Functional Linear Regression for Functional Response via Sparse Basis Selection

Kyunghee Han[1] and Hyejin Shin[2,*]
[1] Seoul National University, Seoul, Republic of Korea
[2] Bell Labs Seoul, Seoul, Republic of Korea
[*] Corresponding author: Hyejin Shin, e-mail: hyejin.shin@alcatel-lucent.com

Abstract

We study a sparse estimation in functional linear regression model for functional response where the regression coefficient function is generated by a finite number of basis functions. In a similar perspective to variable selection, we construct a sparse basis representation for the coefficient function using penalized least squares method. We simultaneously estimate the regression parameters and select basis functions by two-step procedure. For a given basis, we show that our approach consistently identifies true subset of basis so that the resulting estimator achieves $\sqrt{n}$-consistency for the regression parameters.

Keywords: Functional linear regression, basis selection, penalized least squares estimation, group variable selection

1. Introduction

Suppose that a functional response Y is related to a functional covariate X through

$$Y(t) = \mu_Y(t) + \int_0^1 \alpha(s,t)(X(s) - \mu_X(s))\, ds + \epsilon(t), \qquad t \in [0,1], \tag{1.1}$$

where $\mu_Y = \mathbb{E}Y$, $\mu_X = \mathbb{E}X$, α is the regression coefficient function and ϵ is a noise process. Typically, the regression coefficient function α is estimated by the least squares method through functional principal component analysis or penalized least squares method with smoothness-inducing penalty (Ramsay and Silverman, 2005; Antoch et al., 2008)

Let $\{\phi_k\}_{k\geq 1}$ and $\{\psi_m\}_{m\geq 1}$ be two basis systems in $L^2[0,1]$. Now assume that the regression coefficient function admits a sparse basis representation as follows:

$$\alpha(s,t) = \sum_{k\in K}\sum_{m\in M} a_k^m \phi_k(s)\psi_m(t), \qquad s,t \in [0,1], \tag{1.2}$$

for unknown finite index sets $K, M \subset \mathbb{N}$. Under the setting (1.2), it is natural to estimate α by selecting basis functions rather than using a classical estimation in the functional linear regression literature where the consecutive basis functions are used.

In this paper, we propose a penalized least squares (PLS) method to identify K and M and estimate the coefficients a_k^m. For this, we adopt group variable selection techniques (Yuan and Lin, 2006) in a similar manner to Obozinski et al. (2011) for high-dimensional multivariate linear regression model, but we impose 2-directional penalties on both the rows and the columns of the basis coefficient matrix of the regression coefficient function with the consecutive basis functions. The proposed method automatically selects necessary basis functions among the candidate basis system consisting of consecutive basis

[1]Supported by National Research Foundation of Korea Grant funded by the Korean Government (NRF-2011-Global Ph.D. Fellowship Program).
[2]Partially supported by the Seoul Metropolitan Government R&BD Program WR080951

functions. As discussed by James et al. (2009) and Lee and Park (2012), this makes interpretation of the relationship between Y and X much easier as well as improves the predictability of the functional linear regression model when the regression coefficient function is exactly zero over some region of domain. Also, when the ϕ_k and ψ_m are given as B-splines, basis selection for the regression coefficient function corresponds to knot selection problem. In our development, we only consider the case where the regression coefficient function is generated by a finite number of given basis functions such as B-splines or Fourier basis. For finite K and M, we show that our proposed estimator achieves $\sqrt{n}$-rate of convergence for estimating the regression coefficient function α.

The remainder of the paper is organized as follows. Section 2 introduces our method in PLS frameworks and Section 3 presents the asymptotic properties of the resulting estimator. Section 4 then provides simulation study to illustrate the finite sample performance of the proposed method. In Section 5, concluding remarks and further issues are discussed.

2. Methodology

Let X be a square integrable random function satisfying $\mathbb{E}\|X\|^2 < \infty$ with the L^2-norm $\|\cdot\|$. Assume that ϵ is a mean zero process with $\mathbb{E}\|\epsilon\|^2 = \sigma_0^2 < \infty$ and independent of X. Denote (Y_i, X_i), $1 \le i \le n$, to be random copies of (Y, X), generated by the model (1.1). Without loss of generality, basis functions are considered to have unit norms. Abusing notations, for $A \in \mathbb{R}^{|K|,|M|}$, denote $[A]_k^m = a_k^m$ to be the (k, m) element of A, and $\mathbf{a}_k = (a_k^m)_{m\in M}$ and $\mathbf{a}^m = (a_k^m)_{k\in K}$ to be real arrays according to those elements.

2.1. Penalized least squares method

To estimate the sparse coefficient function in (1.2), we consider a minimization problem for the integrated least squares criterion

$$L_n(A) = \sum_{i=1}^{n} \int_0^1 \left(Y_i(t) - \bar{Y}(t) - \sum_{k\in K} \sum_{m\in M} a_k^m \zeta_k^i \psi_m(t) \right)^2 dt$$

with respect to A, where $\zeta_k^i = \langle X_i - \bar{X}, \phi_k \rangle$. For sufficiently large n, one can verify that the minimizer $\breve{A}$ for $L_n(A)$ consists of row vectors $\{\breve{\mathbf{a}}_k\}_{k\in K}$ which are the solution of the system of equations

$$\breve{\mathbf{a}}_k = \left(\sum_{i=1}^{n} (\zeta_k^i)^2 \right)^{-1} \sum_{i=1}^{n} \zeta_k^i \left(\Sigma_M^{-1} \boldsymbol{\eta}^i - \sum_{k'\in K\setminus\{k\}} \zeta_{k'}^i \breve{\mathbf{a}}_{k'} \right), \quad k \in K, \tag{2.1}$$

where $\eta_m^i = \langle Y_i - \bar{Y}, \psi_m \rangle$ and $\Sigma_M \in \mathbb{R}^{|M|,|M|}$ with $[\Sigma_M]_m^{m'} = \langle \psi_m, \psi_{m'} \rangle$. Since K and M are unknown, however, (2.1) is impractical. We use sequences of basis functions $\{\phi_k\}_{k=1}^{J_K}$ and $\{\psi_m\}_{m=1}^{J_M}$ with non-decreasing indices J_K and J_M as $n \to \infty$, but less than n, to construct a working version of system of equations

$$\breve{\mathbf{b}}_k = \left(\sum_{i=1}^{n} (\zeta_k^i)^2 \right)^{-1} \sum_{i=1}^{n} \zeta_k^i \left(\Sigma_{J_M}^{-1} \boldsymbol{\eta}^i - \sum_{k'(\neq k)=1}^{J_K} \zeta_{k'}^i \breve{\mathbf{b}}_{k'} \right), \quad 1 \le k \le J_K. \tag{2.2}$$

This scheme makes the true model be nested in our model for sufficiently large n. Denoting the dimension of $\boldsymbol{\eta}^i$ in (2.1) by $|M|$ and in (2.2) by J_M and a working version of Σ_M by Σ_{J_M}, the joint form of estimators $\{\breve{\mathbf{b}}_k\}_{k=1}^{J_K}$ is given by

$$\breve{B} = \left(ZZ^\top \right)^{-1} ZH^\top \Sigma_{J_M}^{-1} \tag{2.3}$$

where $[\check{B}]_k^m = \check{b}_k^m$, $[Z]_k^i = \zeta_k^i$ and $[H]_m^i = \eta_m^i$ with $\check{B} \in \mathbb{R}^{J_K, J_M}$, $Z \in \mathbb{R}^{J_K, n}$ and $H \in \mathbb{R}^{J_M, n}$. We call $\check{B}$ a working version of the estimated basis coefficient matrix.

The purpose of our study is to construct the sparse representation of coefficient function (1.2) from data. We adopt penalization methods in multivariate linear regression models. In the perspective of selecting basis functions, our problem requires a PLS criterion that induces row-wise and column-wise sparsity in (2.3). Based on group variable selection techniques, we consider the following penalized integrated least squares criterion

$$Q_n(B; \lambda_1, \lambda_2) = n^{-1} L_n(B) \; + \; \sum_{k=1}^{J_K} \hat{\omega}_k \|\mathbf{b}_k\|_2 \; + \; \sum_{m=1}^{J_M} \hat{\theta}_m \|\mathbf{b}^m\|_2 \tag{2.4}$$

where $\|\cdot\|_2$ is the Euclidean norm satisfying $\|\mathbf{b}\|_2^2 = \mathbf{b}^\top \mathbf{b}$ and $(\hat{\omega}_k, \hat{\theta}_m)$ is a pair of non-negative penalty weights, possibly random.

The criterion (2.4) covers various types of PLS methods. It forms, for instance, a compounded criterion of row-wise and column-wise group lasso with fixed $(\hat{\omega}_k, \hat{\theta}_m) = (\lambda_1, \lambda_2)$. The choice of $(\hat{\omega}_k, \hat{\theta}_m) = (\lambda_1 \|\check{\mathbf{b}}_k\|_2^{-\gamma_1}, \lambda_2 \|\check{\mathbf{b}}^m\|_2^{-\gamma_1})$ with some positive $\gamma_1, \gamma_2 > 0$ is an applied version of group penalty based on adaptive lasso (Zou, 2006). Moreover, several non-convex penalties, for example the SCAD (Fan and Li, 2001) and the minimax concave penalties (Zhang, 2010) can be approximated by (2.4). To see this, one may consider a generic form of criterion

$$Q_n(B; \lambda_1, \lambda_2) = n^{-1} L_n(B) \; + \; \sum_{k=1}^{J_K} p_{\lambda_1}(\|\mathbf{b}_k\|_2) \; + \; \sum_{m=1}^{J_M} q_{\lambda_2}(\|\mathbf{b}^m\|_2) \tag{2.5}$$

for some non-negative, monotone increasing and differentiable penalty functions p and q. Note that $p_\lambda(\cdot) = \lambda^2 p(\cdot/\lambda)$ can represent the SCAD and minimax concave (MC) penalties, and (2.5) can be approximated by $\hat{\omega}_k = \lambda p'(\|\check{\mathbf{b}}_k\|_2/\lambda)$ near $\mathbf{b}_k \approx \check{\mathbf{b}}_k$. For theoretical details, see Noh and Park (2010) and Zou and Li (2008).

2.2. Two-step procedure

Technically, optimizing (2.5) with respect to $B \in \mathbb{R}^{J_K, J_M}$ may involve heavy computational challenges including 2-dimensional grid search to find an optimal pair of regularization parameters $(\lambda_1^{opt}, \lambda_2^{opt})$ under non-disjoint penalization effect caused by mixing row-wise and column-wise penalties.

To avoid an exhaustive 2-dimensional grid search, we consider a two-step procedure. In the first step, find the solution $\widetilde{B}$ that minimizes

$$\frac{1}{n} \sum_{i=1}^{n} \int_0^1 \left(Y_i(t) - \bar{Y}(t) - \sum_{k=1}^{J_K} \sum_{m=1}^{J_M} b_k^m \zeta_k^i \psi_m(t) \right)^2 dt \; + \; \sum_{k=1}^{J_K} p_{\lambda_1^*}(\|\mathbf{b}_k\|_2)$$

with suitable λ_1^*. As a result, we have an index set $\widehat{K}$ satisfying $\tilde{b}_k^m = 0$ unless $k \in \widehat{K}$ so that we select basis functions as $\{\phi_k; k \in \widehat{K}\}$. In the second step, find $\widehat{B}$ that minimizes

$$\frac{1}{n} \sum_{i=1}^{n} \int_0^1 \left(Y_i(t) - \bar{Y}(t) - \sum_{k \in \widehat{K}} \sum_{m=1}^{J_M} b_k^m \zeta_k^i \psi_m(t) \right)^2 dt \; + \; \sum_{m=1}^{J_M} q_{\lambda_2^*}(\|\mathbf{b}^m\|_2)$$

with suitable λ_2^*. Similarly, $\widehat{M}$ is given by an index set satisfying $\hat{b}_k^m = 0$ unless $m \in \widehat{M}$ so that we select basis functions as $\{\psi_m; m \in \widehat{M}\}$. We take $\widehat{B}$ as our two-step estimator of B. We use BIC type criterion with approximated degree of freedom proposed by Yuan and

Lin (2006) to choose $(\lambda_1^*, \lambda_2^*)$ in each step. Remark that our two-step procedure may not guarantee an optimal solution of minimizing (2.4). However, as derived in the following sections, the convergence rate of the second step estimator is dramatically improved by conditioning $\widehat{K}$ obtained from the first step.

3. Theory

Let (L_Z, l_Z) be a pair of the largest and smallest eigenvalues of $n^{-1}ZZ^\top$, and $\kappa_Z = L_Z/l_Z$ be the condition number of $n^{-1}ZZ^\top$. Also, let (R, r) be a pair of the largest and smallest eigenvalues of Σ_{J_M}, and $\rho = R/r$ be the condition number of Σ_{J_M}. For the weights $\{\hat{\omega}_k\}_{k=1}^{J_K}$ and $\{\hat{\theta}_m\}_{m=1}^{J_M}$, we assume that

(C1) (i) $\hat{\omega}_k = o_p(n^{-1/2})$ for all $k \in K$,
(ii) $\sup_{k \notin K}\left\{n^{-1/2}\hat{\omega}_k^{-1}J_K^{1/2}J_M^{1/2}\kappa_Z\rho\right\} = o_p(1)$,

(C2) (i) $\hat{\theta}_m = o_p(n^{-1/2})$ for all $m \in M$,
(ii) $\sup_{m \notin M}\left\{n^{-1/2}\hat{\theta}_m^{-1}J_M^{1/2}\rho\right\} = o_p(1)$.

Based on the above conditions, we have the following results.

Theorem 1. *Let $\widetilde{B}$ be the first step estimator, and $\widehat{K}$ be the corresponding row index set. Under (C1), we have that*

$$\|\tilde{\mathbf{b}}_k - \mathbf{b}_k\|_2 = O_p(n^{-1/2}J_K^{1/2}J_M^{1/2}l_Z^{-1}r^{-1}). \tag{3.1}$$

holds uniformly for $k = 1, \cdots, J_K$. Also, $P(\tilde{\mathbf{b}}_k = \mathbf{0}$ for $k \notin K) \to 1$ as $n \to \infty$.

Theorem 2. *Let $\widehat{B}$ be the second step estimator, and $\widehat{M}$ be the corresponding column index set. Under (C1) and (C2), we have that*

$$\|\hat{\mathbf{b}}^m - \mathbf{b}^m\|_2 = O_p(n^{-1/2}J_M^{1/2}r^{-1}). \tag{3.2}$$

holds uniformly for $m = 1, \cdots, J_M$. Also, $P(\hat{\mathbf{b}}^m = \mathbf{0}$ for $m \notin M) \to 1$ as $n \to \infty$.

If we take orthonormal bases $\{\phi_k\}$ and $\{\psi_m\}$, the condition (C1-(ii)) and (C2-(ii)) can be weakened by $\sup_{k \notin K}\{n^{-1/2}\hat{\omega}_k^{-1}\kappa_Z\} = o_p(1)$ and $\sup_{m \notin M}\left\{n^{-1/2}\hat{\theta}_m^{-1}\right\} = o_p(1)$. Consequently, the convergence rates in (3.1) and (3.2) are improved up to $O_p(n^{-1/2}l_Z^{-1})$ and $O_p(n^{-1/2})$, respectively.

Since (2.4) is a linear approximation of (2.5), it would be more reasonable, however, to consider sufficient conditions of penalty function that induce (C1) and (C2). In the case of $p_{\lambda_1} = \lambda_1 p$ and $q_{\lambda_2} = \lambda_2 q$, we assume that

(S1) (a) $p'(u)^{-1} = O(u^\gamma)$ $(u \to 0)$ for some $\gamma > 0$ and $n^{1/2}\lambda_1 \to 0$ $(n \to \infty)$,
(b) $n^{(\gamma+1)/2}\lambda_1 J_K^{-(\gamma+1)/2}J_M^{-(\gamma+1)/2}\kappa_Z^{-1}\rho^{-1}l_Z^\gamma r^\gamma \xrightarrow{p} \infty$,

(S2) (a) $q'(u)^{-1} = O(u^\gamma)$ $(u \to 0)$ for some $\gamma > 0$ and $n^{1/2}\lambda_2 \to 0$ $(n \to \infty)$,
(b) $n^{(\gamma+1)/2}\lambda_2 J_M^{-(\gamma+1)/2}\rho^{-1}r^\gamma \to \infty$ $(n \to \infty)$.

It can be verified that (S1) and (S2) are sufficient conditions for (C1) and (C2), respectively. In the case of orthonormal basis, (S1-(b)) and (S2-(b)) can be also reduced to $n^{(\gamma+1)/2}\lambda_1\kappa_Z^{-1}l_Z^\gamma \to 0$ and $n^{(\gamma+1)/2}\lambda_2 \to 0$. Note that, these reductions coincide to the conditions that the adaptive lasso guarantees the oracle properties in the multiple linear regression setting (Fan and Li, 2001). Similar arguments can be easily extended to the case of $p_\lambda(\cdot) = \lambda^2 p(\cdot/\lambda)$. See Lee and Park (2012).

Based on Theorem 1 and 2, it is natural to have the following corollary which implies that our two-step estimator achieves the parametric rate of convergence.

Corollary 3. *Under (C1) and (C2),* $P(\hat{b}_{km} = 0 \text{ for } (k,m) \notin K \times M) \to 1$ *as* $n \to \infty$. *Denote* $\widehat{A} \in \mathbb{R}^{|\widehat{K}|,|\widehat{M}|}$ *to be a sparse form of* $\widehat{B}$ *satisfying* $\hat{a}_k^m = \hat{b}_k^m$ *for* $(k,m) \in \widehat{K} \times \widehat{M}$. *If* $|K|$ *and* $|M|$ *are bounded, we have that*

$$\|\widehat{A} - A\|_{HS} = O_p(n^{-1/2})$$

holds with probability tending to 1, where $\|\cdot\|_{HS}$ *is the Hilbert-Schmidt norm.*

4. Numerical study

We demonstrate the performance of the proposed method by considering two scenarios of simulation settings: (i) orthogonal basis, (ii) non-orthogonal basis to construct the sparse representation of regression coefficient function. Let $\{f_j\}_{j\geq 1}$ be the normalized Fourier basis on $[0,1]$ except constant function, and denote $\{f_j^{(s)}\}$ and $\{f_j^{(c)}\}$ as follows:

$$f_j^{(s)}(u) = f_{2j}(u) = \sqrt{2}\sin(j\pi u), \qquad f_j^{(c)}(v) = f_{2j+1}(v) = \sqrt{2}\cos(j\pi v)$$

for $j = 1, 2, \ldots$. Let $\{g_j\}_{j\geq 1}$ be normalized cubic B-spline basis on $[0,1]$ with equal knots. The covariate function X is generated by $X = \sum_{j=1}^{200} Z_j f_j$ where $Z_j \sim N(0, j^{-2})$. For $K = \{4, 7, 9, 14\}$ and $M = \{3, 5, 11, 17\}$, let $A \in \mathbb{R}^{4,4}$ be an array of real coefficients whose elements are as follows:

a_k^m	(m=3)	(m=5)	(m=11)	(m=17)
(k=4)	1.5	-1	2	1
(k=7)	-1	2	1.5	-1.5
(k=9)	2	1	-1.5	1
(k=14)	-1	1.5	1	-2

For the orthogonal case, the response function Y is generated by (1.1) and (1.2) with $\mu_Y = \mu_X = 0$, $\phi_k = f_k^{(c)}$ and $\psi_m = f_m^{(s)}$. The noise process ϵ is assumed to be $\epsilon(t) \sim N(0, \sigma_0^2)$ for each $t \in [0,1]$ with $cov(\epsilon(t), \epsilon(t')) = \sigma_0^2 I(t = t')$. We take the same settings for the case of non-orthogonal basis, except $\phi_j = \psi_j = g_j$.

We summarize our simulation results in Table 1 and 2. Based on the two-step procedure described in Section 2.2, we repeat simulations 100 times to calculate true positive (TP) and false positive (FP) basis call numbers for selecting $\{\phi_k\}$ (1st step) and $\{\psi_m\}$ (2nd step), and report those prediction errors for 100 test sets. We take $\gamma = 1$ for the adaptive lasso type penalty (grAL), and fix $\gamma = 3$ for both SCAD and MC type penalties (grSCAD and grMC). One can choose J_K and J_M first in a data driven way, but we set J_K=J_M=30 for simplicity. We do not report here but similar results are obtained under J_K=J_M=50. Using the knowledge of K and M, the oracle estimator is given by the solution of (2.1).

Table 1: Performance comparison for orthogonal basis case

		TP/FP basis call (1st step)			TP/FP basis call (2nd step)			Prediction error ($\times$ 1E-02)			
n	σ_0	grAL	grSCAD	grMC	grAL	grSCAD	grMC	grAL	grSCAD	grMC	Oracle
100	0.5	3.5/0	3.7/0	2.96/0	4/0	4/0	4/0	0.223	0.235	0.422	0.051
	1	1.03/0	1.07/0	0.99/0	4/0	4/0	3.96/0	2.221	2.180	2.251	0.204
200	0.5	4/0	4/0	4/0	4/0	4/0	4/0	0.054	0.054	0.056	0.050
	1	3.11/0	3.01/0	2.96/0	4/0	4/0	4/0	0.443	0.459	0.496	0.201

In Table 1 and 2, we see that true basis functions are selected consistently in each step. Note that, identification of basis functions is dramatically improved in the second step, than the first, in the sense of calling exact basis. It can be partially accounted by the result of Theorem 2. In contrast to the orthogonal case, there are several FP basis

Table 2: Performance comparison for non-orthogonal basis case

		TP/FP basis call (1st step)			TP/FP basis call (2nd step)			Prediction error ($\times$ 1E-04)			
n	σ_0	grAL	grSCAD	grMC	grAL	grSCAD	grMC	grAL	grSCAD	grMC	Oracle
100	0.5	4/1.58	3.92/0.2	4/0.16	4/0	4/0	4/0	1.200	1.396	1.274	1.161
	1	4/1.71	3.73/1	3.92/0.51	4/0	4/0	4/0	1.482	1.689	1.719	1.392
200	0.5	4/1.49	4/0.01	4/0	4/0	4/0	4/0	1.154	1.183	1.161	1.133
	1	4/1.64	4/0.15	4/0.11	4/0	4/0	4/0	1.417	1.390	1.397	1.366

calls in the first step under non-orthogonal settings. We observe that those FP calls are included by nothing but $\{\phi_3, \phi_5, \phi_6, \phi_8, \phi_{10}, \phi_{12}, \phi_{13}, \phi_{15}\}$ and this is a set of basis functions precisely adjacent to $\{\phi_4, \phi_7, \phi_{11}, \phi_{14}\}$. It is eased, however, within decrement of σ_0 related to signal-to-noise ratio for observed values of response functions. We guess that non-orthogonality may raise noise signals to the adjacent components of the estimated coefficient matrix, since the adjacent elements of cubic B-spline basis share a part of domain region each other. Since the estimator is constructed based on inner-product scores $\{\eta^i_m\}$ and $\{\zeta^i_k\}$, orthogonal basis case may avoid this phenomenon in our settings.

5. Discussion

In this paper, we consider a two-step procedure to estimate sparse representation of coefficient function in basis expansion and show that the resulting estimator achieves $\sqrt{n}$-consistency, a parametric rate, when the cardinality of true basis for regression coefficient function is bounded. We note that prediction performances of two-step estimators become close to that of the oracle estimator within increment of n among Table 1 and 2. This results, together with Corollary 3, encourage us further developments on the oracle property as in Lee and Park (2012).

Selected references

ANTOCH, J., PRCHAL, L., DE ROSA, M. and SARDA, P. (2008). Functional linear regression with functional response: Application to prediction of electricity consumption. *Springer*.

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.

JAMES, G. M., WANG, J. and ZHU, J. (2009). Functional linear regression that's interpretable. *The Annals of Statistics*, **37** 2083–2108.

LEE, E. R. and PARK, B. U. (2012). Sparse estimation in functional linear regression. *Journal of Multivariate Analysis*, **105** 1–17.

NOH, H. S. and PARK, B. U. (2010). Sparse varying coefficient models for longitudinal data. *Statistica Sinica*, **20** 1183–1202.

OBOZINSKI, G., WAINWRIGHT, M. J. and JORDAN, M. I. (2011). Support union recovery in high-dimensional multivariate regression. *The Annals of Statistics*.

RAMSAY, J. O. and SILVERMAN, B. W. (2005). Functional data analysis, 2nd ed. *Springer*.

YUAN, M. and LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of Royal Statistical Society: Series B*, **68** 49–67.

ZHANG, C.-H. (2010). Nearly unbiased variable slection under minimax concave penalty. *The Annals of Statistics*, **38** 894–942.

ZOU, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, **101** 1418–1429.

ZOU, H. and LI, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *The Annals of Statistics*, **36** 1509–1533.