



Integrative weighted group lasso and generalized local quadratic approximation



Qing Pan^{a,*}, Yunpeng Zhao^b

^a The Department of Statistics, George Washington University, Washington, DC, United States

^b The Department of Statistics, George Mason University, Fairfax, VA, United States

HIGHLIGHTS

- Weighted group lasso emphasizing extremum selects dynamic biomarker effects.
- Parameters in the loss, penalty and weight functions are estimated simultaneously.
- Local quadratic approximation is generalized to non-convex optimization problems.

ARTICLE INFO

Article history:

Received 13 October 2015

Received in revised form 19 April 2016

Accepted 13 June 2016

Available online 27 June 2016

Keywords:

Adaptive group lasso

Integrative group lasso

Generalized local quadratic approximation

GWAS

Optimization of non-convex function

Varying-coefficient regression

ABSTRACT

Longitudinal clinical outcomes are often collected in genomic studies, where selection methods accounting for dynamic effects of biomarkers are desirable. Biomarker effects can be modeled by nonparametric B-splines and selected by group lasso. A novel weight function is proposed based on the extremum of the biomarker effects over time for the penalty. In addition to the common practice treating weights as adaptive functions depending on some first-stage estimates, an integrative group lasso which treats the loss, penalty and weight functions as an integrative whole is proposed, where parameters in all three are jointly estimated in one step. Generalized local quadratic approximations are developed to optimize the integrative group lasso whose guidelines are applicable in a wide range of non-convex optimization problems. The integrative version has theoretical advantages as it requires weaker assumptions in achieving consistency and sparsistency. Both adaptive and integrative procedures show larger areas under the ROC curves as well as smaller biases and mean square prediction errors over unweighted group lasso in simulation studies. Finally, the proposed method is illustrated on the GWAS from the Epidemiology and Intervention of Diabetes Complication trial. To accommodate more candidate markers, 23 chromosomes are analyzed separately with common tuning parameters.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

In the presence of large number of candidate markers with inherent group structure, group lasso (Yuan and Lin, 2006) or COSSO (Lin and Zhang, 2006) are desirable for their group sparsity property. However, similar to lasso, regression coefficients estimates from group lasso are biased. By contrast, Wei et al. (2011) proved that under some regularity conditions, group lasso with adaptive weights selects the correct subset of variables with probability converging to one. Besides adaptive

* Corresponding author. Fax: +1 202 994 6517.

E-mail address: qpan@gwu.edu (Q. Pan).

weights where parameters in the weight function are estimated in the first stage and plugged into the penalized log-likelihood in the second stage, we propose integrative weights, which are functions of the unknown parameter values. In the adaptive version, the regression coefficient estimates depend on the accuracy of the estimates in the first stage. In the integrative group lasso, the parameters in the weight function are set to be unknown, and hence the loss function and the weighted penalty function are minimized jointly in one step as a unified whole. Such a weight function does not require any prior information on or first-stage estimation of the parameters. As a consequence, the integrative group lasso has theoretical advantages—it achieves sparsistency under milder conditions. However, neither traditional local quadratic approximation (Fan and Li, 2001) nor group LARS (Yuan and Lin, 2006) can solve the integrative group lasso directly due to the unknown parameters in the weights. We develop a generalized local quadratic approximation (GLQA) that gives a convex quadratic approximation of any penalty function and is guaranteed to converge in combination with back-tracking line search (Conway, 2004). The three guidelines of GLQA, which were not found in the literature by our knowledge, can be borrowed into other non-convex optimization problems.

The comparison of the adaptive and integrative weights also provides new insights into the adaptive lasso (Zou, 2006) and the smoothly clipped absolute deviation (SCAD) (Fan and Li, 2001) methods. Although appearing to be different, adaptive lasso and SCAD both modify the lasso penalty to achieve asymptotic consistency in both selection and estimation. The difference lays in the form of their penalty functions—adaptive weights depend on the first-stage estimators while SCAD penalty is a function of the unknown parameters. The same difference lays between the adaptive group lasso and our integrative group lasso. The adaptive and integrative group lasso procedures modify group lasso following the spirits of adaptive lasso and SCAD, respectively, in the following sense: Adaptive group lasso and adaptive lasso share the same essential properties; while both the integrative group lasso and SCAD construct penalties which are complicated functions of the unknown parameters, resulting in selection and estimation procedures with desirable properties such as unbiasedness and sparsity.

Our study is motivated by a genome wide association study (GWAS) with longitudinal clinical outcomes, where the effects of SNPs are presumably time-varying. Various nonparametric splines have been widely employed in literature to model time-varying effects of biomarkers. Take a few examples: Fan et al. (2012) used linear penalized splines to model temporal trends of SNP effects on longitudinal quantitative traits in analyzing the Framingham Heart Study Genetic Analysis Workshop data. Wang et al. (2007) also modeled the effects of transcription factors as B-splines. Chen and Zhang (2008) used multivariate adaptive splines to describe the relationship between the presence of regulatory motifs and gene expression. Chen and Wang (2011) estimated functional mixed effects models in which both the random and fixed effects are modeled as P-splines. Yan and Huang (2002) extended the adaptive group lasso procedure to Cox proportional hazards models and selected variables by maximizing penalized partial likelihood where the two penalties represented time-invariant effects and time-varying effects, respectively.

By using the adaptive and integrative lasso, we propose novel variable selection and coefficient estimation procedures for time-varying effects with an emphasis on effects with large extremum. The effects of the markers over time are modeled as cubic B-splines. And all the coefficients in constructing the same spline are viewed as a group, and selected or unselected jointly by group lasso with adaptive or integrative weights. Furthermore, we design novel weight functions based on the largest absolute values of the spline coefficients in the group. The weights based on extremum are motivated from, but not limited to, the following biological scenarios—when the true marker effects are small values around zero, they are most likely biological fluctuations without serious disease consequences. In such cases, the proposed method will assign large penalties, leading to lower chances of selection. On the contrary, we target at markers with large effects for at least a period of time, which pass the threshold to trigger disease onset or progression. In summary, the proposed method prefers markers with large temporary effects for at least a period of time over markers with consistently small fluctuations.

The rest of the article is organized as follows. The adaptive and integrative versions of the weighted group lasso are introduced in Section 2. Section 3 describes the computation algorithms for both versions. Asymptotic selection consistency and estimation consistency of the parameters are derived in Section 4, with emphasis on the difference between the conditions required by adaptive and integrative weights. Section 5 examines their performances and compares them to the unweighted group lasso through simulation studies. Finally, the motivating data are analyzed by both procedures in Section 6.

2. Weighted group lasso based on extremum

First we introduce the notation. Define $Y = [y_{ik}]$ as an $n \times T$ matrix, where y_{ik} is the outcome for the i th subject at the k th time point. Outcomes are measured at t_1, \dots, t_T . Let $X = [x_{ip}]$ be an $n \times (P + 1)$ matrix. The i th row X_i denotes the covariate vector for the i th subject, where $x_{i0} = 1$ corresponds to the intercept and x_{i1}, \dots, x_{iP} represent candidate biomarkers. Note that the index p starts from 0. We assume $(y_{i1}, \dots, y_{iT}, x_{i1}, \dots, x_{iP})$ are independently and identically distributed for all i . The relationship between Y and X is modeled as,

$$y_{ik} = \sum_{p=0}^P x_{ip} \beta_p(t_k) + \epsilon_{ik}, \quad \epsilon_{ik} \text{ i.i.d. } \sim N(0, \sigma^2),$$

where $\beta_0(t), \dots, \beta_p(t)$ are time-varying regression coefficients. We approximate the time-varying coefficients by cubic B-spline expansions. Let $\mathcal{B}_1(t), \dots, \mathcal{B}_j(t), \dots, \mathcal{B}_J(t)$ be the cubic B-spline basis functions with $J - 4$ uniform internal knots covering the time range. Then $\beta_p(t) = \sum_{j=1}^J \theta_{pj} \mathcal{B}_j(t)$, where $\Theta = [\theta_{pj}]$ is a $(P + 1) \times J$ coefficient matrix.

The sum of squared errors (SSE) is calculated as

$$\text{SSE} = \sum_{i=1}^n \sum_{k=1}^T \left(y_{ik} - \sum_{p=0}^P x_{ip} \beta_p(t_k) \right)^2.$$

Let $B = [b_{jk}]$ be an $J \times T$ matrix where $b_{jk} = \mathcal{B}_j(t_k)$. Then SSE can be written in the following matrix form, $\text{SSE} = \|Y - X\Theta B\|_F^2$, where $\|\cdot\|_F$ indicates the Frobenius norm. In the following discussion, matrices are converted to vectors by column. That is, for an $m \times n$ matrix M , $\text{vec}(M) = [M_{11}, \dots, M_{m1}, \dots, M_{1n}, \dots, M_{mn}]'$. Define $\Theta = \text{vec}(\Theta')$, $\mathbf{Y} = \text{vec}(Y')$, and $\mathbb{X} = X \otimes B'$ where \otimes denotes the Kronecker product. With some simple algebra, we obtain $\text{SSE} = \|\mathbf{Y} - \mathbb{X}\Theta\|^2$.

We adopt the regularization framework to estimate Θ . In general, the objective function for estimating Θ could be written as

$$\arg \min_{\Theta} \left\{ \text{SSE}(\Theta) + \lambda \left(\sum_{p=1}^P \text{Pen}(\Theta_{p,\cdot}) \right) \right\},$$

where $\text{Pen}(\Theta_{p,\cdot})$ indicates a penalty function for the p th coefficient vector $\Theta_{p,\cdot}$. Note that the row index p goes from 1 to P , corresponding to the P covariates but not the coefficients Θ_0 for the intercept function. Therefore, the estimate of the intercept function $\beta_0(t)$ will not shrink. A natural choice of the penalty function is the group lasso penalty (Yuan and Lin, 2006),

$$\arg \min_{\Theta} \left\{ \text{SSE}(\Theta) + \lambda \left(\sum_{p=1}^P \|\Theta_{p,\cdot}\| \right) \right\},$$

where $\|\cdot\|$ denotes the l_2 norm. All the coefficients in constructing the same spline are viewed as a group and estimated to be zero or non-zero simultaneously. Similarly to the lasso (Zou, 2006), a drawback to the group lasso is that it introduces bias in estimating non-zero coefficients. To correct the bias and hence achieve both consistency and sparsistency, we introduce a weighted group lasso with novel weight functions based on the extremum of each coefficient group. For an arbitrary vector $z = [z_1, \dots, z_J]$, define

$$w(z) = \exp \left\{ -\frac{\|z\|_{\infty}}{\sigma} \right\},$$

where $\|z\|_{\infty} = \max\{|z_1|, \dots, |z_J|\}$ is the l_{∞} norm of z , and σ is a scale parameter. The weights are constructed in a way such that the larger the extremum of $\Theta_{p,\cdot}$, the smaller the weighted group lasso penalty. The l_{∞} norm is preferred over the common choices of l_1 or l_2 norm because we target at biomarkers with large effects in at least a period of time. Furthermore, we choose exponential decay form of $\|z\|_{\infty}$ instead of $\|z\|_{\infty}^r$, $r < 0$, because $\|z\|_{\infty}^r$ functions are undefined for vectors of zeros.

We further study two versions of weighted group lasso—adaptive group lasso and integrative group lasso. Adaptive group lasso is a two-stage procedure. In the first stage, we obtain a \sqrt{n} -consistent estimator $\hat{\Theta}^0$, for example, the least square estimator $\hat{\Theta}^{LS}$. Then we plug $\hat{\Theta}^0$ into the weight function $w(z)$ for $p = 1, \dots, P$. The objective function of the adaptive group lasso is

$$\arg \min_{\Theta} \left\{ \text{SSE}(\Theta) + \lambda \left(\sum_{p=1}^P w(\hat{\Theta}_{p,\cdot}^0) \|\Theta_{p,\cdot}\| \right) \right\}. \quad (1)$$

The performance of adaptive group lasso depends on the accuracy of the initial estimator, which can be problematic especially when P is large. To remedy this issue, we propose the following integrative group lasso objective function,

$$\arg \min_{\Theta} \left\{ \text{SSE}(\Theta) + \lambda \left(\sum_{p=1}^P w(\Theta_{p,\cdot}) \|\Theta_{p,\cdot}\| \right) \right\}. \quad (2)$$

The weight function in the integrative group lasso contains the unknown parameter Θ . Therefore, no prior information is needed. On the contrary, parameters in the weights are estimated together with parameters in the loss and penalty functions in one step.

We illustrate the unweighted and integrative weighted group lasso penalties using a simplified example of a coefficient vector z with two elements z_1 and z_2 . The weighted penalties $\lambda w(z) \|z\|$ are plotted side by side in Fig. 1. The ranges of z_1 and z_2 are restricted to $(-10, 10)$. To compare the shapes, different λ values are used so that the ranges of the four functions are all $[0, 1]$. It can be seen that the weights result in a bounded penalty function, compared to the unweighted penalty which always increases when either z_1 or z_2 increases. The plot also helps us understand the extra tuning parameter σ in addition

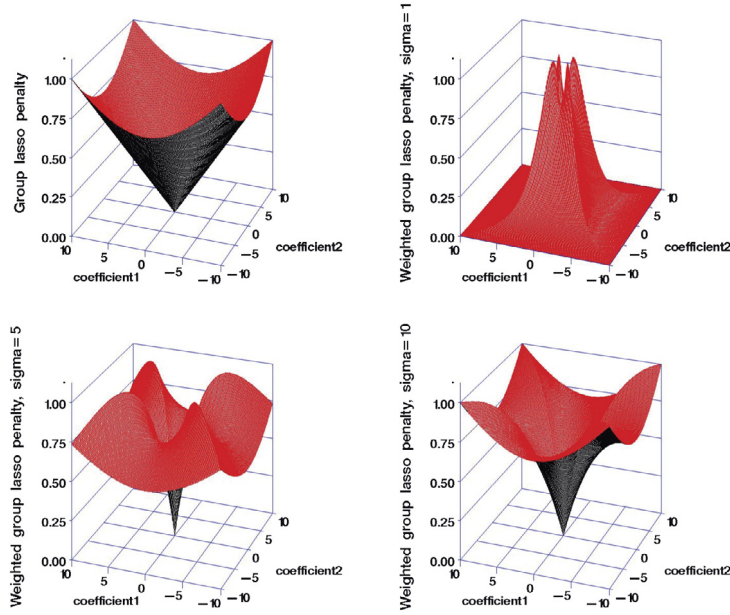


Fig. 1. Comparison of the unweighted and weighted penalty functions for a two-dimensional coefficient vector.

to λ , which is set to regulate the locations of the peaks of $\lambda w(z) \|z\|$. In the cases in Fig. 1, for any direction $z_1 = rz_2$, it can be shown that the integrative group penalty peaks at $\frac{1+r}{1+r^2} \sigma$.

3. Estimating procedures

3.1. LQA for adaptive group lasso

In the adaptive group lasso, the parameter estimates minimizing (1) can be solved by a direct application of local quadratic approximation (LQA). Following Fan and Li (2001), let $\hat{\Theta}^{(m)}$ be the solution at the m th iteration. The LQA goes as follows,

$$\|\Theta_p\| \approx \|\hat{\Theta}_p^{(m)}\| + \frac{\sum_{j=1}^J (\theta_{pj}^2 - (\hat{\theta}_{pj}^{(m)})^2)}{2 \|\hat{\Theta}_p^{(m)}\|}. \tag{3}$$

The group lasso penalty is approximated by an l_2 penalty and the optimization of the objective function is reduced to a ridge regression.

$$\begin{aligned} \mathbf{SSE}(\Theta) + \lambda \left(\sum_{p=1}^P w(\hat{\Theta}_p^0) \|\Theta_p\| \right) &\approx \mathbf{SSE}(\Theta) + \frac{\lambda}{2} \left(\sum_{p=1}^P \frac{w(\hat{\Theta}_p^0) \sum_{j=1}^J \theta_{pj}^2}{\|\hat{\Theta}_p^{(m)}\|} \right) + \frac{\lambda}{2} \left(\sum_{p=1}^P w(\hat{\Theta}_p^0) \|\hat{\Theta}_p^{(m)}\| \right) \\ &= \|\mathbf{Y} - \mathbb{X}\Theta\|^2 + \lambda \Theta' \Phi^{A(m)} \Theta + \frac{\lambda}{2} \left(\sum_{p=1}^P w(\hat{\Theta}_p^0) \|\hat{\Theta}_p^{(m)}\| \right), \end{aligned} \tag{4}$$

where $\Phi^{A(m)} = \text{diag} \left(V_0, \frac{w(\hat{\Theta}_1^0)}{2\|\hat{\Theta}_1^{(m)}\|} I_J, \dots, \frac{w(\hat{\Theta}_p^0)}{2\|\hat{\Theta}_p^{(m)}\|} I_J \right)$, and V_0 and I_J are $J \times J$ zero and identity matrices, respectively. The minimization of (4) has a closed form solution,

$$\hat{\Theta}^{(m+1)} = (\mathbb{X}'\mathbb{X} + \lambda \Phi^{A(m)})^{-1} \mathbb{X}'\mathbf{Y}.$$

The solution $\hat{\Theta}^{(m+1)}$ can be rewritten as a matrix $\hat{\Theta}^{(m+1)}$ where $\hat{\Theta}_p^{(m+1)} = [\hat{\Theta}_{pJ+1}^{(m+1)}, \dots, \hat{\Theta}_{(p+1)J}^{(m+1)}]$. If $\|\hat{\Theta}_p^{(m+1)}\| < 10^{-8}$, we set $\hat{\Theta}_p = 0$ and remove the p th covariate from the candidate list. We repeat the procedure above until convergence.

The adaptive group lasso can be solved using readily usable computation packages for group lasso, such as the “grplasso” package (Meier, 0000) in R, following the transformation in Zou (2006) to rewrite a weighted group lasso in the form of an unweighted group lasso.

3.2. GLQA for integrative group lasso

Since weights in the integrative group lasso are also functions of unknown Θ , more complicated quadratic approximation needs to be developed. We begin with calculating the derivative of the penalty function $\mathbf{Pen}(\Theta_{p.})$ in the integrative group lasso where $\mathbf{Pen}(\Theta_{p.}) = w(\Theta_{p.})\|\Theta_{p.}\|$.

$$\frac{\partial \mathbf{Pen}(\Theta_{p.})}{\partial \theta_{pj}} = d_{pj}\theta_{pj},$$

where

$$d_{pj} = \begin{cases} w(\Theta_{p.})(\|\Theta_{p.}\|)^{-1} & \text{if } |\theta_{pj}| \neq \|\Theta_{p.}\|_{\infty}, \\ w(\Theta_{p.})(\|\Theta_{p.}\|)^{-1} - w(\Theta_{p.})\|\Theta_{p.}\|(|\theta_{pj}|\sigma)^{-1} & \text{if } |\theta_{pj}| = \|\Theta_{p.}\|_{\infty}. \end{cases} \tag{5}$$

Again, let $\hat{\Theta}_{p.}^{(m)}$ be the solution from the m th iteration. By replacing θ_{pj} with $\hat{\theta}_{pj}^{(m)}$ in (5), we obtain $d_{pj}^{(m)}$. Had we applied the LQA method in (3), $\mathbf{Pen}(\Theta_{p.})$ would be approximated by the following function,

$$\mathbf{Pen}(\Theta_{p.}) \approx \mathbf{Pen}(\hat{\Theta}_{p.}^{(m)}) + \frac{1}{2} \sum_{j=1}^J d_{pj}^{(m)} \{\theta_{pj}^2 - (\hat{\theta}_{pj}^{(m)})^2\}. \tag{6}$$

However, a key difference between (6) and (3) is that sometimes $d_{pj}^{(m)} < 0$ when $|\theta_{pj}| = \|\Theta_{p.}\|_{\infty}$. As a consequence, the right hand side of (6) is not guaranteed to be convex. Therefore, we propose a set of general guidelines to design convex quadratic approximations for any penalty function, called generalized local quadratic approximation (GLQA). A quadratic function $G(\Theta_{p.})$ is a generalized local quadratic approximation of $\mathbf{Pen}(\Theta_{p.})$ at $\hat{\Theta}_{p.}^{(m)}$, if it satisfies the following three conditions:

1. $G(\Theta_{p.})$ is convex.
2. $G(\hat{\Theta}_{p.}^{(m)}) = \mathbf{Pen}(\hat{\Theta}_{p.}^{(m)})$.
3. For all p and j ,

$$\left. \frac{\partial G(\Theta_{p.})}{\partial \theta_{pj}} \right|_{\theta_{pj}=\hat{\theta}_{pj}^{(m)}} = \left. \frac{\partial \mathbf{Pen}(\Theta_{p.})}{\partial \theta_{pj}} \right|_{\theta_{pj}=\hat{\theta}_{pj}^{(m)}}.$$

For adaptive group lasso, GLQA reduces to the standard LQA, since (3) satisfies the three conditions above.

For integrative group lasso, the classical LQA $\mathbf{Pen}(\hat{\Theta}_{p.}^{(m)}) + \frac{1}{2} \sum_{j=1}^J d_{pj}^{(m)} \{\theta_{pj}^2 - (\hat{\theta}_{pj}^{(m)})^2\}$ satisfies the second and third condition. However, it may not satisfy the first condition since some $d_{pj}^{(m)} < 0$.

We propose the following approximation,

$$\mathbf{Pen}(\Theta_{p.}) \approx \mathbf{Pen}(\Theta_{p.}^{(m)}) + \sum_{j=1}^J Q(\theta_{pj}),$$

where $Q(\theta_{pj}) = \frac{1}{2} |d_{pj}^{(m)}| [(\theta_{pj} + c_1)^2 + c_2]$,

with c_1 and c_2 being constants not including θ_{pj} .

$\mathbf{Pen}(\Theta_{p.}^{(m)}) + \sum_{j=1}^J Q(\theta_{pj})$ satisfies the first condition since the second derivative of $Q(\theta_{pj})$ equals $|d_{pj}^{(m)}|$ which is always nonnegative. We need to select the proper values for c_1 and c_2 in order to satisfy the second and third condition, which leads to the following function

$$Q(\theta_{pj}) = \frac{1}{2} |d_{pj}^{(m)}| \left[[\theta_{pj} - \{1 - \text{sgn}(d_{pj}^{(m)})\} \hat{\theta}_{pj}^{(m)}]^2 - (\hat{\theta}_{pj}^{(m)})^2 \right].$$

Here $\text{sgn}(\cdot)$ is the sign function, i.e.,

$$\text{sgn}(z) = \begin{cases} 1 & \text{if } z > 0, \\ -1 & \text{otherwise.} \end{cases}$$

GLQA is solved as follows. Define a $(P + 1) \times J$ matrix $D^{(m)} = [\tilde{d}_{pj}^{(m)}]$, where

$$\tilde{d}_{pj}^{(m)} = \begin{cases} |d_{pj}^{(m)}| & \text{if } p \geq 1, \\ 0 & \text{if } p = 0, \end{cases}$$

and let $\Phi^{I(m)} = \text{diag}[\text{vec}(D^{(m)'})]$. Define another $(P + 1) \times J$ matrix $C^{(m)} = [c_{pj}^{(m)}]$, where

$$c_{pj}^{(m)} = \begin{cases} |d_{pj}^{(m)}| \{1 - \text{sgn}(d_{pj}^{(m)})\} \hat{\theta}_{pj}^{(m)} / 2 & \text{if } p \geq 1, \\ 0 & \text{if } p = 0, \end{cases}$$

and let $\mathbf{C}^{(m)} = \text{vec}(C^{(m)'})$. At the $(m + 1)$ th iteration, we minimize the following quadratic function

$$\text{SSE}(\Theta) + \lambda \sum_{p=1}^P \sum_{j=1}^J Q_{pj} = \|\mathbf{Y} - \mathbb{X}\Theta\|^2 + 2\lambda\Theta' \Phi^{I(m)} \Theta - \lambda \mathbf{C}^{(m)'} \Theta. \tag{7}$$

The solution of (7) goes as follows

$$\hat{\Theta}^{(m*)} = (\mathbb{X}'\mathbb{X} + \lambda\Phi^{I(m)})^{-1} (\mathbb{X}'\mathbf{Y} + \lambda\mathbf{C}^{(m)}).$$

Let $\hat{\Theta}^{(m*)}$ be the corresponding matrix form of $\hat{\Theta}^{(m*)}$. Since the approximation is only accurate near $\hat{\Theta}^{(m)}$, $G(\hat{\Theta}^{(m+1)})$ may give a very poor approximation of $\text{Pen}(\hat{\Theta}^{(m+1)})$ when $\hat{\Theta}^{(m*)}$ is far away from $\hat{\Theta}^{(m)}$. We thereby employ the back-tracking line search algorithm to guarantee monotone decreasing updates of the objective function at each iteration. One can refer to Chapter 9 of Conway (2004) for details. The parameter estimate $\hat{\Theta}_{pj}^{(m)}$ is updated by a value between itself and $\hat{\Theta}_{pj}^{(m*)}$. The updated estimate $\hat{\Theta}_{pj}^{(m+1)}$ is the one giving the longest step length among the parameter values on the straight line connecting $\hat{\Theta}_{pj}^{(m)}$ and $\hat{\Theta}_{pj}^{(m*)}$ that satisfy $\text{SSE}(\hat{\Theta}^{(m+1)}) + \lambda \sum_{p=1}^P \text{Pen}(\hat{\Theta}_p^{(m+1)}) \leq \text{SSE}(\hat{\Theta}^{(m)}) + \lambda \sum_{p=1}^P \text{Pen}(\hat{\Theta}_p^{(m)})$.

Similar to Section 3.1, at each iteration, if $\|\hat{\Theta}_p^{(m+1)}\| < 10^{-8}$, we set $\hat{\Theta}_p = 0$ and remove the p th covariate from the candidate list. With an initial value $\hat{\Theta}^{(0)}$, we iterate the procedure above until the gradient of the objective function converges to zero. It is worth pointing out that the GLQA procedure may converge to local minimal points, which can be remedied by trying different starting values of the parameters.

Remark 3.1. Zou and Li (2008) proposed an algorithm based on local linear approximation (LLA). Using our notation, $\text{Pen}(\Theta_p)$ can be approximated by

$$\text{Pen}(\Theta_p) \approx \text{Pen}(\hat{\Theta}_p^{(m)}) + \sum_{j=1}^J d_{pj}^{(m)} \{|\theta_{pj}| - |\hat{\theta}_{pj}^m|\}. \tag{8}$$

However, similarly to LQA, the LLA algorithm requires all weights to be non-negative. Therefore, the approximation above cannot be directly used to solve the objective function for the integrative lasso since $d_{pj}^{(m)} < 0$ sometimes. Moreover, LLA cannot be reformatted to make all weights non-negative while satisfying conditions 2 and 3 due to the inflexibility of the linear approximation. In fact, the proposed GLQA (6) is a much more flexible algorithm compared to existing ones because it does not require the weighted penalty function to be nonconcave.

4. Asymptotic properties

We establish the asymptotic properties of the adaptive and integrative group lasso estimators. The tuning parameters λ and σ are denoted by λ_n and σ_n in this section to reflect the fact that they depend on the sample size n .

Theorem 4.1 (Theorem 1). Let $\hat{\Theta}^A$ and $\hat{\Theta}^I$ be the adaptive and integrative group lasso estimator, respectively, and $\hat{\Theta}^0$ be a \sqrt{n} -consistent initial estimator for Θ . Furthermore, we assume $(y_{i1}, \dots, y_{iT}, x_{i1}, \dots, x_{ip})$ having finite fourth moment for $i = 1, \dots, n$. Let $\mathcal{P} = \{p : \|\Theta_p\| \neq 0\}$ and \mathcal{P}^c be the complementary set of \mathcal{P} . Then we have the following conclusions.

1. $\hat{\Theta}^A \xrightarrow{P} \Theta$ and $\mathbb{P}(\hat{\Theta}_{\mathcal{P}^c}^A = 0) \rightarrow 0$, if $\lambda_n/n \rightarrow 0$, $\lambda_n/\sqrt{n} \rightarrow \infty$, $\sigma_n \rightarrow 0$ and $\sigma_n\sqrt{n} \rightarrow \infty$, as $n \rightarrow \infty$.
2. $\hat{\Theta}^I \xrightarrow{P} \Theta$ and $\mathbb{P}(\hat{\Theta}_{\mathcal{P}^c}^I = 0) \rightarrow 0$, if $\lambda_n/n \rightarrow 0$, $\lambda_n/\sqrt{n} \rightarrow \infty$ and $\sigma_n \rightarrow 0$, as $n \rightarrow \infty$.

The proof of Theorem 1 is given in Appendix A. In Appendix A, we prove that $\hat{\Theta}^A$ and $\hat{\Theta}^I$ satisfy the following optimality conditions, respectively,

$$\|\mathbb{X}_p\mathbf{Y} - \mathbb{X}_p\mathbb{X}\text{vec}(\hat{\Theta}^A)\| \leq \lambda_n w(\hat{\Theta}_p^0), \quad \text{for } \hat{\Theta}_p^A = \mathbf{0}, \tag{9}$$

$$\|\mathbb{X}_p\mathbf{Y} - \mathbb{X}_p\mathbb{X}\text{vec}(\hat{\Theta}^I)\| \leq \lambda_n, \quad \text{for } \hat{\Theta}_p^I = \mathbf{0}, \tag{10}$$

where $\mathbf{0}$ is a $J \times 1$ vector of zeros. To prove sparsistency, we construct estimators with the correct non-zero set which asymptotically satisfies (9) or (10). Condition (9) of the adaptive group lasso is similar to the corresponding optimality condition in group lasso (Yuan and Lin, 2006). Following the proof of the adaptive lasso (Zou, 2006), $w(\hat{\Theta}_p^0) \rightarrow 1$ is required for the correct non-zero set to satisfy condition (9), which is guaranteed if $\lambda_n/\sqrt{n} \rightarrow \infty$. By contrast, the proof for the

sparsistency is quite different when parameters in weights take their true values because condition (10) does not require $w(\hat{\Theta}_p^0) \rightarrow 1$. In fact, the right hand side of (10) does not depend on σ_n because $w(\Theta_p) = 0$ automatically holds when $\Theta_p = \mathbf{0}$. Therefore, $\hat{\Theta}^l$ is theoretically superior because it does not require $\sigma_n \sqrt{n} \rightarrow \infty$ in the proof of sparsistency.

We now derive the covariance estimator of the non-zero rows of $\hat{\Theta}^l$, following the sandwich estimator in Fan and Li (2001). Without loss of generality, suppose at the m th iteration the first $a + 1$ rows of $\hat{\Theta}^{l(m)}$ are nonzero, denoted by $\hat{\Theta}_{[a+1]}^{l(m)}$. Let $\mathbb{X}_{[a+1]}$ be the first $a + 1$ columns of \mathbb{X} . Define $\Phi^l(\hat{\Theta}_{[a+1]}^{l(m)})$ and $\mathbf{C}_{[a+1]}^{(m)}$ by replacing $\hat{\Theta}^{l(m)}$ with $\hat{\Theta}_{[a+1]}^{l(m)}$ in the definitions of $\Phi^{l(m)}$ and $\mathbf{C}^{(m)}$. Through GLQA, $\hat{\Theta}_{[a+1]}^{l(m)}$ is updated by,

$$\hat{\Theta}^{l(m*)} = \left(\mathbb{X}'_{[a+1]} \mathbb{X}_{[a+1]} + \lambda \Phi^l(\hat{\Theta}_{[a+1]}^{l(m)}) \right)^{-1} \left(\mathbb{X}'_{[a+1]} \mathbf{Y} + \lambda \mathbf{C}_{[a+1]}^{(m)} \right).$$

At the time of convergence, $\hat{\Theta}_{[a+1]}^{l(m)} \approx \hat{\Theta}_{[a+1]}^{l(m*)} = \hat{\Theta}_{[a+1]}^l$. Therefore, the estimated covariance matrix is

$$\text{Cov}(\hat{\Theta}_{[a+1]}^l) = \hat{\zeta}^2 \left(\mathbb{X}'_{[a+1]} \mathbb{X}_{[a+1]} + \lambda \Phi^l(\hat{\Theta}_{[a+1]}^l) \right)^{-1} \mathbb{X}'_{[a+1]} \mathbb{X}_{[a+1]} \left(\mathbb{X}'_{[a+1]} \mathbb{X}_{[a+1]} + \lambda \Phi^l(\hat{\Theta}_{[a+1]}^l) \right)^{-1},$$

where $\hat{\zeta}^2$ is the variance estimated from the least square estimators without penalty. The covariance estimation for adaptive group lasso is derived similarly and omitted here.

5. Simulation studies

In this section, we compare the performance of unweighted group lasso, adaptive group lasso and integrative group lasso through simulations. Each dataset contains 200 independent observations. The outcome Y is generated as $y_{ik} = \sum_{p=0}^{30} \beta_p(t_k) x_{ip} + \epsilon_{ik}$, $k = 1, \dots, 20$, where the error term ϵ_{ik} is independent from x_{ip} . Furthermore, the error vector $(\epsilon_{i1}, \dots, \epsilon_{i20})$ follows multivariate normal distribution with mean zero, standard deviation 0.5 and auto-regressive correlation matrix with $\rho(\epsilon_{ik_1}, \epsilon_{ik_2}) = 0.5^{|k_1 - k_2|}$ for $k_1 = 1, 2, \dots, 20$ and $k_2 = 1, 2, \dots, 20$. Covariates x_{i1}, \dots, x_{i30} also follow multivariate normal distribution with auto-regressive covariance matrix $\rho(x_{ip_1}, x_{ip_2}) = 0.5^{|p_1 - p_2|}$ for $p_1 = 1, 2, \dots, 30$ and $p_2 = 1, 2, \dots, 30$. Correspondingly, there are 31 coefficient functions $\beta_0(t), \beta_1(t), \beta_2(t), \dots, \beta_{30}(t)$. Regression coefficient functions $\beta_{11}(t), \dots, \beta_{30}(t)$ are all zero. The true values of the eleven nonzero coefficient functions are listed below,

$$\begin{aligned} \beta_0(t) &= 1/4, \\ \beta_1(t) &= 1/4 \cos(t), \\ \beta_2(t) &= 1/4 \sin(t), \\ \beta_3(t) &= 1/4 \cos(t) I(0 \leq t \leq \pi), \\ \beta_4(t) &= 1/4 \sin(t) I(0 \leq t \leq \pi), \\ \beta_5(t) &= 1/4 \cos(t/2) I(0 \leq t \leq 2\pi), \\ \beta_6(t) &= 1/4 \sin(t/2) I(0 \leq t \leq 2\pi), \\ \beta_7(t) &= 1/4 \cos(t/3) I(0 \leq t \leq 3\pi), \\ \beta_8(t) &= 1/4 \sin(t/3) I(0 \leq t \leq 3\pi), \\ \beta_9(t) &= 1/4 \cos(t/4) I(0 \leq t \leq 4\pi) \text{ and} \\ \beta_{10}(t) &= 1/4 \sin(t/4) I(0 \leq t \leq 4\pi). \end{aligned}$$

In each setup, 500 replicates are generated.

Because it is computationally infeasible to fine-tune the optimal knots for thousands of candidate genomic markers with different regression coefficient functions, we follow the general guidelines given in Wold (1974) to choose knots for B-splines, ensuring that there are at least 4–5 time points and no more than one extremum or inflexion points per interval. Five evenly based interval knots are employed for the simulation study with 20 time points.

Firstly, we compare the performance of unweighted group lasso, adaptive group lasso and integrative group lasso when the tuning parameters are set to the same values in Figs. 2 and 3. The patterns of the mean squared prediction errors (MSPEs) and extended Bayesian Information Criterion (BIC) (Chen and Chen, 2008) for the two weighted group lasso procedures are plotted over a series of pre-specified λ and σ values in Fig. 2. The objective function of extended BIC is

$$\text{SSE}(\Theta) + \nu \log n + \log \binom{P_j}{\nu}, \quad (11)$$

where ν is the number of non-zero $\hat{\theta}_{pj}$ selected by the method under evaluation. The calculation of MSPEs is based on validation datasets independent of the datasets used in variable selection and coefficient estimation. To see the comparison of sensitivities/specificities and MSPEs more clearly, we further compare the three procedures by fixing $\sigma = 5$ and allowing λ to change in Fig. 3. We first plot the receiver operating characteristic (ROC) curves to compare the performances in terms

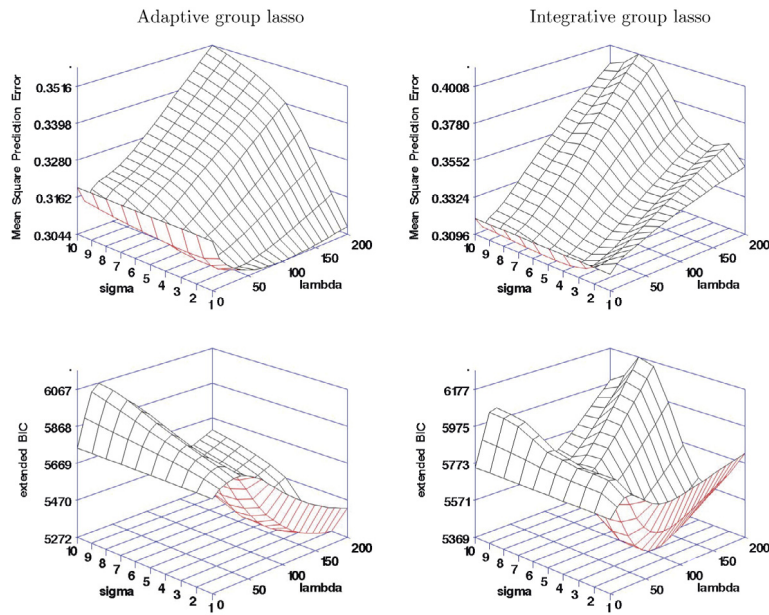


Fig. 2. Adaptive group lasso vs. integrative group lasso vs. (λ, σ) values.

Table 1
Comparison of five lasso procedures.

Method	True Positive	False Positive	MSPE	Bias ²	Variance	MSE
1	0.4432	0.0672	0.4307	–	–	–
2	0.7392	0.0267	0.3468	0.0030	0.0001	0.0032
3	0.9128	0.1019	0.3168	0.0021	0.0005	0.0026
4	0.8940	0.0549	0.3287	0.0020	0.0012	0.0032
5	1	0	0.2975	0.0011	0.0005	0.0016

Method 1: Lasso. Method 2: Group lasso. Method 3: Adaptive group lasso. Method 4: Integrative group lasso. Method 5: Oracle.

of variable selection. Both weighted group lasso procedures show similar performance according to ROC curves, and both of them have larger areas under the curve (AUCs) than unweighted group lasso. We further plot the MSPEs versus false positive rates in Fig. 3. The MSPEs for both weighted procedures are smaller than the MSPEs for the unweighted procedure at the same false positive level. Moreover, the MSPEs for the integrative group lasso is larger than the adaptive lasso. As mean square errors can be decomposed into bias square and variance, we plot the average squared biases and variances of $\hat{\beta}$ from the three procedures. The integrative group lasso procedure gives the smallest biases but the largest variances, which may result from the true parameter values in the weights of integrative group lasso which require the more complicated GLQA algorithm.

Secondly, we compare the performance of the three group lasso procedures plus the lasso, which ignores the time-varying effects of $\beta_p(t)$, at the optimal tuning parameter values selected by minimizing extended BIC. We report in Table 1 the average true positive rates, false positive rates and MSPEs from the four methods. We also report the average mean square errors, average squared biases and variances of $\hat{\beta}$ estimated by the unweighted, adaptive and integrative group lasso. We also include the MSPE by the oracle procedure, which fits time-varying coefficients in the form of cubic B-splines for x_0, \dots, x_{10} and excludes x_{11}, \dots, x_{30} . As shown in Table 1, the lasso method ignoring the time-varying property of $\beta_p(t)$ performs worst with the highest MSPEs and the lowest true positive rates. Even though the unweighted group lasso gives slightly lower false positive rates, its true positive rates are much lower and MSPEs are higher than the weighted methods. The MSPEs of the adaptive and integrative group lasso procedures are comparable. Moreover, the integrative group lasso has the smallest bias but the largest variance at the optimal tuning parameters. And the adaptive group lasso gives the best MSE. These observations are consistent with Fig. 3.

Fig. 4 plots mean $(\hat{\beta}_1(t), \hat{\beta}_5(t), \hat{\beta}_9(t), \hat{\beta}_{13}(t))$ and $(\hat{\beta}_2(t), \hat{\beta}_6(t), \hat{\beta}_{10}(t), \hat{\beta}_{14}(t))$ at the optimal λ and σ values. The estimated regression coefficient curves $\hat{\beta}_1(t), \hat{\beta}_2(t), \hat{\beta}_5(t), \hat{\beta}_6(t)$ obtained by integrative group lasso are less biased than the estimates by the other two methods. For $\beta_9(t)$ and $\hat{\beta}_{10}(t)$, all three methods give similar estimates. And all the methods correctly identify the zero coefficient function $\beta_{13}(t)$ and $\hat{\beta}_{14}(t)$.

In summary, the proposed weighted methods provide smaller bias and MSPE as well as higher true positive rates at the same false positive values and at the optimal (λ, σ) value.

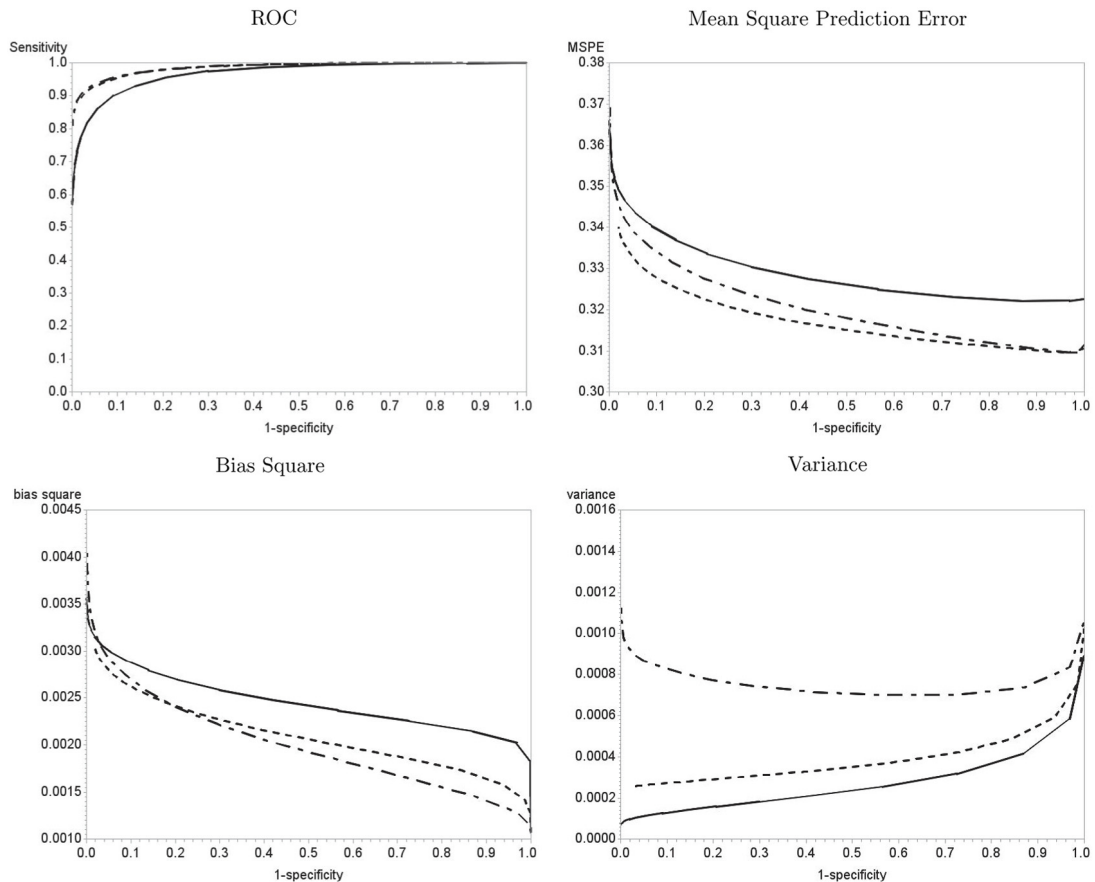


Fig. 3. Comparison of unweighted group lasso (solid), adaptive group lasso (dash) and integrative group lasso (dot-dash) when $\sigma = 5$.

6. Application to GWAS data

The fast advancements of sequencing techniques have provided rich information in genome, epigenome, transcriptome and proteome. However, the number of clinically meaningful biomarker discoveries are less than what researchers expected and the discovered biomarkers account for small percentages of the total heritable phenotype variation (Maher, 2008). One of the challenges is the efficient extraction of phenotypic information representative of the underlying complex diseases. Binary case and control data ignore the dynamic patterns of clinical symptoms. Therefore, studies using such outcomes have limited power to detect biomarkers with strong dynamic effects. Recently, more and more genomic studies are employing longitudinal clinical and biochemical measurements. Such outcomes provide opportunities to capture the time-varying effects of biomarkers.

The Diabetes Control and Complications Trial (DCCT) and Epidemiology of Diabetes Intervention and Complication (EDIC) trial are multi-center randomized clinical trials designed to assess the effects of an intensive glucose monitoring intervention on the incidence of microvascular complications among Type 1 diabetic patients. At the enrollment of DCCT, patients were randomized to either the intensive or conventional therapy. In total, 1441 Type 1 diabetic patients enrolled in DCCT from 1983 to 1989. DCCT ended in 1993 when significant reduction in the risk of microvascular complications was found in the intensive therapy group (The DCCT Research Group, 1993). EDIC study started at the end of DCCT where 1394 of the 1441 DCCT participants continued being followed and intensive therapy was provided to all participants. We are particularly interested in one kind of macrovascular complications—nephropathy, which is commonly measured by urine glomerular filtration rates (GFR). GFR has been measured annually in DCCT/EDIC for up to 26 years. Logarithm transformation is employed for the right-skewed GFR values and used as the outcome. In the real data, some GFR values are missing, especially at the end of follow-up. Missing values in the response are handled conveniently through vectorization. After transforming Y and X into \mathbf{Y} and \mathbf{X} , we remove \mathbf{Y}_l and the l th row of \mathbf{X} when \mathbf{Y}_l is missing.

In total, 1 039 611 candidate SNPs are genotyped in the EDIC GWAS study. We exclude 173 571 SNPs with minor allele frequency less than 0.01. Among the remaining 866 040 markers, 2217 fail the Hardy–Weinberg equilibrium test with unadjusted P -values less than 0.001 and 863 823 passed. We also deleted SNPs with more than 50% missing. SNPs in linkage disequilibrium (>0.5) with other SNPs in a sliding window of 50 SNPs along the chromosomes are pruned out. The data

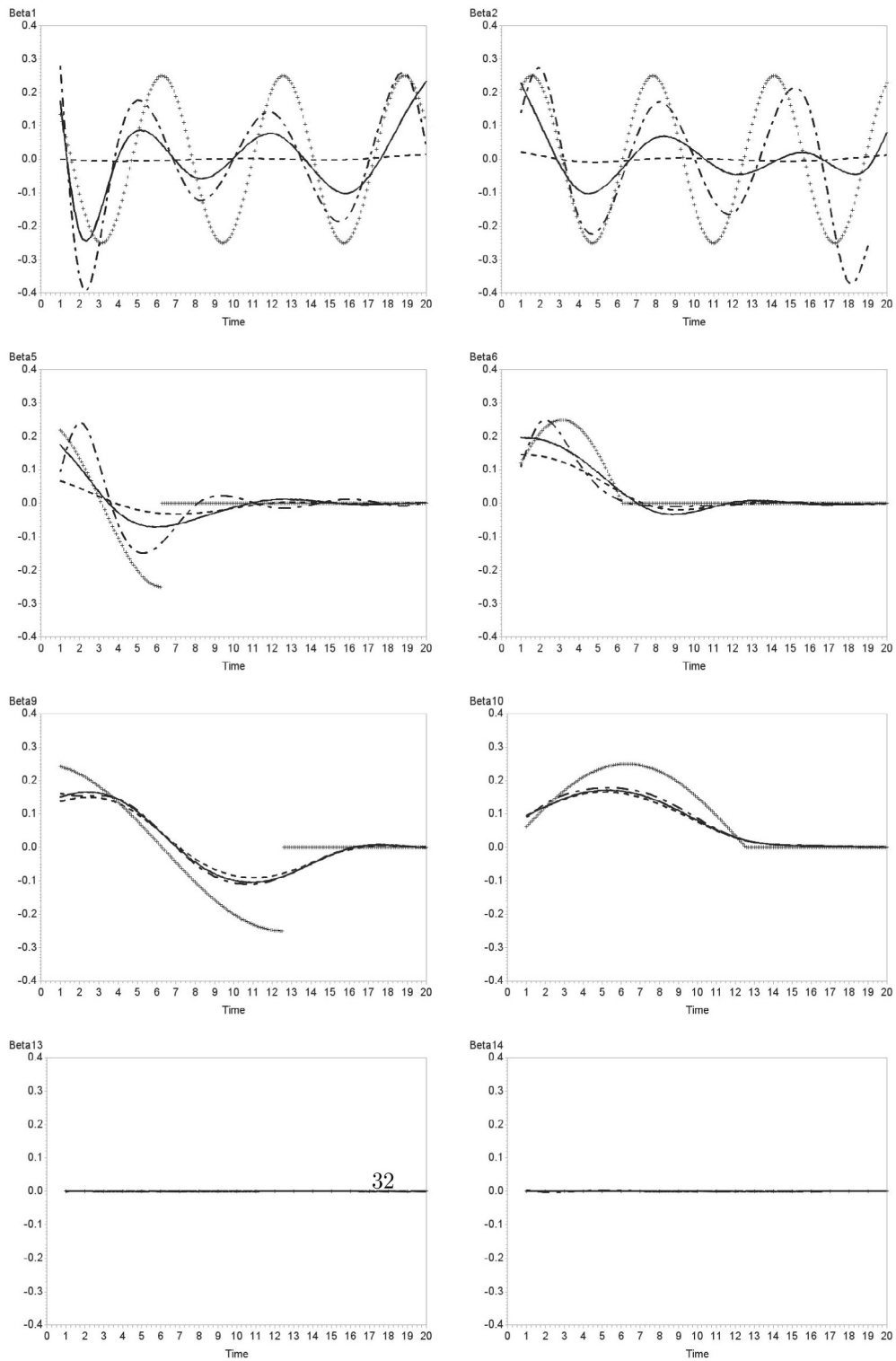


Fig. 4. True (plus sign) and estimated regression coefficient functions from unweighted group lasso (short dash), adaptive group lasso (solid) and integrative group lasso (dot-dash) over time.

cleaning procedure results in 293 469 candidate SNPs. Then we estimate time-varying coefficients in the form of cubic B-spline for the remaining SNPs on $\log(\text{GFR})$ one by one, adjusting for age at randomization, gender, treatment and duration

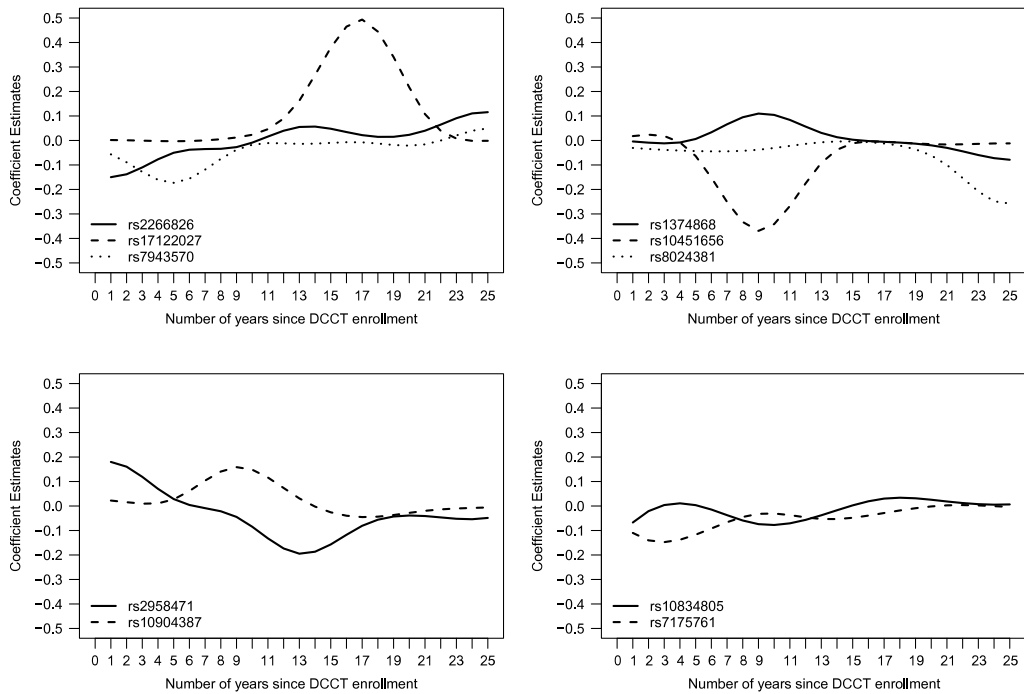


Fig. 5. Top ten SNP effects over time.

of diabetes at enrollment. Five evenly spaced internal nodes (year 4, 8, 12, 16, 20) are employed. Each spline coefficient vector has eight elements, and the smallest P -value from the eight corresponding P -values is taken as the significance level of the spline as we are interested in the extremum effects. Then, the 293 469 marginal P -values resulted from the one-at-a-time regressions are ranked from the smallest to the largest. One by one regressions estimate marginal effects which could be “pseudo” effects resulting from collinearity or confounding of other correlated biomarkers. We employ the proposed group lasso procedures to further select and estimate biomarker effects conditional on other markers. Due to the computation limitation on the number of SNPs that can be handled by group lasso, we set the top one percentile of all 293 469 marginal P -values as our threshold, which is 0.000005673. With such a threshold, 2864 SNPs get in the prescreened candidate set for the proposed group lasso procedures.

However, it takes $2868 \times 8 = 22\,944$ coefficients to specify the nonparametric effects of 2864 SNPs and four adjustment covariates. This number is still too large given current computation ability. We propose a chromosome-wise approximation procedure to accommodate large number of genome-wide biomarkers. As chromosomes are distributed into haploid cells in meiosis independently, we assume that genotypes of SNPs locating on different chromosomes are independent. When two sets of SNPs are independent, the estimated regression coefficients in the linear regression of one set are irrelevant to the other set. At the same time, note that the independence assumption is only approximately true as we ignore recombinations across chromosomes. The same tuning parameters λ and σ are used for all the 23 chromosomes and mitochondria to guarantee consistent selection stringency. We select the optimal tuning parameters by the following procedure. First, the coefficients of SNPs on chromosome 1 to 23 and mitochondria are estimated by the proposed weighted group lasso methods separately, with the same tuning parameters. Then we pool coefficient estimates of SNPs in all 24 regressions and calculate predictions of the outcome. The extended BIC is further calculated based on the predicted outcomes. Finally, tuning parameters are selected by minimizing the overall extended BIC through a grid search.

Adaptive group lasso selected 280 SNPs while integrative group lasso selected 243 SNPs, of which 191 overlaps. The SNPs selected by both weighted group lasso procedures are scattered on all chromosomes as listed in Table 2. Their relationship to GFR or retinopathy is worth further biological and clinical investigation.

In Fig. 5, we plotted the unpenalized least square estimates of the top ten regression coefficient functions, which are defined as the ten $\hat{\beta}(t)$ functions with the largest $\max_j(|\hat{\theta}_{pj}|)$ values using the integrative weights. All the ten SNPs are also selected by adaptive group lasso. It is worth noting that these top SNPs selected by the proposed methods do not overlap with another set of 22 SNPs known to be associated with severe nephropathy or persistent microalbuminuria in the DCCT/EDIC cohort (Al-Kateb et al., 2008) possibly because the previous publication treated regression coefficients of SNPs as scalars, ignoring potentially time-varying SNP effects over time. Besides, the 22 SNPs were ranked top using marginal p -values from one SNP at a time regressions.

Table 2
Selected SNPs for the log(GFR) outcome in DCCT/EDIC.

CHR	Selected by both procedures	Selected by only one
1	rs6426025 rs10493186 rs1570727 rs1323723 rs877343 rs704859 rs4915219 rs2799085	rs17162990 rs16825658 rs4926747 rs12409037 rs4920564 rs4654861 rs6702689 rs1949009 rs536014 rs2243188 rs10746361 rs11585386 rs11210966 rs3011207 rs11810829 rs17301783 rs312965 rs10201541 rs10207139 rs13385731 rs1428571 rs4952725
2	rs7582948 rs6709878 rs1400153 rs12470501 rs13384029 rs4246586 rs6545276 rs805312 rs2272429 rs11681842 rs6732834 rs13420256 rs17044762 rs745907 rs6430686 rs17384549 rs10497114 rs288077 rs4141835 rs4571049 rs17863775	rs11123744 rs163660 rs12620150 rs6716449 rs6706564 rs2430425 rs6753711 rs11894267 rs9808399
3	rs2686698 rs9311438 rs6443126 rs17781377 rs17042373 rs10510577 rs2041376 rs35845142 rs9854395 rs11711534 rs2027815 rs10511167 rs704574	rs13061928 rs161898 rs11715608 rs1642756 rs3852027 rs9289584 rs7624985 rs415926 rs17008142
4	rs2946405 rs4861106	rs4689705 rs4689157 rs13442 rs4689166 rs4696662 rs3828557 rs13114204 rs2973271 rs13119998 rs10019259 rs2298756
5	rs13178673 rs6420053 rs883535 rs3797954 rs16882600 rs297218 rs35354255 rs2029108	rs275460 rs2463506 rs10521011 rs12188200 rs1901509 rs12659828 rs8180405
6	rs1612811 rs1286044 rs179996 rs17771861 rs7775617 rs991974 rs12212942	rs9328033 rs3763236 rs1158059
7	rs4720377 rs12056266 rs293184 rs2107755 rs395158 rs11764487 rs7800244 rs921630 rs34768413 rs4727283	rs17133176 rs1709592 rs1156381 rs17169954 rs3801204 rs1978326 rs17241640 rs2113849 rs2351383 rs6976662
8	rs2912268 rs1504769 rs17855458 rs6601540 rs9792165 rs2683300 rs17524988 rs508419 rs6992462 rs4294223 rs7837124 rs6994178 rs261565	rs1616976 rs7820311 rs17055923
9	rs12238888 rs12351930 rs2381526 rs6477360 rs10961156 rs4144932 rs2888998 rs10867972 rs2277175 rs7044827 rs6478489 rs10820780	rs10973251 rs884886 rs4361859
10	rs17158243 rs10795733 rs12769311 rs12360332 rs10509433 rs1931144	rs7087615 rs12258781 rs12415542 rs6482487 rs2460543 rs10508899
11	rs1354803 rs7102266 rs7104985 rs1590287 rs2447524 rs3751031 rs5743248 rs12804886 rs12807005 rs12802587 rs12418496 rs11227644 rs4945343 rs1939616	rs12806058 rs3849994 rs7945325 rs11607491 rs1447576 rs4945357 rs7119735 rs11034732
12	rs4763561 rs9783464 rs3847860 rs2070875 rs17546579 rs10506416 rs4761378 rs17043605 rs11106355	rs6488591 rs7301316 rs12827160 rs10506202 rs7971372 rs2286646 rs4246224 rs2231696 rs11107199 rs3858659
13	rs7997997 rs3794387 rs9535625 rs10507602	rs551714 rs9578870 rs9600846 rs11616606 rs6562141 rs9284246
14	rs10148442 rs17126349 rs1676300 rs1959462 rs17126543 rs7152495	rs8005345 rs4904117
15	rs4779705 rs4438268 rs9652429 rs1711493 rs7180597 rs435137 rs8025780	rs7175761 rs4932438
16	rs2516787 rs12927811 rs13330694 rs8061112 rs1032618	rs12448481 rs4843349 rs4987018 rs17596071 rs9302384 rs11861871
17	rs16950266 rs11655238 rs11657599 rs7217851 rs3751928 rs7224144	rs471943 rs17821506
18	rs11665524 rs16945001 rs11081082 rs4796915 rs1284402 rs12608110 rs1469593 rs7243138 rs1460537 rs8094401 rs17205646 rs9947297 rs8087543	rs12960602 rs8084112 rs1371726 rs9945821 rs7229701
19	rs4807441 rs12427 rs1317877 rs11084046	rs12373539 rs8112223 rs10419308 rs7245988 rs380442
20	rs188205 rs1573166 rs225146 rs13038175 rs6097667	rs1047527 rs4811948 rs6061772 rs6010965
21	rs2832956 rs7279465 rs2410348 rs8133951 rs647688	rs2254583
22	rs5763914 rs4820864 rs2283956	rs9619738 rs6002232
23	rs7885992 rs5944670 rs2529554 rs7890762 rs12849033 rs5918554 rs1320317 rs17283948 rs5974525 rs909439	rs36073189 rs7061285 rs16994333 rs1554916 rs6628658

7. Conclusion

In this article, we model time-varying biomarker effects as nonparametric functions over time. Furthermore, we propose the weighted group lasso procedures based on extreme values to select markers with large maximum absolute effects at some time point. Novel generalized local quadratic approximation is developed for non-convex minimization of the integrative objective function. The integrative group lasso guarantees that the weight function equals one for truly zero coefficients and hence requires milder conditions in the proof of sparsistency. In simulation studies, the weighted group lasso procedures have larger AUC, smaller bias and MSPE compared to the unweighted group lasso. The proposed procedure is applied to the GWAS data from the DCCT/EDIC trial of type 1 diabetic patients. Three hundred and thirty-two SNPs prognostic of GFR values during follow-up and potentially related to nephropathy development are identified.

Both adaptive and integrative weights are designed to reduce the estimation bias, but they differ in the treatment of the parameters. Adaptive weights employ some first-stage estimates, which can be solved by tricking existing group lasso computing algorithms, but requires stronger condition in theoretical analysis. Integrative weights employ unknown parameter values and requires the generalized local quadratic approximation, but has better theoretical properties. This comparison provides new insights into the understanding of adaptive lasso and SCAD.

The idea of the adaptive and integrative weights can also be applied in other contexts. For example, the time-varying effects can be modeled by nonparametric regression methods other than B-splines. In addition, the groups may not be the vectors of spline coefficients. For example, in the screening of biomarkers for correlated outcomes measures of the same disease, the coefficients of the same biomarker for different outcomes can be viewed as a group. Finally, the choice of weights can also be flexible. Various weights can be developed to down-weight the penalty of markers with desirable characteristics. For example, in Yan and Huang (2002), the first element of their coefficient vector represents the time-invariant effect and the rest elements capture the time-varying effect. We could design corresponding adaptive and integrative procedures for emphasizing either the time-invariant or time-varying effects.

Acknowledgments

This publication was supported by Award Numbers UL1TR000075 and KL2TR000076 from the NIH National Center for Advancing Translational Sciences. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Center for Advancing Translational Sciences or the National Institutes of Health.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <http://dx.doi.org/10.1016/j.csda.2016.06.004>.

References

- Al-Kateb, H., Boright, A.P., Mirea, L., Xie, X., Sutradhar, R., Mowjoodi, A., Bharaj, B., Liu, M., Bucks, J.M., Arends, V.L., Steffes, M.W., Cleary, P.A., Sun, W., Lachin, J.M., Thorner, P.S., Ho, M., McKnight, A.J., Maxwell, A.P., Savage, D.A., Kidd, K.K., Kidd, J.R., Speed, W.C., Orchard, T.J., Miller, R.G., Sun, L., Bull, S.B., Paterson, A.D., 2008. Multiple superoxide dismutase 1/splicing factor serine alanine 15 variants are associated with the development and progression of diabetic nephropathy: The diabetes control and complications trial/epidemiology of diabetes interventions and complications genetics study. *Diabetes* 57 (1), 218–228.
- Chen, J., Chen, Z., 2008. Extended Bayesian information criteria for model selection with large model spaces. *Biometrika* 95 (3), 759–771.
- Chen, H., Wang, Y., 2011. A penalized spline approach to functional mixed effects model analysis. *Biometrics* 67, 861–870.
- Chen, Z., Zhang, H., 2008. Modelling gene regulation networks via multivariate adaptive splines. *Cancer Genomics Proteomics* 5, 55–62.
- Conway, D., 2004. *Convex Optimization*. Cambridge University Press, Cambridge, UK.
- Fan, J., Li, R., 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* 96 (456), 1348–1360.
- Fan, R., Zhang, Y., Albert, P., Liu, A., Wang, Y., Xiong, M., 2012. Longitudinal genetic analysis of quantitative traits. *Genet. Epidemiol.* 36, 856–869.
- Lin, Y., Zhang, H., 2006. Component selection and smoothing in multivariate nonparametric regression. *Ann. Statist.* 34 (5), 2272–2297.
- Maher, B., 2008. Personal genomes: The case of the missing heritability. *Nature* 456 (7218), 18–21.
- Meier, L., grplasso: Fitting user specified models with group lasso penalty, <http://CRAN.R-project.org/package=grplasso>.
- The DCCT Research Group, 1993. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N. Engl. J. Med.* 329, 977–986.
- Wang, L., Chen, G., Li, H., 2007. Group scad regression analysis for microarray time course gene expression data. *Bioinformatics* 23, 1486–1494.
- Wei, F., Huang, J., Li, H., 2011. Variable selection and estimation in high-dimensional varying-coefficient models. *Statist. Sinica* 21, 1515–1540.
- Wold, S., 1974. Spline functions in data analysis. *Technometrics* 16 (1), 1–11.
- Yan, J., Huang, J., 2002. Model selection for cox models with time-varying coefficients. *Biometrics* 68, 419–428.
- Yuan, M., Lin, Y., 2006. Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B* 68, 49–67.
- Zou, H., 2006. The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* 101 (476), 1418–1429.
- Zou, H., Li, R., 2008. One-step sparse estimates in nonconcave penalized likelihood models. *Ann. Statist.* 36 (4), 1509–1533.