Statistica Sinica **21** (2011), 1901-1911
doi:http://dx.doi.org/10.5705/ss.2008.117

ON THE IDENTIFIABILITY OF ADDITIVE INDEX MODELS

Ming Yuan

Georgia Institute of Technology

Abstract: In this paper, we investigate the identifiability of the additive index model, also known as projection pursuit regression. Although a flexible regression tool, additive index models can be hard to interpret in practice due to a lack of identifiability. As noted by Horowitz (1998), "it is an open question whether there are identifying restrictions that yield useful forms", in reference to additive index models that differ from the known structured nonparametric regression models such as additive models and single index models. We provide an affirmative answer to this question: the additive index model is identifiable as long as the projection indices are linearly independent and there is at most one quadratic ridge function. Furthermore, we show that when there are multiple quadratic ridge functions, the identifiability can still be ensured for the non-quadratic ridge functions and their corresponding projection indices, whereas it is not possible to identify the quadratic ridge functions. Such an identifiability result enables us to check if a more restrictive nonparametric model such as the additive model can be adopted as opposed to the more general additive index model.

Key words and phrases: Additive index model, additive model, identifiability, projection pursuit regression, single index model.

1. Introduction

In an additive index model, a predictor $\mathbf{x} \in \mathbf{R}^p$ and a response y are related through $E(y|\mathbf{x}) = f(\mathbf{x})$, where

$$f(\mathbf{x}) = \mu + h_1(\alpha_1'\mathbf{x}) + h_2(\alpha_2'\mathbf{x}) + \ldots + h_M(\alpha_M'\mathbf{x}) \tag{1.1}$$

for some M, $\alpha_1, \alpha_2, \ldots, \alpha_M$, and $h_1, h_2, \ldots, h_M$. To remove trivial ambiguity, it is commonly assumed that (a) $h_j(0) = 0$ for $j = 1, \ldots, M$; (b) $\|\alpha\| = 1$; and (c) the first nonzero entry of α_j is positive for any $j = 1, \ldots, M$. The additive index model is closely related to the popular projection pursuit regression introduced by Friedman and Stuetzle (1981). Following their terminology, we refer to α_j's as projection indices and h_j's as ridge functions. Several structured nonparametric regression models are special cases of the additive index model through various restrictions on M, α's, and h's. When $M = p$ and the projection indices matrix $A = (\alpha_1, \alpha_2, \ldots, \alpha_p)$ is a permutation matrix, the additive index model becomes the additive model (Hastie and Tibshirani (1990)). If $M = 1$, (1.1) reduces to the single index model (Duan and Li (1991); Härdle, Hall and Ichimura (1993); Ichimura (1993)). Horowitz (1998) discusses an extension of the single index model that is also a special case of the additive index model; the so-called multiple index model amounts to assuming that the sets $\mathcal{A}_j = \{k : \alpha_{jk} \neq 0\}$, $j = 1, \ldots, M$, are disjoint.

It is known that the additive index model is much more flexible than these special cases. In particular, it has been be shown that any square integrable function can be approximated to arbitrary precision by a function of form (1.1) (Diaconis and Shahshahani (1984)). The flexibility, unfortunately, comes at the cost of interpretability since (1.1) is not identifiable if M, α's, and h's are left unrestricted. For example, even if the true regression function follows an additive model, the estimate obtained from the projection pursuit regression may provide little evidence to confirm it. It is therefore of great interest to seek a class of functions that retain both the flexibility of the projection pursuit and interpretability of other common structured nonparametric regression models. However, as noted by Horowitz (1998, p.14), 'it is an open question whether there are identifying restrictions that yield useful forms of (1.1) that are not single-index, multiple-index or additive models." This question motivated the study here, and we show that the additive index model is identifiable under mild conditions.

To avoid ambiguity, it is natural to assume that all projection indices in (1.1) are distinct and ridge functions are nonzero. We show that the dimensionality of the additive index model, M, as well as all the projection indices and their corresponding ridge functions are identifiable if the indices are linearly independent in that the projection indices matrix A is of column full rank, and no more than one ridge function is a polynomial of order two or less. The identifiability of the additive index model has been studied earlier by Diaconis and Shahshahani (1984), Chiou and Müller (2004), and Lin and Kulasekera (2007). Diaconis and Shahshahani (1984) focus on the case when $p = 2$ and M is known. They showed that if a projection index is not identifiable then its corresponding ridge function has to be a polynomial of order $2M - 2$ or less. They also remark on the difficulty of extending such a result to $p > 2$. The identifiability conditions provided by Chiou and Müller (2004) are much more restrictive than ours, they require the ridge functions be monotone. They also assume that the level set of f, $\{\mathbf{x} : f(\mathbf{x}) = c\}$, contains at least two points $\mathbf{x}_1$ and $\mathbf{x}_2$ such that $\alpha_j'\mathbf{x}_1 = \alpha_j'\mathbf{x}_2$ if and only if $j \neq j_0$ for some j_0. The latter condition is not intuitive and is hard to check. To illustrate the difficulty in identifying the additive index model, Lin and Kulasekera (2007) consider the special case where all ridge functions are quadratic and conjectured that it might be necessary to require

$p > (2M-1)/2 + \sqrt{(2M-1)^2 + 8M^2}/2$ to ensure identifiability. We show that such conditions cannot be sufficient; no matter how big p is, we cannot identify an additive index model with two or more quadratic ridge functions.

The main results of the paper are presented in the next section. The proof of the main theorem is relegated to Section 4. Section 3 presents an application of the identifiability result in checking additivity of the regression function.

2. Identifiability

Assume that there exist two sets of projection indices and ridge functions such that

$$\mu + h_1(\alpha_1'\mathbf{x}) + h_2(\alpha_2'\mathbf{x}) + \ldots + h_q(\alpha_q'\mathbf{x}) = \nu + g_1(\beta_1'\mathbf{x}) + g_2(\beta_2'\mathbf{x}) + \ldots + g_l(\beta_l'\mathbf{x}), \quad (2.1)$$

for all $\mathbf{x}$ in an open set of $\mathbf{R}^p$. Without loss of generality, assume that the set contains the origin. The additive index model is identifiable if (2.1) implies that

(a) the intercepts agree, $\mu = \nu$;

(b) the dimensionality agrees, $q = l$;

(c) there exists a permutation $\pi(1), \pi(2), \ldots, \pi(q)$ of $1, \ldots, q$ such that

$$\alpha_j = \beta_{\pi(j)}, \qquad g_j = h_{\pi(j)}, \quad (2.2)$$

for $j = 1, \ldots, q$.

Our main result shows that the identifiability can be ensured under mild conditions.

Theorem 1. *Assume that*

(I) *there is at most one linear ridge function; if h_j is linear, then $\alpha_j'\alpha_k = 0$ for all $k \neq j$;*

(II) *there is at most one quadratic ridge function;*

(III) *the projection indices matrix A is of column full rank.*

Then the additive index model is identifiable.

The proof of Theorem 1 is relegated to Section 4. We now examine the implications and necessity of the assumptions.

Assumption (I) is in place to deal with linear ridge functions. It is clear that the additive index model is unidentifiable when there is more than one linear ridge function, because

$$\alpha_1'\mathbf{x} + \alpha_2'\mathbf{x} = \beta_1'\mathbf{x} + \beta_2'\mathbf{x} \quad (2.3)$$

as long as there exists a 2×2 invertible matrix T such that $(\alpha_1, \alpha_2) = (\beta_1, \beta_2)T$. Furthermore, even when there is only one linear ridge function, its projection index still cannot be identified without further restrictions because

$$h_1(\alpha_1'\mathbf{x}) + b\alpha_2'\mathbf{x} = \left\{h_1(\alpha_1'\mathbf{x}) - a\alpha_1'\mathbf{x}\right\} + \|a\alpha_1 + b\alpha_2\| \cdot \left(\frac{a\alpha_1 + b\alpha_2}{\|a\alpha_1 + b\alpha_2\|}\right)'\mathbf{x} \quad (2.4)$$

for any scalars a and $b \neq 0$. Note that the aforementioned argument only suggests that the identifiability of linear ridge functions is lost. Generally the nonlinear ridge functions may still be identifiable. To this end, we assume that $\alpha_j'\alpha_k = 0$ for all $k \neq j$.

Now consider Assumption (II). Its necessity is justified by the following proposition:

Proposition 1. *If there are two quadratic ridge functions, then their corresponding projection indices are not identifiable.*

Proof. Without loss of generality, assume that $h_1(u) = au^2$ and $h_2(u) = bu^2$ for some constant $a, b \neq 0$. Let $U = a\alpha_1\alpha_1' + b\alpha_2\alpha_2'$. It follows that $h_1(\alpha_1'\mathbf{x}) + h_2(\alpha_2'\mathbf{x}) = \mathbf{x}'U\mathbf{x}$. It is not hard to check that for any $0 < c < 1$, $g_1(\beta_1'\mathbf{x}) + g_2(\beta_2'\mathbf{x}) = \mathbf{x}'U\mathbf{x}$ where

$$\beta_1 = \frac{\left(c\sqrt{a}\alpha_1 + \sqrt{1-c^2}\sqrt{b}\alpha_2\right)}{\|c\sqrt{a}\alpha_1 + \sqrt{1-c^2}\sqrt{b}\alpha_2\|};$$
$$\beta_2 = \frac{\left(\sqrt{1-c^2}\sqrt{a}\alpha_1 - c\sqrt{b}\alpha_2\right)}{\|\sqrt{1-c^2}\sqrt{a}\alpha_1 - c\sqrt{b}\alpha_2\|};$$
$$g_1(u) = \|c\sqrt{a}\alpha_1 + \sqrt{1-c^2}\sqrt{b}\alpha_2\|^2 u^2;$$
$$g_2(u) = \|\sqrt{1-c^2}\sqrt{a}\alpha_1 - c\sqrt{b}\alpha_2\|^2 u^2.$$

Therefore, α_1 and α_2 are not identifiable.

To illustrate the necessity of the requirement that the projection indices are linearly independent, consider $h_1(x_1) + h_2(x_2) + (x_1 + x_2)^2$. Because

$$h_1(x_1) + h_2(x_2) + (x_1 + x_2)^2 = \left\{h_1(x_1) + 2x_1^2\right\} + \left\{h_2(x_2) + 2x_2^2\right\} - (x_1 - x_2)^2, \quad (2.5)$$

for any functions h_1, h_2, one cannot determine whether the last projection direction is $(x_1 + x_2)/\sqrt{2}$ or $(x_1 - x_2)/\sqrt{2}$. More generally, if $h_3(u) = u^2$ and $\alpha_3 = (\alpha_1 + c\alpha_2)/\|\alpha_1 + c\alpha_2\|$, then we cannot identify the additive index model. A direct consequence of this condition is $M \leq p$. However, we do not require orthogonality among the projection indices.

The additive index model is closely related to a model that is often used for the purpose of sufficient dimension reduction (Li (2000); Hristache et al. (2001); Cook (2007)):

$$f(\mathbf{x}) = h(\alpha_1'\mathbf{x}, \ldots, \alpha_M'\mathbf{x}), \tag{2.6}$$

where $\alpha_1, \ldots, \alpha_M$ form an orthonormal basis. The projection indices of this model are generally unidentifiable and the main focus is to reveal the linear space spanned by $\alpha_1, \ldots, \alpha_M$, which is known to be possible under the so-called linearity condition (Li (2000)). Assuming an additive structure in terms of the factors allow us to identify the projection indices as well as the ridge functions in the additive index model.

When f represents a log-density function of a p-dimensional random vector $\mathbf{x}$, the additive index model leads to the so-called independent component analysis (Hyvärinen, Karhunen and Oja (2001); ICA, for short) that has recently gained considerable popularity in such fields as signal processing and image analysis. In this case, our result implies that ICA is identifiable as long as there is at most one Gaussian component. This result has been obtained earlier by Comon (1994). Our result, however, is more general in several crucial aspects. First, we do not require that $\exp(f)$ be a density function. The condition is not as vacuous as it appears to be. For example, according to Marcinkiewicz's Theorem, whenever a probability distribution has only a finite number of non-vanishing cumulants, it must be Gaussian, and that every cumulant of order greater than two vanishes. In other words, if $\exp(f)$ is a density function whose corresponding characteristic function can be written as $\exp(P(\cdot))$ for some polynomial P, then both f and P are necessarily quadratic. This property of density functions plays a critical role in proving the identifiability of ICA (Comon (1994)). Second, in ICA the number of components, M, always equals p and needs not to be identified. In contrast, in regression allowing M to be smaller than p is not only possible but of great importance from a modeling perspective because of its potential in dimension reduction.

A careful examination of the proof of Theorem 1 reveals that when Assumption (I) is violated, all nonlinear ridge functions and their projection indices can still be identified. Similarly if there are multiple quadratic ridge functions, all non-quadratic ridge functions and their corresponding projection indices remain identifiable. Furthermore, the sum of the quadratic components is also uniquely determined. The only loss of identifiability is in the individual quadratic components. Taking this into account, we can re-write (1.1) as

$$f(\mathbf{x}) = \beta_0 + \beta' x + \mathbf{x}' A\mathbf{x} + \sum_{j=1}^{M} h_j(\alpha_j'\mathbf{x}) \tag{2.7}$$

and impose side condition

$$h_j(0) = h'_j(0) = h''_j(0) = 0, \qquad j = 1, \ldots, M. \tag{2.8}$$

Following the proof of Theorem 1, we have

Theorem 2. *Under Assumptions* (I)−(III), *the alternative formulation of the additive index model* (2.7) *is identifiable under condition* (2.8) *in that, if*

$$\beta_0 + \beta' x + \mathbf{x}' A \mathbf{x} + \sum_{j=1}^{M} h_j(\alpha'_j \mathbf{x}) = \tilde{\beta}_0 + \tilde{\beta}' x + \mathbf{x}' \tilde{A} \mathbf{x} + \sum_{j=1}^{\tilde{M}} \tilde{h}_j(\tilde{\alpha}'_j \mathbf{x}), \tag{2.9}$$

then $\beta_0 = \tilde{\beta}_0$, $\beta = \tilde{\beta}$, $A = \tilde{A}$, $M = \tilde{M}$, *and there exists a permutation* $(\pi(1), \ldots, \pi(M))$ *of* $\{1, \ldots, M\}$ *such that* $h_j = \tilde{h}_{\pi(j)}$ *and* $\alpha_j = \tilde{\alpha}_{\pi(j)}$.

3. Application

The identifiability of the additive index model has important practical implications. One is in checking if a more restrictive model could be adopted. As discussed before, various nonparametric regression models such as the additive model are special cases of the additive index model. Without identifiability, making such a decision can be difficult. In the light of our results, this becomes much more feasible. One can fit a additive index model and then simply check whether or not the projection indices are the columns of an identity matrix. To illustrate such a strategy, we consider the analysis of two real world data examples.

Consider a data set taken from the 1994 Canadian Survey of Labor and Income Dynamics, for the province of Ontario. The question of interest concerns the relationship between wages and a person's age and educational background. We consider only the English-speaking males, which includes a total of 1,608 subjects. The additive model has

$$wage = h_1(age) + h_2(education) + \epsilon, \tag{3.1}$$

which is a special case of the additive index model when the projection indices form a permutation matrix. Although a powerful modeling tool in many applications, the additive model sometimes can be too restrictive. For comparison, we also fit the additive index model.

We use an estimation procedure similar to that of Chen (1991). Instead of modeling the ridge functions by regression splines as in Chen (1991), we consider the smoothing spline type estimate:

$$\hat{f} = \arg\min_{f} \frac{1}{n} [y_i - f(\mathbf{x}_i)]^2 + \sum_{j=1}^{M} \lambda_j \int (h''_j)^2. \tag{3.2}$$

The estimate $\hat{f}$ can be computed in an iterative fashion. When the projection indices are known, this becomes the usual smoothing spline estimate for additive models and can be solved as in the standard additive models (Hastie and Tibshirani (1990)); once the ridge functions h_js are updated, evaluating the projection indices becomes a nonlinear optimization problem that can be solved using the Newton-Ralphson algorithm. The tuning parameters are chosen using five-fold cross validation. Using this approach, we estimate the two projection indices to be $\alpha_1 = (0.9202, 0.3914)'$ and $\alpha_2 = (0.6274, -0.7787)'$. To assess the significance of the departure from an additive model, one hundred bootstrap samples were simulated from the additive model fit. Specifically, for each sample, we replaced the response y_i with $\hat{f}(\mathbf{x}_i) + \epsilon_i$ where $\hat{f}$ is the additive fit, $\mathbf{x}_i$ is the predictors for the ith observation, and ϵ_i was simulated from the empirical residual distribution. We ran the additive index model for each simulated data set and recorded the distance between the estimated projection direction $A = (\alpha_1, \alpha_2)$ and an identity matrix I measured by the so-called Amari metric (Hyvärinen, Karhunen and Oja (2001))

$$d(I, A) = \frac{1}{4}\left(\frac{\min\{|\alpha_{11}|, |\alpha_{12}|\}}{\max\{|\alpha_{11}|, |\alpha_{12}|\}} + \frac{\min\{|\alpha_{21}|, |\alpha_{22}|\}}{\max\{|\alpha_{21}|, |\alpha_{22}|\}} + \frac{\min\{|\alpha_{11}|, |\alpha_{21}|\}}{\max\{|\alpha_{11}|, |\alpha_{21}|\}} + \frac{\min\{|\alpha_{12}|, |\alpha_{22}|\}}{\max\{|\alpha_{12}|, |\alpha_{22}|\}}\right).$$

This metric is invariant to the permutation of the columns of A and change of signs of the columns. Here the Amari metric of the projection indices matrix estimated from the full data is 7.05, which is larger than the distances obtained from any of the 100 simulated data sets. This suggests that the additive model may not be an appropriate choice for the data set.

We now look at a second example where the additive model may be more appropriate. The diabetes data set of Hastie and Tibshirani (1990) comes from a study aimed at describing the relationship between the concentration of C-peptide and two predictor variables, Age and Base.Deficit, a measure of acidity for $n = 43$ insulin-dependent diabetes mellitus children (Sockett et al. (1987)). Hastie and Tibshirani (1990) consider the following additive model:

$$\log(C - peptide) = h_1(Age) + h_2(Base.Deficit) + \epsilon. \tag{3.3}$$

More generally, we fit the additive index model. Using the estimating procedure described above, we estimate the two projection directions as $\alpha_1 = (0.0170, -0.9999)$ and $\alpha_2 = (0.9730, -0.2308)$, respectively. The corresponding Amari metric is 0.126. To gain further insight, we again conducted boostrap to evaluate the significance of the difference between the two models. Among 100 simulated data sets from the additive model fit, 81 yielded larger Amari metrics. This suggests that the additive model may be an appropriate choice for the diabetes data.

4. Proof

4.1. Auxiliary lemmas

Our proof relies on the two auxiliary lemmas. The first is a result from Khatri and Rao (1968).

Lemma 1 (Khatri and Rao (1968)). *Consider the functional equation*

$$\phi_1(\alpha_1' t) + \ldots + \phi_r(\alpha_r' t) = \xi_1(t_1) + \ldots + \xi_p(t_p) \tag{4.1}$$

defined for $|t_i| \le R$, $i = 1, \ldots, p$, *where* $t = (t_1, \ldots, t_p)'$ *and* $\alpha_1, \ldots, \alpha_r$ *are the column vectors of a* $p \times r$ *matrix* A. *Let* A *be of column full rank such that each column has at least two non-zero entries. Then* $\phi_1, \ldots, \phi_r$ *and* $\xi_1, \ldots, \xi_p$ *are all quadratic functions.*

The second lemma concerns a characterization of linear functions that goes back to Cauchy.

Lemma 2. *If*

$$g(x) + g(y) = g(x + y) \tag{4.2}$$

holds for $x, y, x + y \in O$ *where* O *is an open set containing the origin, then* g *is linear.*

4.2. Proof of Theorem 1

We are now in position to prove Theorem 1. Setting $\mathbf{x} = 0$, we obtain $\mu = \nu$ from (2.1) which can now be rewritten as

$$h_1(\alpha_1'\mathbf{x}) + h_2(\alpha_2'\mathbf{x}) + \ldots + h_q(\alpha_q'\mathbf{x}) = g_1(\beta_1'\mathbf{x}) + g_2(\beta_2'\mathbf{x}) + \ldots + g_l(\beta_l'\mathbf{x}). \tag{4.3}$$

Without loss of generality, assume that $q \le l$. Because $\beta_1, \beta_2, \ldots, \beta_l$ are linearly independent, there exists a $p \times p$ full rank matrix B whose first l columns are $\beta_1, \beta_2, \ldots, \beta_l$. Write $\mathbf{z} = B'\mathbf{x}$ and $\Gamma = B^{-1}A$ where $A = (\alpha_1, \alpha_2, \ldots, \alpha_q)$. Then (4.3) becomes

$$h_1(\gamma_1'\mathbf{z}) + h_2(\gamma_2'\mathbf{z}) + \ldots + h_q(\gamma_q'\mathbf{z}) = g_1(z_1) + g_2(z_2) + \ldots + g_l(z_l), \tag{4.4}$$

where γ_j is the jth column of Γ. It is clear that Γ is of column full rank because B is of full rank and A is of column full rank. The rest of the proof proceeds by induction. First consider the case with $q = 1$. Then

$$h_1(\gamma_1'\mathbf{z}) = g_1(z_1) + \ldots + g_l(z_l). \tag{4.5}$$

For any $j = 1, 2, \dots, p$, setting $z_k = 0$ for all $k \neq j$ yields

$$h_1(\gamma_{1j} z_j) = g_j(z_j) \qquad j \leq l; \tag{4.6}$$
$$h_1(\gamma_{1j} z_j) = 0 \qquad j > l, \tag{4.7}$$

where $\gamma_1 = (\gamma_{11}, \gamma_{12}, \dots, \gamma_{1p})'$. If $\gamma_{1j} \neq 0$ for any $j > l$, then (4.7) implies that $h_1 = 0$ which is impossible. Hence, there is at least one $j \leq l$ such that $\gamma_{1j} \neq 0$. Without loss of generality, assume that $\gamma_{11} \neq 0$. Next we argue that $\gamma_{1j} = 0$ for any $1 < j \leq l$. Otherwise, assume that γ_{12} is also nonzero. From (4.5) and (4.6), we have

$$h_1(\gamma_{11} z_1 + \gamma_{12} z_2) = g_1(z_1) + g_2(z_2) = h_1(\gamma_{11} z_1) + h_1(\gamma_{12} z_2) \tag{4.8}$$

which, by Lemma 2, implies that h_1 is linear. Consequently g_1 and g_2 are also linear, which contradicts Assumption (I). Now that $\gamma_1 = (1, 0, \dots, 0)'$, (4.5) implies that

$$h_1(z_1) = g_1(z_1) + \dots + g_l(z_l). \tag{4.9}$$

Therefore, $l = 1 (= q)$ and $g_1 = h_1$.

Now assume that (4.4) implies that $q = l$ and $\alpha_j = \beta_j$, $h_j = g_j$ for all $j = 1, \dots, q$. Consider the functional equation

$$h_1(\gamma_1' \mathbf{z}) + h_2(\gamma_2' \mathbf{z}) + \dots + h_q(\gamma_q' \mathbf{z}) + h_{q+1}(\gamma_{q+1}' \mathbf{z}) = g_1(z_1) + g_2(z_2) + \dots + g_l(z_l), \tag{4.10}$$

where $l \geq q + 1$. Recall that $\gamma_1, \dots, \gamma_{q+1}$ are linearly independent. From Lemma 1, if in addition each γ_j has at least two nonzero entries, then all h_j's and g_j's have to be quadratic functions. By Assumption (II), this is not possible. Whence there exists at least one γ_j that has only one nonzero component. Without loss of generality, assume that γ_{q+1} has only one nonzero component; it has to be a column vector of an identity matrix I. With a permutation of coordinates, we can further take γ_{q+1} as the lth column of an identity matrix.

To invoke induction, let

$$\eta_j = \gamma_j - \gamma_{jl} \cdot \gamma_{q+1}, \qquad j = 1, \dots, q \tag{4.11}$$

and γ_{jl} be the lth entry of γ_j. Such a transformation essentially reduces the role of z_l. Because $\gamma_1, \dots, \gamma_{q+1}$ are linearly independent, so are $\eta_1, \dots, \eta_q$. Setting $z_l = 0$ in (4.10), we have

$$h_1(\eta_1' \mathbf{z}) + \dots + h_q(\eta_q' \mathbf{z}) = g_1(z_1) + \dots + g_{l-1}(z_{l-1}) \tag{4.12}$$

which, by induction, implies that $q = l - 1$ and there exists a permutation $\pi(1), \dots, \pi(l-1)$ such that $\eta_{\pi(j)} / \eta_{\pi(j),j}$ is the jth column of an identity matrix and $h_{\pi(j)}(\eta_{\pi(j),j}\cdot) = g_j(\cdot)$ for $j = 1, \ldots, l-1$. Without loss of generality, assume that $\pi(j) = j$. Then (4.10) can be written as

$$h_1(\gamma_{11}z_1 + \gamma_{1l}z_l) + h_2(\gamma_{22}z_2 + \gamma_{2l}z_l) + \ldots + h_q(\gamma_{qq}z_q + \gamma_{ql}z_l) + h_l(z_l)$$
$$= h_1(\gamma_{11}z_1) + h_2(\gamma_{22}z_2) + \ldots + h_q(\gamma_{qq}z_q) + g_l(z_l).$$

Recall that $q+1 = l$ and $\gamma_{jj} = \eta_{jj}$ for $j = 1, \ldots, q$. Setting $z_j = 0$ for all $j \neq 1, l$, yields

$$h_1(\gamma_{11}z_1 + \gamma_{1l}z_l) = h_1(\gamma_{11}z_1) + g_l^*(z_l), \tag{4.13}$$

where

$$g_l^*(z_l) = g_l(z_l) - h_2(\gamma_{2l}z_l) - \ldots - h_q(\gamma_{ql}z_l) - h_l(z_l). \tag{4.14}$$

If $\gamma_{1l} \neq 0$, this implies that both h_1 and g_l^* are linear which violates Assumption (I) that $\gamma_1'\gamma_l = \gamma_{1l} = 0$. Therefore $\gamma_1 = (1, 0, \ldots, 0)'$ and $h_1 = g_1$. Similarly, γ_j is the jth column of an identity matrix and $h_j = g_j$ for $j = 2, \ldots, q$. Finally, $h_l = g_l$ by setting $z_j = 0$ for all $i \neq l$. The proof is now completed by induction.

Acknowledgement

This research was supported in part by NSF grants DMS-MPSA-0624841 and DMS-0846234 and a grant from Georgia Cancer Coalition.

References

Chen, H. (1991). Estimation of a projection-pursuit type regression model. *Ann. Statist.* **19**, 142-157.

Chiou, J. and Müller, H. (2004). Quasi-likelihood regression with multiple indices and smooth link and variance functions. *Scand. J. Statist.* **31**, 367-386.

Comon, P. (1994). Independent component analysis, a new concept?. *Signal Processing* **36**, 287-314.

Cook, D. (2007). Fisher lecture: dimension reduction in regression. *Statist. Sci.* **22**, 1-26.

Diaconis, P. and Shahshahani, M. (1984). On nonlinear functions of linear combinations. *SIAM J. Schientific Computing* **5**, 175-191.

Duan, N. and Li, K. (1991). Slicing regression: a link-free regression method. *Ann. Statist.* **19**, 505-530.

Friedman, J. and Stuetzle, W. (1981). Projection pursuit regression. *J. Amer. Statist. Assoc.* **76**, 817-823.

Härdle, W., Hall, P. and Ichimura, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.* **21**, 157-178.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models.* Chapman and Hall, London.

Horowitz, J. (1998). *Semiparametric Methods in Econometrics.* Springer, New York.

Hristache, M., Juditsky, A., Polzehl, J. and Spokoiny, V. (2001). Structure adaptive approach for dimension reduction. *Ann. Statist.* **29**, 1537-1566.

Hyvärinen, A., Karhunen, J. and Oja, E. (2001). *Independent Component Analysis*. Wiley, New York.

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *J. Econometrics* **58**, 71-120.

Khatri, C. and Rao, C. (1968). Solutions to some functional equations and their applications to characterization of probability distributions. *Sankhyā A* **30**, 167-180.

Li, K. (2000). *Dimension Reduction and Data Visualization*. Unpublished lecture notes, available at http://www.stat.ucla.edu/~kcli/.

Lin, W. and Kulasekera, K. (2007). Identifiability of single-index models and additive-index models. *Biometrika* **94**, 496-501.

Sockett, E., Daneman, D., Clarson, C. and Ehrich, R. (1987). Factors affecting and patterns of residual insulin secretion during the first year of type I (insulin dependent) diabetes millitus in Children. *Diabetologia* **30**, 453-459.

H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, 755 Ferst Dr NW, Atlanta, GA 30332-0205. USA.

E-mail: myuan@isye.gatech.edu

(Received April 2008; accepted April 2010)