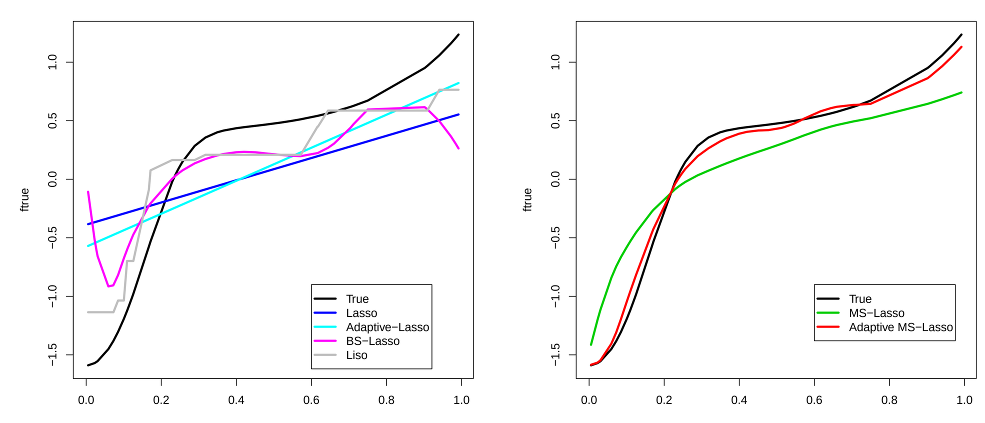
NONLINEAR MONOTONE REGRESSION FOR HIGH-DIMENSIONAL DATA

Kukatharmini Tharmaratnam, Linn Cecilie Bergersen and Ingrid K. Glad
kukathat@math.uio.no
Department of Mathematics, University of Oslo, Norway

Introduction

In recent years, high-dimensional regression problems have become one of the most active research area. The standard penalized regression methods (Lasso) limit the search to linear relationships between the covariates and the response. We focus on the special case of nonlinearity as nonlinear *monotone* effects on the response, as is often a natural assumption in medicine and biology. The additive components in the model are represented by monotone spline (I-spline) basis functions and the component selection becomes that of selecting the groups of coefficients. We use a recent procedure called Cooperative Lasso to select sign-coherent groups, that is selecting the groups with either non-negative or non-positive coefficients.



True monotone function and fitted regression curves from Lasso, Adaptive Lasso, BS-Lasso, LISO, MS-Lasso and Adaptive MS-Lasso estimators.

Regression Model

Consider the additive model,

$$y_i = \beta_0 + \sum_{j=1}^{P} g_j(x_{ij}) + \varepsilon_i, \quad i = 1, \ldots, n$$

where y_i is the response, $\boldsymbol{x}_i = (x_{i1}, ..., x_{ip})^t$ the vector of covariates, β_0 is the intercept, g_j's are unknown functions to be estimated and ϵ_i is the unobserved random error with mean 0 and variance σ^2. g_j's as a linear combination of m monotone basis functions;

$$g_j(\boldsymbol{x}_j) = \sum_{k=1}^{m} \beta_{jk} I_k^{(2)}(\boldsymbol{x}_j), \quad 1 \leq j \leq P.$$

where $I_k^{(2)}(\cdot)$ is monotone I-spline basis function of order 2 and β_{jk} is k^{th} mototone spline coefficient for the j^{th} covariate.
Monotone spline functions, I-splines are integrated version of M-splines introduced by Ramsay(1988). I-splines can be used as basis splines for regression analysis and data transformation when monotonicity is desired.

Monotone Splines Lasso (MS-Lasso)

- The Monotone splines lasso combines the idea of I-splines with the Cooperative lasso (Chiquet et al.,2012) in high-dimensional setting $(p >> n)$.
- We constitute a group in the Cooperative lasso by letting each covariate represented via an I-splines basis.
- We need more flexible than the linear lasso method but more restrictive than the general nonlinear methods. MS-Lasso is more flexible than Liso, obtaining a smooth representation of the functions.

We define P groups, each group corresponding to the basis functions and the paprameters for each covariate with equal group size m.
Let $\boldsymbol{\beta} = (\boldsymbol{\beta_1}, \ldots, \boldsymbol{\beta_P})^t \in \mathbb{R}^{Pm}$. We define $\boldsymbol{\beta}_{\mathcal{G}_j} \in \mathbb{R}^m$ as the vector $(\boldsymbol{\beta_j})_{j \in \mathcal{G}_j}$, $\boldsymbol{\beta_j} = (\beta_{j1}, \ldots, \beta_{jm})^t$

The group lasso norm for the groups $\{\mathcal{G}_j\}_{j=1}^P$ is

$$||\boldsymbol{\beta}||_{group} = \sum_{j=1}^{P} w_j ||\boldsymbol{\beta}_{\mathcal{G}_j}||.$$

where $w_j > 0$ are fixed weights for each covariate and it is used to adapt the amount of penalty for each group.

Monotone Splines Lasso (MS-Lasso)

The MS-Lasso estimates of $\boldsymbol{\beta}^{MS}$ are defined as

$$\hat{\boldsymbol{\beta}}^{MS} = \underset{\beta \in \mathbb{R}}{\operatorname{argmin}} \left\{ \frac{1}{2} ||\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\beta}||^2 + \lambda ||\boldsymbol{\beta}||_{coop} \right\},$$

where $\boldsymbol{Z}$ is $n \times (pm)$ design matrix where all covariates are represented by a centered I-spline basis. $\boldsymbol{y} = (y_1 - \bar{y}, ..., y_n - \bar{y})^t$ is a vecotor of length n and $\lambda \geq 0$ is a common tuning parameter to all groups.
The cooperative lasso norm is

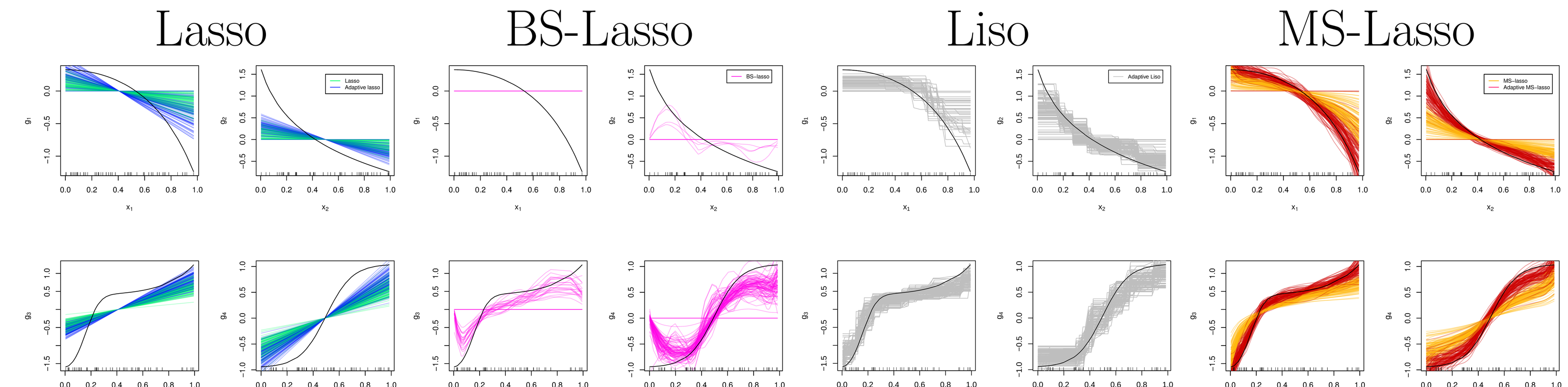
$$||\boldsymbol{\beta}||_{coop} = ||\boldsymbol{\beta}^+||_{group} + ||\boldsymbol{\beta}^-||_{group} = \sum_{j=1}^{P} w_j \left(||\boldsymbol{\beta}^+_{\mathcal{G}_j}|| + ||\boldsymbol{\beta}^-_{\mathcal{G}_j}|| \right),$$

where $\boldsymbol{\beta}^+$ and $\boldsymbol{\beta}^-$ are the componentwise positive and negative part of $\boldsymbol{\beta}$, that is, $\boldsymbol{\beta_j}^+ = \max(0, \beta_j)$ and $\boldsymbol{\beta_j}^- = \max(0, -\beta_j)$ respectively.

Simulation Results

We generate $P = 1000$ independent covariates and error ε_i with $SNR \approx 4$, the number of replication is 100 and the sample size $n = 50$. The response variable generated from $\boldsymbol{y} = g_1(x_1) + g_2(x_2) + g_3(x_3) + g_4^a(x_4) + \boldsymbol{\epsilon}$.

True and fitted curves from all simulation runs



Lasso and Adaptive Lasso The proportion of the method selects each component in the true model, together with the average number of true positives and false positives.

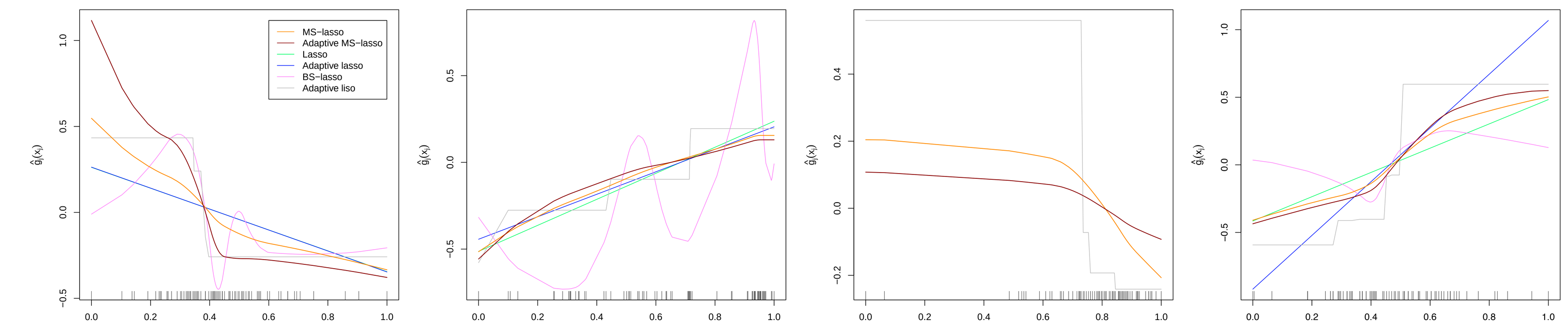
Selection Methods	g_1	g_2	g_3	g_4	TP	FP
MS-Lasso	1.00	0.89	1.00	1.00	3.89 (0.31)	17.72 (9.84)
Adaptive MS-Lasso	0.98	0.87	1.00	1.00	3.85 (0.41)	2.97 (3.05)
Lasso	0.85	0.72	1.00	1.00	3.57 (0.62)	25.01(11.85)
Adaptive Lasso	0.81	0.68	1.00	1.00	3.49 (0.67)	18.40 (7.72)
Adaptive Liso	0.39	0.98	1.00	1.00	3.37 (0.51)	5.81 (2.46)
BS-Lasso	0.00	0.04	0.23	0.93	1.20 (0.62)	1.09 (1.87)

The MS-Lasso is able to select all four components in the true model in most of the simulation runs. The adaptive step, reduces the number of false posives.

Bone Mineral Data Example

To illustrate proposed method, use Bone Mineral data (Reppe et al., 2010) and considered 84 women who had a trans-iliacal bone biopsy. Study the relationship between the bone mineral density and expression of the 2000 genes.

Estimated functions in Bone Mineral Data



References:

J. Chiquet, Y. Grandvalet, and C.Charbonnier. 2012. Sparsity with sign-coherent groups of variables via the cooperative-lasso. *Annals of Applied Statistics*, 6(2):795-830.

J. O. Ramsay. 1988. Monotone regression splines in action. *Statistical Science*, 3(4):425-461.

S. Reppe, et al. 2010. Eight genes are highly associated with BMD variation in postmenopausal Caucasian women. *Bone*, 46, 604-612.