$ are equivalent under i.i.d. and finite second moment assumptions on $\{\varepsilon_i\}_1^{\infty}$. For the rest of the paper, $b'M_n^{-1}b \to 0$ will be interpreted as the strong or weak consistency of $b'\hat{\theta}$, depending on which of the above assumptions is imposed on $\{\varepsilon_i\}$.

In Section 2, the above problem is reformulated as the following mathematical problem:

Given a sequence of $p \times 1$ vectors $\{v_i\}_{i=1}^{\infty}$ such that $\Sigma_{i=1}^{n} v_i v_i'$ is nonsingular for some n (therefore, for all $N \geqslant n$), characterize the linear space

$$\left\{ v : v'(\Sigma_{i=1}^{n} v_i v_i')^{-1} v \to 0 \qquad \text{as } n \to \infty \right\}. \tag{1.3}$$

The proposed characterization is in terms of the limiting properties of $\{v_i\}_{i=1}^{\infty}$ and has a very natural statistical interpretation. The characterization result of Section 2 is applied to give necessary and sufficient conditions for the asymptotic consistency of any best linear unbiased estimator in Section 3. Another application of the characterization result is in D_s-optimal design theory (Wynn 1976). Using the n.a.s. conditions, we find a "paradoxical" phenomenon which occurs with some linear models. The paradox says that if you take infinite observations in the vicinity of a certain point, you can only predict without error (in the sense of asymptotic consistency) the outcome of the experiment at this point and some (ranging from one to countably many) other points. In Section 4, we investigate the consistency problem for the polynomial regression case in full detail. Special properties of the polynomial system are utilized to give simple and verifiable conditions.

Drygas (1976) has given a characterization of the space (1.3) from a different viewpoint. Let $X_n(b)^{\perp} = \{z : z = X_n b_1 \text{ for a } p \times 1 \text{ vector } b_1 \text{ with } b'b_1 = 0\}$ and P_F = the projection matrix onto the subspace F. Then he proved that b is a consistent direction iff $\|(I - P_{X_n(b)^{\perp}})X_n b\| \to \infty$ as $n \to \infty$. Since this involves the p column vectors of X_n rather than the sequence $\{x_i\}$ or $\{f(x_i)\}$, it looks unlikely that it can achieve the same purpose of the present paper.

2. Main results. Given a sequence of $p \times 1$ vectors $\mathbf{v} = \{v_i\}_{i=1}^{\infty}$ with the assumption that $M_n = \Sigma_{i=1}^{n} v_i v_i'$ is nonsingular for some n, we want to characterize the linear space

$$S(\mathbf{v}) = \left\{ u : u'M_n^{-1}u \to 0 \qquad \text{as } n \to \infty \right\}. \tag{2.1}$$

Note that $u'M_n^{-1}u$ is nonincreasing in n.

THEOREM 1. *For* $M_n = \Sigma_{i=1}^n v_i v_i'$, $u'M_n^{-1}u \to 0$ *if and only if* $\Sigma_{i=1}^\infty (w'v_i)^2 = \infty$ *for all* $w'u \neq 0$.

PROOF. For any positive definite $p \times p$ matrix M and $p \times 1$ vector u,

$$\max\left\{ \frac{(w'u)^2}{w'Mw} : w'u \neq 0 \right\} = u'M^{-1}u$$

follows from Cauchy-Schwarz inequality. Therefore,

$$u'M_n^{-1}u \to 0$$

if and only if

$$\max\left\{ \frac{(w'u)^2}{w'M_n w} : w'u \neq 0 \right\} \to 0,$$

if and only if

$$w'M_n w = \Sigma_{i=1}^\infty (w'v_i)^2 = \infty \qquad \text{for all} \quad w'u \neq 0.$$

The last equivalence follows by applying Dini's Theorem (see Dieudonne,1960) to the nonincreasing functions $g_n(w) = (w'u)^2/w'M_n w$ on the compact set $|w| = 1$. □

When the infinite sequence $\mathbf{v} = \{v_i\}_{i=1}^\infty$ gives "enough" projections onto w, i.e., $\Sigma_{i=1}^\infty (w'v_i)^2 = \infty$, we call w a *good direction* for $\mathbf{v}$; otherwise, we call it a *bad direction* for $\mathbf{v}$. Theorem 1 can then be restated as; $u'M_n^{-1}u \to 0$ *if and only if all the bad directions for* $\mathbf{v}$ *are orthogonal to* u. A statistical interpretation of good or bad direction will be given in the next section.

COROLLARY 1. *If* $\Sigma_{i=1}^\infty (w'v_i)^2 = \infty$ *for all nonzero* w, *then* $S(\mathbf{v}) = R^p$.

COROLLARY 2. *For any nonzero* w, *let* $\theta_i^{(w)}$ *be the angle between* w *and* v_i. *If*

$$\liminf_{i\to\infty} |\cos \theta_i^{(w)}| > 0 \qquad \textit{for all nonzero } w,$$

and $\mathbf{v}$ *is an unbounded sequence, then* $S(\mathbf{v}) = R^p$.

PROOF. Since $\limsup_{i\to\infty} |v_i| = \infty$ and $\liminf_{i\to\infty} |\cos \theta_i^{(w)}|^2 > 0$, $\Sigma_{i=1}^\infty (w'v_i)^2 = \Sigma_{i=1}^\infty |w|^2 |v_i|^2 |\cos \theta_i^{(w)}|^2 = \infty$. The result follows from Corollary 1.

Corollary 2 has a natural geometric interpretation. When the unbounded sequence $\mathbf{v}$ eventually stays at least a positive angle away from any direction, $u'M_n^{-1}u \to 0$ for any u.

If there are infinitely many v_i's which are all equal to a fixed vector u_0, the matrix M_n becomes larger and larger along direction u_0 and we naturally expect $u_0'M_n^{-1}u_0 \to 0$. This idea can be further exploited to give a refinement of Theorem 1. First we have to introduce a new concept of convergence, the Q-limit points of a sequence.

DEFINITION 1. Let $\mathbf{v} = \{v_i\}_{i=1}^\infty$ be an infinite sequence in R^p; $y \varepsilon R^p$ is called a Q-limit point of $\mathbf{v}$ if and only if there exists an infinite subsequence $\{n_i\}_{i=1}^\infty$ and scalars $|\lambda_{n_i}| \leqslant 1$ such that $\lambda_{n_i} v_{n_i} \to y$ as $i \to \infty$.

When $\lambda_{n_i} = 1$, this reduces to the usual concept of limit point (or called cluster point, point of accumulation). The relationship between Q-limit point and limit point is given in Proposition 1.

PROPOSITION 1. *For a bounded sequence* $\mathbf{v}, y$ *is a* Q-*limit point of* $\mathbf{v}$ *if and only if there exists a limit point* z *of* $\mathbf{v}$ *and a scalar* $|\lambda| \leq 1$ *such that* $y = \lambda z$. *In particicluar, the linear space spanned by the* Q-*limit points of* $\mathbf{v}$ *is the same as the linear space spanned by the limit points of* $\mathbf{v}$.

PROOF. If z is a limit point of $\mathbf{v}$ and $|\lambda| \leq 1$, λz is obviously a Q-limit point of $\mathbf{v}$. Conversely, if there exists an infinite subsequence $\{n_i\}_{i=1}^{\infty}$ and scalars $|\lambda_{n_i}| \leq 1$ with $\lambda_{n_i} v_{n_i} \to y$, from the compactness of $[-1, 1]$, we can find an infinite subsequence $\{m_i\}_{i=1}^{\infty}$ of $\{n_i\}$ with λ_{m_i} converging to a λ in $[-1, 1]$. Therefore, we have $\lambda_{m_i} v_{m_i} \to y$ and $\lambda_{m_i} \to \lambda$. If $\lambda = 0$, then $y = 0$ follows from the boundedness of $\mathbf{v}$. Taking any limit point of $\mathbf{v}$ (which exists) as $z, y = \lambda z$ holds. If $\lambda \neq 0$, $v_{m_i} \to \lambda^{-1} y$, a limit point of $\mathbf{v}$. $y = \lambda(\lambda^{-1} y)$ also holds. □

If $\mathbf{v}$ is not bounded, the equivalence relation may not hold as the following counterexample shows. Let $v_i = i(\cos i\theta, \sin i\theta)$, $i = 1, 2, \cdots \infty$, θ is an irrational number. $\mathbf{v}$ has no limit point. But the limit points of $\{i^{-1} v_i = (\cos i\theta, \sin i\theta) : i = 1, 2, \cdots, \infty\}$ covers the whole unit circle. The set of the Q-limits points of $\mathbf{v}$ is thus the whole R^2-plane.

DEFINITION 2. For any linear subspace V of R^p, its orthogonal complement $V^{\perp}$ is defined as

$$\{w \in R^p : w'u = 0 \quad \text{for all } v \in V\}.$$

THEOREM 2. *For the infinite sequence* $\mathbf{v} = \{v_i\}_{i=1}^{\infty}$, *the linear space* $S(\mathbf{v})$ *in* (2.1) *is equal to*

$$A(\mathbf{v}) \oplus B(\mathbf{v}) \tag{2.2}$$

where $\oplus$ *means the direct sum of two vector spaces,* $A(\mathbf{v})$ *is the linear subspace spanned by the* Q-*limit points of* $\mathbf{v}$, *and* $B(\mathbf{v})$ *is equal to*

$$\left\{u \in A(\mathbf{v})^{\perp} : \Sigma_{i=1}^{\infty}(w'v_i)^2 = \infty \quad \textit{for all} \quad w \in A(\mathbf{v})^{\perp} \quad \textit{and} \quad w'u \neq 0\right\}. \tag{2.3}$$

PROOF. It is easy to show that $B(\mathbf{v})$ is indeed a vector subspace.

(i) $A(\mathbf{v}) \subseteq S(\mathbf{v})$: For any $u \in A(\mathbf{v})$, from Theorem 1, we want to show that $\Sigma_{i=1}^{\infty}(w'v_i)^2 = \infty$ for all $w'u \neq 0$. Since $w'u \neq 0$ and $u \in A(\mathbf{v})$ imply $w \notin A(\mathbf{v})^{\perp}$, it is sufficient to show that $\Sigma_{i=1}^{\infty}(w'v_i)^2 = \infty$ for all $w \notin A(\mathbf{v})^{\perp}$. If $w \notin A(\mathbf{v})^{\perp}$, there exists a Q-limit point y of $\mathbf{v}$ with $w - y \neq 0$. From the definition of Q-limits point, there exists an infinite subsequence $\{n_i\}_{i=1}^{\infty}$ and scalars $|\lambda_{n_i}| \leq 1$ such that $|\lambda_{n_i} v_{n_i} - y| \to 0$. Therefore, $|\lambda_{n_i} w' v_{n_i}| \geq \frac{1}{2}|w'y| > 0$ for i sufficiently large and $\Sigma_{i=1}^{\infty}(w'v_i)^2 \geq \Sigma_{i=1}^{\infty}(\lambda_{n_i} w' v_{n_i})^2 = \infty$ is established.

(ii) $B(\mathbf{v}) = S(\mathbf{v}) \cap A(\mathbf{v})^{\perp}$ which implies $S(\mathbf{v}) = A(\mathbf{v}) \oplus B(\mathbf{v})$:

$$u \in S(\mathbf{v}) \cap A(\mathbf{v})^{\perp}$$

$$\Leftrightarrow u \in A(\mathbf{v})^{\perp} \text{ and } \Sigma_{i=1}^{\infty}(w'v_i)^2 = \infty \qquad \text{for all } w'u \neq 0$$

$$\Leftrightarrow u \in A(\mathbf{v})^{\perp} \text{ and } \Sigma_{i=1}^{\infty}(w'v_i)^2 = \infty \qquad \text{for all } w \in A(\mathbf{v})^{\perp} \text{ and } w'u \neq 0$$

$$\Leftrightarrow u \in B(\mathbf{v}).$$

The first equivalence follows from Theorem 1 and the second equivalence follows from the fact, proved in (i), that $\Sigma_{i=1}^{\infty}(w'v_i)^2 = \infty$ for all $w \notin A(v)^{\perp}$. □

For a bounded sequence **v**, Wynn (1976) proved that all the limit points of **v** are in $S(\mathbf{v})$. This kind of result has been utilized in providing sufficient conditions for design sequences converging to a D_s-optimal design (Wynn, 1976).

COROLLARY 3. *If the infinite sequence* **v** *has at least p linearly independent* Q-*limit points, then* $S(\mathbf{v}) = R^p$.

PROOF. The assumption implies that $A(\mathbf{v})$ in Theorem 2 is equal to R^p.

A necessary and sufficient condition for $S(\mathbf{v}) = R^p$ is that $(\Sigma_{i=1}^{n} v_i v'_i)^{-1}$ converges to the zero matrix. Typically, when the Q-limits points of **v** can be obtained easily, it is considerably simpler to verify the sufficient condition in Corollary 3 than to invert the matrix and check whether it converges to zero.

3. Necessary and sufficient conditions for the asymptotic consistency of best linear unbiased estimators.

Let $\hat{\theta}$ be the least squares estimator (1.2), based on observations $y_1, y_2, \cdots, y_n$ taken at $x_1, x_2, \cdots, x_n$. $b'\hat{\theta}$ is the best linear unbiased estimator of $b'\theta$. In this section, we want to investigate the asymptotic behavior of $b'\hat{\theta}$ for each b.

To study the asymptotic consistency of any best linear unbiased estimator $b'\hat{\theta}$ for model (1.1), define the *space of consistent directions*

$$(3.1) \qquad S(\mathbf{f}(x)) = \{b : b'(\Sigma_{i=1}^{n} f(x_i)f(x_i)')^{-1}b \to 0 \quad \text{as } n \to \infty\},$$

where $\mathbf{f}(x) = \{f(x_i)\}_{i=1}^{\infty}$. If the data are collected with a careful design (choice of $\{x_i\}_{i=1}^{\infty}$), consistency will hold for any b. But the characterization problem (3.1) has still not been answered. For prediction or control purpose, we are interested in the *consistency region* defined as

$$(3.2) \quad C(\mathbf{f}(x)) = \{x : x \in \mathcal{X}, f(x)'(\Sigma_{i=1}^{n} f(x_i)f(x_i)')^{-1}f(x) \to 0 \quad \text{as } n \to \infty\}$$
$$= \{x : x \in \mathcal{X}, f(x) \in S(\mathbf{f}(x))\}.$$

By identifying $f(x_i)$ with v_i in Theorem 2, we immediately obtain a characterization of $S(\mathbf{f}(x))$.

THEOREM 3. *The space of consistent directions* $S(\mathbf{f}(x))$ *in* (3.1) *is equal to* $A(\mathbf{f}(x)) \oplus B(\mathbf{f}(x))$ *where* $A(\mathbf{f}(x))$ *and* $B(\mathbf{f}(x))$ *are defined in Theorem* 2 *and* $\mathbf{f}(x) =$

$\{f(x_i)\}_{i=1}^{\infty}$. *The consistency region* $C(\mathbf{f}(X))$ *in* (3.2) *is equal to*

$$\{x : x \in \mathfrak{X}, f(x) \in A(\mathbf{f}(x)) \oplus B(\mathbf{f}(x))\}.$$

Statistical interpretations of good or bad directions (defined in Section 2) and limit points of $\{f(x_i)\}$ are now in order. Each observation y_i taken at $f(x_i)$ not only gives information about $f(x_i)'\theta$ but also gives partial information about $w'\theta$, $w'f(x_i) \neq 0$. A good direction w, i.e., $\Sigma_{i=1}^{\infty}(w'f(x_i))^2 = \infty$, is a direction on which one may gain much information from the data $\{y_i\}$ collected at $\{f(x_i)\}$. b is a consistent direction for $\hat{\theta}$ (i.e., $b'\hat{\theta}$ is asymptotically consistent for estimating $b'\theta$) if and only if all the bad directions are orthogonal to b. In any neighborhood of a limit point $f(x^*)$ of $\{f(x_i)\}$, infinitely many observations are taken, or in Huber's term (page 289, 1975), $f(x^*)$ has infinite approximate replications. Consistency is certainly expected.

COROLLARY 4. *For a bounded sequence* $\{f(x_i)\}_{i=1}^{\infty}$,

$$f(x_n)'M_n^{-1}f(x_n) \to 0 \quad \text{as} \quad N \to \infty, \tag{3.3}$$

where $M_n = \Sigma_{i=1}^{n} f(x_i)f(x_i)'$.

PROOF. For any infinite subsequence S_1, we want to show that there exists an infinite subsequence S_2 of S_1 such that

$$\lim_{n\to\infty,\, n\in S_2} f(x_n)'M_n^{-1}f(x_n) \to 0. \tag{3.4}$$

Since $\{f(x_i)\}$ is bounded, there exists an infinite subsequence of S_2 of S_1 and a $p \times 1$ vector z_0 such that

$$\lim_{n\to\infty,\, n\in S_2} f(x_n) = z_0. \tag{3.5}$$

Now that z_0 is a limit point of $\{f(x_i)\}$, $z'_0M_n^{-1}z_0 \to 0$ follows from Theorem 3. From the monotonicity of M_n, M_n^{-1} is bounded from above. (3.4) now follows from (3.5) and $z'_0M_n^{-1}z_0 \to 0$. □

This result was obtained by Pázman (1973) and Wynn (1976). Their main use is in establishing sufficient conditions for the generations of D_s-optimal designs.

As a final remark, if $f(x)$ is continuous in x, all the limit points of $\{x_i\}_{i=1}^{\infty}$ are in the consistency region $C(\mathbf{f}(x))$. For the polynomial regression model discussed in the next section, we can show that $C(\mathbf{f}(x))$ is either $\mathfrak{X}(= R^1)$ or the set of limit points of $\{x_i\}_{i=1}^{\infty}$. This is certainly not true in general as the following example shows. Consider the linear model

$$y = \theta_0 e^x + \theta_1 x + \theta_2 x^2 + \varepsilon = f(x)'\theta + \varepsilon \tag{3.6}$$

where $f(x)' = (e^x, x, x^2)$, $\{y_i\}$ are observed at $\{x_i\}$ with

$$x_i \to 2,\ \Sigma_{i=1}^{\infty}(x_i - 2)^2 = \infty \quad \text{and} \quad \Sigma_{i=1}^{\infty}(x_i - 2)^4 < \infty. \tag{3.7}$$

We will show later that for this choice of model and design sequence the space of consistent directions is

$$S(\mathbf{f}(x)) = \{af(2) + bf^{(1)}(2) : -\infty < a, b < \infty\}, \tag{3.8}$$

where $f^{(1)}(2)$ is the derivative of $f(x)$ at 2. Therefore, x is in $C(\mathbf{f}(x))$ if and only if there exist a and b such that

$$\begin{aligned} e^x &= ae^2 + be^2 \\ x &= 2a + b \\ x^2 &= 4a + 4b. \end{aligned} \tag{3.9}$$

Equations (3.9) are solvable if and only if $4e^{x-2} = x^2$, i.e. if and only if $x = 2$ or -0.5569. We may summarize our findings as follows:

For model (3.6) with x_i chosen according to (3.7), $f(2)'\hat{\theta}$ and $f(-0.5569)'\hat{\theta}$ are the only two consistent estimators among the best linear unbiased estimators $\{f(x)'\hat{\theta} : -\infty < x < \infty\}$.

A layman's version of this result is the following:

> For a highly twisted model like $f(x) = (x, x^2, e^x)'$ if you keep on taking observations in the vicinity of Madison so that the sum of the squared distances between location i and Madison is infinite but the sum of the fourth powers is finite, then you gain "enough information" (in terms of consistency) only about two locations, Madison (denoted by $x = 2$) and Berkeley (denoted by $x = -0.5569$), but nowhere else. Even a suburban town of Madison, say, $x = 1.9$ or 2.1, won't do it.

This example suggests that the consistency behavior depends on the geometric structure of the $f(x)$ in the linear model (1.1), while most of the work on linear estimation deals with the X or $X'X$ matrix. Our interpretation to this puzzling phenomenon is purely geometric, namely, the relevant linear space intersects the curve $\{(e^x, x, x^2) : -\infty < x < \infty\}$ at two points. But a statistical interpretation is still lacking.

But if the x_i are chosen according to

$$x_i \to 2, \ \Sigma_{i=1}^{\infty}(x_i - 2)^2 = \Sigma_{i=1}^{\infty}(x_i - 2)^4 = \infty,$$

we can show that $S(\mathbf{f}(x))$ is spanned by $f(2)$, $f^{(1)}(2)$ and $f^{(2)}(2)$ and that $f(x)'\hat{\theta}$ is consistent for any x.

The above counter-intuitive phenomenon about the consistency region can be made even more pronounced. A general way of constructing such examples is to choose a sequence $\{x_i\}_{i=1}^{\infty}$ converging to a point such that the corresponding $S(\mathbf{f})$ is of lower dimension (than p) and that $S(\mathbf{f}) \cap \{f(x) : -\infty < x < \infty\}$ has more than one point. If the function $f(x)$ is highly curved, this intersection may even have countably many points.

PROOF OF (3.8). Since $f(2)$ is the only limit point, $A(\mathbf{f}(x))$ in Theorem 3 is equal to $\{uf(2): -\infty < u < \infty\}$. $A(\mathbf{f}(x))\perp$ is equal to

$$\{(-u2e^{-2} - v4e^{-2}, u, v): -\infty < u, v < \infty\}. \tag{3.10}$$

All the vectors in $A(\mathbf{f}(x))^{\perp}$ can be represented as the scalar multiples of the set of vectors

$$\{v_\lambda = (-2e^{-2} - 4\lambda e^{-2}, 1, \lambda),\ -\infty < \lambda < \infty,\ v_\infty = (-4e^{-2}, 0, 1)\}.$$

Let $f^{(1)}(2)$ be the derivative of $f(x)$ at $x = 2$. Denote the cross-product of $f(2)$ and $f^{(1)}(2)$ by $f(2) \times f^{(1)}(2) = (4, 0, -e^2) = -e^2 v_\infty$. $f(2) \times f^{(1)}(2)$ is in $A(\mathbf{f}(x))^\perp$ and is orthogonal to the projection of $f^{(1)}(2)$ onto $A(\mathbf{f}(x))^\perp$. Let $\delta_i = x_i - 2$. The following can be verified easily.

$$\begin{aligned} f(x_i)'v_\lambda &= e^{x_i-2}(-2-4\lambda) + x_i + \lambda x_i^2 = -\delta_i - (1+\lambda)\delta_i^2 + 0(\delta_i^3) \\ &= \delta_i(-1 + 0(\delta_i)), \\ f(x_i)'v_\infty &= 4e^{x_i-2} + x_i^2 = -\delta_i^2 + 0(\delta_i^3) = \delta_i^2(-1 + 0(\delta_i)), \end{aligned}$$

where $0(\delta_i$ is the big 0-notation. (3.7) implies that

$$\Sigma_{i=1}^\infty (f(x_i)'v_\lambda)^2 = \infty \qquad \text{for } -\infty < \lambda < \infty$$

and

$$\Sigma_{i=1}^\infty (f(x_i)'v_\infty)^2 < \infty.$$

Therefore, $B(\mathbf{f}(x))$ in Theorem 3 is equal to

$$\{uv_{\lambda^*},\ -\infty < u < \infty\},$$

where $\mathbf{v}_{\lambda^*}$ is orthogonal to v_∞. Since v_∞ is proportional to $f(2) \times f^{(1)}(2)$, we can easily see that v_{λ^*} is the projection direction of $f^{(1)}(2)$ onto $A(\mathbf{f}(x))^\perp$. $S(\mathbf{f}(x))$ is thus equal to

$$\{af(2) + bv_{\lambda^*} : -\infty < a, b < \infty\} = \{af(2) + bf^{(1)}(2) : -\infty < a, b < \infty\}. \ \square$$

4. The polynomial regression case. In this section the general results developed before will be applied to the polynomial regression model:

$$y_j = \Sigma_{i=0}^{p-1}\theta_i x_j^i + \varepsilon_j = f(x_j)'\theta + \varepsilon_j, \tag{4.1}$$

where $f(x)' = (1, x, \cdots, x^{p-1})$. Necessary and sufficient conditions for asymptotic consistency will be given in terms of the limiting properties of $\{f(x_j)\}_{j=1}^\infty$. Some useful properties of the polynomial system are stated below.

PROPOSITION 2. *Let $f^{(i)}(x)$ be the* ith *derivative of $f(x) = (1, x, \cdots, x^{p-1})'$. The p vectors*

$$\{f^{(i)}(x_j) : 0 \leqslant i \leqslant m_j,\ 1 \leqslant j \leqslant k,\ \Sigma_{j=1}^k (m_j + 1) = p\}, \tag{4.2}$$

where $\{x_j\}_{j=1}^k$ are k distinct points, are linearly independent.

PROOF. This follows from the fact that the polynomial system $\{x^i\}_{i=0}^{p-1}$ is an extended Tchebycheff system (for definition and proof, see Karlin and Studden, 1966, page 8). The result can also be derived by successive differentiations of the Vandermonde determinant.

COROLLARY 5. *The intersection of the linear space spanned by*

$$\{f^{(i)}(x_j) : 0 \leqslant i \leqslant m_j,\ 1 \leqslant j \leqslant k,\ \Sigma_{j=1}^k (m_j + 1) = p - 1\} \tag{4.3}$$

and the curve $\{f(x) : -\infty < x < \infty\}$ is still $\{f(x_j) : 1 \leqslant j \leqslant k\}$.

PROOF. Let the polynomial $r(x)$ be the determinant of the matrix with column vectors $\{f(x), f^{(i)}(x_j), 0 \leqslant i \leqslant m_j, 1 \leqslant j \leqslant k, \Sigma_{j=1}^{k}(m_j + 1) = p - 1\}$. $r(x)$ is a nonzero polynomial of degree at most $p - 1$, since the $p - 1$ vectors in (4.3) are linearly independent by Proposition 2. It is easily seen that $r(x)$ has root x_j with multiplicity $m_j + 1$(i.e., $r(x_j) = r^{(1)}(x_j) = \cdots = r^{(m_j)}(x_j) = 0$. $\Sigma_{j=1}^{k}(m_j + 1) = p - 1$ implies that $\{x_1, \cdots, x_k\}$ are the only roots of $r(x)$. Since $r(x) = 0$ is equivalent to $f(x)$ belonging to the linear space spanned by (4.3), the desired result follows. □

The above two results play a key role in determining the consistency region for polynomial regression model.

For an unbounded sequence $\{x_i\}_{i=1}^{\infty}$, the characterization problem is easy to solve.

PROPOSITION 3. *If* $\limsup_{i\to\infty}|x_i| = \infty$, *then* $S(\mathbf{f}(x))$ *and* $C(\mathbf{f}(x))$ *in Theorem* 3 *are, respectively*, R^p *and* R^1.

PROOF. For any nonzero vector $w' = (w_0, \cdots, w_{p-1})$,

$$\limsup_{i\to\infty}(w'f(x_i))^2 = \limsup_{i\to\infty}\left(\Sigma_{j=0}^{p-1}w_j x_i^j\right)^2 = \begin{cases}\infty & \text{if } w_j \neq 0 \text{ for some } j > 0,\\ = w_0^2 & \text{otherwise.}\end{cases}$$

Both imply that $\Sigma_{i=1}^{\infty}(w'f(x_i))^2 = \infty$. The desired result follows from Corollary 1.

For the rest of the section, we *assume that the sequence* $\{x_i\}_{i=1}^{\infty}$ *is bounded*. For a bounded $\{x_i\}_{i=1}^{\infty}$, from Proposition 1, the linear space spanned by the Q-limit points of $\{f(x_i)\}_{i=1}^{\infty}$ is the same as the linear space spanned by the limit points of $\{f(x_i)\}_{i=1}^{\infty}$. It is thus sufficient to work on the limit points of $\{f(x_i)\}_{i=1}^{\infty}$, or equivalently $\{x_i\}_{i=1}^{\infty}$. If $\{x_i\}_{i=1}^{\infty}$ has at least p distinct limit points $\{a_i\}_{i=1}^{p}$, then the corresponding $\{f(a_i)\}_{i=1}^{p}$ are linearly independent from Proposition 2. The linear space $A(\mathbf{f}(x))$ in Theorem 3 is equal to R^p. We have the following result.

PROPOSITION 4. *If the* $\{x_i\}_{i=1}^{\infty}$ *has at least p distinct limit points, then* $S(\mathbf{f}(x))$ *and* $C(\mathbf{f}(x))$ *in Theorem* 3 *are, respectively*, R^p *and* R^1.

DEFINITION 3. Let T be any set of $p \times 1$ vectors. The linear space generated by T is denoted by $L(T)$.

Suppose $\mathbf{x} = \{x_i\}_{i=1}^{\infty}$ has k limit points $\{a_i\}_{i=1}^{k}$, $k \leqslant p - 1$. For any $\mathbf{m} = (m_1, \cdots, m_k)$, define $g(\mathbf{m}) = \Sigma_{i=1}^{\infty}\Pi_{j=1}^{k}(x_i - a_j)^{2(1+m_j)}$. Define $\mathbf{m} \geqslant \mathbf{m}'$ if and only if $m_i \geqslant m_i'$ for $1 \leqslant i \leqslant k$. Define

$$\mathcal{G} = \{\mathbf{m}: m_j \geqslant 0, \Sigma_{j=1}^{k}(m_j + 1) \leqslant p - 1, g(\mathbf{m}) < \infty \text{ and } g(\mathbf{m}') = \infty \text{ for all } \mathbf{m}' \neq \mathbf{m} \text{ and } \mathbf{m} \geqslant \mathbf{m}'\}, \tag{4.4}$$

and, for every $\mathbf{m} \in \mathcal{G}$,

$$E(\mathbf{m}) = \{f^{(i)}(a_j): 0 \leqslant i \leqslant m_j, 1 \leqslant j \leqslant k\}. \tag{4.5}$$

THEOREM 4. *For the polynomial regression model* (4.1), *if* $\{x_i\}_{i=1}^{\infty}$ *is bounded and has k limit points* $\{a_i\}_{i=1}^{k}(k \leqslant p - 1)$, *then* $S(\mathbf{f}(x))$ *is equal to*

$$\bigcap_{\mathbf{m}\in\mathcal{G}} L(E(\mathbf{m})) \tag{4.6}$$

where $E(\mathbf{m})$ *and* $\mathcal{G}$ *are defined in* (4.5) *and* (4.4), *and* (4.6) *is taken to be* R^p *if* $\mathcal{G}$ *is* *empty. If* $S(\mathbf{f}(x)) = R^p$, *the consistency region* $C(\mathbf{f}(x)) = R^1$; *otherwise*, $C(\mathbf{f}(x)) = \{a_i\}_{i=1}^k$.

PROOF. If $\mathbf{x} = \{x_i\}_{i=1}^\infty$ has k limit points $\{a_i\}_{i=1}^k$, the space $A(\mathbf{f}(x))$ in Theorem 3 is equal to $L(f(a_i), 1 \leqslant i \leqslant k)$. Since $\mathbf{x}$ is bounded, $k \geqslant 1$. It remains to characterize $B(\mathbf{f}(x))$ in Theorem 3. For any nonzero $w \in A(\mathbf{f}(x))^\perp$, $w'f(x)$ is a nonzero polynomial in x of degree q, $q \leqslant p - 1$. From orthogonality, $w'f(a_i) = 0$ for $1 \leqslant i \leqslant k$. Therefore the degree of $w'f(x)$ is at least k. Let $a_1, \cdots, a_k, b_{k+1}, \cdots, b_q$ be the roots of $w'f(x)$, i.e., $w'f(x) = \Pi_{j=1}^k(x - a_j)\Pi_{j=k+1}^q(x - b_j)$. If b_j is not equal to one of $a_1, \cdots, a_k$, or equivalently not a limit point of $\mathbf{x}$,

$$0 < \liminf_{i\to\infty}|x_i - b_j| \leqslant \limsup_{i\to\infty}|x_i - b_j| < \infty.$$

Therefore the convergence or divergence of $\Sigma_{i=1}^\infty(w'f(x_i))^2 = \Sigma_{i=1}^\infty\Pi_{j=1}^k(x_i - a_j)^2\Pi_{j=k+1}^q(x_i - b_j)^2$ is independent of terms $(x_i - b_j)^2$ with b_j not equal to one of the a_i's. In order to study the asymptotic behavior of $\Sigma_{i=1}^\infty(w'f(x_i))^2$ for $w \in A(\mathbf{f}(x))^\perp$, it suffices to study the infinite series $\Sigma_{i=1}^\infty\Pi_{j=1}^k(x_i - a_j)^{2(1+m_j)}$ where m_j is a nonnegative integer and $\Sigma_{j=1}^k(m_j + 1) \leqslant p - 1$. Since $\mathbf{x}$ is bounded, it is easy to see that $g(\mathbf{m}') < \infty$ implies $g(\mathbf{m}) < \infty$ for all $\mathbf{m} \geqslant \mathbf{m}'$. This suggests that, in studying the asymptotic behavior of $\Sigma_{i=1}^\infty(w'f(x_i))^2$ for $w \in A(\mathbf{f}(x))^\perp$, it is enough to work on $g(\mathbf{m})$ for $\mathbf{m} \in \mathcal{G}$. If $\mathcal{G}$ is an empty set, i.e., $\Sigma_{i=1}^\infty\Pi_{j=1}^k(x_i - a_j)^{2(1+m_j)} = \infty$ for all $m_j \geqslant 0$ and $\Sigma_{j=1}^k(m_j + 1) \leqslant p - 1$, then $\Sigma_{i=1}^\infty(w'f(x_i))^2 = \infty$ for all nonzero w. From Corollary 1, $S(f(x)) = R^p$ and $C(f(x)) = R^1$. A necessary and sufficient condition for $\Sigma_{i=1}^\infty(w'f(x_i))^2 < \infty$, $w \in A(\mathbf{f}(x))^\perp$, is that $w'f(x)$ is divisible by the polynomial $\Pi_{j=1}^k(x - a_j)^{1+m_j}$ for some $\mathbf{m} \in \mathcal{G}$. The latter polynomial has roots a_j with multiplicity $m_j + 1$, which implies $w'f^{(i)}(a_j) = 0$ for $0 \leqslant i \leqslant m_j$, $1 \leqslant j \leqslant k$. ($f^{(i)}(a_j)$ is the ith derivative of $f(a_j)$). The above n.a.s. condition can now be restated as: w is orthogonal to the linear space spanned by $E(\mathbf{m})$ for some $\mathbf{m} \in \mathcal{G}$. The linear space $B(\mathbf{f}(x))$ in Theorem 3 consists of all vectors in $A(\mathbf{f}(x))^\perp$ orthogonal to the set

$$\{w: w \in A(\mathbf{f}(x))^\perp, \Sigma_{i=1}^\infty(w'f(x_i))^2 < \infty\}. \tag{4.7}$$

From the above n.a.s. condition on w in (4.7), $B(\mathbf{f}(x))$ can be rewritten as the collection of vectors in $A(\mathbf{f}(x))^\perp$ orthogonal to

$$\bigcup_{\mathbf{m}\in\mathcal{G}} L(E(\mathbf{m}))^\perp$$

or equivalently, orthogonal to

$$\Sigma_{\mathbf{m}\in\mathcal{G}} L(E(\mathbf{m}))^\perp.$$

This sum is equal to (see Halmos, 1958, page 134.)

$$\left(\bigcap_{\mathbf{m}\in\mathcal{G}} L(E(\mathbf{m}))\right)^\perp.$$

Note that $L(E(\mathbf{m}))$ is the linear space spanned by $E(\mathbf{m})$. Thus we obtain

$$B(\mathbf{f}(x)) = \bigcap_{\mathbf{m}\in\mathcal{G}} L(E(\mathbf{m})) \bigcap A(\mathbf{f}(x))^\perp.$$

Since $A(\mathbf{f}(x))$ is a subspace of $\bigcap_{\mathbf{m}\in\mathcal{G}} L(E(\mathbf{m}))$, by writing

$$\bigcap_{\mathbf{m}\in\mathcal{G}} L(E(\mathbf{m})) = A(\mathbf{f}(x)) \oplus N \text{ with } N \subseteq A(\mathbf{f}(x))^\perp,$$

we easily obtain

$$A(\mathbf{f}(x)) \oplus B(\mathbf{f}(x)) = A(\mathbf{f}(x)) \oplus \left[\bigcap_{\mathbf{m}\in\mathcal{G}} L(E(\mathbf{m})) \bigcap A(\mathbf{f}(x))^{\perp} \right]$$
$$= A(\mathbf{f}(x)) \oplus N = \bigcap_{\mathbf{m}\in\mathcal{G}} L(E(\mathbf{m})).$$

This establishes the first part of the Theorem.

It remains to prov ; the second part. If $S(\mathbf{f}(x)) \subsetneq R^p$, there exists an $\mathbf{m}_0 \in \mathcal{G}$ with $L(E(\mathbf{m}_0)) \subsetneq R^p$. From Corollary 5, $L(E(\mathbf{m}_0)) \cap \{f(x): -\infty < x < \infty\} = \{f(a_i), 1 \leqslant i \leqslant k\}$, which implies $S(\mathbf{f}(x)) \cap \{f(x): -\infty < x < \infty\} \subseteq \{f(a_i), 1 \leqslant i \leqslant k\}$. It is obvious that $\{f(a_i), 1 \leqslant i \leqslant k\} \subseteq A(\mathbf{f}(x)) \subseteq S(\mathbf{f}(x))$. $C(\mathbf{f}(x)) = \{a_i, 1 \leqslant i \leqslant k\}$ is thus proved. □

The above result on the consistency region excludes the puzzling phenomenon exhibited in the example in Section 3. For that example, $f(x)'\hat{\theta}$ can be consistent on part of the set of nonlimit points and inconsistent on the other part.

SOME CONSEQUENCES OF THEOREM 4.

1. Theorem 4 is true without assuming $k \leqslant p - 1$. In fact, when $k \geqslant p$, all the $L(E(\mathbf{m}))$ are equal to R^p and the result coincides with Proposition 4.

In the following discussion we assume $k \leqslant p - 1$.

2. If $\Sigma_{i=1}^{\infty}\Pi_{j=1}^{k}(x_i - a_j)^{2(1+m_j)} = \infty$ for all $m_j \geqslant 0$ and $\Sigma_{j=1}^{k}(1 + m_j) = p - 1$, then $S(\mathbf{f}) = R^p$ and $C(\mathbf{f}) = R^1$. The condition is just a restatement of $\mathcal{G} = \phi$.

3. If $\Sigma_{i=1}^{\infty}\Pi_{j=1}^{k}(x_i - a_j)^2 < \infty$, then $S(\mathbf{f}) = L\{f(a_i), 1 \leqslant i \leqslant k\}$ and $C(\mathbf{f}) = \{a_i\}_{i=1}^{k}$.

4. $k = p - 1$; if $\Sigma_{i=1}^{\infty}\Pi_{j=1}^{k}(x_i - a_j)^2 < \infty$, the result follows from 3; if $\Sigma_{i=1}^{\infty}\Pi_{j=1}^{k}(x_i - a_j)^2 = \infty$, then $S(\mathbf{f}) = R^p$ and $C(\mathbf{f}) = R^1$.

For $k = p - 1$, $\mathcal{G} = \{(0, \cdots, 0)\}$ or ϕ which corresponds to the convergence or divergence of the above series.

5. $k = p - 2$; let $K = \{j': \Sigma_{i=1}^{\infty}\Pi_{j=1}^{p-2}(x_i - a_j)^2(x_i - a_{j'})^2 < \infty\}$.

(i) If $\Sigma_{i=1}^{\infty}\Pi_{j=1}^{p-2}(x_i - a_j)^2 < \infty$, the result follows from 3.

(ii) If $K = \phi$, the result follows from 2.

(iii) $\Sigma_{i=1}^{\infty}\Pi_{j=1}^{p-2}(x_i - a_j)^2 = \infty$; if K has only one element j_0, then $S(\mathbf{f}) = L(f(a_1), \cdots, f(a_{p-2}), f^{(1)}(a_{j_0}))$ and $C(\mathbf{f}) = \{a_i\}_{i=1}^{p-2}$; if K has at least two elements, then $S(\mathbf{f}) = L(f(a_i), 1 \leqslant i \leqslant p - 2)$ and $C(\mathbf{f}) = \{a_i\}_{i=1}^{p-2}$.

PROOF OF (iii). If K has only one element, the result follows from Theorem 4. If there exist $j_1 \neq j_2$ in K, the p vectors $\{f(a_i), 1 \leqslant i \leqslant p - 2, f^{(1)}(a_{j_1}), f^{(1)}(a_{j_2})\}$ are linearly independent from Proposition 2. This implies that $L(f(a_i), 1 \leqslant i \leqslant p - 2, f^{(1)}(a_{j_1})) \cap L(f(a_i), 1 \leqslant i \leqslant p - 2, f^{(1)}(a_{j_2})) = L(f(a_i), 1 \leqslant i \leqslant p - 2)$. Therefore, $\bigcap_{\mathbf{m}\in\mathcal{G}} L(E(\mathbf{m}))$ in Theorem 4 is equal to $L(f(a_i), 1 \leqslant i \leqslant p - 2)$.

6. $k = 1$, if $\Sigma_{i=1}^{\infty}(x_i - a_1)^{2(p-1)} = \infty$, the result follows from 2. If $\Sigma_{i=1}^{\infty}(x_i - a_1)^{2m_1} = \infty$ and $\Sigma_{i=1}^{\infty}(x_i - a_1)^{2(1+m_1)} < \infty$, $0 \leqslant m_1 \leqslant p - 2$, then $S(\mathbf{f}) = L(f(a_1), f^{(1)}(a_1), \cdots, f^{(m_1)}(a_1))$ and $C(\mathbf{f}) = \{a_1\}$.

Using Theorem 4 and its consequences, simple and workable n.a.s. conditions for asymptotic consistency are given below for cubic regression $f(x)' = (1, x, x^2, x^3)$. The results for other polynomial regression models can be derived similarly.

(i) $\mathbf{x}$ has at least four limit points. $S(\mathbf{f}) = R^4$ and $C(\mathbf{f}) = R^1$.

(ii) $\mathbf{x}$ has three limit points a_1, a_2, a_3.

$$\Sigma_{i=1}^{\infty}\Pi_{j=1}^{3}(x_i - a_j)^2 < \infty : S(\mathbf{f}) = L(f(a_i), 1 \leqslant i \leqslant 3) \text{ and } C(\mathbf{f}) = \{a_i\}_{i=1}^{3}.$$

$$\Sigma_{i=1}^{\infty}\Pi_{j=1}^{3}(x_i - a_j)^2 = \infty : S(\mathbf{f}) = R^4 \text{ and } C(\mathbf{f}) = R^1.$$

(iii) $\mathbf{x}$ has two limit points a_1, a_2.

$$\Sigma_{i=1}^{\infty}(x_i - a_1)^2(x_i - a_2)^2 < \infty : S(\mathbf{f}) = L(f(a_1), f(a_2)) \text{ and } C(\mathbf{f}) = \{a_1, a_2\}.$$

$$\Sigma_{i=1}^{\infty}(x_i - a_1)^2(x_i - a_2)^2 = \Sigma_{i=1}^{\infty}(x_i - a_1)^2(x_i - a_2)^4 = \infty \text{ and}$$
$$\Sigma_{i=1}^{\infty}(x_i - a_1)^4(x_i - a_2)^2 < \infty : S(\mathbf{f}) = L(f(a_1), f(a_2), f^{(1)}(a_1))$$
$$\text{and } C(\mathbf{f}) = \{a_1, a_2\}.$$

(A similar result holds for the case with a_1 and a_2 interchanged).

$$\Sigma_{i=1}^{\infty}(x_i - a_1)^2(x_i - a_2)^2 = \infty, \Sigma_{i=1}^{\infty}(x_i - a_1)^4(x_i - a_2)^2 < \infty \text{ and}$$
$$\Sigma_{i=1}^{\infty}(x_i - a_1)^2(x_i - a_2)^4 < \infty : S(\mathbf{f}) = L(f(a_1), f(a_2))$$
$$\text{and } C(\mathbf{f}) = \{a_1, a_2\}.$$

If all the above three series diverge: $S(\mathbf{f}) = R^4$ and $C(\mathbf{f}) = R^1$.

(iv) $\mathbf{x}$ has one limit point a.

$$\Sigma_1^{\infty}(x_i - a)^2 < \infty : S(\mathbf{f}) = L(f(a)) \text{ and } C(f) = \{a\}.$$

$$\Sigma_{i=1}^{\infty}(x_i - a)^2 = \infty, \Sigma_{i=1}^{\infty}(x_i - a)^4 < \infty : S(\mathbf{f}) = L(f(a), f^{(1)}(a))$$
$$\text{and } C(\mathbf{f}) = \{a\}.$$

$$\Sigma_{i=1}^{\infty}(x_i - a)^4 = \infty \text{ and } \Sigma_{i=1}^{\infty}(x_i - a)^6 < \infty : S(\mathbf{f}) = L(f(a), f^{(1)}(a), f^{(2)}(a))$$
$$\text{and } C(\mathbf{f}) = \{a\}.$$

$$\Sigma_{i=1}^{\infty}(x_i - a)^6 = \infty : S(\mathbf{f}) = R^4 \text{ and } C(\mathbf{f}) = R^1.$$

Some of the results obtained in this section and (3.8) actually hold for a general one-dimensional regression model, i.e., $f(x)$ is a smooth function of x in R^1. This will be treated elsewhere.

Acknowledgment. The author wishes to thank a referee for pointing out an error in the earlier draft of the paper and to Professor Rupert Miller and the Associate Editor for comments which improve the style of the paper.

REFERENCES

Dieundonne, J. (1960). *Foundations of Modern Analysis*. Academic, New York.

Drygas, H. (1976). Weak and strong consistency of the least squares estimates in regression models. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*. **34** 119–127.

Halmos, P.R. (1958). *Finite Dimensional Vector Spaces*. 2nd Ed. Van Nostrand, Princeton.

Huber, P.J. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1** 799–821.

HUBER, P.J. (1975). Robustness and designs. In *A Survey of Statistical Design and Linear Models*. (J.N. Srivastava, ed.) 287–301. North-Holland, Amsterdam.

JENNRICH, R.I. (1969). Asymptotic properties of nonlinear least squares estimators. *Ann. Math. Statist.* **40** 633–643.

KARLIN, S. and STUDDEN, W.J. (1966). Tchebycheff systems: with applications in analysis and statistics. Wiley, New York.

LAI, T.L., ROBBINS, H. and WEI, C.Z. (1978). Strong consistency of least squares estimates in multiple regression. *Proc. Nat. Acad. Sci.* **75** 3034–3036.

PÁZMAN, A. (1974). A convergence theorem in the theory of D-optimum experimental designs. *Ann. Statist.* **2** 216–218.

WU, C.F. (1979). Asymptotic theory of nonlinear least squares estimation. Technical Report No. 561, Univ. Wisconsin.

WYNN, H.P. (1976). The limiting variance function for design sequences. *Math. Operationsforch. Statist.* **7** 69–74.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN
1210 WEST DAYTON STREET
MADISON, WISCONSIN 53706