The Annals of Statistics
1983, Vol. 11, No. 4, 1257–1262

FURTHER RESULTS ON THE CONSISTENT DIRECTIONS OF LEAST SQUARES ESTIMATORS[1]

BY SONG-GUI WANG AND C. F. J. WU

Chinese University of Science and Technology, Hefei, Anhui; and University of Wisconsin, Madison

Wu (1980) defined the consistent directions of the least squares estimator in a linear model as the linear combinations of parameter estimates that are asymptotically consistent. For the polynomial regression model, a characterization of the space of consistent directions S was obtained in terms of the convergence rates of the corresponding design sequence to its limit points. By employing a more general and yet simpler approach, we obtain here a similar result for any regression model with one independent variable and smooth regression function. When $f_i(x)$ is an extended Tchebycheff system, the above characterization is further refined and the consistency region C is shown to be either the set of limit points of the design sequence or the whole real line.

1. Introduction. Recently there has been a revived interest in the consistency properties of the least squares estimators in the linear model (e.g., Chen, 1979; Lai, Robbins and Wei, 1979; Chen, Lai and Wei, 1981; and others in the references of Wu, 1980). A basic result is that the least squares estimator $\hat{\theta} \to \theta$ a.s. (or in probability) if and only if $(X'_n X_n)^{-1} \to 0$ when $\{e_i\}_{i=1}^{\infty}$ are i.i.d. (or uncorrelated) with mean zero and variance σ^2 (Notations defined in (1.1) and (1.2) below). Wu (1980) observed the following interesting fact: the best linear unbiased estimator $\mathbf{b}'\hat{\theta}$ of a linear combination $\mathbf{b}'\theta$ can be consistent for $\mathbf{b}'\theta$ for some vectors $\mathbf{b}$, even if the vector-valued estimator $\hat{\theta}$ is not. Such a vector $\mathbf{b}$ is called a *consistent direction* of the least squares estimator $\hat{\theta}$. Wu gave a general characterization of the space of consistent directions of $\hat{\theta}$. For the one dimensional polynomial regression, he gave a more refined characterization in terms of the convergence rates of some design subsequences. More recently Li (1982) has extended some of the results to regression models with infinite parameters. The results of Wu (1981) and Li (1982) show that for finite parameters the consistency of the least squares estimator is equivalent to the existence of a consistent estimator. Therefore the consideration of linear estimators herein is not restrictive.

The consideration of a design sequence converging to a finite number of limit points is relevant in some statistical contexts. In stochastic approximation, the design sequence, if properly chosen, converges to the desired location with probability one. An excellent recent work is Lai and Robbins (1979). In sequential generation of optimal designs for nonlinear situations, the design sequence tends to cluster around the (finite number of) points in the support of an optimal design. A particular case was studied by Ford and Silvey (1980). Our results are not directly applicable since these sequences are dependent due to the nature of sequential generation.

It was discovered later that, by employing a simple and unified approach, a similar characterization can be obtained for any one-dimensional regression problem with smooth function f. Before spelling out the details, we shall give the basic framework and result in the following.

Received June 1982; revision January 1983.

[1] Research supported by the National Science Foundation Grant No. MCS-7901846. The paper was completed while S. G. Wang was visiting the Department of Statistics, University of Wisconsin, on the UW exchange program with PRC.

AMS 1980 *subject classifications.* Primary 62J05; secondary 62E20.

Key words and phrases. Asymptotic consistency, consistent direction, consistency region, extended Tchebycheff system, least squares estimators, polynomial regression.

A linear model is given by

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

where $\boldsymbol{\theta}$ and $\mathbf{f}(x)$ are $p \times 1$ vectors. We assume that x is a scalar, the random error ε has mean zero, variance σ^2 and errors corresponding to different observations are uncorrelated or independent. If y_i is observed at x_i, $i = 1, \cdots, n$, and $X'_n = (\mathbf{f}(x_1), \cdots, \mathbf{f}(x_n))$ is of full rank, the least squares estimate of $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\theta}} = (X'_n X_n)^{-1} X'_n \mathbf{y}, \tag{1.2}$$

where $\mathbf{y}' = (y_1, \cdots, y_n)$.

The following theorem plays a key role in the determination of consistent directions.

THEOREM A (Wu, 1980). $\mathbf{u}$ *is a consistent direction if and only if*

$$\sum_{i=1}^{\infty} (\mathbf{w}'\mathbf{f}(x_i))^2 = \infty \quad \textit{for all} \quad \mathbf{w}'\mathbf{u} \neq 0. \tag{1.3}$$

Denote the *space of consistent directions* and the *consistency region* by $S(\mathbf{f})$ and $C(\mathbf{f})$, respectively, i.e.,

$$S(\mathbf{f}) = \{\mathbf{b}: \mathbf{b}'(X'_n X_n)^{-1}\mathbf{b} = \mathbf{b}'(\textstyle\sum_{i=1}^{n} \mathbf{f}(x_i)\mathbf{f}(x_i)')^{-1}\mathbf{b} \to 0 \quad \text{as} \quad n \to \infty\},$$

$$C(\mathbf{f}) = \{x: \mathbf{f}(x) \in S(\mathbf{f}),\ x \in \mathscr{X}\},$$

where $\mathscr{X}$ is a set on which linear model (1.1) is valid.

The main result for models with one independent variable x and smooth function $\mathbf{f}(x)$ is stated as Theorem 1. When $\mathbf{f}(x)$ is an extended Tchebycheff system, the decomposition of $S(\mathbf{f})$ in Theorem 1 is further refined and a simple characterization of $C(\mathbf{f})$ is given in Theorem 2. For smooth regression models with several independent variables, extensions of Proposition 2 are given in Wang and Wu (1982).

2. General consistency results for one independent variable. In the process of establishing the characterization of $S(\mathbf{f})$ for the polynomial regression model, Wu (1980) employed special algebraic properties of the polynomial system. It will be shown that the use of these algebraic properties is not necessary. The simple approach given below involves examining the dominating terms of the Taylor series expansion of $\mathbf{f}(x)$. It is generally applicable for any smooth $\mathbf{f}(x)$.

Let $L(\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_t)$ be the subspace spanned by vectors $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_t$ and $C^r(a, b)$ the set of functions with rth continuous differential in (a, b). In particular, $C^0(a, b)$ is the set of continuous functions in (a, b). Define $\mathbf{f}(x) \in C^r(a, b)$ if $f_i(x) \in C^r(a, b)$ for $i = 0, 1, \cdots, p - 1$. Throughout this paper we assume the design sequence $\{x_i\}_{i=1}^{\infty}$ is bounded. If there is an infinite subsequence n_i such that $x_{n_i} \to a$ as $i \to \infty$, we say that a is a limit point of $\{x_i\}_{i=1}^{\infty}$.

PROPOSITION 1. *If* $\mathbf{f}(x) \in C^0(a - \delta, a + \delta)$ *for some* $\delta > 0$, *then* $\mathbf{f}(a)$ *is a consistent direction for any limit point* a *of the sequence* $\{x_i\}_{i=1}^{\infty}$.

It follows immediately from Theorem 2 of Wu (1980).

PROPOSITION 2. *If* $x_{n_i} \to a$ as $i \to \infty$, $\sum_{i=1}^{\infty} (x_{n_i} - a)^{2r} = \infty$, $\mathbf{f}(x) \in C^{r+1}(a - \delta, a + \delta)$ *for some* $\delta > 0$, *then* $\mathbf{f}^{(t)}(a)$ *are consistent directions for* $t = 1, 2, \cdots, r$, *where*

$$\mathbf{f}^{(t)}(a) = \left(\frac{d^t f_0}{dx^t}, \cdots, \frac{d^t f_{p-1}}{dx^t}\right)' \Bigg|_{x=a}.$$

(Li (1982) obtained a similar result for nonparametric regression models.)

PROOF. We first prove that $\mathbf{f}^{(r)}(a)$ is a consistent direction. From Theorem A, it is sufficient to prove $\sum_{i=1}^{\infty} [\mathbf{w}'\mathbf{f}(x_{n_i})]^2 = \infty$ for any $\mathbf{w}'\mathbf{f}^{(r)}(a) \neq 0$.

By the assumption, for any $0 < \varepsilon < \delta$, there exists an N_ε such that $|x_{n_i} - a| < \varepsilon$ for $i > N_\varepsilon$. Consider the Taylor series expansion of $\mathbf{f}(x_{n_i})$ at a:

$$\mathbf{w}'\mathbf{f}(x_{n_i}) = \sum_{k=0}^{r} (x_{n_i} - a)^k \mathbf{w}'\mathbf{f}^{(k)}(a)/k! + (x_{n_i} - a)^{(r+1)} \mathbf{w}'\mathbf{f}^{(r+1)}(\xi_i)/(r+1)!,$$

where $\xi_i = a + \theta_i(x_{n_i} - a)$ and $0 \leq \theta_i \leq 1$ from the mean-value theorem. Thus

$$\sum_{i=1}^{\infty} [\mathbf{w}'\mathbf{f}(x_{n_i})]^2 = \sum_{i=1}^{\infty} (x_{n_i} - a)^{2r} \cdot \left[\frac{\mathbf{w}'\mathbf{f}(a)}{(x_{n_i} - a)^r} + \frac{\mathbf{w}'\mathbf{f}^{(1)}(a)}{(x_{n_i} - a)^{r-1}} + \cdots + \frac{\mathbf{w}'\mathbf{f}^{(r)}(a)}{r!} + \frac{\mathbf{w}'\mathbf{f}^{(r+1)}(\xi_i)}{(r+1)!}(x_{n_i} - a)\right]^2. \tag{2.1}$$

Since $\mathbf{f}(x) \in C^{r+1}(a - \delta, a + \delta)$, $\delta > 0$, there exists a constant $M > 0$ such that $|f_j^{(r+1)}(\xi_i)| \leq M$ for sufficiently large i and for all j. Therefore the last term inside the square bracket of (2.1) converges to zero as $i \to \infty$.

Let i_0 be the first i with $\mathbf{w}'\mathbf{f}^{(i)}(a) \neq 0$. From $\mathbf{w}'\mathbf{f}^{(r)}(a) \neq 0$, $i_0 \leq r$. The terms inside the square bracket of (2.1) are dominated by the leading term $\mathbf{w}'\mathbf{f}^{(i_0)}(a)/(x_{n_i} - a)^{r-i_0}$, which is bounded away from zero as $i \to \infty$. Therefore $\sum_{i=1}^{\infty} [\mathbf{w}'\mathbf{f}(x_{n_i})]^2 = \infty$ follows from $\sum_{i=1}^{\infty} (x_{n_i} - a)^{2r} = \infty$.

That $\mathbf{f}^{(t)}(a)$, $t = 1, 2, \cdots, r-1$, are consistent directions follows from the fact that $\sum_{i=1}^{\infty} (x_{n_i} - a)^{2r} = \infty$ implies $\sum_{i=1}^{\infty} (x_{n_i} - a)^{2t} = \infty$ for $t < r$. □

By combining the previous results for each limit point a_j of the sequence $\{x_i\}_{i=1}^{\infty}$, we obtain the following main theorem.

Denote by $V^{\perp}$ and $V_1 \oplus V_2$, respectively, the orthogonal complement of subspace V and the direct sum of subspaces V_1 and V_2 (Shephard, 1966).

THEOREM 1. *Suppose* $\{x_{n_i}^{(j)}\}_{i=1}^{\infty}$, $j = 1, 2, \cdots, k$, *are subsequences of* $\{x_i\}_{i=1}^{\infty}$ *such that* $x_{n_i}^{(j)} \to a_j$ *as* $i \to \infty$, *and* $\mathbf{f}(x) \in C^{(r_j+1)}(a_j - \delta, a_j + \delta)$ *for some* $\delta > 0$, $j = 1, \cdots, k$, *where* $r_j = \max\{l: \sum_{i=1}^{\infty} (x_{n_i}^{(j)} - a_j)^{2l} = \infty$, *l positive integer.) Then, the space of consistent directions*

$$S(\mathbf{f}) = \sum_{j=1}^{k} A_j(\mathbf{f}) \oplus B_{k+1}(\mathbf{f}),$$

where $A_j(\mathbf{f}) = L\{\mathbf{f}^{(r)}(a_j), 0 \leq r \leq r_j\}$, $j = 1, 2, \cdots, k$,

$$\sum_{j=1}^{k} A_j(\mathbf{f}) = \{\mathbf{u}: \mathbf{u} = \sum_{j=1}^{k} \mathbf{u}_j,\ \mathbf{u}_j \in A_j(\mathbf{f})\},$$

$$B_{k+1}(\mathbf{f}) = \{\mathbf{u} \in [\textstyle\sum_{j=1}^{k} A_j(\mathbf{f})]^{\perp}: \sum_{i=1}^{\infty} [\mathbf{w}'\mathbf{f}(x_i)]^2 = \infty \text{ for any } \mathbf{w} \in [\sum_{j=1}^{k} A_j(\mathbf{f})]^{\perp} \text{ with } \mathbf{w}'\mathbf{u} \neq 0\}.$$

PROOF. From Propositions 1 and 2, for each limit point a_j, $A_j(\mathbf{f}) \subset S(\mathbf{f})$ $j = 1, \cdots, k$. Therefore $\sum_{j=1}^{k} A_j(\mathbf{f}) \subset S(\mathbf{f})$. Since $B_{k+1}(\mathbf{f}) = S(\mathbf{f}) \cap [\sum_{j=1}^{k} A_j(\mathbf{f})]^{\perp}$ by using the same argument as in Theorem 2 of Wu (1980), the theorem is proved. □

We should note that there is no loss of generality in assuming finite k in Theorem 1. In fact, if the design sequence $\{x_i\}_{i=1}^{\infty}$ has infinitely many limit points $\{a_j\}_{j=1}^{\infty}$, then there exists a sufficiently large k such that $A_l(\mathbf{f}) \subset \sum_{j=1}^{k} A_j(\mathbf{f})$ for all $l \geq k+1$. In other words the two sets $\{a_j\}_{j=1}^{k}$ and $\{a_j\}_{j=1}^{\infty}$ are equivalent in the sense of producing the same space of consistent directions.

COROLLARY 1. *Under the conditions of Theorem* 1, *if the dimension of* $L\{\mathbf{f}^{(r)}(a_j), r = 0, 1, \cdots, r_j, j = 1, \cdots, k\}$ *is* p, *then* $S(\mathbf{f}) = R^p$, $C(\mathbf{f}) = R^1$.

In Corollary 1, an alternative sufficient condition for the asymptotic consistency of the least squares estimator $\hat{\theta}$ is provided. Unlike the more familiar condition $(X'_nX_n)^{-1} \to 0$ which involves matrix inversion, our conditions are directly expressed in terms of the "convergence rates" of the different design subsequences to their respective limit points. There are many examples for which our conditions are easier to verify than the usual one. We should also point out that Theorem 1 covers a broad class of $\mathbf{f}(x)$ functions including those considered in Wu (1980).

3. Consistency results for the extended Tchebycheff system. For the polynomial regression model, Wu (1980) expressed $S(\mathbf{f})$ as the intersection of some subspaces. Since the more general Extended Tchebycheff system (ET-system) shares all the required properties of the polynomial system in establishing the characterization of $S(\mathbf{f})$, we will obtain, via Theorem 1, a simpler characterization of $S(\mathbf{f})$ for model (1.1) with $\mathbf{f}(x)$ being an ET-system. For simplicity, we label such model as Model (T).

DEFINITION 1. The functions $\{f_i(x)\}_{i=0}^{p-1}$ are called an ET-system of order $q+1$ on $[a, b]$, if $f_i(x) \in C^q[a, b]$, $i = 0, 1, \cdots, p-1$, and the determinant of the $p \times p$ matrix with column vectors $\{\mathbf{f}^{(i)}(x_j),\ 0 \le i \le m_j \le q,\ 1 \le j \le k,\ \sum_{j=1}^k (m_j+1) = p\}$, where $\mathbf{f}(x) = (f_0(x), \cdots, f_{p-1}(x))'$, is positive for any choice $a \le x_1 < \cdots < x_k \le b$. Further, if $\{f_i(x)\}_{i=0}^r$ is an ET-system on $[a, b]$ for each $r = 1, 2, \cdots, p-1$, then $\{f_i(x)\}_{i=0}^{p-1}$ is called an Extended Complete Tchebycheff system (ECT-system) on $[a, b]$. The extensions of the ET-system or ECT-system to open or infinite intervals are obvious.

According to Theorem 4.3 of Karlin and Studden (1966), the ET-system or ECT-system includes a wide class of functions, e.g., the power functions $\{x^i\}_{i=0}^{p-1}$, the Tchebycheff polynomial system $\{T_i(x)\}_{i=0}^{p-1}$ (i.e., $T_0(x) = 1$, $T_1(x) = x$ and $T_{r+1}(x) = 2x\,T_r(x) - T_{r-1}(x)$, $r = 1, 2, \cdots, p-2$), and more generally any polynomial system $\{\Phi_i(x)\}_{i=0}^{p-1}$, where $\Phi_i(x)$, $i = 0, 1, \cdots, p-1$, are polynomials of degree at most $p-1$. Further, from elementary properties of the ECT-system (Karlin and Studden, 1966, Theorem 1.2), it is easy to verify that the following are ECT-system on the corresponding intervals: (i) $\{e^{\alpha_i x}\}_i$ on $(0, \infty)$, (ii) $\{1, e^x, xe^x\}$ on $(-\infty, \infty)$, (iii) $\{1, \ln x, x\}$ on $(0, \infty)$.

The following two properties of ET-system will play a key role in determining $S(\mathbf{f})$ and $C(\mathbf{f})$ for Model (T).

If $\{f_i(x)\}_{i=0}^{p-1}$ is an ET-system of order p on $[a, b]$ and $\{x_j\}_{j=1}^k$ are k distinct points in $[a, b]$, then it follows from the definition that the vectors

$$\{\mathbf{f}^{(r)}(x_j),\ 0 \le r \le m_j,\ 1 \le j \le k,\ \textstyle\sum_{j=1}^k (m_j+1) = p\} \tag{3.1}$$

are linearly independent.

The second property is stated as a proposition.

PROPOSITION 3. *For an* ET-*system* $\{f_i(x)\}_{i=0}^{p-1}$ *of order* p *on* $[a, b]$,

$$L\{\mathbf{f}^{(r)}(x_j),\ 0 \le r \le m_j,\ 1 \le j \le k,\ \textstyle\sum_{j=1}^k (m_j+1) = p-1\} \cap \{\mathbf{f}(x),\ x \in [a, b]\} = \{\mathbf{f}(x_j),\ 1 \le j \le k\},$$

where $\{x_j\}_{j=1}^k$ *are* k *distinct points in* $[a, b]$.

PROOF. The idea of proof is similar to that of Corollary 5 of Wu (1980). Any function of the form $F(x) = \mathbf{w}'\mathbf{f}(x) = \sum_{i=0}^{p-1} w_i f_i(x)$, where w_i, $i = 0, 1, \cdots, p-1$, are real numbers, is called a *generalized* polynomial. Consider a special generalized polynomial $F(x)$ defined by the determinant

$$F(x) = |\,\mathbf{f}^{(0)}(x_1), \cdots, \mathbf{f}^{(m_1)}(x_1), \mathbf{f}^{(0)}(x_2), \cdots, \mathbf{f}^{(m_2)}(x_2), \cdots, \mathbf{f}^{(0)}(x_k), \cdots, \mathbf{f}^{(m_k)}(x_k), \mathbf{f}(x)\,|.$$

We note that x_j, $j = 1, \cdots, k$, are zeros of $F(x)$ and their multiplicities are, respectively,

$m_j + 1, j = 1, \cdots, k$. According to Theorem 4.3 of Karlin and Studden (1966), $F(x)$ has no other zeros. From (3.1), $\mathbf{f}^{(i)}(x_j)$, $0 \le i \le m_j$, $1 \le j \le k$, are linearly independent. Therefore

$$F(x) = 0 \text{ iff } \mathbf{f}(x) \in L\{\mathbf{f}^{(i)}(x_j),\ 0 \le i \le m_j,\ 1 \le j \le k,\ \textstyle\sum_{j=1}^{k} (m_j + 1) = p\}.$$

The proposition is proved. □

We are now in the position to establish the characterization of $S(\mathbf{f})$ for Model (T). The notations are the same as in Section 2.

THEOREM 2. *For model (T), i.e., model (1.1) with $\{f_i(x)\}_{i=0}^{p-1}$ being an ET-system of order p on $[a, b]$,*

$$S(\mathbf{f}) = \oplus_{j=1}^{k} A_j(\mathbf{f}). \tag{3.2}$$

This differs from Theorem 1 in that $B_{k+1}(\mathbf{f}) = 0$.

Further, its consistency region takes two extreme forms,

$$\text{(i) for } \textstyle\sum_{j=1}^{k} (r_j + 1) \ge p, \quad S(\mathbf{f}) = R^p,\ C(\mathbf{f}) = R^1, \tag{3.3}$$

$$\text{(ii) for } \textstyle\sum_{j=1}^{k} (r_j + 1) \le p - 1, \quad C(\mathbf{f}) = \{a_i\}_{i=1}^{k}. \tag{3.4}$$

PROOF. According to Theorem 1, $\sum_{j=1}^{k} A_j(\mathbf{f}) \subset S(\mathbf{f})$, and from (3.1), the bases of the subspaces $A_j(\mathbf{f})$, $j = 1, \ldots, k$ are linearly independent. We may therefore replace $\sum_{j=1}^{k} A_j(\mathbf{f})$ by $\oplus_{j=1}^{k} A_j(\mathbf{f})$ in Theorem 1, i.e.,

$$S(\mathbf{f}) = \oplus_{j=1}^{k} A_j(\mathbf{f}) \oplus B_{k+1}(\mathbf{f}).$$

It remains to prove $B_{k+1}(\mathbf{f}) = \{0\}$. To this end, it suffices to prove $\sum_{i=1}^{\infty} [\mathbf{w}'\mathbf{f}(x_i)]^2 < \infty$ for any $\mathbf{w} \in [\oplus_{j=1}^{k} A_j(\mathbf{f})]^{\perp}$ because of Theorem A.

For any $\mathbf{w} \in [\oplus_{j=1}^{k} A_j(\mathbf{f})]^{\perp}$ and $\mathbf{w} \ne 0$, we have $\mathbf{w}'\mathbf{f}^{(i)}(a_j) = 0$ for $j = 1, \cdots, k$, and $i = 0, 1, \cdots, r_j$. Therefore, the generalized polynomial $\mathbf{w}'\mathbf{f}(x)$, $x \in [a, b]$, has zeros $a_1, \cdots, a_k$ with multiplicities $r_1 + 1, \cdots, r_k + 1$, respectively. By Theorem 2.3 of Karlin and Studden (1966), $\mathbf{w}'\mathbf{f}(x)$ has $l(\le p - \sum_{j=1}^{k} (r_j + 1))$ other zeros, denoted by $b_1, \cdots, b_l$. Thus, for any nonzero $\mathbf{w} \in [\oplus_{j=1}^{k} A_j(\mathbf{f})]^{\perp}$, we may write

$$\mathbf{w}'\mathbf{f}(x) = c \textstyle\prod_{j=1}^{k} (x - a_j)^{(r_j+1)} \prod_{h=1}^{l} (x - b_h), \quad x \in [a, b]$$

where c is constant, and

$$\textstyle\sum_{i=1}^{\infty} [\mathbf{w}'\mathbf{f}(x_i)]^2 = c^2 \sum_{i=1}^{\infty} \prod_{j=1}^{k} (x_i - a_j)^{2(r_j+1)} \prod_{h=1}^{l} (x_i - b_h)^2. \tag{3.5}$$

Because $a_j, j = 1, \cdots, k$, are limit points of $\{x_i\}_{i=1}^{\infty}$, $b_h \ne a_j$, for any $h = 1, \cdots, l$, $j = 1, \cdots, k$. Therefore, the convergence or divergence of (3.5) is independent of the factor $\prod_{h=1}^{l} (x_i - b_h)^2$. It remains to prove the convergence of $g(\mathbf{r}) = \sum_{i=1}^{\infty} \prod_{j=1}^{k} (x_i - a_j)^{2(r_j+1)}$, where $\mathbf{r}' = (r_1, \cdots, r_k)$. Since $| x_i - a_j | \le b - a$, for any i, j, we have

$$0 \le g(\mathbf{r}) \le M \textstyle\sum_{j=1}^{k} \sum_{i=1}^{\infty} (x_{n_i}^{(j)} - a_j)^{2(r_j+1)}, \tag{3.6}$$

where $M = \max\{1, (b - a)^{2\sum_{j=1}^{k}(r_j+1)}\}$. The convergence of $g(\mathbf{r})$ follows from (3.6) and the definition of r_j in Theorem 1. The conclusions (3.3) and (3.4) follow easily from (3.1) and Proposition 3, respectively. The proof is completed. □

Since the polynomial system is an ET-system, Theorem 2 applies to the general polynomial regression models. For this special case Theorem 4 of Wu (1980) characterized $S(\mathbf{f})$ in terms of the intersection of some subspaces, while our Theorem 2 does it by the direct sum of $A_j(\mathbf{f})$. Obviously, the latter is much simpler.

The results obtained so far provide us better insights to the following example given in

Wu (1980, page 794):

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

where $\mathbf{f}(x) = (e^x, x, x^2)'$, $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2)'$, y_i is observed at x_i.

There are two cases to be considered.

1. On $[-2, 2]$, $\{e^x, x, x^2\}$ is not an ET-system. It follows by taking $x_1 = -2$, $x_2 = 2$, $m_1 = 0$, $m_2 = 1$ in Definition 1.

   (i) If $x_i \to 2$, $\sum_{i=1}^{\infty} (x_i - 2)^2 = \infty$, $\sum_{i=1}^{\infty} (x_i - 2)^4 < \infty$, by Theorem 1,

   $$S(\mathbf{f}) = L\{\mathbf{f}(2), \mathbf{f}^{(1)}(2)\}.$$

   From straightforward calculation, we have $C(\mathbf{f}) = \{2, -0.5569\}$. It is a puzzling phenomenon that consistency is achieved at $x = -0.5569$ whose neighborhood contains no data at all. For an ECT-system this will not happen. Compare the case 2(i) below.

   (ii) If $x_i \to 2$, $\sum_{i=1}^{\infty} (x_i - 2)^2 = \sum_{i=1}^{\infty} (x_i - 2)^4 = \infty$, by Theorem 1,

   $$S(\mathbf{f}) = L\{\mathbf{f}(2), \mathbf{f}^{(1)}(2), f^{(2)}(2)\} = R^3 \quad \text{and} \quad C(\mathbf{f}) = R^1.$$

2. On $(-\infty, 1)$, $\{e^x, x, x^2\}$ is an ECT-system. It follows from Theorem 1.1 of Karlin and Studden (1966). Thus by Theorem 2,

   (i) If $x_i \to 0$, $\sum_{i=1}^{\infty} x_i^2 = \infty$, $\sum_{i=1}^{\infty} x_i^4 < \infty$, then

   $$S(\mathbf{f}) = L\{\mathbf{f}(0), \mathbf{f}^{(1)}(0)\}, \quad C(\mathbf{f}) = \{0\}.$$

   (ii) If $x_i \to 0$, $\sum_{i=1}^{\infty} x_i^2 = \sum_{i=1}^{\infty} x_i^4 = \infty$, then

   $$S(\mathbf{f}) = L\{\mathbf{f}(0), \mathbf{f}^{(1)}(0), \mathbf{f}^{(2)}(0)\} = R^3, \quad C(\mathbf{f}) = R^1.$$

REFERENCES

CHEN, G. J., LAI, T. L. and WEI, C. Z. (1981). Convergence system and strong consistency of least squares estimates in regression models. *J. Multivariate Anal.* **11** 319–333.

CHEN, XI-RU, (1979). On consistency of least squares estimates in linear models. *Sci. Sinica.* special issue, 162–176.

FORD, I. and SILVEY, S. D. (1980). A sequentially constructed design for estimating a nonlinear parametric function. *Biometrika* **67** 381–388.

KARLIN, S. and STUDDEN, W. J. (1966). *Tchebycheff System: with Applications in Analysis and Statistics.* Wiley, New York.

LAI, T. L. and ROBBINS, H. (1979). Adaptive design and stochastic approximation. *Ann. Statist.* **7** 1196–1221.

LAI, T. L., ROBBINS, H. and WEI, C. Z. (1979). Strong consistency of least squares estimates in multiple regression. *J. Multivariate Anal.* **9** 343–361.

LI, K. C. (1984). Regression models with infinitely many parameters: consistency of bounded linear functionals. *Ann. Statist.*, to appear.

SHEPHARD, G. C. (1966). *Vector Space of Finite Dimension.* Wiley, New York.

WANG, S. G. and WU, C. F. J. (1982). Consistent directions of the least squares estimators in linear models. Technical report no. 2421, Mathematics Research Center, Madison.

WU, C. F. (1980). Characterizing the consistent directions of least squares estimates. *Ann. Statist.* **8** 789–801.

WU, C. F. (1981). Asymptotic theory of nonlinear least squares estimation. *Ann. Statist.* **9** 501–513.

DEPARTMENT OF MATHEMATICS
CHINESE UNIVERSITY OF SCIENCE
AND TECHNOLOGY
HEFEI, ANHUI
PEOPLE'S REPUBLIC OF CHINA

DEPARTMENT OF STATISTICS
1210 W. DAYTON STREET
MADISON, WISCONSIN 53706