

Group selection in high-dimensional partially linear additive models

Fengrong Wei

University of West Georgia

Abstract. We consider the problem of simultaneous variable selection and estimation in partially linear additive models with a large number of grouped variables in the linear part and a large number of nonparametric components. In our problem, the number of grouped variables may be larger than the sample size, but the number of important groups is “small” relative to the sample size. We apply the adaptive group Lasso to select the important groups, using spline bases to approximate the nonparametric components and the group Lasso to obtain an initial consistent estimator. Under appropriate conditions, it is shown that, the group Lasso selects the number of groups which is comparable with the underlying important groups and is estimation consistent, the adaptive group Lasso selects the correct important groups with probability converging to one as the sample size increases and is selection consistent. The results of simulation studies show that the adaptive group Lasso procedure works well with samples of moderate size. A real example is used to illustrate the application of the proposed penalized method.

1 Introduction

Let (Y_i, X_i, T_i) , $i = 1, \dots, n$, be random vectors that are independently and identically distributed as (Y, X, T) , where Y is a response variable, $X = (X_1, X_2, \dots, X_p)'$ is a covariate vector with X_k being an $d_k \times 1$ vector corresponding to the k th group in the linear part and $T = (T_1, \dots, T_J)$ is a J -dimensional covariate vector corresponding to the nonparametric additive components. Consider the partially linear additive model with grouped variables (GPLAM)

$$Y_i = \sum_{k=1}^p X_{ik} \beta_k + \sum_{j=1}^J g_j(T_{ij}) + \varepsilon_i, \quad (1.1)$$

where X_{ik} is a $d_k \times 1$ covariate vector representing the k th group, β_k is the $d_k \times 1$ vector of corresponding regression coefficients, g_j 's are unknown functions of T_{ij} , and ε_i is an unobserved random variable with mean zero and finite variance σ^2 . In this model, the response variable Y is linearly related to covariates X , while its relation with covariates T is not specified up to any finite number of parameters.

Key words and phrases. Adaptive group Lasso, group selection, high-dimensional data, selection consistency, semiparametric regression.

Received December 2009; accepted August 2010.

This model combines the linear models with grouped variables and the flexibility of nonparametric additive models. When the relation between Y and X is of main interest and can be approximated by a linear function, it offers more interpretability than a purely nonparametric model. We proposed a penalized method for simultaneous variable selection and estimation in GPLAM when p is large. We allow the possibility that p is much larger than the sample size n , which we represent by letting p increase as n increases. But the number of important variables is small relative to the sample size n . We show that under appropriate conditions, the proposed penalized method can correctly select the important groups in the parametric part with high probability.

There has been much work on penalized methods for variable selection and estimation with high-dimensional data. Several approaches have been proposed, including the least absolute shrinkage and selection operator (Lasso, Tibshirani (1996)), the smoothly clipped absolute deviation (SCAD) penalty (Fan and Li (2001), Fan and Peng (2004)), the elastic net penalty (Zou and Hastie (2006)), and the minimum concave penalty (Zhang (2007)). Much progress has been made in understanding the statistical properties of these methods in both fixed p and $p \gg n$ settings. In particular, many authors have studied the variable selection, estimation and prediction properties of the Lasso in both low- and high-dimensional settings. See, for example, Knight and Fu (2000), Greenshtein and Ritov (2004), Meinshausen and Bühlmann (2006), Zhao and Yu (2006), van de Geer (2008) and Zhang and Huang (2008), among others. All these authors assume a linear or nonparametric model. In many applications, however, some variables are linear related to the response variable, but some variables whose effects on the response variable can be nonlinear. Then semiparametric model needs to be considered. Partially linear model (PLM) is a basic and one of the most studied semiparametric models. It is a special case of GPLAM, which has only individual variables in the linear part and only one nonparametric component g .

In PLM, for finite dimensional β , several approaches have been proposed to estimate β and g . Examples include the partial spline estimator (Wahba (1984), Engle et al. (1986) and Heckman (1986)) and the partial residual estimator (Robinson (1988), Speckman (1988) and Chen (1988)). When p is large in the sense that $p \rightarrow \infty$ as the sample size $n \rightarrow \infty$, but $p < n$, some penalized methods have been proposed to estimate β and g , see, for example, SCAD penalized estimator (Xie and Huang (2009)). They showed that under some regularity conditions, consistency in terms of variable selection and estimation can be achieved simultaneously for the linear and nonparametric component. However, all these studies did not discuss variable selection and estimation in high-dimensional setting, in the sense that $p \gg n$. Moreover, all these studies are concerned with only individual variable selection in the linear part, group structure of variables is not considered.

In this paper, we consider a partially linear additive model with a large number of grouped variables in the linear part and a large number of nonparametric components. We use the group Lasso method for variable selection in the linear

part in GPLAM based on a polynomial spline approximation of the nonparametric components. With this spline approximation, each nonparametric component is represented by a linear combination of spline basis functions. Consequently, by using partial residual method with B-spline bases, the problem of variable selection and estimation in GPLAM becomes that of selecting and estimating of grouped variables in a linear model. It is natural to apply the group Lasso method, since it is taking into account of the group structure in the model. To achieve selection consistency, we apply the group Lasso iteratively as follows. First, we use the group Lasso to obtain an initial rate consistent estimator and reduce the dimension of the problem. Then we use the adaptive group Lasso to select the final sets of grouped variables. This approach follows the idea of the adaptive Lasso (Zou (2006)) in the context of variable selection in linear regression. They considered a combination of Lasso and adaptive Lasso steps, and more generally, a multi-step Lasso procedure.

We show that the group Lasso selects the number of variables in the parametric part has the same order as the underlying model and is estimation consistent under a sparse Riesz condition which is a relatively mild condition compared with the strong irrepresentable condition (Meinshausen and Bühlmann (2006), Zhao and Yu (2006) and Wainwright (2006)). Bach (2008) proved that the group Lasso is selection consistent under the strong condition which requires the maximum of ℓ_2 norm of an off-diagonal designed matrix from the covariate matrix to be uniformly less than 1. This strong condition is quite restrictive even for moderately large group number. In this paper, we prove that by using the group Lasso result as the initial estimator for the adaptive group Lasso, the adaptive group Lasso selects the correct variables with high probability and is selection consistent. An important aspect of our results is that p can be much larger than n . And the variables in the parametric part do not necessarily to be individual variables, they can be grouped variables, categorical or continuous.

The remainder of the paper is organized as the follows. Section 2 describes the group Lasso and the adaptive group Lasso for variable selection and estimation. Section 3 presents the asymptotic properties of these methods in “large p , small n ” settings. Technical proofs of the results are given in Section 6. Section 4 presents the results of simulation studies to evaluate the finite sample performance of these methods and an illustrative application. Section 5 gives a summary and discussion.

2 Penalized estimation in GPLAM with the adaptive group Lasso penalty

In this section, we describe a two-step approach using the group Lasso for simultaneous consistent variable selection and estimation of β , then estimating the nonparametric components g_j 's using the partial residual method based on the estimation of β and a spline approximation. In the first step, we use the group Lasso

to achieve an initial estimation consistent estimator which gives a reduction of the dimension of the model. In the second step, we use the adaptive group Lasso to achieve consistent selection.

We use the polynomial splines to approximate each nonparametric component $g_j, j = 1, \dots, J$. Suppose that each T_j takes values in a compact interval $[a, b]$, where $a < b$ are finite numbers. Let $a = \xi_0 < \xi_1 < \dots < \xi_M < \xi_{M+1} = b$ be a partition of $[a, b]$ into $M + 1$ subintervals $I_{Mt} = [\xi_t, \xi_{t+1}), t = 0, \dots, M - 1$ and $I_{MM} = [\xi_M, \xi_{M+1}]$, where $M \equiv M_n = n^\nu$ with $0 < \nu < 0.5$ is a positive integer such that $h_n \triangleq \max_{1 \leq m \leq M+1} |\xi_m - \xi_{m-1}| = O(n^{-\nu})$. Let S_n be the space of polynomial splines of degree $l \geq 1$ with simple knots at the points ξ_1, \dots, ξ_M . This space consists of all functions s satisfying:

- (1) The restriction of s to any interval $I_{Mt} (0 \leq t \leq M)$ is a polynomial of degree l .
- (2) For $l \geq 2, s$ is $l - 2$ times continuously differentiable on $[a, b]$.

According to Corollary 4.10 in Schumaker (1981), there exists a normalized B-spline basis $\{B_w, 1 \leq w \leq m_n\}$ for S_n , where $m_n = M + l$ is the dimension of S_n . Thus, for any function $s \in S_n$, we can write

$$s(t) = \sum_{w=1}^{m_n} \alpha_w B_w(t).$$

We try to find the s in S_n that is close to g_j . Under reasonable smoothness assumptions, the g_j 's can be well approximated by functions in S_n . Thus, the problem of estimating g_j 's becomes that of estimating coefficient vector α .

Let $\|\mathbf{b}\|_2 \equiv (\sum_{j=1}^d |b_j|^2)^{1/2}$ denote the ℓ_2 norm of any vector $\mathbf{b} \in \mathbb{R}^d$. Let $\alpha_{nj} = (\alpha_{j1}, \dots, \alpha_{jm_n})'$ and $\alpha_n = (\alpha'_{n1}, \dots, \alpha'_{nJ})'$. Let $\omega_n = (\omega_{n1}, \dots, \omega_{np})'$ be a given vector of weights, where $0 \leq \omega_{nk} \leq \infty, 1 \leq k \leq p$. Consider the adaptive group Lasso (AGL) penalized least square criterion

$$L_n(\beta_n, \alpha_n; \lambda_n) = \sum_{i=1}^n \left[Y_i - \sum_{k=1}^p X_{ik} \beta_{nk} - \sum_{j=1}^J \sum_{w=1}^{m_n} \alpha_{jw} B_w(T_{ij}) \right]^2 + \lambda_n \sum_{k=1}^p \omega_{nk} \|\beta_{nk}\|_2,$$

where λ_n is a penalty parameter.

Let $Z_{ij} = (B_1(T_{ij}), \dots, B_{m_n}(T_{ij}))'$, so Z_{ij} consists of values of the basis functions at the i th observation for the j th nonparametric function g_j . Let $Z_j = (Z_{1j}, \dots, Z_{nj})'$ be $n \times m_n$ "design" matrix corresponding to the j th function g_j . The total "design" matrix is $Z = (Z_1, \dots, Z_J)$ which is $n \times Jm_n$ -dimensional matrix. With this notation, we can write

$$L_n(\beta_n, \alpha_n; \lambda_n) = \|Y - X\beta_n - Z\alpha_n\|_2^2 + \lambda_n \sum_{k=1}^p \omega_{nk} \|\beta_{nk}\|_2. \tag{2.1}$$

Let

$$(\hat{\beta}_n, \hat{\alpha}_n) = \arg \min_{\beta_n, \alpha_n} L_n(\beta_n, \alpha_n; \lambda_n), \tag{2.2}$$

subject to the constraint that

$$\sum_{i=1}^n \sum_{w=1}^{m_n} \alpha_{jw} B_w(T_{ij}) = 0, \quad 1 \leq j \leq J. \tag{2.3}$$

These centering constraints are to ensure unique identification of the g_j 's which are analogs of the restriction $Eg_j(T_j) = 0, 1 \leq j \leq J$. We can convert (2.2) and (2.3) to an unconstrained optimization problem by centering the response and the basis functions. For simplicity and without causing confusion, we still use (2.2) to denote our unconstrained optimization problem. Then the AGL-GPLAM estimator of β and $g_j, j = 1, \dots, J$ are $\hat{\beta}_n$ and $\hat{g}_{nj}(t) = \sum_{w=1}^{m_n} B_w(t)\hat{\alpha}_{jw}$. Unlike the basis pursuit in nonparametric regression, no penalty is imposed on the estimator of the nonparametric part since our interest lies in the variable selection with regard to the linear part.

For any β_n , the α_n that minimizes L_n necessarily satisfies

$$Z'Z\alpha_n = Z'(Y - X\beta_n). \tag{2.4}$$

Assume the number of nonparametric components is not too large and $Jm_n \leq n$, $Z'Z$ is invertible (see Lemma 1). Let $P_Z = Z(Z'Z)^{-1}Z'$ be the projection matrix of the column space of Z . Then

$$L_n(\beta_n; \lambda_n) = \|(I - P_Z)(Y - X\beta_n)\|^2 + \lambda_n \sum_{k=1}^p \omega_{nk} \|\beta_{nk}\|_2. \tag{2.5}$$

We now describe the two-step approach for variable selection and estimation in GPLAM.

Step 1. Compute the group Lasso estimator. Let

$$L_{n1}(\beta_n; \lambda_{n1}) = \|(I - P_Z)(Y - X\beta_n)\|^2 + \lambda_{n1} \sum_{k=1}^p \|\beta_{nk}\|_2.$$

This objective function is a special case of (2.5) that is obtained by setting $\omega_{nk} = 1, 1 \leq k \leq p$. The group Lasso estimator is $\tilde{\beta}_n \equiv \tilde{\beta}_n(\lambda_{n1}) = \arg \min_{\beta_n} L_{n1}(\beta_n; \lambda_{n1})$. From (2.4), we have $\tilde{\alpha}_n = (Z'Z)^{-1}Z'(Y - X\tilde{\beta}_n)$.

Step 2. Use the group Lasso estimator $\tilde{\beta}_n$ to obtain the weights by setting

$$\omega_{nk} = \begin{cases} \|\tilde{\beta}_{nk}\|_2^{-1}, & \text{if } \|\tilde{\beta}_{nk}\|_2 > 0, \\ \infty, & \text{if } \|\tilde{\beta}_{nk}\|_2 = 0. \end{cases}$$

The adaptive group Lasso objective function is

$$L_{n2}(\beta_n; \lambda_{n2}) = \|(I - P_Z)(Y - X\beta_n)\|^2 + \lambda_{n2} \sum_{k=1}^p \omega_{nk} \|\beta_{nk}\|_2.$$

Here we define $0 \cdot \infty = 0$. Thus, the variables not selected by the group Lasso are not included in Step 2. The adaptive group Lasso estimator is $\hat{\beta}_n \equiv \hat{\beta}_n(\lambda_{n2}) = \arg \min_{\beta_n} L_{n2}(\beta_n; \lambda_{n2})$, $\hat{\alpha}_n = (Z'Z)^{-1}Z'(Y - X\hat{\beta}_n)$.

Finally, the adaptive group Lasso estimator (AGL-GPLAM) of β and g_j are

$$\hat{\beta}_n, \hat{g}_{nj}(t) = \sum_{w=1}^{m_n} B_w(t)\hat{\alpha}_{jw}, \quad j = 1, \dots, J.$$

3 Asymptotic properties of the AGL-GPLAM estimator

In this section, we state the results of the asymptotic properties of the estimators defined in Steps 1 and 2 of Section 2.

In GPLAM, without loss of generality, we suppose that the first q grouped variables are nonzero which are important, that is $\|\beta_k\|_2 \neq 0, 1 \leq k \leq q$, but $\|\beta_k\|_2 \equiv 0, q + 1 \leq k \leq p$. Let $A_1 = \{1, \dots, q\}$, $A_0 = \{q + 1, \dots, p\}$. Then $\sum_{k \in A_0} \|\beta_k\|_2 = 0$. Let $|A|$ denote the cardinality of any set $A \subset \{1, \dots, p\}$,

$$\beta_A = (\beta'_k, k \in A)' \quad \text{and} \quad X_A = (X_k, k \in A).$$

Here β_A is a $\sum_{k \in A} d_k \times 1$ coefficient vector and X_A is a $n \times \sum_{k \in A} d_k$ sub-covariate matrix. Define $\|g\|_2 = [\int_a^b g^2(x) dx]^{1/2}$ for any function g , whenever the integral exists.

We assume the following conditions.

(A1) There exist absolute constants $c > 0$ and $a \in (0, 1]$ such that

$$|g_j^{(k)}(s) - g_j^{(k)}(t)| \leq c|s - t|^a \quad \text{for } s, t \in [a, b],$$

where $0 \leq k \leq l - 1$ and $E(g_j(t)) = 0$ for $j = 1, \dots, J$. Let $s_g = k + a > 0.5$.

(A2) There is a constant $b_n > 0$ such that $\min_{1 \leq k \leq q} \|\beta_{nk}\|_2 \geq b_n$.

(A3) Sparse Riesz Condition (SRC): there exist some constants $q_x^* > 0, c_* > 0$ and $c^* > 0$ where $0 < c_* < c^* < \infty$ such that

$$c_* \leq \frac{\|X_A v\|_2^2}{n\|v\|_2^2} \leq c^* \quad \forall A \text{ with } |A| = q_x^*, v \in \mathbb{R}^{\sum_{k \in A} d_k}.$$

(A4) The random variables $\varepsilon_1, \dots, \varepsilon_n$ are independent and identically distributed with $E\varepsilon_i = 0$ and $\text{Var}(\varepsilon_i) = \sigma^2$. Furthermore, their tail probability satisfy $P(|\varepsilon_i| > x) \leq K \exp(-Cx^2), i = 1, \dots, n$, for all $x \geq 0$ and for some constants C and K .

Let $\Sigma_A = X'_A(I - P_Z)X_A/n$. When $A = \{1, \dots, p\}$, we simply write $\Sigma = X'(I - P_Z)X/n$. Define

$$\gamma_{\min}(m) = \min_{|A|=m} \min_{\|v\|=1} v' \Sigma_A v, \quad \gamma_{\max}(m) = \max_{|A|=m} \max_{\|v\|=1} v' \Sigma_A v.$$

So if we take $A = \{1, \dots, p\}$, $\gamma_{\min}(p)$ and $\gamma_{\max}(p)$ are the smallest and largest eigenvalues of Σ , respectively.

3.1 Estimation consistency of the group Lasso

Let $d^* = \max_{1 \leq k \leq p} d_k$, $d_* = \min_{1 \leq k \leq p} d_k$, $d = d^*/d_*$ and $N = \sum_{k=1}^p d_k$. Let q^* be a fixed integer such that $\gamma_{n^*} = \gamma_{\min}(q^*)$, $\gamma_n^* = \gamma_{\max}(q^*)$. Define

$$\bar{\gamma} = \gamma_n^*/\gamma_{n^*} \quad \text{and} \quad M_1 = 2 + 4d\bar{\gamma}, \tag{3.1}$$

we have

$$(2 + 4\bar{\gamma})q + 1 \leq q^*. \tag{3.2}$$

Below, for any two sequences $\{a_n, b_n, n = 1, 2, \dots\}$, we write $a_n \asymp b_n$ if there are constants $0 < c_1 < c_2 < \infty$ such that $c_1 \leq a_n/b_n \leq c_2$ for all n sufficiently large, and write $a_n \asymp_p b_n$ if this inequality holds with probability converging to one.

Define

$$\lambda_{n,p} = 2\sigma \sqrt{8(1 + c_0)d^*d^2q^*\bar{\gamma}\gamma_n^*n \log N},$$

where $c_0 > 0$. Note that for fixed q^* , $\lambda_{n,p} \asymp_p \sqrt{n \log N}$. Let $\tilde{A}_1 = \{k : \|\tilde{\beta}_{nk}\|_2 \neq 0, 1 \leq k \leq p\}$ which is the set of indices of the groups selected by the group Lasso.

Theorem 3.1. *Suppose that conditions (A1) to (A4) and (3.2) hold and that $Jm_n \leq n$, $\lambda_{n1} \geq \lambda_{np}$, then*

- (i) *With probability converging to 1, $|\tilde{A}_1| \leq M_1|A_1| = M_1q$ for M_1 defined in (3.1).*
- (ii) *All the nonzero β_k with $\|\beta_k\|_2 \geq M_2\lambda_{n1}/n$ are selected with probability converging to one, where $M_2 = ([2/3 + 4d\bar{\gamma}(7 + 4\bar{\gamma})]q/(\gamma_{n^*}\gamma_n^*))^{1/2}d_*$.*

Theorem 3.1 says that the group Lasso selects the number of groups is a constant multiple of the number of underlying nonzero groups, regardless of the large number of the zero groups. The dimension of the selected model has the same order as the underlying model. Furthermore, part (ii) of Theorem 3.1 implies that all the groups with coefficients whose ℓ_2 norm are greater than the threshold given in Theorem 3.1 are selected with high probability.

Theorem 3.2. *Suppose all the conditions in Theorem 3.1 hold. Then the following assertions hold with probability converging to 1,*

$$\begin{aligned} \text{(i)} \quad & \sum_{k=1}^p \|\tilde{\beta}_{nk} - \beta_k\|_2^2 \\ & = O_p \left(\frac{qd^*\bar{\gamma} \log N}{n\gamma_{n^*}} + \frac{Jm_n^{-2s_g}}{\gamma_{n^*}} + \frac{\lambda_{n1}^2 q}{n^2\gamma_{n^*}^2} \right), \end{aligned}$$

$$\begin{aligned}
\text{(ii)} \quad & \sum_{j=1}^J \|\tilde{g}_{nj} - g_j\|_2^2 \\
& = O_p \left(Jm_n^{-2s_g} + \frac{Jm_n \log(Jm_n)}{n} + \frac{Jm_n q^2 (d^*)^2 c^* \bar{\gamma} \log N}{n\gamma_{n^*}} \right. \\
& \quad \left. + \frac{J^2 m_n^{1-2s_g} q d^*}{\gamma_{n^*}} + \frac{Jm_n q^2 d^* c^* \lambda_{n1}^2}{n^2 \gamma_{n^*}^2} \right).
\end{aligned}$$

Theorem 3.2 is stated for a general result when all the conditions (A1) to (A4) and (3.2) are satisfied. Part (i) of the theorem gives the rate of convergence of β which is determined by three terms: the stochastic error in estimating the parametric part (the first term), the spline approximation error (the second term) and the bias due to penalization (the third term). Part (ii) states the rate of convergence of the group Lasso estimator for the nonparametric components.

Immediately from Theorem 3.2, we have the following corollary.

Corollary 3.1. *Suppose all the conditions in Theorem 3.2 hold. Let $\{\sigma, c_0, \bar{\gamma}, d\}$ be fixed and $1 \leq q \leq n \leq p \rightarrow \infty$. If $\lambda_{n1} \asymp \sqrt{n \log N}$, then the following assertions hold with probability converging to 1,*

$$\begin{aligned}
\text{(i)} \quad & \sum_{k=1}^p \|\tilde{\beta}_{nk} - \beta_k\|_2^2 = O_p \left(\frac{q \log N}{n} + Jm_n^{-2s_g} \right), \\
\text{(ii)} \quad & \sum_{j=1}^J \|\tilde{g}_{nj} - g_j\|_2^2 \\
& = O_p \left(Jm_n^{-2s_g} (1 + Jm_n q) + \frac{Jm_n}{n} [\log(Jm_n) + q^2 \log N] \right).
\end{aligned}$$

Remark 3.1. If J is fixed, then the condition $Jm_n \leq n$ in Theorem 3.1 follows from $M = o(n^\nu)$ for some $\nu \in (0, 0.5)$. In particular, if $p = 0$ and J, m_n are fixed, Corollary 3.1 implies the well-known result in nonparametric regression:

$$\sum_{j=1}^J \|\hat{g}_{nj} - g_j\|_2^2 = O_p(1/n).$$

Remark 3.2. If $J = 0$, GPLAM is simplified to a linear model with grouped variables. Theorem 3.1 and Theorem 3.2 are generalization of Corollary 2.1 and Theorem 2.2 on the selection properties of the group Lasso obtain by Wei and Huang (2010). In particular, when $J = 0$,

$$\sum_{k=1}^p \|\tilde{\beta}_{nk} - \beta_{nk}\|_2^2 = O_p(q \log N/n),$$

which is the same as the result of Corollary 2.1 of [Wei and Huang \(2010\)](#).

Remark 3.3. If $d_1 = \dots = d_p = 1$, the grouped variables in GPLAM simplifies to the individual variables, and Theorem 3.1, Theorem 3.2 and Corollary 3.1 is a direct result for individual variable selection in high-dimensional partial linear additive models.

3.2 Selection consistency of the adaptive group Lasso

As shown in the Section 3.1, the first step of the group Lasso selects a model with the same order of dimension as that of the underlying model for the linear part. However, there is still a chance of irrelevant variables being selected. In order to achieve improved variable selection accuracy, we propose a second step, which we call the adaptive group Lasso, in the spirit of the adaptive Lasso ([Zou \(2006\)](#)). It adjusts the penalty on each term according to the consistent parameter estimation from the first step. In this section, we first state a general asymptotic result concerning the selection consistency of the adaptive group Lasso, under the assumption that an initial rate consistent estimator is available. We then apply to the special case when the group Lasso is used as the initial estimator.

In addition to conditions (A1) to (A4), we assume the following conditions.

(B1) The initial estimators $\tilde{\beta}_{nk}$ are consistent at zero with rate r_n if

$$\max_{k \in A_0} \|\tilde{\beta}_{nk}\|_2 = o_p(1), \quad r_n \max_{k \in A_0} \|\tilde{\beta}_{nk}\|_2 = O_p(1), \quad r_n \rightarrow \infty$$

and there exists a constant $c_b > 0$ such that

$$P\left(\min_{k \in A_1} \|\tilde{\beta}_{nk}\|_2 \geq c_b b_n\right) \rightarrow 1$$

for n sufficiently large.

(B2) Let $s_n = p - q$ be the number of zero groups. Suppose that

$$\begin{aligned} \text{(a)} \quad & \frac{(d^*)^{1/2} (\log q)^{1/2}}{n^{1/2} b_n} + \frac{\lambda_{n2} (d^*)^{3/2} q}{n b_n^2} = o(1), \\ \text{(b)} \quad & \frac{n^{1/2} d^{1/2} \log s_n}{r_n \lambda_{n2}} + \frac{(d^*)^{5/2} q^2}{r_n b_n d_*^{1/2}} = o(1). \end{aligned}$$

Theorem 3.3. *Suppose that conditions (B1), (B2) and (A1)–(A4) hold and $Jm_n \leq n$. Then*

$$P(\|\hat{\beta}_{nk}\|_2 > 0, k \in A_1 \text{ and } \|\hat{\beta}_{nk}\|_2 = 0, k \in A_0) \rightarrow 1.$$

Therefore, the adaptive group Lasso is selection consistent if an initial estimation consistent estimator is available and the conditions in Theorem 3.3 hold. Condition (B1) assumes that an initial zero consistent estimator exists. It assumes that we can consistently differentiate between important and nonimportant grouped variables. For fixed p and d_k , the ordinary least square estimator can be used

as the initial estimator. However, when $p > n$, the least squares estimator is no longer feasible. By Theorem 3.1 and Corollary 3.1, the group Lasso estimator is zero consistent with rate $\sqrt{n/(q \log N)} + m_n^{s_g}/\sqrt{J}$.

Theorem 3.4. *Suppose that all the conditions in Theorem 3.3 hold. Then the following assertions hold with probability converging to 1,*

$$\begin{aligned}
 \text{(i)} \quad & \sum_{k=1}^p \|\hat{\beta}_{nk} - \beta_k\|_2^2 = O_p\left(\frac{qd^* \bar{\gamma} \log(qd^*)}{n\gamma_{n^*}} + \frac{Jm_n^{-2s_g}}{\gamma_{n^*}} + \frac{\lambda_{n2}^2 q}{n^2 \gamma_{n^*}^2}\right), \\
 \text{(ii)} \quad & \sum_{j=1}^J \|\hat{g}_{nj} - g_j\|_2^2 \\
 & = O_p\left(Jm_n^{-2s_g} + \frac{Jm_n \log(Jm_n)}{n} + \frac{Jm_n q^2 (d^*)^2 c^* \bar{\gamma} \log(qd^*)}{n\gamma_{n^*}}\right. \\
 & \quad \left. + \frac{J^2 m_n^{1-2s_g} q d^*}{\gamma_{n^*}} + \frac{Jm_n q^2 d^* c^* \lambda_{n2}^2}{n^2 \gamma_{n^*}^2}\right).
 \end{aligned}$$

This theorem is concerned with the rate of convergence for both the parametric and nonparametric parts. Condition (B2) can be further simplified if we have d, b_n be bounded and $r_n \asymp \sqrt{n/(q \log N)} + m_n^{s_g}/\sqrt{J}$ in the initial estimator, for example, the group Lasso with $\lambda_{n1} \asymp \sqrt{n \log N}$. In this case, (B2) becomes

$$\frac{(\log q)^{1/2}}{n^{1/2}} + \frac{q\lambda_{n2}}{n} = o(1) \quad \text{and} \quad \frac{\log s_n}{r_n \lambda_{n2}} + \frac{q^2}{r_n} = o(1). \tag{3.3}$$

We often have $\lambda_{n2} = n^\tau$ for some $0 < \tau < 1/2$. In this case, the number of non-important groups can be as large as $\exp(n^{2\tau}/(q \log q))$ with the number of important groups satisfying $q^5 \log q/n \rightarrow 0$ and the numbers J, m_n being fixed. From the above discussion, we have the following corollary.

Corollary 3.2. *Let the group Lasso estimator $\tilde{\beta}_n \equiv \tilde{\beta}_n(\lambda_{n1})$ be the initial estimator in the adaptive group Lasso. Suppose that all the conditions in Theorem 3.1 hold. If $\lambda_{n2} \sim O(n^\tau)$ for some $0 < \tau < 1/2$ and satisfies (3.3), then the adaptive group Lasso consistently selects the nonzero groups in (1.1), that is, Theorem 3.3 holds. In addition,*

$$\begin{aligned}
 \text{(i)} \quad & \sum_{k=1}^p \|\hat{\beta}_{nk} - \beta_k\|_2^2 = O_p\left(\frac{q \log q}{n} + Jm_n^{-2s_g}\right), \\
 \text{(ii)} \quad & \sum_{j=1}^J \|\hat{g}_{nj} - g_j\|_2^2 \\
 & = O_p\left(Jm_n^{-2s_g} (1 + Jm_n q) + \frac{Jm_n}{n} [\log(Jm_n) + q^2 \log q]\right).
 \end{aligned}$$

This corollary follows directly from Theorems 3.3 and 3.4. It shows that the iterated group Lasso procedure that uses a combination of the group Lasso and the adaptive group Lasso is selection consistent. Moreover, the convergence rate is improved compared with that of the group Lasso by choosing appropriate penalty parameter λ_{n2} .

4 Numerical study

In this section, we use simulation to evaluate the finite sample performance of the group Lasso and the adaptive group Lasso with regard to the variable selection and estimation and use a real data example to illustrate the application of the proposed method. We also compare them with the Lasso. The Lasso estimator is defined as the value that minimizes

$$\|Y - X\beta_n - Z\alpha_n\|_2^2 + \lambda_n \sum_{k=1}^p \sum_{j=1}^{d_k} |\beta_{kj}|.$$

So the Lasso estimator does not take into account of the group structure of the variables in the linear part compared with the group Lasso and the adaptive group Lasso estimators. The penalty parameters for the group Lasso, the adaptive group Lasso and the Lasso methods are all selected by EBIC (Chen and Chen (2008)) which is defined to minimize

$$\log \text{RSS} + df \frac{\log n}{n} + \tau \cdot df \frac{\log p}{n},$$

where RSS represents residual sum of squares, df is the number of selected variables and τ is a tuning parameter which is chosen to be 0.5 according to Chen and Chen (2008).

4.1 Simulation studies

The generating model of our simulation is

$$Y_i = \sum_{k=1}^p X_{ik} \beta_k + \sum_{j=1}^J g_j(T_{ij}) + \varepsilon_i, \quad i = 1, \dots, n,$$

where $n = 100$, $p = 500$ and $J = 4$. In this simulation example, there are 500 groups and each group consists of 3 variables. To generate the covariates X , we first simulate 1,500 random variables $R_1, \dots, R_{1,500}$ independently from $N(0, 1)$. Then we generate Z_j , $j = 1, \dots, 500, 501, \dots, 504$, from a multivariate normal distribution with the mean zero and $\text{cov}(Z_{j1} - Z_{j2}) = 0.6^{|j_1 - j_2|}$. The covariates $X_1, \dots, X_{1,500}$ are generated as

$$X_{3(k-1)+j} = \frac{Z_k + R_{3(k-1)+j}}{\sqrt{2}}, \quad 1 \leq k \leq 500, 1 \leq j \leq 3.$$

Then to generate the covariates T , we let $T_1 = Z_{501}$, and

$$T_j = \rho T_{j-1} + (1 - \rho^2)^{1/2} Z_{p+j}, \quad j = 2, \dots, 4,$$

where $\rho = 0.5$. For the nonparametric additive components, we consider the following four functions,

$$\begin{aligned} g_1(t) &= \cos(2\pi t), \\ g_2(t) &= \sin(2\pi t), \\ g_3(t) &= -3(t - 0.5)^2, \\ g_4(t) &= t(1 - t) \sin\left(\frac{2\pi(1 + 2^{(9-4s)/5})}{x + 2^{(9-4s)/5}}\right), \quad s = 3. \end{aligned}$$

The response variable Y is generated from $y = \sum_{k=1}^{500} X_k \beta_k + \sum_{j=1}^4 g_j(T_j) + \varepsilon$, where the regression coefficients for covariates X are

$$\begin{aligned} \beta_1 &= (0.5, 1, 1.5), & \beta_2 &= (1, 0, -1), & \beta_3 &= (-1.5, 1, 2), \\ \beta_4 &= (1, 1, 1), & \beta_5 &= \dots = \beta_{500} = (0, 0, 0). \end{aligned}$$

The random error $\varepsilon \sim N(0, 2.84^2)$ to give a signal to noise ratio 3 : 1.

Besides $n = 100$, we also consider the cases for $n = 50$ and $n = 200$ respectively when $p = 500$ and $J = 4$. The results are given in Table 1 based on 400 replications. The columns in the table include the average number of groups (NG) being selected, model error (ER), percentage of occasions on which correct groups are included in the selected model (%IN) and percentage of occasions on which the exactly correct groups are selected (%CS) with standard error in parentheses.

Several observations can be made from Table 1. The adaptive group Lasso has higher percentage of occasions on which correct models are selected than the group

Table 1 Simulation study. NG, number of selected groups; ER, model error; IN%, percentage of occasions on which the correct variables are included in the selected model; CS%, percentage of occasions on which exactly correct variables are selected, averaged over 400 replications. Enclosed in parentheses are the corresponding standard errors

	Results for high dimension cases, $p = 500$											
	Adaptive group Lasso				Group Lasso				Lasso			
	NG	ER	IN%	CS%	NG	ER	IN%	CS%	NG	ER	IN%	CS%
$n = 200$	4.08 (0.31)	0.17 (0.10)	90 (0.28)	83 (0.40)	4.27 (0.51)	0.20 (0.10)	90 (0.28)	76 (0.43)	6.54 (3.25)	1.53 (0.87)	83 (0.38)	23 (0.42)
$n = 100$	3.70 (0.52)	2.33 (0.95)	71 (0.46)	64 (0.48)	3.94 (0.74)	2.47 (0.99)	71 (0.46)	51 (0.50)	5.28 (1.75)	2.56 (1.35)	72 (0.45)	19 (0.39)
$n = 50$	3.52 (0.69)	2.60 (1.07)	51 (0.50)	34 (0.48)	3.87 (0.95)	2.67 (1.22)	51 (0.50)	33 (0.47)	5.83 (1.47)	2.68 (1.01)	45 (0.51)	15 (0.36)

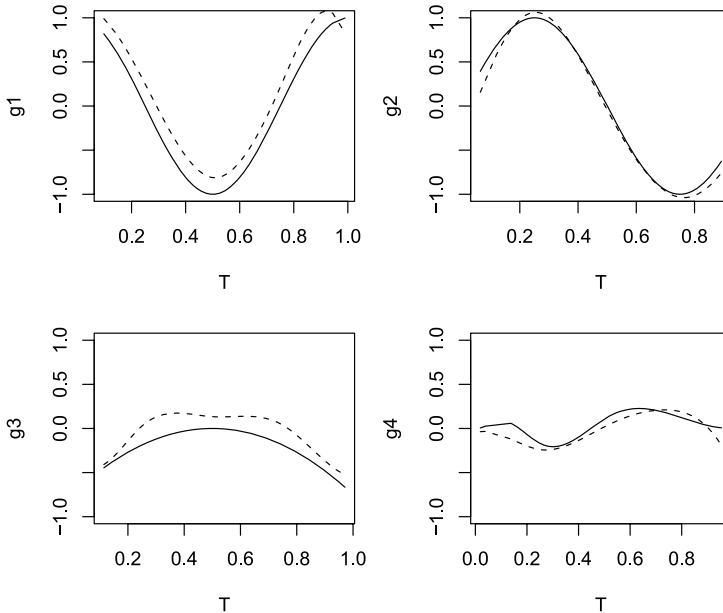


Figure 1 Adaptive group Lasso method. The estimated nonparametric components (dashed line) and true component (solid line) functions in one run when $n = 100$.

Lasso and the Lasso. The group Lasso which gives the initial estimator for the adaptive group Lasso includes the correct groups with high probability. When the sample size decreases, the performance of all the methods becomes worse. This is to be expected since selection in models with a small number of observations is more difficult. Finally, the models selected by the group Lasso and the adaptive group Lasso have similar model error to those selected by the Lasso when the sample size is small, but have higher percentage of correct selection. This shows that it is important to take into account of the group structure when we consider the problems of variable selection with grouped variables. The estimated nonparametric components are plotted along with the true function components in Figure 1. Figure 1 is the result from the adaptive group Lasso method in one run when $n = 100$.

These simulation results suggest that both the adaptive group Lasso and the group Lasso are effective for variable selection in sparse, high-dimensional partial linear additive models with grouped variables and the adaptive group Lasso can considerably improve the selection results over the group Lasso.

4.2 Real data example

We use the data set reported in Scheetz et al. (2006) to illustrate the application of the adaptive group Lasso in high-dimensional partial linear additive models. In this data set, F1 animals were intercrossed and 120 twelve-week-old male offspring

were selected for tissue harvesting from the eyes and microarray analysis. The microarray used to analyze the RNA from the eyes of these F2 animals contain over 31,042 different probe sets (Affymetric GeneChip Rate Genome 230 2.0 Array). The intensity values were normalized using the RMA (robust multi-chip averaging, Irizzary et al. (2003)) method to obtain summary expression values for each probe set. Gene expression levels were analyzed on a logarithmic scale. For the 31,042 probe sets on the array, we first excluded probes that were not expressed in the eye or that lacked sufficient variation. The definition of expressed were based on the empirical distribution of RMA normalized values. For a probe to be considered expressed, the maximum expression value observed for that probe among the 120 F2 rates were required to be greater than the 25th percentile of the entire set of RMA expression values. For a probe to be considered “sufficiently variable,” it had to exhibit at least 2-fold variation in expression level among the 120 F2 animals. A total of 18,976 probes met these two criteria.

We are interested in finding the genes whose expression are correlated with that of gene TRIM32. This gene was recently found to cause Bardet–Biedl syndrome (Chiang et al. (2006)), which is a genetically heterogeneous disease of multiple organ systems including the retina. The probe from TRIM32 is 1389163_at, which is one of the 18,976 probes that are sufficiently expressed and variate. We use the proposed penalized approach to find the probes among the remaining 18,975 probes that are most related to TRIM32. Here the sample size $n = 120$ (i.e., there are 120 arrays from 120 rats), and the number of probes is 18,975. It is expected that only a few genes are related to TRIM32. We first standardize the probes so that they have mean zero and standard deviation 1. We than do the following steps:

1. Select 5,000 probes with the largest variance;
2. Compute the marginal correlation coefficients of the 5,000 probes with the probe corresponding to TRIM32, select the top 504 covariates with the largest correlation coefficients. This is equivalent to selecting the covariates based on the univariate regression, since the covariates are standardized. We use X_1, \dots, X_{504} to represent the top 504 covariates with the correlation coefficients with the probe TRIM32 in decreasing order, that is

$$\text{corr}(X_1, \text{TRIM32}) \geq \dots \geq \text{corr}(X_{504}, \text{TRIM32}).$$

3. Let Y_i and X_{ik} , $i = 1, \dots, 120$, $1 \leq k \leq 504$, denote the gene expression values of TRIM32 and the corresponding top 504 covariates, respectively. We assume the following partial linear additive model

$$Y_i = \sum_{k=5}^{504} (X_{ik}^3 \beta_{k1} + X_{ik}^2 \beta_{k2} + X_{ik} \beta_{k3}) + \sum_{j=1}^4 g_j(X_{ij}), \quad i = 1, \dots, 120,$$

where we approximate the effect of covariates X_5, \dots, X_{504} on TRIM32 by a three order polynomials, while the effect of covariates X_1, \dots, X_4 on TRIM32 are unknown nonparametric functions.

Table 2 Probes selected by the group Lasso and the adaptive group Lasso methods where \checkmark means probes being selected

Probes	Group Lasso	Adaptive group Lasso
1383110_at	\checkmark	\checkmark
1389584_at	\checkmark	\checkmark
1383673_at	\checkmark	\checkmark
1386683_at	\checkmark	\checkmark
1379971_at	\checkmark	\checkmark
1374106_at	\checkmark	
1382517_at	\checkmark	
1393817_at	\checkmark	\checkmark
1380091_at	\checkmark	
1384466_at	\checkmark	
1391039_at	\checkmark	
1384204_at	\checkmark	\checkmark
1379597_at	\checkmark	\checkmark
1380033_at	\checkmark	\checkmark
1374131_at	\checkmark	\checkmark
1382835_at	\checkmark	\checkmark
1383996_at	\checkmark	\checkmark
1371194_at	\checkmark	\checkmark
1398594_at	\checkmark	
1368271_at	\checkmark	
1382633_at	\checkmark	\checkmark

The results are summarized in Table 2 to Table 4. Table 2 lists the probes that are selected by the group Lasso and the adaptive group Lasso including the covariates X_1, \dots, X_4 for both group Lasso and the adaptive group Lasso. Table 3 summarizes the number of selected probes and residual sum of squares for the adaptive group Lasso, the group Lasso and the Lasso. The group Lasso selects 7 more probes than the adaptive group Lasso, Lasso selects much more probes than the group Lasso and the adaptive group Lasso. To evaluate the performance of the adaptive group Lasso relative to the group Lasso and the Lasso, we use 6-fold cross validation and compare their model error and prediction error. Table 4 gives the results when the number of covariates $p = 100, 200$ and 500 , respectively. We

Table 3 No. of probes means the number of probes being selected, RSS is the residual sum of squares

	No. of probes	RSS
Adaptive group Lasso	10	0.12
Group Lasso	17	0.14
Lasso	52	0.41

Table 4 Prediction results using cross-validation. NP means the average of how many probes being selected, RSS means the average of residual sum of squares in the training set, PE is the average of prediction mean square errors for the test set with standard deviation in the parentheses

6-fold	Adaptive group Lasso			Group Lasso			Lasso		
	NP	RSS	PE	NP	RSS	PE	NP	RSS	PE
$p = 100$	11 (0.80)	0.11 (0.05)	1.63 (1.24)	19 (0.93)	0.13 (0.04)	1.71 (1.53)	48 (2.23)	0.24 (0.09)	2.18 (0.80)
$p = 200$	10 (0.70)	0.11 (0.05)	1.84 (1.66)	16 (0.86)	0.13 (0.04)	1.89 (2.01)	58 (2.31)	0.46 (1.55)	2.20 (1.65)
$p = 500$	15 (0.62)	0.11 (0.10)	1.89 (1.04)	20 (0.69)	0.14 (0.09)	1.90 (0.73)	68 (3.15)	0.55 (1.29)	2.31 (1.86)

randomly partition the data set into 6 subsets, consider 5 of them as a training set and the rest subset as a test set. Then the training set consists of 5/6 observations and the test set consists of 1/6 observations. We then follow step 3 above to fit the model with the training set, calculate the prediction mean square error for the testing set. For every partition, we repeat this process 6 times considering every subset as a testing set for one time. We then repeat this whole process 400 times, each time a new partition is made. The results in Table 4 are from 400 random partitions. In the table, NP is the average number of probes being selected, RSS is the average model error, and PE is the average prediction error with standard deviation in the parentheses. Overall, we can see that adaptive group Lasso performs better than the group Lasso and the Lasso in terms of model error and prediction error. Notably, the number of probes selected by the adaptive group Lasso and the group Lasso are fewer than that selected by the Lasso, yet the model error and prediction error are smaller. This implies that the effect of genes correlated to the gene TRIM 32 may not be linear. It is important to take into account of the nonlinearity in the regression model.

5 Concluding remarks

In this paper, we have studied the asymptotic properties of the group Lasso and the adaptive group Lasso for variable selection and estimation in partially linear additive models when p is large. Two important conditions required in our results are that the number of important groups is small and the number of nonparametric additive components J is not too big relative to the sample size n . While these conditions are often satisfied in applications, there are important settings in which they are violated. For example, in studies with microarray data as covariate measurements, the number of genes (covariates) is typically much greater than the sample size, and among those genes, a lot of them may have nonlinear effect on some disease. It is an interesting topic of future research to identify conditions under which

the AGL-GPLAM estimator achieves consistent variable selection and estimation even when $p \gg n$ and $J \gg n$.

Moreover, we have only considered the partially linear additive models which can be considered as a particular case of a generalized additive model (Hastie and Tibshirani (1986, 1990)). The adaptive group Lasso can be applied to a regular generalized additive model and other classes of semiparametric models. However, more work is needed to understand the properties of this approach in those more complicated models.

6 Proofs

We first prove the following lemmas.

Lemma 1. *Let $\Xi_Z = n^{-1}Z'Z$ and $\rho_{\min}(\Xi_Z)$, $\rho_{\max}(\Xi_Z)$ be the smallest and biggest eigenvalues of Ξ_Z . Let $m_n = O(n^\nu)$ where $0 < \nu < 0.5$ and $h \equiv h_n \asymp m_n^{-1}$. Then under conditions (A1) and (A3), with probability converging to one,*

$$c_1 h_n \leq \rho_{\min}(\Xi_Z) \leq \rho_{\max}(\Xi_Z) \leq c_2 h_n,$$

where c_1 and c_2 are two positive constants and

$$c_* \leq \gamma_{\min}(q^*) < \gamma_{\max}(q^*) \leq c^* \quad \text{for } 1 \leq m \leq q^*.$$

Proof. Let $\Xi_j = n^{-1}Z'_j Z_j$. By Lemma 6.2 of Zhou, Shen and Wolf (1998), we have

$$c_3 h \leq \rho_{\min}(\Xi_j) \leq \rho_{\max}(\Xi_j) \leq c_4 h, \quad j = 1, \dots, J.$$

Let $a = (a'_1, \dots, a'_J)'$, where $a_j \in R^{m_n}$. $Z = (Z_1, \dots, Z_J)$, by Lemma 3 of Stone (1985),

$$\|Z_1 a_1 + \dots + Z_J a_J\|_2 \geq c_5 (\|Z_1 a_1\|_2 + \dots + \|Z_J a_J\|_2)$$

for a certain constant $c_5 > 0$. By the triangle inequality,

$$c_5 (\|Z_1 a_1\|_2 + \dots + \|Z_J a_J\|_2) \leq \|Za\|_2 \leq \|Z_1 a_1\|_2 + \dots + \|Z_J a_J\|_2.$$

Therefore,

$$c_5^2 (\|Z_1 a_1\|_2^2 + \dots + \|Z_J a_J\|_2^2) \leq \|Za\|_2^2 \leq 2 (\|Z_1 a_1\|_2^2 + \dots + \|Z_J a_J\|_2^2).$$

Since $\Xi = n^{-1}Z'Z$, it follows that

$$c_5^2 (a'_1 \Xi_1 a_1 + \dots + a'_J \Xi_J a_J) \leq a' \Xi a \leq 2 (a'_1 \Xi_1 a_1 + \dots + a'_J \Xi_J a_J).$$

Therefore,

$$\frac{a'_1 \Xi_1 a_1}{\|a\|_2^2} + \dots + \frac{a'_J \Xi_J a_J}{\|a\|_2^2} \geq \rho_{\min}(\Xi_1) \frac{\|a_1\|_2^2}{\|a\|_2^2} + \dots + \rho_{\min}(\Xi_J) \frac{\|a_J\|_2^2}{\|a\|_2^2} \geq c_3 h.$$

Similarly,

$$\frac{a'_1 \Xi_1 a_1}{\|a\|_2^2} + \dots + \frac{a'_J \Xi_J a_J}{\|a\|_2^2} \leq c_4 h.$$

Thus, we have

$$c_5^2 c_3 h \leq \frac{a' \Xi a}{a' a} \leq 2c_4 h.$$

Since $P_Z = Z(Z'Z)^{-1}Z'$, $(I - P_Z)$ is idempotent, then the eigenvalues of $(I - P_Z)$ are 0 or 1 and the rank of $(I - P_Z)$ is Jm_n . Then there exists a constant q^* , such that $q^* \leq q_x^*$,

$$Jm_n \leq \sum_{k \in A} d_k \quad \text{for } |A| = q^*.$$

Then from condition (A3), we have

$$c_* \leq \gamma_{\min}(q^*) < \gamma_{\max}(q^*) \leq c^*.$$

This complete the proof of Lemma 1. □

Lemma 2. *Suppose that nonparametric function g satisfies condition (A1). If we choose $m_n = O(n^{1/(2s_g+1)})$, then*

$$\|g_n - g\|_2 = O_p(m_n^{-s_g}).$$

Lemma 2 can be obtained from Lemma 9 of Stone (1986). We omit the proof here. □

Lemma 3. *Suppose that conditions (A1), (A3) and (A4) hold. Let $T_{kj} = n^{-1/2} X'_{kj} (I - P_Z) \varepsilon_n$, $1 \leq k \leq p$, $1 \leq j \leq d_k$. Let $T_n = \max_{1 \leq k \leq p, 1 \leq j \leq d_k} T_{kj}$. Then $E(T_n) = O(\gamma_n^* \log N)$.*

Proof. Let

$$s_{nkj}^2 = n^{-1} X'_{kj} (I - P_Z) X_{kj} \quad \text{and} \quad s_n^2 = \max_{1 \leq k \leq p, 1 \leq j \leq d_k} s_{nkj}.$$

Conditional on X_{kj} 's, $T_{jk} \sim N(0, s_{nkj}^2)$. By the maximal inequality for sub-Gaussian random variables,

$$E\left(\max_{1 \leq k \leq p, 1 \leq j \leq d_k} |T_{ij}|\right) \leq C_1 s_n \sqrt{\log N}$$

for some constant $C_1 > 0$. Since $s_n \leq \sqrt{\gamma_n^*}$, then $E(T_n) = O(\sqrt{\gamma_n^* \log N})$. This completes the proof of Lemma 3. □

Lemma 4. *Suppose that conditions (A1), (A3) and (A4) hold. Let $G_j(t) = B(t)'(Z_j'Z_j)^{-1}Z_j'\varepsilon_n$ and $U_j(t) = B(t)'(Z_j'Z_j)^{-1}Z_j'X$, $j = 1, \dots, J$. Then*

$$E(G_j(t)) = O\left(\sqrt{\frac{m_n \log(Jm_n)}{n}}\right) \quad \text{and} \quad E(U_j(t)) = O(\sqrt{m_n q d^* c^*})$$

for $j = 1, \dots, J$.

Proof. Since

$$\|G_j(t)\|_2^2 = \|B(t)'(Z_j'Z_j)^{-1}Z_j'\varepsilon_n\|_2^2$$

from Lemma 6.1 of Zhou, Shen and Wolfe (1998) and Lemma 1,

$$\|G_j(t)\|_2^2 \leq C_2 m_n^{-1} \cdot n^{-2} m_n^2 \|Z_j'\varepsilon_n\|_2^2$$

for some constant $C_2 > 0$. Similar as the proof of Lemma 3, we have

$$\|G_j(t)\|_2^2 \leq C_2 m_n n^{-2} m_n n m_n^{-1} \log(Jm_n) = O\left(\frac{m_n \log(Jm_n)}{n}\right).$$

Similar as the proof for $G_j(t)$,

$$\begin{aligned} \|U_j(t)\|_2^2 &\leq C_2 m_n n^{-2} \|Z_j'X\|_2^2 \leq C_2 m_n n^{-2} m_n q \max_{1 \leq j \leq J, 1 \leq k \leq m_n, 1 \leq i \leq q} \|Z_{jk}'X_i\|_2^2 \\ &\leq C_2 m_n^2 n^{-2} q d^* \max_{1 \leq j \leq J, 1 \leq k \leq m_n, 1 \leq i \leq q, 1 \leq l \leq d_i} |Z_{jk}'X_{il}|^2 \\ &\leq C_2 m_n^2 n^{-2} q d^* n^2 m_n^{-1} c^* = O(m_n q d^* c^*). \end{aligned}$$

This completes the proof of Lemma 4. □

Proof of Theorem 3.1. The proof of part (i) essentially follows the proof of Wei and Huang (2010). The only change that must be made here is that we need to consider the spline approximation error of the regression functions g_j 's.

Specifically, let $\delta_n = \varepsilon_n + \rho_n$, where $\rho_n = (\rho_{n1}, \dots, \rho_{nn})'$ with $\rho_{ni} = \sum_{j=1}^J (g_j(T_{ij}) - g_{nj}(T_{ij}))$. Since $\|g_j - g_{nj}\|_2 = O(n^{-s_g/(2s_g+1)})$ for $m_n = n^{1/(2s_g+1)}$, we have

$$\|\rho_n\|_2 \leq C_3 \sqrt{n J^2 m_n^{-2s_g}} = C_3 J n^{1/4s_g+2}$$

for some constant $C_3 > 0$. For any integer t , let

$$\chi_m = \max_{|A|=m} \max_{\|U_{A_k}\|_2=1, 1 \leq k \leq m} \frac{\delta_n' V_A}{\|V_A\|_2} \quad \text{and} \quad \chi_m^* = \max_{|A|=m} \max_{\|U_{A_k}\|_2=1, 1 \leq k \leq m} \frac{\varepsilon_n' V_A}{\|V_A\|_2},$$

where $V_A = \mathbb{X}_A(\mathbb{X}'_A \mathbb{X}_A)^{-1} S_A - (I - P_A)\mathbb{X}\beta$, $\mathbb{X}_A = (I - P_Z)X_A$, $P_A = \mathbb{X}_A(\mathbb{X}'_A \mathbb{X}_A)^{-1} \mathbb{X}'_A$ for $|A| = q_1 = m \geq 0$, $S_A = (S'_{A_1}, \dots, S'_{A_m})'$, $S_{A_k} = \lambda \sqrt{d_{A_k}} U_{A_k}$ and $\|U_{A_k}\| = 1$.

For a sufficiently large constant $C_4 > 0$, define

$$\begin{aligned} \Omega_{m_0} &= \{(\mathbb{X}, \varepsilon_n) : \chi_m \leq \sigma C_4 \sqrt{(m \vee 1)d_* \log(pd^*)}\} & \forall m \geq m_0, \\ \Omega_{m_0}^* &= \{(\mathbb{X}, \varepsilon_n) : \chi_m^* \leq \sigma C_4 \sqrt{(m \vee 1)d_* \log(pd^*)}\} & \forall m \geq m_0, \end{aligned}$$

where $m_0 \geq 0$.

As in the proof of Theorem 1 of [Wei and Huang \(2010\)](#),

$$(\mathbb{X}, \varepsilon_n) \in \Omega_q \implies |\tilde{A}_1| \leq M_1(\lambda_{n1})q.$$

By the triangle and Cauchy–Schwarz inequalities,

$$\frac{|\delta'_n V_A|}{\|V_A\|_2} = \frac{|\varepsilon'_n V_A + \rho'_n V_A|}{\|V_A\|_2} \leq \frac{|\varepsilon'_n V_A|}{\|V_A\|_2} + \|\rho_n\|_2.$$

In the proof of Theorem 1 of [Wei and Huang \(2010\)](#), it is shown that $P((\mathbb{X}, \varepsilon_n) \in \Omega_0^*) \rightarrow 1$.

Since

$$\frac{|\rho'_n V_A|}{\|V_A\|_2} \leq \|\rho_n\|_2 \leq C_3 J n^{1/(4s_g+2)}$$

and $m_n = O(n^{1/(2s_g+1)})$, then we have for all $m \geq 0$ and n sufficiently large,

$$\|\rho_n\|_2 \leq C_1 J n^{1/2(2s_g+1)} \leq \sigma C_4 \sqrt{(m \vee 1)m_n \log(p)}.$$

It follows from that $P((\mathbb{X}, \varepsilon_n) \in \Omega_0) \rightarrow 1$.

This completes the proof of Theorem 3.1. □

Proof of Theorem 3.2. Let $\mathbb{Y} = (I - P_Z)Y$ and $\mathbb{X} = (I - P_Z)X$, then by the definition of $\tilde{\beta}_n$,

$$\|\mathbb{Y} - \mathbb{X}\tilde{\beta}_n\|_2^2 + \lambda_{n1} \sum_{k=1}^p \|\tilde{\beta}_{nk}\|_2 \leq \|\mathbb{Y} - \mathbb{X}\beta_n\|_2^2 + \lambda_{n1} \sum_{k=1}^p \|\beta_{nk}\|_2.$$

Let $A_2 = \{k : \|\beta_{nk}\|_2 \neq 0 \text{ or } \|\tilde{\beta}_{nk}\|_2 \neq 0\}$ and $q_{n2} = |A_2|$. From part (i), $q_{n2} = O_p(q)$. By the definition of A_2 ,

$$\|\mathbb{Y} - \mathbb{X}_{A_2}\tilde{\beta}_{nA_2}\|_2^2 + \lambda_{n1} \sum_{k \in A_2} \|\tilde{\beta}_{nk}\|_2 \leq \|\mathbb{Y} - \mathbb{X}_{A_2}\beta_{nA_2}\|_2^2 + \lambda_{n1} \sum_{k \in A_2} \|\beta_{nk}\|_2.$$

Let $\delta_n = \mathbb{Y} - \mathbb{X}\beta_n = Y - X\beta_n - Z\alpha_n$. Write

$$\mathbb{Y} - \mathbb{X}_{A_2}\tilde{\beta}_{nA_2} = \mathbb{Y} - \mathbb{X}\beta_n - \mathbb{X}_{A_2}(\tilde{\beta}_{nA_2} - \beta_{nA_2}) = \delta_n - \mathbb{X}_{A_2}(\tilde{\beta}_{nA_2} - \beta_{nA_2}).$$

We have

$$\|\mathbb{Y} - \mathbb{X}_{A_2}\tilde{\beta}_{nA_2}\|_2^2 = \|\mathbb{X}_{A_2}(\tilde{\beta}_{nA_2} - \beta_{nA_2})\|_2^2 - 2\delta'_n \mathbb{X}_{A_2}(\tilde{\beta}_{nA_2} - \beta_{nA_2}) + \delta'_n \delta_n.$$

We can rewrite

$$\begin{aligned} & \|\mathbb{X}_{A_2}(\tilde{\beta}_{nA_2} - \beta_{nA_2})\|_2^2 - 2\delta'_n \mathbb{X}_{A_2}(\tilde{\beta}_{nA_2} - \beta_{nA_2}) \\ & \leq \lambda_{n1} \sum_{k \in A_1} \|\beta_{nk}\|_2 - \lambda_{n1} \sum_{k \in A_1} \|\tilde{\beta}_{nk}\|_2. \end{aligned}$$

Now

$$\begin{aligned} \left| \sum_{k \in A_1} \|\beta_{nk}\|_2 - \sum_{k \in A_1} \|\tilde{\beta}_{nk}\|_2 \right| & \leq \sqrt{|A_1|} \cdot \|\tilde{\beta}_{nA_1} - \beta_{nA_1}\|_2 \\ & \leq \sqrt{|A_2|} \cdot \|\tilde{\beta}_{nA_2} - \beta_{nA_2}\|_2. \end{aligned}$$

Let $v_n = \mathbb{X}_{A_2}(\tilde{\beta}_{nA_2} - \beta_{nA_2})$,

$$\|v_n\|_2^2 - 2\delta'_n v_n \leq \lambda_{n1} \sqrt{|A_1|} \cdot \|\tilde{\beta}_{nA_2} - \beta_{nA_2}\|_2.$$

Let $\delta_n^* = \mathbb{X}_{A_2}(\mathbb{X}'_{A_2} \mathbb{X}_{A_2})^{-1} \mathbb{X}_{A_2} \delta_n$, by the Cauchy–Schwarz inequality,

$$2|\delta'_n v_n| \leq 2\|\delta_n^*\|_2 \cdot \|v_n\|_2 \leq 2\|\delta_n^*\|_2^2 + \frac{1}{2}\|v_n\|_2^2.$$

We have

$$\|v_n\|_2^2 \leq 4\|\delta_n^*\|_2^2 + 2\lambda_{n1} \sqrt{|A_1|} \cdot \|\tilde{\beta}_{nA_2} - \beta_{nA_2}\|_2.$$

$\gamma_*(q_{n2})$ be the smallest eigenvalue of $\mathbb{X}'_{A_2} \mathbb{X}_{A_2}/n$. Since $\|v_n\|_2^2 \geq n\gamma_*(q_{n2})\|\tilde{\beta}_{nA_2} - \beta_{nA_2}\|_2^2$ and $2ab \leq a^2 + b^2$,

$$\begin{aligned} n\gamma_*(q_{n2})\|\tilde{\beta}_{nA_2} - \beta_{nA_2}\|_2^2 & \leq 4\|\delta_n^*\|_2^2 + \frac{(2\lambda_{n1} \sqrt{|A_1|})^2}{2n\gamma_*(q_{n2})} \\ & \quad + \frac{1}{2}n\gamma_*(q_{n2})\|\tilde{\beta}_{nA_2} - \beta_{nA_2}\|_2^2. \end{aligned}$$

It follows that

$$\|\tilde{\beta}_{nA_2} - \beta_{nA_2}\|_2^2 \leq \frac{8\|\delta_n^*\|_2^2}{n\gamma_*(q_{n2})} + \frac{4\lambda_{n1}^2 |A_1|}{n^2 \gamma_*^2(q_{n2})}. \tag{6.1}$$

Let $g(T_i) = \sum_{j=1}^J g_j(T_{ij})$. Write

$$\delta_i = \mathbb{Y}_i - \mathbb{X}_i \beta_{nA_2} = Y_i - X_i \beta_{nA_2} - g(T_i) + g(T_i) - Z_i \alpha_n = \varepsilon_i + g(T_i) - g_n(T_i).$$

Since $\|g_j - g_{nj}\|_\infty = O(m_n^{-s_g})$ from Lemma 2, we have

$$\|\delta_n^*\|_2^2 \leq 2\|\varepsilon_n^*\|_2^2 + O(nJm_n^{-2s_g}),$$

where ε_n^* is the projection of ε_n to the span of \mathbb{X}_{A_2} . We have

$$\|\varepsilon_n^*\|_2^2 = \|(\mathbb{X}'_{A_2} \mathbb{X}_{A_2})^{-1/2} \mathbb{X}'_{A_2} \varepsilon_n\|_2^2 \leq \frac{1}{n\gamma_*(q_{n2})} \|\mathbb{X}_{A_2} \varepsilon_n\|_2^2.$$

Now

$$\begin{aligned} \max_{A:|A|\leq qn_2} \|\mathbb{X}'_A \varepsilon_n\|_2^2 &= \max_{A:|A|\leq qn_2} \sum_{k \in A} \|\mathbb{X}'_k \varepsilon_n\|_2^2 \\ &\leq qn_2 d^* n \max_{1 \leq k \leq p, 1 \leq j \leq d_k} |n^{-1/2} X'_{kj} (I - P_Z) \varepsilon_n|^2. \end{aligned}$$

By Lemma 3, we have

$$\max_{1 \leq k \leq p, 1 \leq j \leq d_k} |n^{-1/2} X'_{kj} (I - P_Z) \varepsilon_n|^2 = O_p((\gamma_n^*) \log N).$$

Then from (6.1), we have

$$\|\tilde{\beta}_{nA_2} - \beta_{nA_2}\|_2^2 = O_p\left(\frac{qd^* \bar{\gamma} \log N}{n \gamma_{n^*}}\right) + O_p\left(\frac{Jm_n^{-2s_g}}{\gamma_{n^*}}\right) + O_p\left(\frac{\lambda_{n1}^2 q}{n^2 \gamma_{n^*}^2}\right).$$

The nonparametric component g_j 's at a point $t \in [a, b]$ is estimated by

$$\tilde{g}_{nj}(t) = \sum_{w=1}^{m_n} B_w(t) \tilde{\alpha}_{jw} = B'(t) (Z'_j Z_j)^{-1} Z'_j (Y - X \tilde{\beta}_n).$$

With probability going to 1,

$$\begin{aligned} \tilde{g}_{nj}(t) - g_j(t) &= B'(t) (Z'_j Z_j)^{-1} Z'_j (Y - X \tilde{\beta}_n) - g_j(t) \\ &= B'(t) (Z'_j Z_j)^{-1} Z'_j (Y - X \beta_n) - g_j(t) \\ &\quad - B'(t) (Z'_j Z_j)^{-1} Z'_j X (\tilde{\beta}_n - \beta_n) \\ &= B'(t) (Z'_j Z_j)^{-1} Z'_j g(T) - g_j(t) \\ &\quad + B'(t) (Z'_j Z_j)^{-1} Z'_j \varepsilon_n \\ &\quad - B'(t) (Z'_j Z_j)^{-1} Z'_j X (\tilde{\beta}_n - \beta_n) \\ &\triangleq I_{n1} + I_{n2} + I_{n3}. \end{aligned}$$

Consider $\|\tilde{g}_{nj} - g_j\|_T^2 = \int [\tilde{g}_{nj}(t) - g_j(t)]^2 f_T(t) dt$. By Lemma 2,

$$\|I_{n1}\|_T^2 = O_p(m_n^{-2s_g}).$$

By Lemma 4,

$$\|I_{n2}\|_2^2 = O_p(m_n \log(Jm_n)/n).$$

$$\begin{aligned} \|I_{n3}\|_2^2 &\leq \|B'(t) (Z'_j Z_j)^{-1} Z'_j X\|_2^2 \|\tilde{\beta}_n - \beta_n\|_2^2 \\ &\leq [O_p(m_n q d^* c^*)] \left[O_p\left(\frac{qd^* \bar{\gamma} \log N}{n \gamma_{n^*}}\right) + O_p\left(\frac{Jm_n^{-2s_g}}{\gamma_{n^*}}\right) + O_p\left(\frac{\lambda_{n1}^2 q}{n^2 \gamma_{n^*}^2}\right) \right]. \end{aligned}$$

To sum up, we have

$$\begin{aligned} & \|\tilde{g}_n - g\|_2^2 \\ &= O_p\left(Jm_n^{-2s_g} + \frac{Jm_n \log(Jm_n)}{n} + \frac{Jm_n q^2 (d^*)^2 c^* \bar{\gamma} \log N}{n\gamma_{n^*}}\right. \\ & \quad \left. + \frac{J^2 m_n^{1-2s_g} q d^*}{\gamma_{n^*}} + \frac{Jm_n q^2 d^* c^* \lambda_{n1}^2}{n^2 \gamma_{n^*}^2}\right). \end{aligned}$$

This complete the proof the Theorem 3.2. □

Proofs of Theorems 3.3 and 3.4. Theorem 3.3 can be obtained directly from Theorem 3 of Wei and Huang (2010), thus we omit the proof here.

As in the proof of Theorem 3.2, we let $\delta_{n1}^* = \mathbb{X}_{A_1} (\mathbb{X}'_{A_1} \mathbb{X}_{A_1})^{-1} \mathbb{X}_{A_1} \delta$, and ε_{n1}^* be the projection of ε_n to the span of \mathbb{X}_{A_1} . We have

$$\begin{aligned} \|\delta_{n1}^*\|_2^2 &\leq 2\|\varepsilon_{n1}^*\|_2^2 + O(nJm_n^{-2s_g}), \\ \|\varepsilon_{n1}^*\|_2^2 &\leq \frac{1}{n\gamma_{n^*}(q)} \|\mathbb{X}'_{A_1} \varepsilon_n\|_2^2 = O_p\left(\frac{qd^* \gamma_{n^*} \log(\sum_{k=1}^q d_k)}{\gamma_{n^*}}\right). \end{aligned}$$

Then we get

$$\begin{aligned} \|\hat{\beta}_{nA_1} - \beta_{nA_1}\|_2^2 &= O_p\left(\frac{qd^* \bar{\gamma} \log(\sum_{k=1}^q d_k)}{n\gamma_{n^*}}\right) + O_p\left(\frac{Jm_n^{-2s_g}}{\gamma_{n^*}}\right) + O_p\left(\frac{\lambda_{n2}^2 q}{n^2 \gamma_{n^*}^2}\right), \\ \|\tilde{g}_n - g\|_T^2 &= O_p\left(Jm_n^{-2s_g} + \frac{Jm_n \log(Jm_n)}{n}\right. \\ & \quad \left. + \frac{Jm_n q^2 (d^*)^2 c^* \bar{\gamma} \log(\sum_{k=1}^q d_k)}{n\gamma_{n^*}}\right. \\ & \quad \left. + \frac{J^2 m_n^{1-2s_g} q d^*}{\gamma_{n^*}} + \frac{Jm_n q^2 d^* c^* \lambda_{n2}^2}{n^2 \gamma_{n^*}^2}\right). \end{aligned}$$

This complete the proof the Theorem 3.4. □

Acknowledgments

The author is grateful to Professor Jian Huang for sharing his insights into the problem and related topics, and wishes to thank the Editor and referees for their helpful comments.

References

- Bach, F. R. (2008). Consistency of the group lasso and multiple kernel learning. *Journal of Machine Learning Research* **9**, 1179–1225. [MR2417268](#)
- Chen, H. (1988). Convergence rates for parametric components in partly linear model. *The Annals of Statistics* **16**, 136–146. [MR0924861](#)
- Chen, J. H. and Chen, Z. H. (2008). Extended BIC for small- n -large- p sparse GLM. Technical Report 241, Dept. Statistics, Univ. British Columbia.
- Chiang, A. P., Beck, J. S., Yen, H.-J., Tayeh, M. K., Scheetz, T. E., Swiderski, R., Nishimura, D., Braun, T. A., Kim, K.-Y., Huang, J., Elbedour, K., Carmi, R., Slusarski, D. C., Casavant, T. L., Stone, E. M. and Sheffield, V. C. (2006). Homozygosity mapping with SNP arrays identifies a novel gene for Bardet–Biedl syndrome (BBS10). *Proceedings of the National Academy of Sciences of the USA*, **103**, 6287–6292.
- Engle, R. F., Granger, C. W., Rice, J. and Weiss, A. (1986). Semiparametric estimate of the relation between weather and electricity sales. *Journal of the American Statistical Association* **81**, 310–320.
- Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96**, 1348–1360. [MR1946581](#)
- Fan, J. and Peng, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *The Annals of Statistics* **32**, 928–961. [MR2065194](#)
- Greenshtein, E. and Ritov, Y. (2004). Persistence in high-dimensional linear predictor selection and the virtue of overparametrization. *Bernoulli* **10**, 971–988. [MR2108039](#)
- Hastie, T. and Tibshirani, R. (1986). Generalized additive model. *Statistical Science* **1**, 297–318. [MR0858512](#)
- Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Model*. London: Chapman & Hall/CRC. [MR1082147](#)
- Heckman, N. E. (1986). Spline smoothing in a partly linear model. *Journal of the Royal Statistical Society, Ser. B* **48**, 224–248. [MR0868002](#)
- Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U. and Speed, T. P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264.
- Knight, K. and Fu, W. J. (2000). Asymptotics for lasso-type estimators. *The Annals of Statistics* **28**, 1356–1378. [MR1805787](#)
- Meinshausen, N. and Bühlmann, P. (2006). High dimensional graphs and variable selection with the Lasso. *The Annals of Statistics* **34**, 1436–1462. [MR2278363](#)
- Robinson, P. M. (1988). Root- n -consistent semiparametric regression. *Econometrica* **56**, 931–954. [MR0951762](#)
- Scheetz, T. E., Kim, K.-Y. A., Swiderski, R. E., Philpl, A. R., Braun, T. A., Knudtson, K. L., Dorrance, A. M., DiBona, G. F., Huang, J., Casavant, T. L., Sheild, V. C. and Stone, E. M. (2006). Regulation of gene expression in the mammalian eye and its relevance to eye disease. *Proceedings of the National Academy of Sciences of the USA* **103**, 14429–14434.
- Schumaker, L. (1981). *Spline Functions: Basic Theory*. New York: Wiley. [MR0606200](#)
- Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society, Ser. B* **50**, 413–436. [MR0970977](#)
- Stone, C. J. (1985). Additive regression and other nonparametric models. *The Annals of Statistics* **13**, 689–705. [MR0790566](#)
- Stone, C. J. (1986). The dimensionality reduction principle for generalized additive models. *The Annals of Statistics* **14**, 590–606. [MR0840516](#)
- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Ser. B* **58**, 267–288. [MR1379242](#)

- van de Geer, S. (2008). High-dimensional generalized linear models and the Lasso. *The Annals of Statistics* **36**, 614–645. [MR2396809](#)
- Wahba, G. (1984). Partial spline models for the semiparametric estimation of functions of several variables. In *Analyses for Time Series, Japan–US Joint Seminar* 319–329. Tokyo: Institute of Statistical Mathematics.
- Wainwright, M. (2006). Sharp thresholds for high-dimensional and noisy recovery of sparsity. Technical report, Dept. Statistics, UC Berkeley.
- Wei, F. and Huang, J. (2010). Consistent group selection in high-dimensional linear regression. *Bernoulli* **16**, 1369–1384.
- Xie, H.-L. and Huang, J. (2009). SCAD-penalized regression in high-dimensional partially linear models. *The Annals of Statistics* **37**, 673–696.
- Zhang, C.-H. (2007). Penalized linear unbiased selection. Technical Report 2007-003, Dept. Statistics, Rutgers Univ.
- Zhang, C.-H. and Huang, J. (2008). Model-selection consistency of the LASSO in high-dimensional linear regression. *The Annals of Statistics* **36**, 1567–1594. [MR2435448](#)
- Zhao, P. and Yu, B. (2006). On model selection consistency of LASSO. *Journal of Machine Learning Research* **7**, 2541–2563. [MR2274449](#)
- Zhou, S., Shen, X. and Wolfe, D. A. (1998). Local asymptotics for regression splines and confidence regions. *The Annals of Statistics* **26**, 1760–1782. [MR1673277](#)
- Zou, H. (2006). The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association* **101**, 1418–1429. [MR2279469](#)
- Zou, H. and Hastie, T. (2006). Regularization and variable selection via the elastic net. *The Annals of Statistics* **67**, 301–320. [MR2137327](#)

Department of Mathematics
University of West Georgia
1601 Maple Street
Carrollton
Georgia 30118
USA
E-mail: fwei@westga.edu